

(1)

LONGITUDINAL ANALYSIS

Feb 8, 2019

Lecture 1 → MissedLecture 2 Review of ANOVA, 2 different ways

- hypothesis test
- regression model → more useful

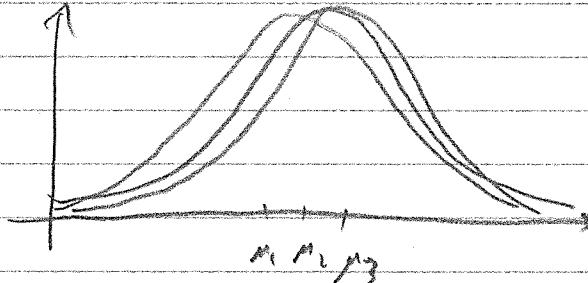
One-way ANOVA { 1 response var (continuous)

$k = \# \text{ groups}$ { 1 explanatory variable (categorical)

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$, $H_a: \text{at least one differs}$.

variance among groups > within groups

Ex H_0 true $k = 3$



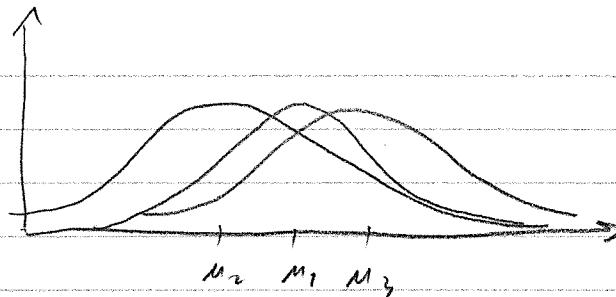
Compose 2 types of variability → between groups > within groups

- (1) Between groups → How spread out the group means are
 - (2) Var. within groups
↳ how much observations vary around group means?
- μ_1, μ_2, μ_3 from μ

In this case $\Rightarrow \mu_1 \approx \mu_2 \approx \mu_3$

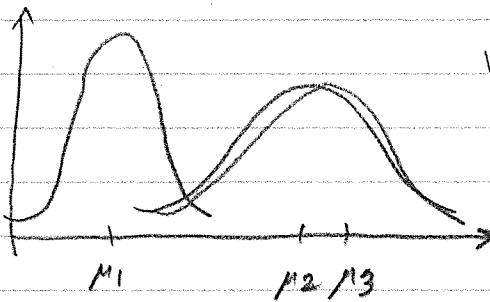
If (1) > (2) → evidence against H_0
 (1) < (2) → evidence for H_0

Ex H_0 true again



→ There's a lot of overlap → might not be able to tell if H_0 is true.

Ex H_0 true



Within var < between var

How to estimate ① - ②? (under aspt. that H_0 true)

- ② Consider var within group. If we assume $\sigma_i^2 = \sigma_j^2 + \epsilon_{ij}$, then an unbiased estimate of within group var would pool σ_i^2 together

$$\hookrightarrow \text{Var}_{\text{(within)}} \underset{\text{between groups}}{\cancel{\text{Var}}} \rightarrow \sum_{\text{groups}} (n_j - 1) s_j^2 \sim \chi^2 \quad \text{weight for sample size}$$

- ① Var between groups:

$$\sum_{\text{groups}} n_j (\bar{Y}_j - \bar{Y})^2 \sim \chi^2 \quad \text{overall mean}$$

$N = \# \text{ total observations}$

ANOVA table

Source of var	SS	df	MS	F	p-value
Between G's	SSG	$k-1$	$SSG/k-1$	MSB/MSE	
Within G's	SSE	$N-k$	$SSE/N-k$		
Total	SS_{tot}	$N-1$	$SS_{\text{tot}}/(N-1)$		

ANOVA Conditions

↳ Representative Sample (SRS) → sensitive

Very
Sensitive

- Equal Variance $\sigma_1^2 \approx \dots \approx \sigma_k^2 \rightarrow \max < 2 \times \min$

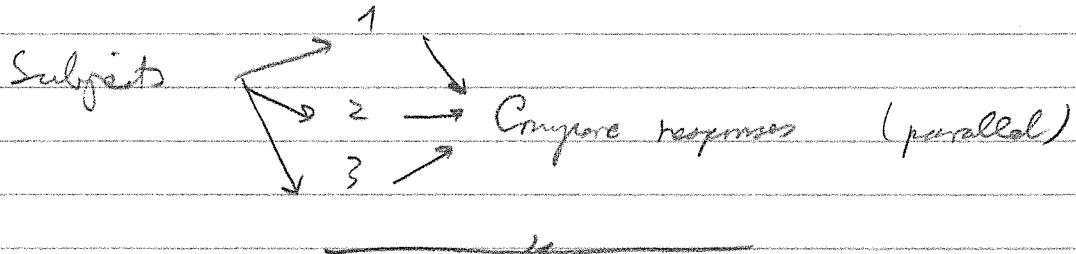
- Independent observations → sensitive

- Each group is normally dist or all n is large

not sensitive
robust

↳ look at Anova \$ residual

Ex R script weight loss - diet $k=3$ association?



ANOVA for regression

Y_{ij} = individual i in grp j 's response

$$Y_{ij} = \alpha + \beta_j + \epsilon_{ij}$$

↑ ↑ ↑
 reference effect error
 group

or

$$Y_{ij} = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \epsilon_{ij}$$

↑ ↑
 grp 1 grp 2
 indicator moderate

grp 1
indicator

grp 2
moderate

(4)

Repeated measures ANOVA

or Single-sample repeated measures
ANOVA

Feb 11, 2019

Consider N objects each measured at n balanced time periods Y_{ij} is the response from obj. i at time j $\left\{ \begin{array}{l} i = 1, 2, \dots, N \\ j = 1, 2, \dots, n \end{array} \right.$ "balanced" \rightarrow each object is measured at the same times

<u>Data</u>	<u>Subj</u>	<u>Data</u>	1	2	...	<u>n</u>
-------------	-------------	-------------	---	---	-----	-----------------------

1	Y_{11}	Y_{12}	.	.	Y_{1n}
---	----------	----------	---	---	----------

2	Y_{21}	.	.	.	Y_{2n}
---	----------	---	---	---	----------

:	:
---	---	---	---	---	---

<u>N</u>	Y_{N1}	.	.	.	Y_{Nn}
-----------------------	----------	---	---	---	----------

Model

$$Y_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij}$$

grand subject time error
mean effect effect

Note Two random terms on RHS

$$(1) \quad \varepsilon_{ij} \rightarrow \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

subject effect \rightarrow (2) $\pi_i \rightarrow N(0, \sigma_\pi^2)$ \rightarrow we call this "random effect"
 \rightarrow we consider this random

$$(3) \quad \tau_j \text{ is fixed} \rightarrow \text{and} \quad \sum \tau_j = 0$$

And so $E(Y_{ij}) = \mu + \tau_j$

$$V(Y_{ij}) = \sigma_\pi^2 + \sigma_\varepsilon^2$$

\hookrightarrow (subj effect is independent of error)

(5)

What about $\text{Cov}(Y_{ij}, Y_{ij'})$?

$\rightarrow \boxed{\text{Cov}(Y_{ij}, Y_{ij'}) = 0} \rightarrow \text{subjects are different at same time point}$

But $\boxed{\text{Cov}(Y_{ij'}, Y_{ij}) = \sigma^2_\pi}$

and $\boxed{\text{Corr}(Y_{ij}, Y_{ij'}) = \frac{\sigma^2_\pi}{\sigma^2_\pi + \sigma^2_\epsilon}}$ \rightarrow Intra-class correlation coefficient (ICC)

Note

We have an individual-specific random effect that is assumed to remain constant for all responses from the same individual

\rightarrow induces structure on correlation within-subjects

\hookrightarrow within-subject covariance matrix

$$\sum_i = \begin{bmatrix} \text{Var}(Y_{ii}) & & & \\ \text{Cov}(Y_{ii}, Y_{ii'}) & \text{Cov}(Y_{ii}, Y_{in}) & & \\ \vdots & \vdots & \ddots & \\ & & & \text{Var}(Y_{in}) \end{bmatrix} \nearrow n \times n$$

subject i

$$\sum = \begin{bmatrix} \sigma^2_\pi + \sigma^2_\epsilon & \sigma^2_\pi & \dots & \sigma^2_\pi \\ \sigma^2_\pi & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ \sigma^2_\pi & \dots & \sigma^2_\pi & \sigma^2_\pi \\ & \sigma^2_\pi & \sigma^2_\pi & \sigma^2_\pi + \sigma^2_\epsilon \end{bmatrix}$$

(6)

Note Σ symmetric $\Rightarrow \Sigma$ has a compound symmetry structure

$\hookrightarrow \left\{ \begin{array}{l} \text{All variances are equal } \sum_{ii} = \sum_{jj} = \sigma_n^2 + \sigma_e^2 \\ \text{All covariances are equal } \sum_{ij} = \bar{\sigma}_n^2 \quad (i \neq j) \end{array} \right.$

\Rightarrow We can estimate covariance matrix with 2 parameters:
 σ_n^2 and σ_e^2

\hookrightarrow ANOVA table

Source of variability	df	SS	MS	F
Subjects	N-1	(1)	$SS_S / N-1$	MS_S / MS_E
Time	n-1	(2)	$SS_T / n-1$	MS_T / MS_E
(Residual)	$(N-1) \cdot (n-1)$	(3)	$SS_E / (N-1)(n-1)$	
Total	$N \cdot n - 1$	$\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y})^2$		

$$(1) = \boxed{SS_{\text{subj}} = n \cdot \sum_{i=1}^N (\bar{Y}_i - \bar{\bar{Y}})^2} \quad \begin{matrix} \bar{\bar{Y}} \rightarrow \text{grand mean} \\ \langle Y_i \rangle \text{ across } n \text{ observations.} \end{matrix}$$

$\forall i \in I$ then have $\bar{Y}_{\cdot j} = \langle Y_j \rangle$ mean of obs j

$$(2) \quad \boxed{SS_{\text{Time}} = N \sum_{j=1}^n (\bar{Y}_{\cdot j} - \bar{\bar{Y}})^2}$$

$$(3) \quad \boxed{SS_{\text{Error}} = \sum_{i=1}^N \sum_{j=1}^n [y_{ij} - \bar{Y}_i - \bar{Y}_{\cdot j} + \bar{\bar{Y}}]^2}$$

$$(4) \quad \boxed{SS_T = \sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y})^2}$$

(7)

Conditions

① Representative samples

② Independent subjects

③ Normal dist of residuals

④ SPHERICITY → equal variances for all differences

Test using Mauchly's test.

Now Estimates

$$\hat{\sigma}_\pi^2 = \frac{MS_s - MS_\varepsilon}{n}$$

(*) look at R code!

$$\hat{\sigma}_\varepsilon^2 = MS_\varepsilon$$

$$= \text{ICC} = \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\pi^2 + \hat{\sigma}_\varepsilon^2}$$

ANOVA ← aov (Calvin ~ Month + Error (Student/Month))

↳ p-value tells us whether the μ changes over time

→ use library (nlme) → aov (nlme())

"non-linear mixed effect model"

(8)

Feb 13, 2019 \rightarrow test time \rightarrow same measurement through time

↳ Repeated measures ANOVA is a special type of ^a mixed model
 & mixed model has both FIXED & RANDOM terms.

$$Y_{ij} = \mu + \pi_i + \tau_j + e_{ij} \quad i: \text{subject} \\ \begin{matrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \text{fixed} & \text{fixed} & \text{fixed} & \text{random} \end{matrix} \quad j: \text{response}$$

$\rightarrow \pi_i \sim N(0, \sigma^2_\pi) \quad \left. \begin{array}{l} \text{induces compound symmetry} \\ \text{structure on the covariance} \end{array} \right\}$

$$\rightarrow e_{ij} \sim N(0, \sigma^2_e) \quad \text{matrix of } Y_{ij} \\ \Sigma_1 = \begin{pmatrix} \sigma^2_\pi + \sigma^2_e & \sigma^2_\pi & \dots & \sigma^2_\pi \\ \sigma^2_\pi & \ddots & & \sigma^2_\pi \\ \vdots & & \ddots & \sigma^2_\pi \\ \sigma^2_\pi & & & \sigma^2_\pi + \sigma^2_e \end{pmatrix} \quad \begin{matrix} \uparrow \\ n \times n \text{ matrix} \\ n \text{ measures...} \end{matrix}$$

Sphericity \rightarrow assumption that $\text{Var}(Y_{ij} - Y_{ij'})$ constant $\forall i, j$

↳ most common test is Mauchly's Test.

Mauchly's Test \rightarrow very sensitive to sample size

\rightarrow likely to reject when sample is large

\rightarrow likely to not reject for small sample

\rightarrow normal dist of residuals needed

H_0 : sphericity is satisfied

H_a : sphericity is not satisfied

Note input must be

a matrix, not
dataframe

"Contrast" \rightarrow "Linear combination of means"

(x) $\bar{Y}_1 - \bar{Y}_2 \rightarrow$ compare groups 1 + 2

$(\frac{1}{2}\bar{Y}_1 + \frac{1}{2}\bar{Y}_2) - \bar{Y}_3 \rightarrow$ compare (1+2) to 3.

Def A contrast is defined as,

$$L = \tilde{C} \bar{\mu}$$

Sum of coeffs in any contrast must be 0.

contract coeff matrix

$$\text{Ex } L_j = C_j^T \bar{\mu} = \sum_{j=1}^n C_{jj} \bar{Y}_j$$

let's look at change relative to baseline if $n=4$

$$\tilde{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \quad \begin{matrix} 1 \text{ vs } 2 \\ 1 \text{ vs } 3 \\ 1 \text{ vs } 4 \end{matrix} \quad 3 \text{ contrasts} = \underline{\text{columns}}$$

$$\tilde{\mu} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \\ \bar{Y}_4 \end{pmatrix} \quad \text{time points } t=1, 2, 3, 4$$

Note: 2 contrasts are orthogonal if their dot product is 0.

$$L_j = C_j^T \bar{\mu} \quad \text{eg } \vec{e}_1 \cdot \vec{e}_2 = 0$$

$$C = [C_1 \ C_2 \ \dots \ C_n]$$

- If 2 contrasts are orthogonal, then they are independent.

Ex Contrast 1 vs 2. Dot product is $1 = 1+0+0+0$

$$\text{Ex } \tilde{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \bar{a} \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{pmatrix} \rightarrow 1 \text{ vs } 2$$

$$\rightarrow 2 \text{ vs } 3$$

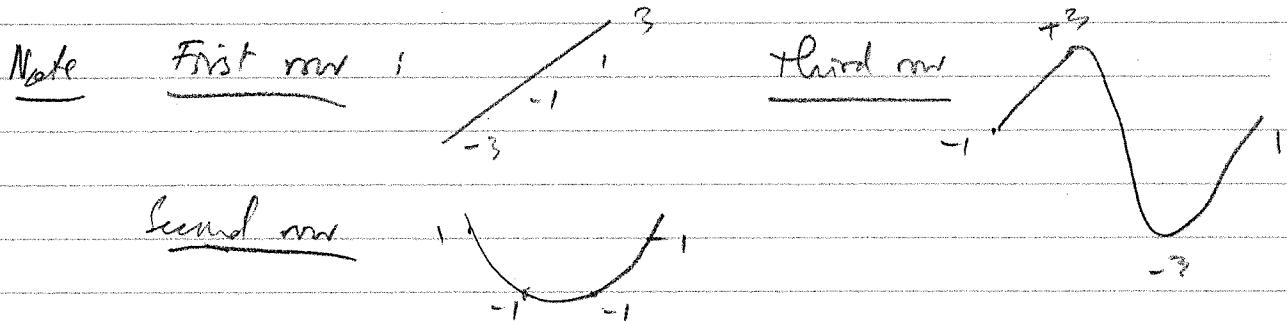
$$\rightarrow 3 \text{ vs } 4$$

Trend Analysis

$n=4$

$$\tilde{C} = \begin{pmatrix} -3/\sqrt{20} & -1/\sqrt{20} & 1/\sqrt{20} & 3/\sqrt{20} \\ +1/\sqrt{4} & -1/\sqrt{4} & -1/\sqrt{4} & 1/\sqrt{4} \\ -1/\sqrt{20} & 3/\sqrt{20} & -3/\sqrt{20} & 1/\sqrt{20} \end{pmatrix} \rightarrow \begin{array}{l} n=4 \\ \text{linear trend} \\ \text{quadratic trend} \\ \text{cubic trend} \end{array}$$

Each row \rightarrow 1 contrast



So \rightarrow denominator for trend analysis $\Rightarrow \sqrt{\sum c_{ij}^2}$ j fixed

Note "Contrasts" = "rows"

Rows are orthogonal = independent.

