LONGITUDINAL ANALYSIS

Feb 8, 2019

Lecture 1 → Missed

Lecture 2 Review of ANOVA, 2 different ways

- Hypothesis test
- Regression model → more useful

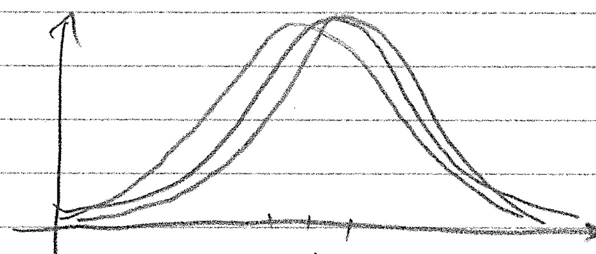
ONEWAY ANOVA { 1 response var (continuous), 1 explanatory variable (categorical) }

k = # groups

H_0: μ_i same , H_a: at least one differs.

→ variances among groups & within groups.

Ex H_0 true $k = 3$



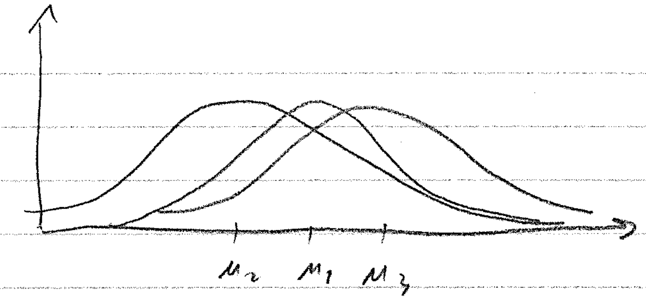
Compare 2 types of variabilities → between groups & within groups

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

In this case ⇒ $\mu_1 \simeq \mu_2 \simeq \mu_3$

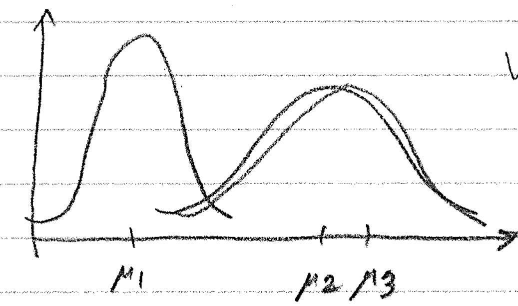
If 1 > 2 → evidence against H_0
1 < 2 → evidence for H_0

Ex H_0 true again



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

Ex H_a true



Within var < between var

How to estimate 1 & 2? (under asspt. that H_0 true)

2 Consider var within group. If we assume $\sigma_i^2 = \sigma_j^2 \; \forall i, j$, then an overall estimate of within group var would pool σ_i^2 ~~together~~

↳ Var ~~between~~ (within) groups → $\boxed{SSE = \sum_{\text{groups}} (n_j - 1) s_j^2}$

$\sim \chi^2$

weigh for sample size

1 Var between groups.

$\boxed{SSG = SSB = \sum n_j (\bar{Y}_j - \bar{\bar{Y}})^2}$

$\sim \chi^2$

overall mean

N = # 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}	$N-1$	$SS_{tot}/N-1$		

ANOVA Conditions

Very Sensitive

- Representative Sample (SRS) → Sensitive
- Equal Variance $\sigma_1^2 \cong \dots \cong \sigma_k^2$ → $\boxed{\max < 2 \times \min}$
- Independent observations → Sensitive
- Each group is normally dist or all n is large

not sensitive robust

look at Anova$residual

<u>Ex</u> R_script Weightloss - diet $k = 3$ association?

Subjects → 1, 2, 3 → Compare responses (parallel)

ANOVA for regression

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

$$\boxed{Y_{ij} = \alpha + \tau_j + \varepsilon_{ij}}$$

reference group (α), offset (τ_j), error (ε_{ij})

<u>OR</u>

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

group 1 indicator (I_1), group 2 indicator (I_2)

Feb 11, 2019

$\boxed{\text{Repeated measures ANOVA}}$ or $\boxed{\text{Single-sample repeated measures ANOVA}}$

Consider N subjects each measured at n balanced time periods

Y_{ij} is the response from subj i at time j $\begin{cases} i = 1, 2 \ldots N \\ j = 1, 2 \ldots n \end{cases}$

"balanced" → each subject is measured at the same times

Data

Subj \ Data	1	2	...	n
1	Y_{11}	Y_{12}	...	Y_{1n}
2	Y_{21}			⋮
⋮		⋱		⋮
N	Y_{N1}	---		Y_{Nn}

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

grand mean, subject effect, time effect, error

Note Two random terms on RHS

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

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

(3) τ_j is fixed → and $\boxed{\sum \tau_j = 0}$

And so $\boxed{\begin{aligned} E(Y_{ij}) &= \mu + \tau_j \\ V(Y_{ij}) &= \sigma_\pi^2 + \sigma_\varepsilon^2 \end{aligned}}$ ← (subj effect is independent of error)

What about $Cov(Y_{ij}, Y_{i'j})$?