We want to test $H_0: L_j = L_{j0}$, $H_a: L_j \neq L_{j0}$

Test Statistic

$$t = \frac{\hat{L}_j - L_{j0}}{\sqrt{MSE \left(\sum_{j=1}^n c_{jj}^2 / N \right)}} \quad df = (N-1)(n-1)$$

But we still have multiple comparisons issue

Summary

$$L_j = c_j^T Y_i \rightarrow \text{formally a list of differences between some combination of means}$$

Feb 15, 2019

Multisample Repeated Measures ANOVA

Model

$$Y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + \epsilon_{hij}$$

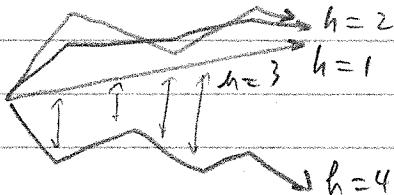
$h = 1, 2, \dots, 5 \rightarrow$ treatment group interaction term

$i = 1, 2, \dots, N \rightarrow$ subjects

$j = 1, 2, \dots, n \rightarrow$ time points.

or among

Note $\gamma\tau \rightarrow$ tests the difference in slopes between groups.



If the $\gamma\tau$ term is significant, we can't summarize the treatment effect with one value. We also can't summarize the time effect with a single value either.

If the interaction term is significant, we can't simplify model.

■ $\mu \rightarrow$ grand mean

$\gamma_h \rightarrow$ group effect

$\tau_j \rightarrow$ time effect

$(\gamma\tau)_{hj} \rightarrow$ interaction of time j \times treatment/group h

$\pi_{i(h)} \rightarrow$ individual different component for subject i in group h

$\epsilon_{hij} \rightarrow$ random errors

$$\sigma_\pi^2, \sigma_\epsilon^2$$

we have said that $\pi_{ij(h)} \sim N(0, \sigma^2_\pi)$

$$\epsilon_{hij} \sim N(0, \sigma^2_\epsilon)$$

average out to be zero

$$\left\{ \begin{array}{l} \sum_h Y_h = 0 \\ \sum_j Y_{..j} = 0 \end{array} \right. \quad \sum_h \sum_j (\bar{Y}_h)_{..j} = 0$$

we need the design to be balanced, i.e. everyone is measured at the same time. Anyone with missing later is excluded (not good)

It's not required that the sample sizes in groups are equal.

+ comes from inclusion-

ANOVA Table

of subj in group h

Source of variability	df	SS	MS	F
γ Group/Treatment	s-1	$n \sum_{h=1}^s N_h (\bar{Y}_h - \bar{\bar{Y}})^2$	$SSG/s-1$	MSG/MSB
τ Time	n-1	$N \sum_{j=1}^n (\bar{Y}_{..j} - \bar{\bar{Y}})^2$	$SST/n-1$	
$\gamma\tau$ Group x Time	(s-1)(n-1)	$n \sum_{h=1}^s \sum_{j=1}^n N_h (\bar{Y}_{h..j} - \bar{\bar{Y}}_{..j})^2$	$SSGT/(s-1)(n-1)$	$MSGT/MSB$
Π Subject	N-s	$n \sum_{h=1}^s \sum_i (\bar{Y}_{i..h} - \bar{Y}_h)^2$	$SSS/N-s$	$S_{(a)}$
ϵ Error	(N-s)(n-1)	$\sum_{h=1}^s \sum_{j=1}^n \sum_i (Y_{hij} - \bar{Y}_{h..j} - \bar{Y}_{i..h} + \bar{Y}_{..})^2$	$SSE/(N-s)(n-1)$	

subjects are confounded

by groups: subject \rightarrow Rnd $\xrightarrow{G1 \rightarrow T1}$ compare subject ~ Groups

Total

N_{n-1}

$$\sum_{h=1}^s \sum_{j=1}^n (Y_{hij} - \bar{\bar{Y}})^2$$

Notes $\bar{Y}_{h..}$ = averaging everybody @ firm j , in group h
 $\bar{Y}_{hi..}$ = averaging one's response over time in group h

$\bar{Y}_{...}$ = grand mean \rightarrow across groups, subjects \times times

$$\bar{Y} = \frac{\sum \text{all response}}{N_n}$$

$\bar{Y}_{h..}$ = mean across all subj-time points in group h .

$\bar{Y}_{hi..}$ = mean across time pts for subj i in group h .

$$\bar{Y}_{...}$$

\rightarrow interaction \rightarrow

Tests \rightarrow the primary interest is Group \times Time.

$H_0: (\bar{Y})_{hj} = 0$ | sphericity conditions
Conditions { Normal dist. within group}

$H_a: (\bar{Y})_{hj} \neq 0$ | sphericity $V(Y_{ij} - Y_{ij'})$ const

Test statistic $F = \frac{MS_{\text{GT}}}{MS_{\text{Error}}} \rightarrow$ error term

If the sphericity condition is violated, we can use MANOVA

MANOVA \rightarrow multivariate ANOVA

G

Useful commands

\hookrightarrow dplyr ... group_by() \rightarrow wrt at $\bar{Y}_{..}$,

ggplot --

Assumed
fixed

(14)

Again if (τ_i) srs, then other terms have no main

Feb 18, 2019 Recall multi-sample repeated measure ANOVA

$$Y_{hij} = \mu + \gamma_h + \tau_j + (\gamma^*)_{hj} + \epsilon_{hij} + \tau_i$$

(Time effect)

(Group Effect)

between subject effects

$$\begin{array}{l} \text{Test } H_0: \tau_1 = \tau_2 = \dots = \tau_n = 0 \\ \text{Test statistic } F = \frac{MS_T}{MS_E} \end{array} \quad \left. \begin{array}{l} \rightarrow \text{within-subject effect} \\ \text{time effect?} \end{array} \right\}$$

$$\left. \begin{array}{l} \text{Test statistics } F = \frac{MSG}{MSS(a)} \\ \rightarrow \text{MS of subject} \end{array} \right\}$$

R line (response ~ group + time + group*time)

model

data = df,

random = ~1/ Subject

cor ComSymm (form ~ Month | Subject),
method = "REML")

Time Subject

rand of x
induces compound symmetry structure of cov. matrix

$\{ \tau_{hi} \sim N(0, \sigma_{\tau}^2) \}$ → σ_{τ}^2
assumes subject and
form is rand

(constant) σ_{ϵ}^2

So that $Cov(Y_{ij}, Y_{ij'}) \neq 0$

σ_{τ}^2 while $Var(Y_{ij}, Y_{ij'}) = \sigma_{\epsilon}^2 + \sigma_{\tau}^2$

Compound Symmetry

$$\left\{ \begin{array}{l} Cov(Y_{ij}, Y_{ij'}) = \sigma_{\tau}^2 \\ Var(Y_{ij}, Y_{ij'}) = \sigma_{\epsilon}^2 \end{array} \right. \rightarrow$$

$$Var(Y_{hij}, Y_{hij'}) = \sigma_{\epsilon}^2 + \sigma_{\tau}^2$$

Q Without random subject $\rightarrow \Sigma = \begin{pmatrix} \sigma^2_{\epsilon} & 0 \\ 0 & \sigma^2_{\epsilon} \end{pmatrix}$ wrong

With random subject

$$\hookrightarrow \Sigma = \begin{pmatrix} \sigma^2_{\alpha} + \sigma^2_{\epsilon} & \sigma^2_{\alpha} \\ \sigma^2_{\alpha} & \sigma^2_{\alpha} + \sigma^2_{\epsilon} \end{pmatrix}$$

PQ Once R runs \rightarrow Anova() \rightarrow anova table.

If δT^* significant \Rightarrow can't care abt γ_2 γ
can check residuals, ...

6

Q Does intra-class correlation matter? How to check?

\hookrightarrow We might want to test $Y_{hij} = \mu + \gamma_h + \tau_j + \pi_i + (\delta T)_{hij} + \epsilon_{hij}$

H_a : test $Y_{hij} = \dots + \pi_i + \dots \rightarrow$ with subject fix

\hat{H}_0 : no, $Y_{hij} = \dots + 0 + \dots \rightarrow$ no subject fix

This is \equiv testing $\sigma^2_{\pi} = 0$

If π_i not significant \rightarrow gets absorbed into μ

\hookrightarrow use gls() model without the random subject term

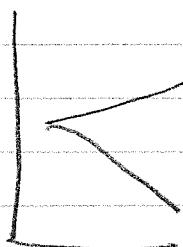
\rightarrow anova(model 1, model 2)

Compare models ...

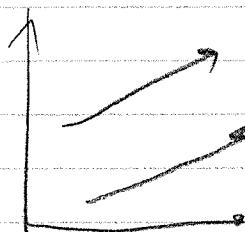
If significant, then H_a , not H_0 , \rightarrow keep π_i in the model

Post hoc comparison of means -- $cd()$ → multiple comparisons
 → pairwise comparison adjustment

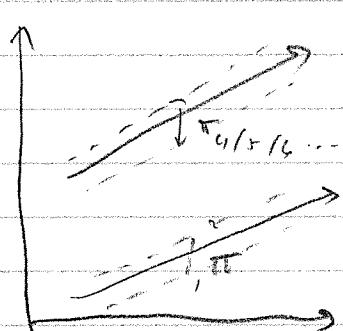
mt



→ δT significant



δT not significant, even though there is
 T effect = δ effect



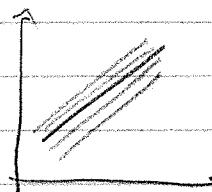
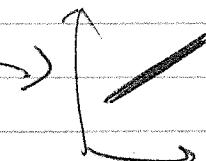
But $\delta_{\bar{x}}^2 = 0 \quad \bar{T} \sim N(0, \sigma^2)$

If $\delta_{\bar{x}}^2 = 0$, then individual
 effects don't vary.



So $\delta_{\bar{x}}^2 = 0 \quad \delta_{\bar{x}}^2 \neq 0$

if they
 test at
 @ the same
 level



If $\delta_{\bar{x}}^2 = 0 \rightarrow$ good \rightarrow more degrees of freedom \rightarrow more power

↳ more df better

Multivariate ANOVA - MANOVA

Feb 19
2019

↳ general class of models for correlated data.

Correlated data? → several responses on same individual at a single time point

$$\text{Ex } \vec{Y} = \begin{pmatrix} \text{SBP} \\ \text{DBP} \\ \text{Rate 1} \\ \text{Rate 2} \end{pmatrix}$$

In second, this is a multivariate response.

For us... $\vec{Y} = \begin{pmatrix} \text{response t}_1 \\ \text{response t}_2 \\ \vdots \\ \text{response t}_n \end{pmatrix} \rightarrow \text{"one response" are commensurate}$
 meaning the same thing on the same measurement scale.
 ↳ Expect strong correlated

MANOVA relates a vector of responses to a matrix of predictors.

MANOVA for longitudinal data is a special case of a procedure called "profile analysis"

Ex Consider a single case of 2 treatments in which subjects are measured at n times each (balanced - same-treat)

↳ 3 questions that we might want to ask:

- 1) Are trends in mean response over time the same in both groups?
- 2) Averaged over the 2 groups, is the overall trend in the mean response over time flat?
- 3) Are the overall mean responses averaged over time points the same for the 2 groups?

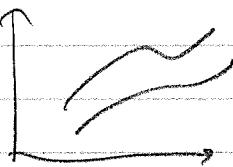
profile analysis

① → group x time interaction (primary interest)

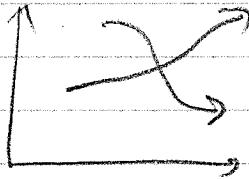
② → time effect ↗?

③ → group effect ?: ?

① Are the groups parallel?



Yes

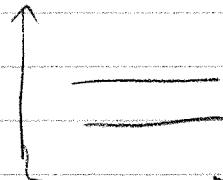


No

② Are the groups flat?

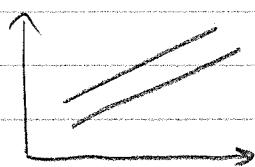


No

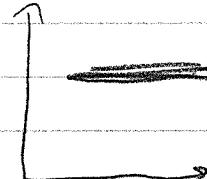


Yes

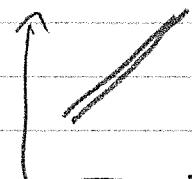
③ Do they overlap?



No



Yes.



Yes

What MANOVA does is constructing a new set of variables to address the 3 questions.

$$\# \text{ new vars} = \# \text{ time points}$$

PSS

Consider another simple example in which we are comparing a treatment to placebo. The response is measured at $n = 3$ times (balanced).

MANOVA needs to construct 3 new variables that we will call $V_{i1}, V_{i2}, V_{i3} \rightarrow i = \text{"subjects"} = 1, 2, \dots, N$.

Let

$$V_{i1} = Y_{i1} + Y_{i2} + Y_{i3} \rightarrow \text{running response over time}$$

$$V_{i2} = Y_{i2} - Y_{i1}$$

$$V_{i3} = Y_{i3} - Y_{i1}$$

Note

$$\begin{pmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = T \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix}$$

$\rightarrow V_{i1} \rightarrow$ tells us abt mean response from time nothing about time trends

$\rightarrow V_{i2} > V_{i3} \rightarrow$ tells us something abt within-subject time trends.

T^{-1} transformation matrix \rightarrow original response \rightarrow new vars.

Note \rightarrow First row of T is always $(1, 1, 1, -1)$

The subsequent rows can be different.

\blacksquare The first row of T addresses the 3rd question about group effect

\blacksquare The remaining rows address change over time. (time effects). There

are many ways to construct the remaining $n-1$ mvs.

Ex If we wanted to look at contrast to test linear & quadratic contrast

$$\hookrightarrow \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

☒ The multivariate statistics for the time trend are invariant to how we choose to look at change over time & how we characterize it.

☒ So, repeated measures MANOVA first takes the $n-1$ derived variables for time trends & analyzes them via the MANOVA process.

q → ① \hookrightarrow In our example, finding no group effect in V_{12} or V_{13} tells us there is no group \times time interaction.

q → ② \rightarrow Whether or not there is a linear trend if the mean of both V_{12} & V_{13} are 0.

q → ③ \rightarrow 3rd q can be addressed by looking at V_{11} to see if means differ across groups.

☒ Repeated measures ANOVA requires

- Balanced design (all measured at same time, not necessarily equally spaced)
- No missing data (quite restrictive)

☒ In the book \rightarrow
$$Y_i = \underline{\mu} + \varepsilon_i$$
 \rightarrow error vector

$\underline{\mu} = \mu + \tau$ \rightarrow $n \times 1$ vect of means for time points
 \rightarrow $n \times 1$ vect of time effects.

(21)

$$\Sigma_i = \sigma_n^2 \mathbf{1}_n \mathbf{1}_n' + \sigma_e^2 I_n = \begin{pmatrix} \sigma_n^2 + \sigma_e^2 & & \\ & \ddots & \\ & & \sigma_n^2 + \sigma_e^2 \end{pmatrix}$$

(covariance matrix) \rightarrow Same compound symmetry.

→ Under (MANOVA in book) transformation matrix

$$\underline{P} \underline{Y}_i = \underline{P}\mu + \underline{P}\epsilon_i$$

\underline{P} could be $\underline{P} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}$

- ↳ overall mean
- ↳ linear
- ↳ quad

→ In R → poly() generates a matrix of polynomial contrasts...

Feb 22
2019

→ MANOVA example

$$\underline{P} = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \quad \begin{matrix} \leftarrow \text{linear trend} \\ \leftarrow \text{quadratic trend} \end{matrix}$$

$$\underline{P}\underline{Y} = \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} & y_{i4} \end{pmatrix} \quad \begin{matrix} \textcircled{1} \text{Are trajectories parallel} \\ \text{for each group} \\ (\text{GxT}) \\ \textcircled{2} \text{Are trajectories flat? (T)} \end{matrix}$$

So $\theta_1 \rightarrow$ $\Omega_2 \rightarrow$ can be formulated differently? $\theta_3 \rightarrow$ group effect are the means the same

are trajectories quadratic?

Note Polynomial contrast basis functions are standard.
(get from R)

Feb 25, 2019

first basis matrix (time 1, time 2, time 3) or $\text{Group} (+1)$

↳ advantage → if you switch groups → oops

if

RANDOM INTERCEPTS MODEL

Special case of a "mixed effects" model. A mixed-effects model has

- ① Fixed effects: age, sex
- ② Random effects: subject, time, etc (known dist)

Names

{ mixed-effects models ≡ hierarchical models
 ≡ Random effects - model
 ≡ Variance-component models
 ≡ Multi-level models
 ≡ Empirical Bayes

Advantage → random effects induce a structure on the within-subject correlation matrix

⇒ Simplest mixed effects model for longitudinal data has a random subject effect for the intercept only.

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_i + \epsilon_{ij}$$

$i \rightarrow$ subject $1 \rightarrow N$ different number of
 $j \rightarrow$ time point $1 \rightarrow n_i$ & time points for each subject.

$n_i \rightarrow$ allows us to have unbalanced designs with true points that are not equally spaced

advantage

$\nu_{oi} \sim N(0, \sigma_v^2)$

$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Note ε_{ij} are conditionally independent, on the random effect.

(each subject has their own random effect)

↳ { 2-level modeling strategy

- subjects → Level 1: $Y_{ij} = b_{0i} + b_{1i} \times t_{ij} + \varepsilon_{ij}$ (within subject)

) Random effects → Level 2: $b_{0i} = \beta_0 + \nu_{oi}$ (between subjects)
 $b_{1i} = R_i \sim \text{constant}$

Variance of Y_{ij} :

$$\text{Var}(Y_{ij}) = \text{Var}(\nu_{oi}) + \text{Var}(\varepsilon_{ij})$$

↳ $\text{Var}(Y_{ij}) = \sigma_v^2 + \sigma_\varepsilon^2$

independent

And $\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_v^2$

$\text{Cov}(Y_{ij}, Y_{ij'}) = 0$

Subjects are independent

Get

→ Compound Symmetry (requires balance)

$$\text{ICC} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2}$$

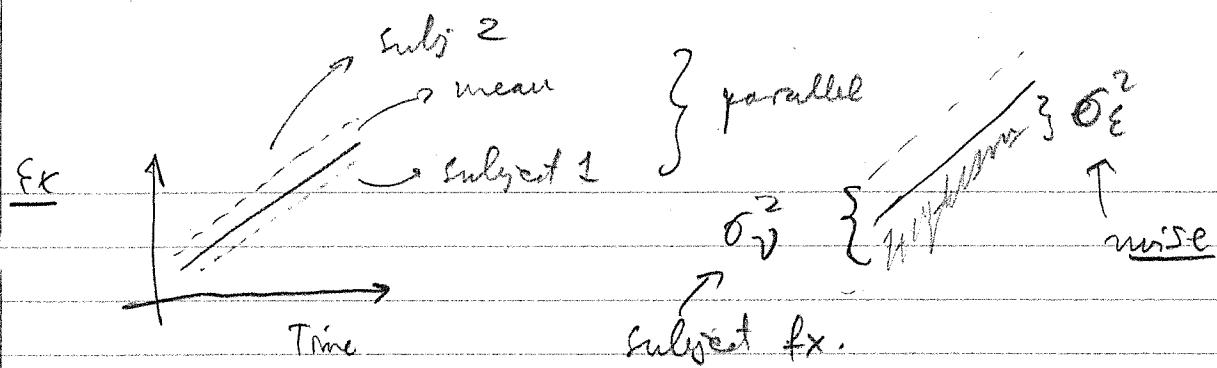
↳ intra-class corr

↳ % of var due to between subjects

ideally, ICC large.

→ model good (var small)

↳ observed differences in subject



Hypothesis Testing for Model Parameters

(1) Parameter estimates are usually tested via Wald tests.

$$z = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \stackrel{H_0}{\sim} N(0, 1)$$

Caveat: if $SE(\hat{\theta})$ is close to 0, these tests are not reliable.

(2) Likelihood ratio tests (LRT) (number latter)

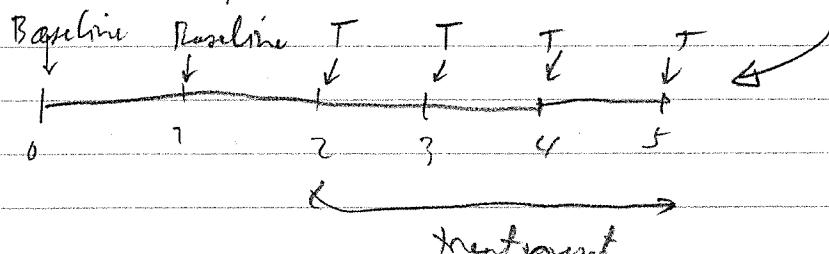
particularly good at testing nested model.

(3) Can also be used to test covariance structure

divide p-value in half to avoid Type 2.

Ex → Dataset in Book (1977)

Took a group of $n=66$ depressed patients and placed them on desipramine for 4 weeks. 2 Baseline measures ~~at~~



Response: Score on Beck Depression Inventory.

Note MANOVA doesn't handle missing data

Compound Symmetry \rightarrow subj corr constant.

Feb 27, 2011

Result Random Intercept Model: $\sim N(0, \sigma_e^2)$

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + V_{oi} + \epsilon_{ij}$$

V_{oi} subject effect $\sim N(0, \sigma_V^2)$

Note

$V_{oi} + \epsilon_{ij}$ are independent $\rightarrow \text{Var} = \text{Var}(V_{oi}) + \text{Var}(\epsilon_{ij})$

■ $E[Y_{ij}|V_{oi}] = \beta_0 + \beta_1 t_{ij} + V_{oi} + 0 \sim E[\epsilon_{ij}|V_{oi}]$

conditional expectation of Y_{ij} given V_{oi} just 0

$$E[Y_{ij}|V_{oi}] \neq E[Y_{ij}] = \beta_0 + \beta_1 t_{ij} + 0 + 0$$

conditional
expectation of
 $Y_{ij}|V_{oi}$

expectation averaged
over population

In fact $E[Y_{ij}] = E[E[Y_{ij}|V_{oi}]]$

Random Intercept Models can handle unbalanced
and missing data
↑ ↗ unlike MANOVA...
(some type of)

Random Intercept Models have an induced compound symmetry structure.

Result $\text{Var}(Y_{ij}) = \text{Var}[\beta_0 + \beta_1 t_{ij} + V_{oi} + \epsilon_{ij}] = \text{Var}(V_{oi} + \epsilon_{ij})$

Now $\text{Var}(Y_{ij}) = \text{Var}(V_{oi} + \epsilon_{ij}) \rightarrow$ by indepen-
-dence
 $= \text{Var}(V_{oi}) + \text{Var}(\epsilon_{ij})$

$$\boxed{\text{Var}(Y_{ij}) = \sigma_v^2 + \sigma_e^2}$$

Variance is the same for every time point

Q What $\text{Cov}(Y_{ij}, Y_{ij'}) \rightarrow$ same subj @ different time pts

$$= \text{Cov}([B_0 + B_1 t_{ij} + V_{oi} + \epsilon_{ij}], [B_0 + B_1 t_{ij'} + V_{oi} + \epsilon_{ij'}])$$

$$= \text{Cov}([V_{oi} + \epsilon_{ij}], [V_{oi} + \epsilon_{ij'}])$$

$$= \underbrace{\text{Cov}(V_{oi} + V_{oi})}_{\text{Var}(V_{oi})} + \underbrace{\text{Cov}(\epsilon_{ij}, V_{oi})}_{0} + \underbrace{\text{Cov}(\epsilon_{ij'}, V_{oi})}_{0} + \underbrace{\text{Cov}(\epsilon_{ij}, \epsilon_{ij'})}_{\text{Var}(\epsilon_{ij}, \epsilon_{ij'})}$$

$$= \text{Var}(V_{oi}) + \underbrace{\text{Var}(\epsilon_{ij}, \epsilon_{ij'})}_{0}$$

$$= \text{Var}(V_{oi}) + 0$$

$$= \text{Var}(V_{oi}) = \tilde{\sigma}_v^2$$

$\therefore \text{Cov}(Y_{ij}, Y_{ij'}) = \tilde{\sigma}_v^2 \leftarrow \text{constant}$

\Rightarrow Compound symmetry in covariance matrix. $\begin{pmatrix} \tilde{\sigma}_v^2 & \tilde{\sigma}_v^2 & \dots \\ \tilde{\sigma}_v^2 & \tilde{\sigma}_v^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$

Q Now, what if R outputs something a covariance matrix Σ that doesn't have compound symmetry?

\hookrightarrow then our model (CompSym) might not work

Side note

↳ In R, when we say "random = id" ↗

id → is a var that uniquely identifies subjects.

id

1

1

1

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

2

<p

What if we code time as

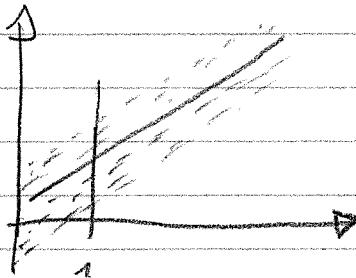
→ interpretation of the intercept depends on what we call $t=0$, - Expected response when $t=0$

↳ so the intercept ($t=0$) is before the study began
→ Extrapolating.

↳ BAD,

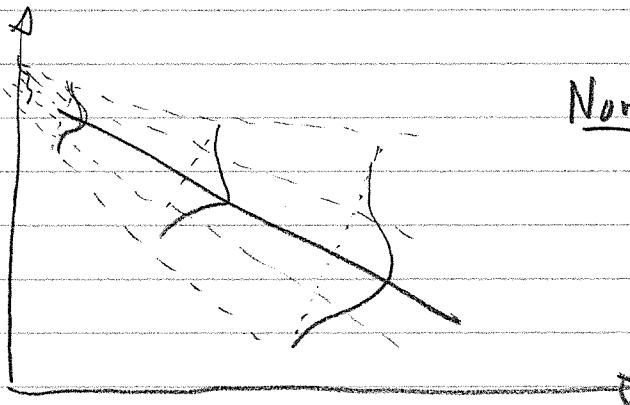
⇒ Change the subject-specific random effects.

↳



→ change estimate, but
not the variance.

⇒ What if we wanted to allow subjects to have both a random intercept and a random slope?



Now if we code time differently
the variability $t=0$ also
changes.

⇒ We always want $t=0$ to be within the time range
of data that we have collected

RANDOM INTERCEPT + SLOPE

Nov 1, 2019

last time → random intercept alone doesn't fit data well.

Reall [2-level model specification.]

note, we're
not really
fitting level
by level

$$\left\{ \begin{array}{l} \text{Level 1} \quad Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij} \\ \text{Level 2} \quad \beta_{0i} = \beta_0 + v_{0i} * \beta_1 \text{ and } \beta_{1i} = \beta_1 + v_{1i} \end{array} \right. \quad \begin{array}{l} \text{slope} \\ \curvearrowright \\ \text{intcept} \end{array}$$

just estimate random subject effect

SD

$$\begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_v \right] \quad \begin{array}{l} \text{Cov}(v_0, v_1) \\ \curvearrowright \end{array}$$

Covariance matrix $\Sigma_v = \begin{bmatrix} \sigma_{v_0}^2 & \sigma_{v_0 v_1} \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 \end{bmatrix}$

↓ [1-level formulation]

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \epsilon_{ij}$$

↳ How many parameters do we estimate?

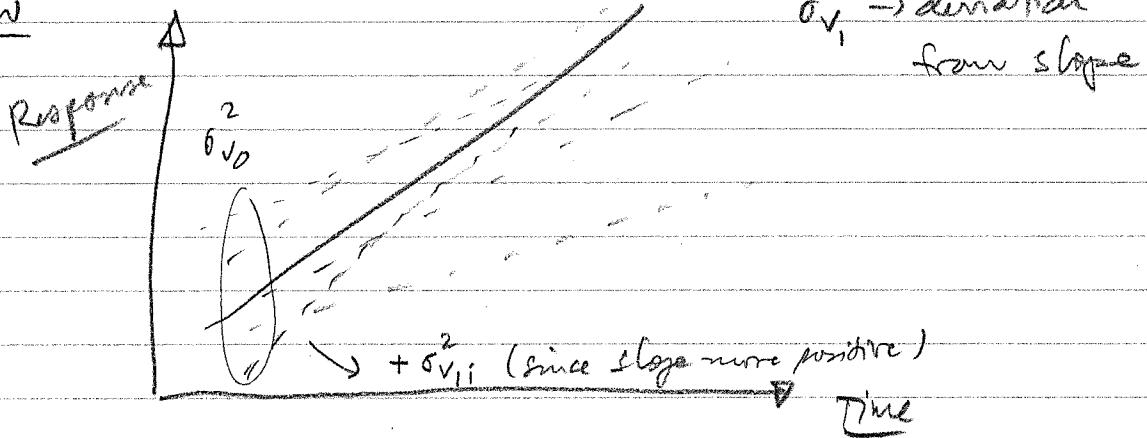
$$\left\{ \beta_0, \beta_1, \sigma_{v_0}^2, \sigma_{v_1}^2, \sigma_{\epsilon}^2, \sigma_{v_0 v_1} \right\} \quad \begin{array}{l} \text{var of intcpt} \\ \curvearrowright \\ \text{var of slope} \end{array} \quad \begin{array}{l} \text{random error} \\ \text{(var of rad error)} \end{array} \quad \begin{array}{l} \text{cov}(v_0, v_1) \\ \curvearrowright \end{array}$$

intcpt slope (subj-specific
(subj-specific slope effect) \oplus
offset at $t=0$)

But if $\sigma_{v_0 v_i}$ is positive it tells us that those with a high intercept tend to have steep slope positive

If $\sigma_{v_0 v_i}$ negative then the opposite occurs.

Now



Note We don't have compound symmetry anymore.

$$\text{Var}(V_{ij}) = \text{Var} \left[\underbrace{\beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij}}_{\text{fixed}} + \varepsilon_{ij} \right]$$

$$= \text{Var}[v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}]$$

$$\boxed{\text{Var}(V_{ij}) = \sigma_{v_0}^2 + \sigma_{v_1}^2 + t_{ij}^2 + \sigma_{\varepsilon_{ij}}^2 + 2t_{ij} \text{Cov}(v_0, v_1)}$$

$\text{Var}(Y_{ij}) \rightarrow$ time-dependent.

$$\text{Consider } \text{Cov}(Y_{ij}, Y_{ij'}) = \text{Cov}[(\beta_0 + \dots + \varepsilon_{ij}), (\beta_0 + \dots + \varepsilon_{ij'})]$$

$$= \text{Cov}(v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}, v_{0i} + v_{1i} t_{ij'} + \varepsilon_{ij'})$$

$$= \text{Var}(v_{0i}) + t_{ij} t_{ij'} \text{Cov}(v_{0i}, v_{1i}) + t_{ij} \text{Cov}(v_0, v_1) + (t_{ij} + t_{ij'}) \sigma_{v_1}^2$$

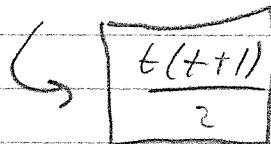
$$= \sigma_{v_0}^2 + \sigma_{v_1}^2 + \sigma_{v_0 v_1} (t_{ij} + t_{ij'})$$

$$\boxed{\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_{v_0}^2 + (t_{ij} + t_{ij'}) \sigma_{v_1}^2 + \sigma_{v_0 v_1} (t_{ij} + t_{ij'})}$$

These mixed effects models allow us to fit a flexible set of covariance structures to data

parameters = 4

of elements in Σ_0 = covariance matrix of Y_{ij}'



↳ instead of $\sim 1/\text{id}$, in R `lmer` " $\sim \text{time}/\text{id}$ "

So, is random intercept + slope better than random intercept?

↳ Random Int + slope: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \epsilon_{ij}$

nestled

Random Int: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + \epsilon_{ij}$

We compare nested models via likelihood ratio test

$$\begin{cases} H_0: \text{rnd int} \\ H_a: \text{rnd int + slope} \end{cases}$$

Test statistic $-2(\ln L_0 - \ln L_1) \sim \chi^2_{df = \# \text{ of parameters being introduced}}$

↳ df = difference between # of parameters of being introduced

Run by test statistic $= -2(\ln(11.44594) - (-109.519))$

$$= 66.95 \sim \chi^2_{df = 6 - 4 = 2}$$

Again,

H₀: estimate 4 params

H_a: estimate 6 params

Near (6, 2019)

Recall model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{oi} + v_{itij} + \varepsilon_{ij}$

Matrix formulation

$$\underline{Y}_i = \underline{X}_i \underline{\beta} + \underline{z}_i \underline{v}_i + \underline{\varepsilon}_i$$

$\begin{matrix} \text{vector of} \\ \text{response for} \\ \text{subject } i \end{matrix} \quad [\underline{n} \times 1] \quad [\underline{n} \times p] \quad [\underline{p} \times 1] \quad [\underline{n} \times r] \quad [\underline{r} \times 1] \quad [\underline{n} \times 1]$

(0) $[\underline{X}_i]$ → design matrix for fixed effects (FIXED EFFECTS)

$$(1) [\underline{X}_i][\underline{\beta}] = \beta_0 + \beta_1 t_{ij}$$

(2) $[\underline{\beta}]$ → vector of population level coeffs that relate fixed effects to response

(3) $[\underline{z}_i]$ → design matrix for RANDOM EFFECTS

(4) $[\underline{v}_i]$ → vector of Random effects coefficients

(5) $[\underline{\varepsilon}_i]$ → Random error errors vector

Note $\left. \begin{array}{l} p \rightarrow \# \text{ predictors (with intercept)} \\ r \rightarrow \# \text{ random components} \\ r \leq p \end{array} \right\}$

In Riesz example

$$[\underline{z}_i] = [\underline{X}_i]$$

$$[\underline{X}_i] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

\uparrow
 $i=1, \dots, 6$

$$[\underline{v}_i] = \begin{bmatrix} v_{o1} \\ v_{t1} \\ \vdots \\ v_{t3} \end{bmatrix}$$

$$\underline{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{i3} \end{pmatrix}$$

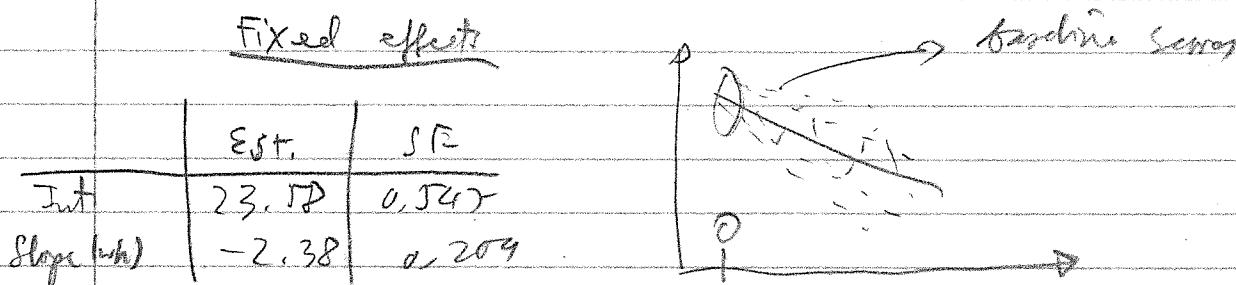
◻ Note X, Z, ε can have different numbers of rows across individuals

◻ Now, think about how type of depression might predict/explain depression scores.

◻ Note $\beta_0, \beta_1 \rightarrow$ fixed effects

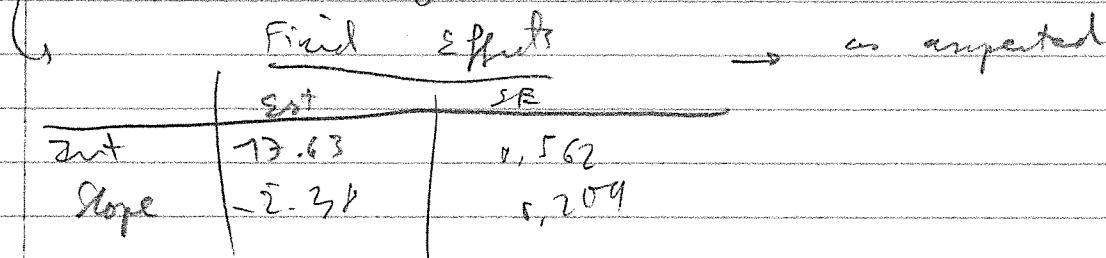
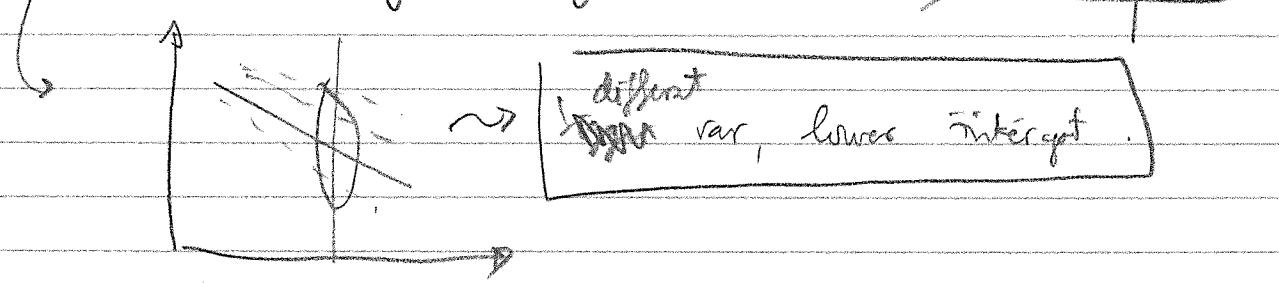
$$\gamma_{0i}, \gamma_{1i} \rightarrow \text{random effects} \sim MVN\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma_\gamma\right)$$

◻ what happens if we code time differently? (Recentered time)



◻ If we define $t=0$ at middle of study, then

$$\gamma_{0j} \rightarrow \beta_{0j} - 2.5$$



Interaction

$$+ \beta_3 D_{X_i} \times t_{ij})$$

Q Now, talk about DIAGNOSIS

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij}}_{\text{fixed}} + \underbrace{\beta_2 D_{X_i}}_{\text{(+) diag}} + v_{oi} + v_{it_{ij}} + \epsilon_{ij}$$

$$\text{fixed}, [X_i] = \begin{pmatrix} 1 & 0 & D_{X_i} \\ 1 & 1 & D_{X_i} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix} \quad (+) \text{ (diag)}$$

But D_{X_i} alone assume fixed \rightarrow add interaction

$$\rightarrow [X_i] = \begin{pmatrix} 1 & 0 & D_{X_i} & D_{X_i}(0) \\ 1 & 1 & D_{X_i} & D_{X_i}(1) \\ 1 & 2 & D_{X_i} & D_{X_i}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (+) \text{ (diag)} \quad (\text{diag} \times t)$$

$\beta_3 D_{X_i} \times t_{ij}$ \rightarrow tests whether the reactions through time
is different by diag rows.

$$b Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 D_{X_i} + \beta_3 D_{X_i} \times t_{ij} + v_{oi} + v_{it_{ij}} + \epsilon_{ij}$$

$$\underline{\text{Level 1}} \quad Y_{ij} = b_0 + b_1 t_{ij} + b_2 D_{X_i} + b_3 D_{X_i} \times t_{ij} + \epsilon_{ij}$$

$$\underline{\text{Level 2}} \quad b_{0i} = \beta_0 + v_{0i} \quad (\text{fixed + random})$$

$$b_{1i} = \beta_1 + v_{1i} \quad (\text{fixed + random})$$

$$b_{2i} = \beta_2$$

$$b_{3i} = \beta_3$$

Note D_{X_i} is constant within individual (between subj)
 t_{ij} not constant within an individual (within-subj)

Last time, we were looking at diagnosis Dx

Mar 6, 2019

Fixed effects: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 D_{xi} + \beta_3 D_{yj} + \epsilon_{ij}$

Random effects: $+ v_{oi} + v_{ii} t_{ij} + \varepsilon_{ij}$

$$\text{In matrix } Y_i = \underbrace{\mathbf{x}_i \beta}_{\substack{\text{fixed} \\ \text{rnd}}} + \underbrace{z_i \gamma}_{\text{rnd}} + \varepsilon_i$$

where

$$\mathbf{x}_i = \begin{pmatrix} 1 & 0 & D_{x1}(0) & D_{x1}(1) \\ 1 & 1 & D_{x2}(0) & D_{x2}(1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & q & D_{xq}(0) & D_{xq}(1) \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

$$z_i = \text{Random} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \quad \gamma_i = \begin{pmatrix} v_{o1} \\ v_{i1} \\ \vdots \\ v_{iq} \end{pmatrix}$$

• Compare (H_a) versus H_0 : the model without diagnosis

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{oi} + v_{ii} t_{ij} + \varepsilon_{ij}$$

Note Smaller model is

NESTED in the bigger model \rightarrow likelihood ratio test

Ruby data ★ Use tapply() for summary statistics --

random ~ week / id \rightarrow slope + intercept random
 $\sim 1 | id$ \rightarrow intercept random

Run line \rightarrow $+ \ln L_A = -1107.45$

trajectory thru time \rightarrow insignificant of $D_{xi} t_{ij}$ coef
 is significant --

Difference between models \rightarrow degrees of freedom.

$$H_{A1}: \beta_0, \beta_1, \beta_2, \beta_3, \sigma^2_{v_o}, \sigma^2_{v_i}, \sigma^2_{v_{ii}}, \sigma^2_{\varepsilon} \parallel H_{A2}: \beta_0, \beta_1, \sigma^2_{v_o}, \sigma^2_{v_i}, \sigma^2_{v_{ii}}, \sigma^2_{\varepsilon}$$

(a) $\sigma^2_{v_{ii}}$

$$z = 8 - 6$$

Ayeris, $\ln(L_0) = -11.09, 519$

Test stat
$$\left[-2 \left(\ln(L_0) - \ln(L_A) \right) \xrightarrow{H_0} \chi^2_{2df} \right]$$

$p = 0.128 \rightarrow$ fail to reject hyp $\rightarrow H_0$ is adequate

Q Singer model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{ij} + v_{j-1} t_{ij} + \epsilon_{ij}$

How many var are being used to estimate the within-subject covariance matrix?

$$\rightarrow 4: \sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \sigma_{v_{j-1}}^2, \sigma_{\epsilon}^2$$

Need to estimate

Cov Matrix

$$\hat{\Sigma}_i = \bar{z}_i \cdot \bar{z}_i^T + \sigma_{\epsilon}^2 \cdot I_{n_i}$$

Σ_i

$$\begin{bmatrix} \sigma_{\beta_0}^2 & \sigma_{v_{j-1}}^2 \\ \sigma_{v_{j-1}}^2 & \sigma_{\epsilon}^2 \end{bmatrix}$$

We know that the "real" covariance matrix within-subjects has $\frac{(6)(7)}{2} = 21$ parameters.

Q We want to know how well $\hat{\Sigma}_i$ does in estimating Σ_i .
There are 6 variances it has to estimate and 15 covariances.

$$\rightarrow \text{Before we derived} \rightarrow \left[\text{Var}(Y_{ij}) = \sigma_{\beta_0}^2 + t_{ij}^2 \sigma_{v_{j-1}}^2 + 2t_{ij} \sigma_{v_{j-1}} \right. \\ \left. + \sigma_{\epsilon}^2 \right]$$

and

$$\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_{\beta_0}^2 + \sigma_{v_{j-1}}^2 (t_{ij} \cdot t_{ij'}) + t_{ij} t_{ij'} \sigma_{\epsilon}^2$$

How to test? \rightarrow generate $\hat{\Sigma}_i$ from σ_{ϵ}^2 , \bar{z}_i
get Cov Matrix get VarCov $\rightarrow \underline{\Sigma_V}$

$$\hat{\Sigma}_i = z^T \Sigma z + \sigma_e^2 I_{n_i}$$

To get actual $\Sigma_i \rightarrow \text{cor}(\cdot)$ (In R make)

\hookrightarrow compare $\hat{\Sigma}_i$ and Σ_i by eye... Actually testing in R can get tricky...

Note: $z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \end{pmatrix}$ only true for balanced.

If time different $\rightarrow \Sigma_i = (\cdot)$ unique to individual

\rightarrow just do it again. But and again for different i .

\rightarrow so we can generalize, so by as # of measurements same

Nov 6, 2019

TIME-VARYING COVARIATES \rightarrow PREDICTOR OR EXP. VARS

- Generally speaking, some common covariates don't change thru time: sex, diagnosis
- but some do \rightarrow risk factors (blood pressure)
(treatment in crossover study)

- We can easily incorporate time-varying covariates into model \rightarrow we can place them into level 2 or we can think of putting them into the design matrix

Ex

	β_0	β_1	β_2
1	1	0	x_{11}
1	1	1	x_{12}
1	1	2	\vdots
1	1	1	\vdots
1	1	1	\vdots

intercept \rightarrow time covariate, time varying

Back to Riesberg data

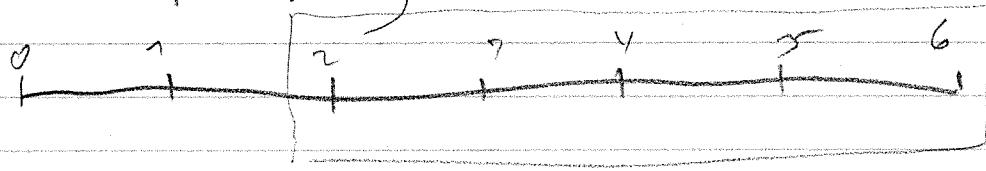
The original data contains 2 varying time covariate

- ① Blood plasma concentration of imipramine ✓ transform
- ② Plasma concentration of desipramine

Note imipramine bio-transforms into desipramine.

We may hypothesize that higher concentration of one or the other is associated with lower depression scores.

- Real study design ↗ 1st measure at all, we expect fx.



↓
drug
administered

→ data set
with plasma
concentration ob-
contains those 4
time points.

- Note that plasma concentration is highly right-skewed, the natural log was taken on both depression

- The responses are gonna be the change from baseline score.

Main Effects Model

↑
change in
depression
score

↑
ini-concentration ↓
desip-
concentration

$$Y_{ij} - Y_{i0} = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln I_{ij} + \beta_3 \ln D_{ij}$$

↑
none baseline

↑
 $\nu_{oi} + \nu_{iti} + \epsilon_{ij}$

Need to code time 2 3 4 5
 $\downarrow \downarrow \downarrow \downarrow$
 0 1 2 3 \rightarrow target meaningful intercept

How do we want to code the plasma concentration?

Reich slope $\beta_0 \rightarrow$ represents the expected change score at $t=0$ and when plasma concentrations are effectively zero.

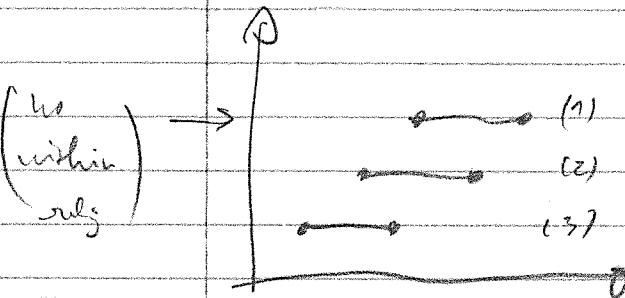
would make more sense to center $\ln(\text{plasma concentration})$

\rightarrow change β_0 . β_0 now denotes the expected change score for someone at $t=0$ with average concentration within levels.

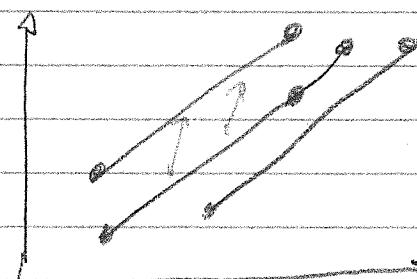
$$\ln(F_{ij}) \rightarrow \ln I_{c_{ij}} = \ln I_{ij} - \ln \bar{I}_j$$

$$\ln(D_{ij}) \rightarrow \ln D_{c_{ij}} = \ln D_{ij} - \ln \bar{D}_j$$

Note when we center a covariate, we are assuming it's between subject & within-subject are the same.



between-subj fx
no within-subj fx



within-subj fx
no between-subj fx

mean across
all subjects

$$\ln T_{ij} = \bar{\ln T}_{ij} + (\ln T_{ij} - \bar{\ln T}_{ij})$$

$$\ln D_{ij} = \bar{\ln D}_{ij} + (\ln D_{ij} - \bar{\ln D}_{ij})$$

between within

- We could separate the between & within parts in the model \rightarrow add terms

$$\begin{aligned} Y_{ij} - Y_{i0} = & \beta_0 + \beta_1 t_{ij} + \beta_2 (\ln T_{ij} - \bar{\ln T}) \\ & + \beta_3 (\ln D_{ij} - \bar{\ln D}) + \beta_4 \ln I + \beta_5 \ln D \\ & + V_{0i} + V_{1i} t_{ij} + \epsilon_{ij} \end{aligned}$$

Mar 13, 2019

Reall Time-vary Covariates

$$\text{⑦ } (Y_{ij} - Y_{i0}) = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln T_{ij} + \beta_3 \ln D_{ij} \\ + V_{0i} + V_{1i} t_{ij} + \epsilon_{ij}$$

deval
from
baseline

Now, $\ln T_{ij}$ } both have been "centered" by subtracting
 $\bar{\ln T}_{ij}$ } the overall mean of $\ln T$ and $\ln D$
 across all subjects / observations.

Between vs Within Effects of Time-varying Covariates

We can separate each chemical's effect into between and within subject components. \rightarrow deviation at time i of subject j from $\bar{\ln T}_i$

BETWEEN
SUBJECT

$$\{ \bar{\ln T}_{ij} = \bar{\ln T}_{ij} + (\ln T_{ij} - \bar{\ln T}_i) \text{ (WITHIN PIECE)}$$

\rightarrow averaged across time, for each individual (i)
 \rightarrow varies. This is the BETWEEN SUBJECT

among groups j (41)

If we include the between \rightarrow within subject prep separately,

$$(Y_{ij} - \bar{Y}_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 (\ln I_{ij} - \bar{\ln I}_{ij}) \\ + \beta_3 (\ln D_{ij} - \bar{\ln D}_{ij}) + \beta_4 \ln I_i \\ + \beta_5 \ln D_i + v_{0i} + v_{1i} t_{ij} + \epsilon_{ij}$$

small effect

$$\text{If } \beta_2 = \beta_4, \text{ then } \beta_2 (\ln I_{ij} - \bar{\ln I}_{ij}) + \beta_4 \ln I_i \\ = \beta_2 (\ln I_{ij} - \bar{\ln I}_i + \bar{\ln I}_i) \\ = \beta_2 \ln I_{ij}$$

\hookrightarrow This assumes that the between and within effect are equal.

How to group by individual & groupby (df, id) % > %

library(dplyr)

Then merge into original df \rightarrow id --- mean

1 replicated { $\frac{1}{n}$

1 replicated { $\frac{1}{n}$

\hookrightarrow create \bar{I} , then create $I_{ij} - \bar{I}_{ij}$

Now $\beta_1, \dots, \beta_5 \rightarrow$ fixed effects \rightarrow no "i" index

$v_{0i}, v_{1i} \dots \rightarrow$ random fx \rightarrow "i" index.

Compare models without splitting (w-13) and with.

Model 1 $\ln L_0 = -749$

$$\left\{ \begin{array}{l} \beta_0, \beta_1, \beta_2, \beta_3, \beta_4 \\ \sigma^2_{v_0}, \sigma^2_{v_1}, \sigma^2_{v_0v_1}, \sigma^2_{\epsilon} \end{array} \right.$$

Model 2 $\ln L_1 = -747$

$$\left\{ \begin{array}{l} \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \\ \sigma^2_{v_0}, \sigma^2_{v_1}, \sigma^2_{v_0v_1}, \sigma^2_{\epsilon} \end{array} \right.$$

Nested F-test $\rightarrow \chi^2 = -2(\ln L_0 - \ln L_1) = 3.08 \xrightarrow{p=0.08} \text{NS}$

p-value = 0.275 \rightarrow null model is adequate

\hookrightarrow separating out within-subjects is probably not something we want to do...

{Interaction model} \rightarrow between time-varying covariates and time itself (int)

$$(Y_{ij} - V_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln I_{ij} + \beta_3 \ln D_{ij} + \beta_4 T_{ij} t_{ij} \\ + \beta_5 \ln D_{ij} t_{ij} + V_{0i} + V_{1i} t_{ij} + \underbrace{e_{ij}}_{\text{random}}$$

week 2

$\boxed{\beta_0}$ (β_0) is the mean change in depression score for patients with average log chemical levels

$\boxed{\beta_1}$ (β_1) is the average weekly change in change scores for patients with average drug levels.

$\boxed{\beta_2}$ (β_2) \rightarrow change score difference for a one-unit T in ln TMC at week 2

$\boxed{\beta_3}$ (β_3) \rightarrow ... in ln DAI at week 2

$\boxed{\beta_4} = (\beta_5) \rightarrow$ indicates per-week change in drug effect on depression change scores

\hookrightarrow compare models \rightarrow interaction model \Rightarrow better

Note $df = 2$ because $\Delta \# \text{parameters} = 2$

Estimation

Just get a general feel about how we estimate β variance parameters, and the individual \rightarrow specific random effects.

Two types of estimation used

(1) Likelihood-based $\rightarrow \beta, \Sigma_v, \sigma^2_\epsilon$

(2) Empirical Bayes $\rightarrow v_i$ (random effects/slopes)

Bayesian estimation

① make a guess about distribution of a parameter.

② look at the data

prior distribution

③ Update the guess about distribution
(posterior distribution)

↳ Involves iterating, even from estimating $\beta, \Sigma_v, \sigma^2_\epsilon$ and Estimating v_i

↳ $\Sigma_{v_i} / \gamma_i \rightarrow$ covariance matrix of random effects.

Iterate until converge on a solution

★ Two potential problems \rightarrow (I) Exceed maximum of allowed iterations and/or (II) Can't get Δ in iterations to be small enough \rightarrow don't converge.

Estimation method**ML versus REML**

ML: Maximum likelihood,

REML: Restricted Maximum likelihood.

ML: produces estimates for variance parameters ($\Sigma_v, \sigma^2_\epsilon$) that are biased. How biased depends on # of independent

and # of parameters estimating.

Q Consider a standard multi regression model -

- $\hat{\sigma}_{ML}^2 = \frac{SSE}{N}$ the intercept. Type I PT
- $\hat{\sigma}_{REML}^2 = \frac{SSE}{N-(p+1)}$ $\Rightarrow p$ # of predictors

- (ML) → gives estimates of the variance that too small -
 → Positive Increase in type I error rate
 → Confidence intervals aren't as "confident" as they should be.

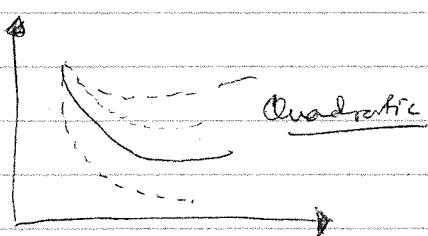
Q Why don't we use REML all the time?

- Because we can't compare nested models using REML
 (Need ML for this --)

Q What we do in practice → use REML for all model building
 and use REML to reestimate final model

Chap 5: CUVILINEAR TREND

Ex,

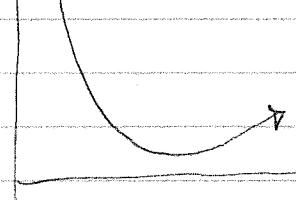


Fixed effect Linear + Quadratic -

Random effect Linear + Quadratic -

High model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2$

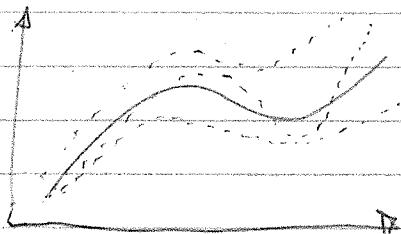
change in direction of trend when
 t_{ij}^{st} derivative is zero.



$\frac{d}{dt}$ model = $\beta_1 + 2\beta_2 t_{ij} = 0$

$t_{ij} = -\beta_1 / 2\beta_2$

Can also have cubic trend



$$\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + v_{oi}$$

Can also estimate random effect, which
can be lin, quad, cubic

Mar 18 2019

Think back to Rieley data. Now, we would like to see if there's quadratic trend in both the fixed and random effects.

Consider model $\rightarrow Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + v_{oi} + v_{ij} t_{ij} + v_{2ij} t_{ij}^2 + \varepsilon_{ij}$

How many parameters do we need to estimate?

↳ $\beta_0, \beta_1, \boxed{\beta_2}, \varepsilon \rightarrow$ fixed effects
 $\sigma^2_\varepsilon, \sigma^2_{v_0}, \sigma^2_{v_1}, \boxed{\sigma^2_{v_2}}, \sigma^2_{v_1 v_2}, \boxed{\sigma_{v_1 v_2}}, \sigma^2_{v_2 v_3} \leftarrow$ rand effs } (10)

$$\Sigma_V = \begin{bmatrix} \sigma^2_{v_0} & & \\ \sigma_{v_0 v_1} & \sigma^2_{v_1} & \\ \sigma_{v_0 v_2} & \sigma_{v_1 v_2} & \sigma^2_{v_2} \end{bmatrix}$$

If we consider only linear model \rightarrow have [6] parameters

↳ χ^2 expansion from df = 4 = 10 - 6.

[P] $\hat{Y} \sim \text{wrech} + I(\text{wrech}^2)$

• The quadratic term for time is just like time interacting with itself $\rightarrow t_{ij} \times t_{ij}$.

↳ If we do have a quadratic component, we need to have its linear component ~~too~~ as well.

[R] \rightarrow merge (df 1, df 2, by = 0) $\xrightarrow{\text{non-by-row}} \text{merge}$

COVARIANCE PATTERN MODEL

Mar 18, 2019

\hookrightarrow Extension of multiple regression

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Problem σ^2 is fixed

$$\rightarrow \Sigma_\epsilon = \sigma^2 \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = \sigma^2 I$$

which assumes independence across individuals and within individuals.

Covariance Pattern Models don't separate variability into within-subj and between-subj process. The mixed models did do this (random and fixed)

\hookrightarrow In the mixed model setting $\rightarrow \underbrace{\sigma_{\text{res}}^2}_{\text{within-subj}} \text{ and } \underbrace{\sigma_{\text{int}}^2}_{\text{between}} + \sigma^2 \epsilon$

Covariance Pattern Models \rightarrow 2 steps

① Modeling covariance

② Modeling the mean response

Potential Problem \rightarrow we might not have enough df to estimate all parameters

□ Think about the most general covariance model pattern.

$$\boxed{Y_i = X_i \beta + \varepsilon_i} \quad \varepsilon_i \sim N[0, \Sigma_i]$$

if \underline{Y}_i is $n_i \times 1$, $i = 1, 2, \dots, N$ obs:

then $\underline{X}_i = n_i \times p$, $\beta = p \times 1$, $\varepsilon_i = n_i \times 1$

Ex
Before $\underline{X}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 5 \end{bmatrix} \rightarrow$ assumes a linear relationship between time and response points and response indicators

int time \nearrow indicators

Now,

$$\underline{X}_i = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \text{design matrix}$$

$$\sum_{\text{int}}^T I(t_{ij}=0) \quad I(t_{ij}=1) \dots \quad I(t_{ij}=4)$$

\Rightarrow imposes ~~that~~ no structure at all on how times of measurement relate to \underline{Y} .

• Before > Now \rightarrow less params to be estimated

• Now > Before \rightarrow doesn't impose time structure

\Rightarrow Can save df by modeling the mean in a more parsimonious way. (provided our model is correct)

Q What about covariance of ϵ_i ?

If we have p measurements taken on each individual, how many variances / covariance are there in Σ_i ?

$$q = p + \binom{p}{2} = \frac{p(p+1)}{2}$$

\uparrow \uparrow
variance covariance

This is what MANOVA does...

when $p=2 \Rightarrow q=3$

$p=4 \Rightarrow q=10$

$p=10 \Rightarrow q=55$

Q If we place no structure on $\Sigma_i =$

$$\begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_2^2 & \ddots & & & \\ \vdots & & \ddots & & \\ \sigma_p^2 & & & \ddots & \end{pmatrix}$$

Compound Symmetry (?)

$\hookrightarrow q=2$

$$\Sigma_i = \begin{pmatrix} \sigma_1^2 + \sigma^2 & & & & & & & \sigma_1^2 \\ \sigma_1^2 & \ddots & & & & & & \\ \vdots & & \ddots & & & & & \\ \sigma_1^2 & & & \ddots & & & & \\ & & & & \ddots & & & \\ & & & & & \ddots & & \\ & & & & & & \ddots & \sigma_1^2 + \sigma^2 \end{pmatrix}$$

\hookrightarrow same df,

but price \Rightarrow all variances

are equal to each, and all covariance (correlation) are equal

This is what repeated measures ANOVA assumes

Note rep. measure ANOVA & MANOVA are at different extremes in terms of placing structure on Σ_i .

and even random-intercept and LMM -

First-order Autoregressive

Covariance between two time points j and j' is given by

$$\sigma_{jj'} = \sigma^2 p^{|j-j'|}$$

correlation

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & p & & p^{n-1} \\ p & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \rightarrow \text{Assumes variances are all equal}$$

$$q = 2 \quad (\text{6 and } p)$$

This implies a decay in the correlation as the time separation gets larger.

Toeplitz structure

$$\Sigma_i = \begin{pmatrix} \theta_0 & & & \theta_n \\ \theta_1 & \theta_0 & & \\ \vdots & \vdots & \ddots & \\ \theta_n & \cdots & \theta_1 & \theta_0 \end{pmatrix}$$

$$q = \frac{\log(\theta_0)}{\log(p)} n$$

measurements

Covariance between 2 time points:

$$\sigma_{jj'} = \theta_k, \quad k = |j'-j| + 1$$

If we assume $\theta_{ij} = 0$ once we reach a certain threshold then

Exponential structure

$$p^{-k} = p^{|j'-j|}$$

$$p^{-k} = p^{-\frac{|j'-j|}{q}}$$

time pts don't have to be equally spaced

$$\Sigma_i = \sigma^2 \begin{pmatrix} \theta_0 & & & & \\ p^{-1} & 1 & & & \\ p^{-2} & & 1 & & \\ p^{-3} & & & \ddots & \\ p^{-n} & & & & 1 \end{pmatrix}$$

q=2

Review Can bring 1 sided sheet of notes

• Fundamentals of longitudinal data (within subject, datasets)

• Visualize longitudinal data ("spaghetti plot")

• Repeated Measures ANOVA (time effect + compound symmetry + no missing data + Pseudo)

• MANOVA → entire set of longitudinal observation at once, transformation matrices, questions how can be answered with profile analysis, no structure or with-subj. M_{cov}

• Mixed models → fixed and random effects, estimates (β_i , random components σ^2_{ϵ} , ...)

→ much more flexible in terms of structure of Σ_i
(ex: random int, and int + slope of higher order etc.)

Date . Estimation: (1) 2-step iteration: between estimates of β_i , Σ_V and V_i (fixed to known basis)

$\epsilon_{ij} \perp V_{0i}$ independent | (2) ML vs REML. ML gives biased est of variance terms, especially σ^2_{ϵ}

$\epsilon_{ij}, \epsilon_{ij}'$ independent | REML is unbiased. Catch: We need ML to conduct likelihood ratio test.

↳ Remember $V(Y_{ij})$, $\text{cov}(Y_{ij}, Y_{ij}')$ derivation...

$$\left\{ \begin{array}{l} E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} \\ E(Y_{ij}|V_{0i}) = \beta_0 + \beta_1 t_{ij} + V_{0i} \end{array} \right| \begin{array}{l} \text{get } V_{0i} \text{ (} I \rightarrow \text{cov matrix)} \\ \text{Var}[\text{corr}] \rightarrow \text{correlation} \end{array}$$

April 1, 2019

Recall Covariance pattern models → (need balanced data)

- ① Model the covariance (use ML) and choose a structure.
- ② Choose the most parsimonious one that still adequately fits the data
- ③ Model the mean. (ML) and choose model
- ④ Restimate using REML.

Some common covariance structures

most extreme → ① Unstructured → #params = $\frac{k(k+1)}{2}$ where $k = \# \text{ time pts}$

most extreme → ② Compound symmetry → #params = 2 → one each subj

↳ assumes constant variance, covariances... variance

③ Autoregressive (1) $q=2$. Assumes constant correlation and decay correlation as time separation ↑

$$f_{ij} = \text{Alt } \rho^{1|i-j|} \quad ? \rightarrow \text{need equally spaced measurements}$$

↳ doesn't make sense if time pts are not equally spaced

(4) Toepitz (Banded) $q = k$

Assumes constant variance + constant correlation for a given time separation.

All nested
within
unstructured
when
compute.

1-2: σ_{12} } can set threshold

7-3: $\sigma_{13} \dots$ $\sigma_{ij} = 0$ for some ij

more flexible.

(5) Exponential $q \geq 2$, Assumes constant variance

Assume the same correlation for each time separation

$$\rho^{|j-i|}$$

does not require equally spaced measurements.

- How do we specify these in R?

Ex Bock data $N = 75$ depressed patients who either received 3 weeks of tricyclic anti-depressant (TCA) followed by 3 weeks of no drug treatment, (OR) 3 wks of no drug treatment followed by 3 wks of TCA.

→ Patients were NOT randomized. Balanced with measurements taken at 6 weekly assessments.

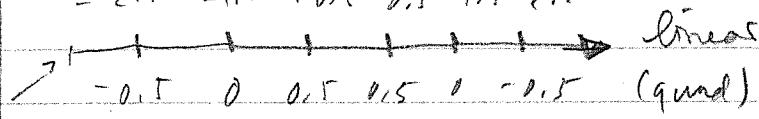
Response → Clinical status measured by Weekly Psychiatric Status Scale (WPS) for episodic disorders (1-6)

Group 0: TCA - No drug → Worse

Group 1: No drug - TCA

week ① ② ③ ④ ⑤ ⑥

-2.5 -1.5 -0.5 0.5 1.5 2.5



start

Model	-ln L	g
Unstructured	-472.9676	21
Comp Symm	-592.9134	2

When testing covariance structures, make sure we use a model for the mean that is not restrictive

$$Y_{ij} = \beta_0 + \beta_1 (\text{linear})_{ij} + \beta_2 (\text{lin change})_{ij} + \beta_3 (\text{order})_{ij} \\ + \beta_4 (\text{linear})_{ij} (\text{order})_{ij} + \beta_5 (\text{lin change})_{ij} (\text{order})_{ij}$$

use nlme → use (glm)

unstructured → corSymm (form = ~ week / id),
weights

form =

needs to be
a numeric cont.
obj.

To get unstructured matrix → corSymm (form = ~ week / id),
weights = varIdent (form = ~ 1 / week)

Allows variance to be
unique at each time pt.

Comp Symm → CorCompSymm (form = ~ week / id), 1

Rock Data	Reall	Model	-ln L	g
		Unstructured	-472.9676	21
		Comp Symm	-592.9134	2

$\rightarrow \begin{pmatrix} \sigma^2 & \rho_{12} & \rho_{13} \\ \rho_{12} & \sigma^2 & \rho_{23} \\ \rho_{13} & \rho_{23} & \sigma^2 \end{pmatrix} \leftarrow \text{Comp Symm w/ Heterogeneous Var}$

Auto regressive (1)

Toepiltz

$\hookrightarrow \rho = \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i) \text{Var}(X_j)} \rightarrow \text{different}$

Unstructured

Comp Symm

-586.569

7

-498.174

2

-4941.433

6

April 3, 2019

Comp Symm with Heteros var

Comp Symm

Hetero. var

Cor Comp Symm (form = ~weak|id),
weights = var Identity (form = n/weak)

Now, compare each structure to unstructured.

All cor structures are nested in unstructured

Note abt Toepitz in R \Rightarrow not part of the built-in R cor structures

But we can make it into fitting one by using ARMA model

cor ARMA (form = ~weak|id, p=5, q=0)

$p \quad q=0$
target time gap

Comp Symm to Unstructured

CS with Hetero var to Unstructured ...

unstructured



Model	- ln L	q	-2 (ln L ₀ - ln L _A)	qA	P
CS	2	2	239.12	21	<<
CS - Hetero	7	7	227.02	21	<<
AR(1)	2	2	50.41	21	0.000
Toepitz	6	6	626.03 ***	21	0.0001

When doing these tests, take p-value and divide by 2
 \rightarrow avoid Type I error.

In each case \rightarrow reject null hypothesis \rightarrow unstructured model necessary
 \rightarrow conclude need unstructured structure

choose unstructured

now back to
model
measur

Done with covariance modeling...

$$Y_{ij} = \beta_0 + \beta_1 (\text{lin})_{ij} + \beta_2 (\text{linchge})_{ij} + \beta_3 (\text{order})_{ij} + \beta_4 (\text{lin})_{ij}^2 + \beta_5 (\text{linchge})_{ij}^2 + \beta_6 (\text{order})_{ij}^2$$

Since high order term significant
→ drop all terms

→ get model → run again with REML → get unbiased estimates!!!

April 5, 2019

Think back to mixed models...

$$\underline{Y}_i = \underbrace{\underline{X}_i \underline{\beta}}_{\text{fixed}} + \underbrace{\underline{Z}_i \underline{V}}_{\text{random}} + \underline{\epsilon}_i \quad \underline{\epsilon}_i \sim N[\underline{0}, \underline{\Sigma}_e]$$

$$\hat{\underline{\Sigma}}_i = \underline{Z}_i \underline{\Sigma}_v \underline{Z}_i^\top + \underbrace{\sigma_e^2 \underline{I}}_{\text{intrinsic to variance or diagonal}}$$

The correlation structure comes from this form

[Chapter 7: MIXED MODELS WITH (AUTO)CORRELATED ERRORS]

→ Same model: $\underline{Y}_i = \underline{X}_i \underline{\beta} + \underline{Z}_i \underline{V} + \underline{\epsilon}_i$

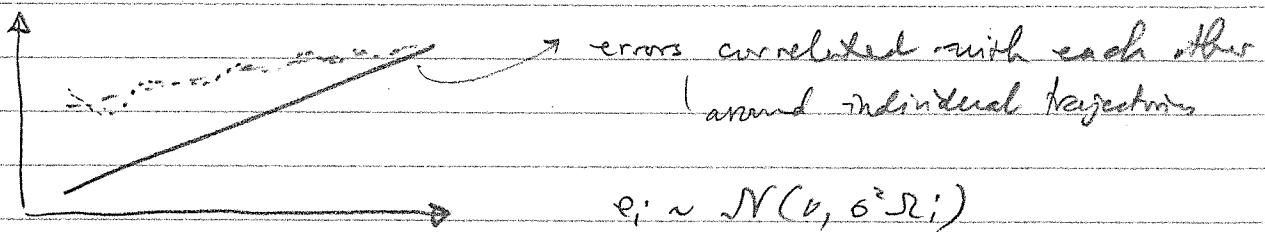
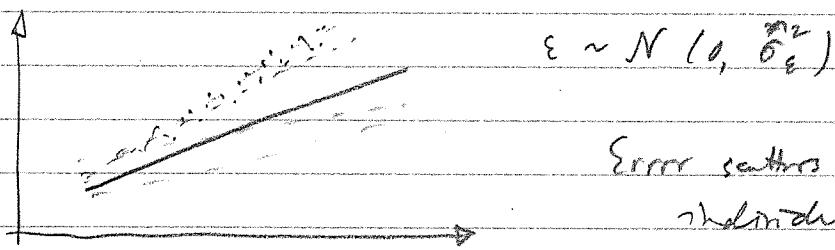
$$\underline{V}_i \sim N[\underline{0}, \underline{\Sigma}_v]$$

$$\underline{\epsilon}_i \sim N[\underline{0}, \underbrace{\sigma_e^2 \underline{R}_i}_{\text{can have any number of different structures...}}]$$

Now,

$$\hat{\underline{\Sigma}}_i = \underline{Z}_i \underline{\Sigma}_v \underline{Z}_i^\top + \underbrace{\sigma_e^2 \underline{R}_i}_{\text{Two lower identity}}$$

Ex Random Intercept - Slope...



Consider AR(1) errors

$$\varepsilon_j = p \varepsilon_{j-1} + \xi_j \text{ where } \xi_j \sim N(0, \sigma^2)$$

$p \rightarrow$ auto correlation coefficient.

Note if we had AR(2), then $\varepsilon_j = p_1 \varepsilon_{j-1} + p_2 \varepsilon_{j-2} + \xi_j$

STATIONARITY $\rightarrow \text{Var}(\varepsilon_j) = \text{Var}(\xi_j)$

Consider AR(1) model...

$$\begin{aligned}\text{Var}(\varepsilon_j) &= \text{Var}(p \varepsilon_{j-1} + \xi_j) \\ &= p^2 \text{Var}(\varepsilon_{j-1}) + \text{Var}(\xi_j) \\ &= p^2 \text{Var}(\varepsilon_{j-1}) + \sigma^2\end{aligned}$$

\Rightarrow Stationarity \Rightarrow variances and covariances (for the same time lag) are independent of j

So $\left| \begin{array}{l} \text{Var}(\varepsilon_j) = p^2 \text{Var}(\varepsilon_j) + \sigma^2 = \frac{\sigma^2}{1-p^2} \\ \text{Cov}(\varepsilon_j, \varepsilon_{j-1}) = \frac{p\sigma^2}{1-p^2} \end{array} \right|$

Theen $\Sigma_v = \frac{\sigma^2}{1-p^2} \begin{pmatrix} 1 & & & p^{n-1} \\ p & \ddots & & \\ & \ddots & \ddots & \\ p^{n-1} & & & 1 \end{pmatrix}$

Before in Ch. 6, we said

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & & & p^{n-1} \\ p & \ddots & & \\ & \ddots & \ddots & \\ p^{n-1} & & & 1 \end{pmatrix}$$

equivalent

\rightarrow scaling or variance is different

↳ We have

Another model \rightarrow moving average...

$$\sigma^2 = \frac{\sigma^2}{1-p^2}$$

↳ we can assume $\varepsilon_j = f_j - \theta f_{j-1}$ where $\{f_j \sim N(0, \sigma^2)$

θ = autocorr.

coeff for moving average process

[MA(1)]

If we assume stationarity at each time point, then

$$\Sigma = \sigma^2 \begin{bmatrix} 1+\theta & -\theta & & & \\ -\theta & 1+\theta & & & \\ & & \ddots & & \\ & & & 1+\theta & -\theta \\ & & & -\theta & 1+\theta \end{bmatrix}$$

- Allow correlations between 2 responses to be zero once a certain lag is achieved.
- ARMA(1,1) model put both AR(1), MA(1) together..

$$\Sigma = \frac{\sigma^2}{1-p^2} \begin{bmatrix} f_0 & & & & p^{n-2} f_1 \\ f_1 & \ddots & & & \\ & \ddots & \ddots & & \\ p f_0 & & & \ddots & \\ & & & & p^{n-2} f_1 \end{bmatrix} \dots \begin{bmatrix} f_0 & & & & p^{n-2} f_1 \\ f_1 & \ddots & & & \\ & \ddots & \ddots & & \\ p f_0 & & & \ddots & \\ & & & & p^{n-2} f_1 \end{bmatrix}$$

An ARMA(1,1) is similar to an AR(1), but with an increase correlation for lag 1. Most useful when lag 1 error correlation is large \Rightarrow remaining lags decrease exponentially

Toeplitz $\Sigma = \sigma^2 \begin{bmatrix} 1 & & & & \\ p_1 & 1 & & & \\ \vdots & & \ddots & & \\ & & & 1 & \\ p_{n-1} & & & & p_1 \end{bmatrix}$

→ We can set time lags above a certain threshold to be 0
 → Banded ...

- Note MAC(1) is a specific Toeplitz structure.

NON-STATIONARY what happens if var not constant
 over time and cov are not constant
 for a particular time lag?

Note can't use nested tests (LLH ratio) to compare
 → AIC

April 8, 2019

We already talked about structures for OLS, AR(1), MAC(1)

- ARMA(1,1) - allows for an arbitrary large correlation for a time gap of 1 that decays rapidly for large gaps
- Toeplitz → in combination of random fx structure → allows an arbitrary structure for correlation that depends only on time separation
 ⇒ Time gap 1 has same corr strategy regardless of the time measurements under consideration.

Ex Consider random int model with Toeplitz structure on Σ

$$\sigma^2 \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ \vdots & & \ddots & & \\ & & & 0 & \\ 0 & & & & 0 \end{bmatrix}$$

Random Effect
 Covariance Symmetry

$$\rho^2 \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

So we can only specify a Toeplitz structure with $b-1=5$ parameters.
at each time point

Non-stationarity

→ we can allow variances to differ

$$\hookrightarrow V(\varepsilon_j) = \rho^j V(\varepsilon_0) + \sigma^2$$

Mansour (1995) gave the following framework...

• Start with $V(\varepsilon_0) = \sigma^2$ (before study began)

$$V(\varepsilon_1) = \sigma^2$$

$$V(\varepsilon_2) = \rho^2 \sigma^2 + \sigma^2 = \sigma^2(1+\rho^2)$$

$$V(\varepsilon_3) = \dots = \text{etc} = \sigma^2(1+\rho^2+\rho^4)$$

As time increases, variances get bigger as long as $\rho \neq 0$

$$\Rightarrow \sigma^2 \Omega_i = \sigma^2 \begin{bmatrix} 1 & & & & \\ \rho & 1+\rho^2 & & & \\ \vdots & \vdots & 1+\rho^2+\rho^4 & & \\ \rho & & \ddots & \ddots & \\ p^{n-1} & p^{n-2} & p^{n-3} & \ddots & \frac{1}{\sigma^2} \rho^{2n} \end{bmatrix}$$

Q

How do we compare models?

Before → used LRTs to compare models... since they were nested

Now → most of these mixed models with correlated error
are not nested...

→ use **AIC** → takes log likelihood - penalizes it
by $\#$ of parameters that we've estimated.

$$\text{AIC} = -2 \ln L + 2p \uparrow \# \text{ params estimated}$$

→ lower is better

(59)

$AIC = -2 \ln L + 2kN$ for not sure what N is
 # observations? & ind. individual
 \rightarrow unclear \rightarrow don't use AIC

$$\sum r_i \text{ i.i.d.} \Rightarrow \sigma^2 \sum r_i = \sigma^2 I_1$$

bad intercept

$$\checkmark r_i. AR(1) \Rightarrow AIC = 1201.2$$

$$\checkmark r_i. har1 \Rightarrow \text{nonstationary} \quad AIC = 1014.9$$

$$\checkmark r_i. har1 \Rightarrow \text{nonstationary} \quad AIC = 1012.7 \dots$$

April 10, 2019

A marginal model for longitudinal data has the following
April 12, 2009 3-part specification

each

① The conditional expectation of response, (or the mean)

$E(Y_{ij}|X_{ij}) = \mu_{ij}$ is assumed to depend on the covariates (predictors) through a known link function
 $g(\mu_{ij}) = \eta_{ij} = \underline{x}_{ij}^T \beta$ (mean / 1-mean)

\underline{x} Identity, ln (odds), ln (mean) = $\beta_0 + \beta_1 X_{ij}$
 (continuous) (binary) (count)

② The conditional variance of each Y_{ij} , given the covariates, is assumed to depend on the mean

$V(Y_{ij}|X_{ij}) = \phi \cdot V(\mu_{ij})$ where $V(\mu_{ij})$ is known variance function and ϕ is a scale parameter.
 ϕ might be known or estimated.

③ The conditional within-subject association among the vector of repeated measures, given the covariates, is assumed to be a function of a set of additional parameters, called α (and also the means, μ_{ij})

→ What do we need to estimate?

$\left\{ \begin{array}{l} \beta's \\ \text{Maybe } \phi \\ \alpha's \quad \text{or nuisance parameters} \end{array} \right.$

→ **MARGINAL MODELS**

→ do not require a distributional assumption for the observations
 → only a regression model for the mean responses

$$\text{Note: } E(Y_{ij}|X_i) = E(Y_{ij}|X_{i1}, \dots, X_{in}) = E(Y_{ij}|X_{ij})$$

This implies that given $X_{ij} \rightarrow$ there's no dependence of Y_{ij} on X_{ik} for $j \neq k$. This is fine for time-invariant records, but not necessarily if time can be considered a random variable itself
 ↗ okay in balanced/experimental settings...

- Y_{ij} can't depend on X_{ij-1}

What abt GEE's?

(in 2t)

- ↳ There's no easy way to specify the distribution of all the responses when responses are discrete.
- ↳ we need special equations to provide us with the parameter in this setting.
 → GEEs provide one way to do this.

GEEs are attractive bcz they provide us with consistent estimates of the β 's even when our model for associations is wrong... (1)

We still need models for $\left\{ \begin{array}{l} \text{mean } \eta_{ij} = X_i \beta \\ \text{variance } V(Y_{ij}|X_i) = \phi u(\mu) \\ \text{correlation } \\ \text{structure } R(\alpha) \approx \text{corr. matrix} \end{array} \right.$

With (2) + (3) and (1) specified we can construct the var-covar matrix of $\hat{\beta}$

$$\Sigma_i = V_i = A_i^{-1/2} R(\alpha) A_i^{1/2}$$

where A_i is a diagonal matrix with $V(Y_{ij}|X_{ij})$
 $= \phi V(x_{i,:})$

And $R_i(\alpha)$ is the correlation matrix (1's on the diagonal)

- $R_i(\alpha)$ can have different structures

(1) Independence $\rightarrow R_i(\alpha) = I$

(2) Exchangeable $\rightarrow R_i(\alpha) = \rho^{|i-j|}$ (same as compound)

(3) AR(1) $\rightarrow R_i(\alpha) = \rho^{|i-j|}$ (same as exponential)

↑
Not autoregressive...

→ like Toeplitz

(4) M-dependent (or Banded): $R_i(\alpha) = \rho^{|i-j|}$

The GEE estimator is the solution to

$$\left[\sum_i^N [D_i^T V(\hat{\alpha})]^{-1} (y_i - \mu_i) \right] = 0 \rightarrow \text{estimating eqn.}$$

where

$$D_i = \frac{\partial \mu_i}{\partial \beta}$$

In continuous case, we have $\mu_i = x_i \beta$ and then
(normal)

$$D_i = x_i, \quad V(\hat{\alpha}) = R_i(\hat{\alpha})$$

in which case,

$$\left[\sum_i [x_i^T (R_i(\hat{\alpha}))^{-1} (y_i - x_i \hat{\beta})] = 0 \right]$$

$$\hat{\beta} = \left[\sum_i x_i^T (R_i(\hat{\alpha}))^{-1} x_i \right]^{-1} \left[\sum_i x_i^T (R_i(\hat{\alpha}))^{-1} y_i \right]$$

This is the same estimator we would get using some version of least squares

→ weighted least squares ...

Obtaining estimator is an iterative process

- ① Given estimates for $R_i(\alpha)$ and ϕ , calculate estimates of β using iterative re-weighted least squares
- ② Given estimates of β , obtain estimates of α and ϕ

Iterate until convergence (\approx w/ same tolerance)

Result is a consistent estimate of β

\Rightarrow what if we want to construct C.I. or do
hyp test \rightarrow ? Trouble since we need a way to estimate the SE's of β 's to do that.

model Based }
} Empirical

-G-

April 15, 2019

GEE provides with a way to get consistent estimates of β 's even if the model for correlation is wrong...
But what we're missing is SEs for β 's.

There are 2 methods that GEEs use to estimate standard errors, of β 's

- ① Model-based (naive) $\rightarrow V(\hat{\beta}) = \left[\sum_{i=1}^N D_i^T \tilde{V}_i^{-1} D_i \right]^{-1}$

- ② Robust (empirical) $\rightarrow V(\hat{\beta}) = M_0^{-1} M_1 M_0^{-1}$

where

$$M_0 = \sum_{i=1}^N D_i^T \tilde{V}_i^{-1} D_i$$

$$M_1 = \sum_{i=1}^N D_i^T \tilde{V}_i^{-1} (y_i - \mu_i)(y_i - \mu_i)^T \tilde{V}_i^{-1} D_i$$

Important to remember \rightarrow If variance of y_i is mis-specified then using model-based SEs will be wrong...

B Coding for group word Helmet contrast...

	H1	H2	H3
Control	-1	0	0
No & show	1/3	-1	0
Tx 1	1/3	1/2	-1
Tx 2	1/3	1/2	+1
	↑	↑	↑
	experimental observational	observational	observational

Null: makes no sense

→ calc. correlation for binary data / cat data → use tetrachoric.

April 17, 2019

GEE & missing data

- MANOVA / RM ANOVA → no missing data is allowed
- Mixed modeling → used ML / REML → can handle missing data that are either MCAR (miss. corr/uncorr) or MAR (miss. corr/uncorr)
- Covariance pattern models ⇒ estimated via ML
→ can handle missing data that are MCAR/MAR
- GEE can only handle missing data that are MCAR (not MAR)

Type of missing data

- MCAR → probability of an observation being missing cannot depend on any characteristics of the individuals observed or unobserved ... (just random)
- MAR → prob. of observation being missing can depend on observed characteristics, but not unobserved ...
- Non-ignorable → prob. of obs. being missing depends on unobserved characteristics

\Rightarrow so most programs default to using robust estimators called the "sandwich" estimator ($\hat{\beta}_0, \hat{V}(\hat{\beta})$)

- The sandwich estimator provides us with consistent estimates for $V(\hat{\beta})$ even if $R_i(\alpha)$ is wrong.

\rightarrow in finite samples, it may be biased and the variance of $\hat{\beta}$ ~~can~~ also be large...

This is a bigger problem for smaller N and large number of measurements.

Example Data obtained from a randomized study on smoking cessation published by Gruber et al. (1993)
People were randomized to receive:

① Control: given access to smoking cessation reading and programming

② Intervention - Discussion group
social support group

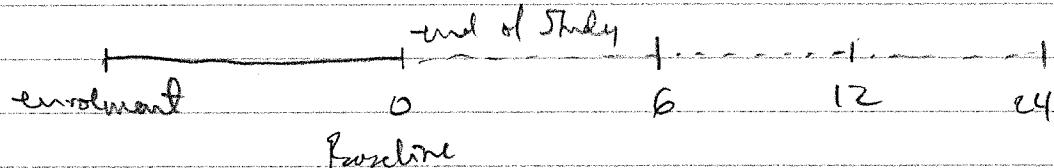
Final grouping has 4 levels: ① Control

② No show: randomized into a group that never showed up

③ Tx 1 (Discussion)

④ Tx 2 (Social Support)

Timing



Response Binary $\rightarrow \begin{cases} 0 : \text{smoking} \\ 1 : \text{Not smoking} \end{cases}$

Covariate $\begin{cases} 0 : \text{white} \\ 1 : \text{non-white} \end{cases}$

Running in R.
 waves = time \rightarrow allows for intermittent missingness
 family = binomial \rightarrow Y/N response
 scale: $Ax \approx T \rightarrow$ for binary response

\rightsquigarrow Wald statistics...

April 19, 2019

GEE \rightarrow only models population...

MIXED MODELING FOR BINARY DATA

How would we use ML to estimate the model parameters from a "regular" logistic regression? (no covar., no longitudinal data)

Our likelihood function is a function of the unknown parameters, p_i .
 ($p_i \rightarrow$ probability that observation i is a "success" = $P(Y_i=1)$)
 $1-p_i \rightarrow$ "failure" = $P(Y_i=0)$
 each Y_i is a Bernoulli trial (Bernoulli r.v.)

We relate the vector of binary responses \underline{Y} to some grouping of covariates \underline{X} . Need to relate \underline{Y} to \underline{X} via logit link.

$$\text{logit } \hat{p}_i = \underline{x}_i^T \underline{\beta} = \ln \left(\frac{p_i}{1-p_i} \right) \rightarrow p_i = \frac{\exp[\underline{x}_i^T \underline{\beta}]}{1 + \exp[\underline{x}_i^T \underline{\beta}]}$$

We're trying to fit $\underline{\beta}$ that maximizes the likelihood...
 $= \frac{1}{1 + \exp[-\underline{x}_i^T \underline{\beta}]}$
 $= \Psi(\underline{x}_i^T \underline{\beta})$

For a single Bernoulli trial,

$$P(Y_i=1) = \Psi_i^{y_i} (1-\Psi_i)^{1-y_i} \quad \text{where } \Psi_i = \Psi(x_i^T \underline{\beta})$$

$$\text{likelihood fn} = \prod_i^N \Psi_i^{y_i} (1-\Psi_i)^{1-y_i}$$

Usually, we take log likelihood $\rightarrow L = \ln(L) = \sum_{i=1}^N [y_i \ln \Psi_i + (1-y_i) \ln(1-\Psi_i)]$

$$\text{then } \frac{\partial l}{\partial \beta} = \sum (y_i - \hat{y}_i) x_i = 0 \quad (1)$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta^T} = -\sum \psi_i(1-\psi_i) x_i x_i^T = 0 \quad (2)$$

We can't solve this analytically \Rightarrow need an iterative approach.
 \rightarrow Use Newton-Raphson algorithm to get estimates of β .

$$\boxed{\beta_{k+1} = \beta_k - \left[\frac{\partial^2 l}{\partial \beta \partial \beta^T} \right]^{-1} \cdot \frac{\partial l}{\partial \beta}}$$

provides us with $\text{Var}(\hat{\beta})$
 \hookrightarrow Take the diagonal elements & they will represent the variances of β_j . (Fisher information matrix)

April 22, 2011 Last time \rightarrow considered a standard logistic regression model for independent observations. Saw how to get estimates for β 's using ML

- Probit regression \rightarrow alternate link function... (slip)
- Threshold concept \rightarrow (slip)
 Latent variable $\dots \rightarrow$  dichotomizing $y^* = 0 \quad \leftarrow \rightarrow y^* = 1$

Mixed effects model for longitudinal binary data

- If we assume heterogeneity in the propensity to respond positively across individuals, this can be captured using RE. We generally assume RE come from multivariate normal dist, and that, conditional on these, the responses for a particular individual are independent observations from a binomial dist.
- This is the "conditional" independence assumption. Consider this in GLM framework. Recall that a GLM formulation requires

- the specification of 3 things:

- ① Distributional assumption
- ② Systematic component
- ③ Link function -

- How do we do this in the "normal" response setting?

① In a linear mixed effects model it is assumed that the conditional distribution of \underline{Y}_{ij} given the random fx is normal, with $\text{Var}(\underline{Y}_{ij} | \underline{V}_i) = \sigma^2 \rightarrow \phi = \sigma^2, V(\mu_{ij}) = 1$

Also, given RE, it is assumed that the \underline{Y}_{ij} are independent.

② Conditional mean of \underline{Y}_{ij} is assumed to depend on both fixed + random fx via

$$\underline{Y}_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{v}_i. \text{ Also, } \underline{v}_i \sim N(0, \Sigma_v)$$

$$③ E[\underline{Y}_{ij} | \underline{V}_i] = \underline{\eta}_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{v}_i$$

The link function is the identity, in the continuous response setting. It's notation

i.e., an individual's response differs from population response, by subject-specific RE and random error, $e_{ij} \sim N(0, \sigma^2)$

$$\underline{Y}_i = \underline{x}_i \underline{\beta} + \underline{z}_i \underline{v}_i + \underline{e}_i$$

$$E[\underline{Y}_i | \underline{V}_i] = \underline{x}_i \underline{\beta} + \underline{z}_i \underline{v}_i, E[\underline{Y}_i] = \underline{x}_i \underline{\beta}$$

Conditional exp of \underline{Y}_i given \underline{V}_i is different than the marginal expectation of \underline{Y}_i .

$\Rightarrow \underline{\beta}$ have the interpretation of being population-averaged, i.e.

how mean responses change over time and how these changes relate to the covariates...

The conditional covariance of \underline{Y}_i given \underline{V}_i is assumed to be a diagonal matrix with

$$\text{cov}(\underline{X}_i | \underline{V}_i) = \text{cov}(\underline{\epsilon}_i) = \sigma^2 I_n \text{ (diagonal)}$$

What about the marginal covariance of \underline{Y}_i ?

$$\text{cov}(\underline{Y}_i) = \underline{Z}_i \sum_v \underline{Z}_i^T + \sigma^2 I_n$$

P

not diagonal. The dependence among the repeated measurements is introduced only through RE.

④ How does this extend to binary logitnormal responses?

① Conditional on a single RE, v_i , the y_{ij} are independent and have a Bernoulli dist with

$$V(y_{ij}|v_i) = \underbrace{E[y_{ij}|v_i]}_p \cdot \underbrace[0.5]{[1-E(y_{ij}|v_i)]}_{q} \text{ with } \phi = 1$$

p

q

② Conditional mean of y_{ij} :

$$\gamma_{ij} = \underline{X}_{ij}^T \beta + \underline{z}_{ij}^T \underline{v}_i = \underline{X}_{ij}^T \beta + \underline{v}_i \text{ where } z_{ij}=1 \text{ if } i=j \dots$$

③ The link function is the logit.

$$\ln\left(\frac{P(Y_{ij}=1|v_i)}{P(Y_{ij}=0|v_i)}\right) = \gamma_{ij} = \underline{X}_{ij}^T \beta + v_i$$

The RE is assumed $N(0, \sigma_v^2)$

→ conditioned on subject random β_k .

- Now, the regression parameter (β) interpretations are subject-specific

→ Consider a simple, mixed effects logistic regression.

$$\text{logit}(\pi_{ij}) = \underline{x_{ij}}^T \underline{\beta} + v_i \quad (\text{random intercept-}) \\ N(0, \sigma^2)$$

- The interpretation of any component of β , say β_k , is in terms of change in any given subject's log odds of response for a unit increase in the within-subject covariate x_{ijk}

$$\text{logit}(\pi_{ij}|x_{ijk} = x^*) = v_i + \beta_1 x_{ij1} + \dots + \beta_k x_{ijk}^* + \dots + \beta_p x_{ijp}$$

$$\text{logit}(\pi_{ij}|X_{ijk} = x^* + 1) = v_i + \beta_1 x_{ij1} + \dots + \beta_k (x^* + 1) + \dots + \beta_p x_{ijp}$$

$$\text{logit}(\pi_{ij}|x_{ijk} = x^*) - \text{logit}(\pi_{ij}|x_{ijk} = x^* + 1) = \beta_k$$

→ β_k is change in log odds associated w/ a 1-unit increase in x_{ijk} , holding all the other x_{ijk} constant, i.e. working with the same subject.

⇒ **subject specific**. Note that this interpretation makes more sense for covariates that vary within an individual if x_k is constant for all time points (within one a subject), then the interpretation is misleading. This is common for gender, treatment, ethnicity, ... things that aren't changing through time.

In these cases, the interpretation of β_k is confounded with $v_{0i} - v_{0i'}$

$$T \quad F \\ x_k = 0 \quad x_k = 1$$

If the link function is not the identity,

$$g[E(Y_{ij} | X_{ij}, v_i)] = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{v}_i$$

$g[E(Y_{ij} | X_{ij})] \neq \underline{x}_{ij}^T \underline{\beta}$ for all $\underline{\beta}$ when averaged over the dist of random effects.

- Inferences for $\underline{\beta}$ in mixed-effects model are subject-specific.
- Inferences for $\underline{\beta}$ in marginal model are at population level.

Ex Consider simple example using 3 subjects A, B, C.

Let $p_{ij} = P(Y_{ij}=1)$ be measured at baseline - post baseline.
Treatment is designed to reduce the probability of disease.

Subj	Baseline prop	Post-Baseline - prop	Difference by	Odds ratio
A	0.8	0.67	-0.13	-0.68
B	0.5	0.33	-0.17	-0.71
C	0.2	0.11	-0.09	-0.70
Pop Avg.	0.5	0.37	-0.13	

- ◻ Treat response as continuous. The differences are subject-specific of treatment. We can produce population effects two different ways.

① Average the subject-specific effects

$$\frac{-0.13 - 0.17 - 0.09}{3} = -0.13$$

identical
if we
treat response
as continuous

② Compare pop-average at baseline to pop. avg.
post-baseline : $-0.37 - 0.70 = -0.13$.

- ◻ Now, treating these as Bernoulli and applying logit link.

① Averaging log odds ratios we get

$$\frac{-0.68 - 0.70 - 0.71}{3} = -0.697.$$

↳ The effect of individuals probability of disease is $e^{-0.697} \approx 0.5$

② At the population level, log odds of disease at baseline is

$$\ln\left(\frac{0.5}{1-0.5}\right) = 0$$

$$\text{At post-treatment, } \ln\left(\frac{0.37}{1-0.37}\right) = -0.532$$

↳ population effect is $-0.532 \neq 0.5$

Q Which of the effect estimates is better? -0.697 or -0.532 ?

↳ They're both reasonable.

{ -0.697 provides a measure of the efficacy expected change in odds of disease for any individual in treatment

↳ There is $1 - e^{-0.697} \approx 0.5$ reduction in odds of disease of most interest to you and your doctor.

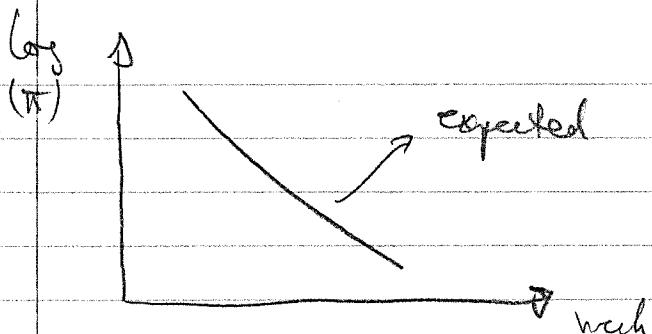
{ -0.532 is a measure of the exposure change in population if everyone were to be treated. There is a $1 - e^{-0.532} \approx 0.40$ reduction in disease in population

↳ different perspective of most interest to public health professionals.

April 26, 2029

Mixed-effect logistic regression model example

- Missing data: Mixed effect model - can handle both MCAR, MAR
GEE can't.
- Linearity of model: \rightarrow assumption



- Simple logistic regression \rightarrow ignores correlation.
- + Residual Deviance \rightarrow tells us about the fit.
- $-2 \log L$
- + what went wrong? \rightarrow standard error is way off.

Add random intercepts

- \rightarrow need library (lme4) to do mixed effect model for binary data
- \rightarrow random intercept are taken care of by $(1/\text{id})$
- \rightarrow heterogeneity \rightarrow histogram of Intercept
- \rightarrow If the model becomes too complicated
 - \rightarrow not happy
 - \rightarrow Firth's likelihood \rightarrow MASS

$\xrightarrow{\text{---}}$

April 29, 2019

glmer ($y \sim x_1 + x_2 + \dots + (\#1/\text{id})$), family = binomial ...

\uparrow
random intercept term

Random slopes $\sim \dots$ (week/ id) $\sim \dots$

glmm PQL \curvearrowright penalizing quasi-likelihood \leftarrow needs library (MASS)

	PQL	ML
Int	0.70 (0.47)	5.29 (0.62)
Shrub	-1.37 (0.21)	-1.43 (0.28)
dt	0.124 (0.52)	-0.016 (0.63)
frt.week	-1.03 (0.24)	-1.04 (1.33)

↗ same model, different estimation routine

Missingness Mechanisms

MCAR, MAR, nonignorable (NAR)

MCAR \rightarrow Data is MCAR if the probability that the responses are missing is unrelated to either specific values that, in principle, should have been obtained, or the set of observed responses.

Let R_{ij} be an indicator of whether we observed Y_{ij}

$$R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed} \\ 0 & \text{if not} \end{cases}$$

warranted

$$P(R_{ij}=1 | Y_{i1}, Y_{i2}, Y_{ij}, Y_{in}, X_j) = P(R_{ij}=1 | X_j)$$

← ↑ potentially
 observed missing
 responses responses

Probability of missingness might depend on predictors.

Simple case \rightarrow only 2 responses Y_{i1}, Y_{i2} .

Assume that Y_{i1} is fully observed

$$P(R_{i2}=1) = 1$$

$$P(R_{i2}=1 | Y_{i1}, X_{i2}, X_j) = P(R_{i2}=1 | X_j)$$

warranted

An example Let's say we have a study of air pollution's effects on lung function of children.
(six cities study)

If parents move out of the cities for reasons unrelated to lung function of children \rightarrow missing data will be MCAR

$$\text{odds} = \frac{p}{q} = \frac{p}{1-p}$$

$$p = 1.27(1-p)$$

$$= 1.27 - 1.27p$$

MISSINGNESS - cont

May 1, 2019

getting a response

MCAR

$$P(R_i = 1 | Y_i^0, Y_i^M, X_i)$$

↓ ↓ ↓
observed missing covariate
responses responses

⇒ observed data is a random sample of all potential data we might observe

⇒ all estimation routines are valid on MCAR as long as all covariates that predict missingness are included in model.

Example In a drug trial, dropout might be predicted by side effects

Inferences are still valid as long as all information about side effects recorded

MAR

→ Missing at random. The probability of a response being missing depends on the observed responses, but it is conditionally unrelated to specific missing values that, in principle, should have been observed

$$P(R_i = 1 | Y_i^0, Y_i^M, X_i) = P(R_i = 1 | Y_i^0, X_i)$$

↑
depends on observed

Back to binary response example

Y_{i1} → always observed

Y_{i2} → might be missing depending on what we observe on Y_{i1}

Ⓐ $P(R_{i2} = 1 | Y_{i1}, Y_{i2}, X_i) = P(R_{i2} = 1 | Y_{i1}, X_i)$

(Ex)

Six Cities study

\Rightarrow if children drop out once lung function reached a certain level (which we observed) then MAR missingness.

(Ex) MAR says that if we stratify on similar values of Y_{ij} , the missingness pattern of Y_{ij} would be random.

$\left\{ \begin{array}{l} \text{ML, in general, provides us with valid inferences} \\ \text{GEE does NOT provide valid inferences} \end{array} \right\}$ might be some bias...

$\left\{ \begin{array}{l} \text{PQL} \Rightarrow \text{gives biased estimates...} \end{array} \right\}$

Non-ignorable missingness (not missing at random NMAR)

$P(R_i | Y_i^0, Y_i^m, X_i)$ depends on at least some elements of the response we didn't observe (Y_i^m)

(Ex) Depression Study where subjects drop out due to depression severity, where we didn't observe their depression score, is non-ignorable.

\hookrightarrow No estimation methods provide valid estimations...

\Rightarrow Multiple imputation is one possible solution.

Missingness example on model

Generate observations based on $Y_{ij} = 5 + 0.25t_j + \varepsilon_{ij}$

$\varepsilon \sim N(0, \Sigma)$, AR structure with $V(\varepsilon_{ij}) = 1$

$$\text{Cor}(\varepsilon_{ij}, \varepsilon_{ij'}) = (0.7)^{|j-j'|}$$

→ model to predict missingness with

at $t=j$

logit (Probability of dropout $\hat{Y}_{ij} \leftarrow \text{dropout at } t=j$)

$$= \theta_1 + \theta_2 (Y_{ij=1}, Y_{ij}) + \theta_3 (Y_{ij} - y_{ij})$$

◻ If $\theta_1 \neq 0, \theta_2 = \theta_3 = 0 \Rightarrow \boxed{\text{MCAR}}$ ie $P(\text{dropout})$ independent of responses

◻ If $\theta_2 \neq 0, \theta_3 = 0 \sim \boxed{\text{MAR}}$ ie. missingness @ $t=j$
depends on $t=j-1$

◻ If $\theta_3 \neq 0$, then $\boxed{\text{NMAR}} \rightarrow$ missing @ $t=j$ depend on what
I should have seen @ $t=j$.

MCAR

$$\begin{cases} \text{ML} & \{ P_0 = 4.97 (0.015) \\ & P_1 = 0.25 (0.016) \end{cases}$$

$$\begin{cases} \text{GEE} & \{ 4.97 (0.048) \\ & 0.262 (0.010) \end{cases}$$

MAR

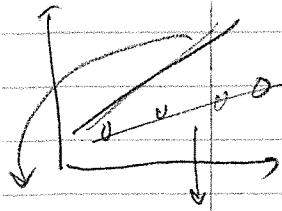
$$\begin{cases} \text{ML} & \{ 5.03 (0.040) \\ & 0.21 (0.016) \end{cases}$$

→ But the

$$\text{GEE}, \{ 5.14 (0.042)$$

$$10.136 (0.0175)$$

→ not so good, Should be 0.25



Based on what we saw, since we can't specify
correlation structure!

ML
within
gees

NMAR

both ML & GEE are worse and none are working
properly...

Recall missingness $\Rightarrow E(Y_{ij}) = \theta_0 + \theta_1 S_i + \theta_2 T_{ij}$

May 6, 2019

ML or GEE

be with missingness

Use missingness to predict missingness \rightarrow ML

MAR

[MAR]

\rightarrow GEE based
ML not



$$\text{logit}(\text{drop}_{ij}) = \theta_0 + \theta_1 Y_{i,-j} + \theta_2 Y_{ij}$$

\uparrow MCAR \uparrow MAR \uparrow (NMAR)

(mostly random) (dep on previous) dep on current value --

help to put cov structure on ML
ML correctly follows population well
GEE doesn't \rightarrow not robust @ MAR.

nonignorable

Non-ignorable

\rightarrow population (actual)

ML

GEE

GEE worse than ML, but ML still pretty bad.

* PQL saves as GEE, fits for MAR and nonignorable
 \uparrow
use for mixed model library.

MODELING COUNT DATA

→ issues same with
binary data... due to link
function

lot of similarity with modeling binary data in
logistical data

→ same problems with estimation.

→ often use GEE for estimation, because specifying
likelihood is difficult, and algorithms often
have convergence problems in ML.

→ Also have interpretation issues in mixed-model
setting, where interpretation is conditioned on
random effects...

Responses? → count of # of occurrences of something
/ some events...

Need to consider standardizing count. to take into
account how long we waited → count / total.

one solution → Rate $E\left(\frac{Y_i}{T_i}\right)$

$$E\left(\frac{Y_i}{T_i}\right) = \frac{\mu_i}{T_i}$$

Count Rate → link function → use log link function

↳ we assume $Y_i \sim \text{Poisson}(\lambda)$

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\text{Var}(u_i) = u_i \quad ; \quad \phi = 1 \quad \text{since} \quad \text{Var}(u_i) = u_i$$

Often, $\phi > 1$ (over dispersion) (not really Poisson)

↳ consequence \Rightarrow might need to estimate ϕ

Model : $\ln\left(\frac{u_i}{T_i}\right) = \beta_0 + \beta_1 x$

$$\Rightarrow \ln(u_i) = \beta_0 + \beta_1 x + \ln(T_i)$$

we call this an "offset",
not estimated

We need to include the offset if we wait different amounts of time for different individuals

In longitudinal setting, we may need to include an offset if data aren't balanced.

(What is the interpretation of β_1 ?)

Let $x=0$, then $\ln(u_i) = \ln(T_i) + \beta_0 + 0$

Let $x=1$, then $\ln(u_i) = \ln(T_i) + \beta_0 + \beta_1$

$\Rightarrow \hat{\beta}_1$ estimates the difference in the log of count for a unit increase in x , holding everything else constant.

$\Rightarrow e^{\hat{P}_i}$ estimates the relative rate of occurrence for

$$\boxed{\partial \hat{P}_i = \frac{\mu_i}{\mu_i'}}$$
 as unit increase in x ,
 holding the rest constant.

② Differences GEE $\left. \begin{array}{l} \text{offset} \\ \text{interpretation} \\ \text{having to estimate } \phi \end{array} \right\}$

③ Same as GEE $\left. \begin{array}{l} \text{estimation} \dots \end{array} \right\}$

`geglm(... same ... family = poisson, ... contr ...)`

$\phi \rightarrow$ "scale parameter"

If $\phi > 1 \rightarrow$ overdispersion ...

recall

$$\ln Y_{ij} = \beta_0 + \beta_1 \cdot \text{time} + \beta_2 \cdot \text{trt} + \beta_3 \cdot \text{time} \cdot \text{trt}$$

↑ offset of observation time differed across individuals

$$\text{Variance } \text{Var}(Y_{ij}) = \phi \mu_{ij} \quad (\text{by Poisson})$$

↑ estimated. Usually is 1, but if this is estimated to be > 1

\Rightarrow **over dispersion**

\rightarrow more common for $\phi > 1$ than $\phi < 1$

64

FINAL EXAM

Mixed model \rightarrow Fixed + Random

population each id has their own random effect, with particular characteristic and can be tricky with predictions...

Why not mixed model for count - binary? $\text{Var}(Y_{ij}) = \phi \mu_{ij}$

(1) problem with likelihood. constant

Normal: $V(Y_{ij}) = \sigma^2 \cdot 1$ \rightarrow function of mean

Binary: $V(Y_{ij}) = \phi \mu_{ij} (1-\mu_{ij})$

Poisson: $V(Y_{ij}) = \phi \mu_{ij}$

(non-normal)

problem with ML estimation for binary data

\rightarrow likelihood hard to model

(2) convergence issues in the computation...

* Don't worry too much abt mixed-models with autocorrelated errors...

* Don't worry about PQL

Standard error for GEE: naive ~ robust

Empirical \rightarrow consistent estimates

requires large sample size

by GEE, provided

GEE bad when MAR or NMAR ...

no bad missing data

$$V_i \sim N(0, \Sigma_V), \quad e_i \sim N(0, \Sigma) \quad \begin{matrix} \uparrow \\ \text{RE} \end{matrix}$$

Correlation
optimal line()

$$V(Y_{ij}) = f(\Sigma, \Sigma)$$

$$V(Y_{ij}) = \Sigma \Sigma_V \Sigma^T + \sigma^2 I$$

$$E(Y_{ij}) = \underline{x}_{ij} \beta$$

$$E(Y_{ij}|v_i) = \underline{x}_{ij} \beta + \underline{z}_{ij} v_i$$

Note Covariation pattern models \rightarrow needs balanced
 \rightarrow no random effects at all
errors are model \sim