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

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

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

Note

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

↳ Induces structure on correlation within-subjects

So within-subject covariance matrix

$$\underset{\text{subject } i}{\Sigma_i} = \begin{bmatrix} Var(Y_{i1}) & & & \\ Cov(Y_{i1}, Y_{i2}) & \ddots & & Cov(Y_{i1}, Y_{in}) \\ & & \ddots & \\ & & & Var(Y_{in}) \end{bmatrix} \rightarrow n \times n$$

$$\Sigma = \begin{bmatrix} \sigma_\pi^2 + \sigma_\epsilon^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ \sigma_\pi^2 & \ddots & & \ddots & \\ & & \ddots & & \\ \sigma_\pi^2 & & & \ddots & \sigma_\pi^2 \\ & & & & \\ \sigma_\pi^2 & \cdots & \sigma_\pi^2 & & \sigma_\pi^2 + \sigma_\epsilon^2 \end{bmatrix}$$

<u>Note</u> Σ symmetric $\Rightarrow$ Σ has a <u>compound symmetry structure</u>

$\hookrightarrow$ $\begin{cases} \text{All variances are equal } \Sigma_{ii} = \Sigma_{jj} = \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 \\ \text{All covariances are equal } \Sigma_{ij} = \sigma_{\pi}^2 \quad (i \neq j) \end{cases}$

$\Rightarrow$ We can estimate covariance matrix with 2 parameters: σ_{π}^2 and σ_{ε}^2

$\hookrightarrow$ <u>ANOVA table</u>

	Source of variability	df	SS	MS	F
	Subjects	$N-1$	(1)	$SS_S / N-1$	MS_S / MS_ε
	Time	$n-1$	(2)	$SS_T / n-1$	MS_T / MS_ε
(Residual)	Error	$(N-1)\cdot(n-1)$	(3)	$SSE/(N-1)(n-1)$	
	Total	$N \cdot n - 1$	$\sum_i^N \sum_j^n (Y_{ij} - \bar{\bar{Y}})^2$ (4)		

(1) $= \boxed{SS_{subj} = n \cdot \sum_{i=1}^{N} (\bar{Y}_{i\cdot} - \bar{\bar{Y}})^2 =}$ $\quad \bar{\bar{Y}} \rightarrow$ grand mean

$\bar{Y}_{i\cdot}$: $\langle Y_i \rangle$ across n observations.

we'll then have $\bar{Y}_{\cdot j} = \langle Y_j \rangle$ mean of obs j

(2) $\boxed{SS_{Time} = N \sum_1^n (\bar{Y}_{\cdot j} - \bar{\bar{Y}})^2}$

(3) $\boxed{SS_{Error} = \sum_1^N \sum_1^n \left[Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{\bar{Y}} \right]^2}$

(4) $\boxed{SS_T = \sum_1^N \sum_1^n (Y_{ij} - \bar{\bar{Y}})^2}$

Conditions

1. Representative samples
2. Independent subjects
3. Normal dist of residuals
4. $\boxed{\text{SPHERICITY}}$ → equal variances for all differences

Test using Mauchly's test.

<u>Now</u> <u>Estimates</u>

$$\begin{cases} \hat{\sigma}_\pi^2 = \dfrac{MS_S - MS_\varepsilon}{n} \\ \hat{\sigma}_\varepsilon^2 = MS_\varepsilon \end{cases}$$

⊛ Look at <u>R code</u>!

$$\Rightarrow \widehat{ICC} = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\varepsilon^2}$$

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

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

→ use library(nlme) → anova(nlme())

"non-linear mixed effect model"

Feb 13, 2019 Last time $\longrightarrow$ same measurement through time

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

$$Y_{ij} = \underset{\text{fixed}}{\mu} + \underset{\text{rnd}}{\pi_i} + \underset{\text{fixed}}{\tau_j} + \underset{\text{rnd}}{e_{ij}}$$

i : subject
j : response

$\rightarrow \pi_i \sim N(0, \sigma_\pi^2)$
$\rightarrow e_{ij} \sim N(0, \sigma_e^2)$
} $\rightarrow$ induces Compound symmetry structure on the covariance matrix of Y_{ij}

$$\Sigma_i = \begin{pmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ \sigma_\pi^2 & \ddots & & \\ \vdots & & \ddots & \sigma_\pi^2 \\ \sigma_\pi^2 & & & \sigma_\pi^2 + \sigma_e^2 \end{pmatrix} \rightarrow n \times n \text{ matrix}$$

n measures ...

//

$\boxed{\text{Sphericity}}$ $\longrightarrow$ assumption that $Var(Y_{ij} - Y_{ij'})$ 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

$\boxed{\text{"Contrast"}}$ → "linear combination of means"

Ex $\quad \bar{Y}_1 - \bar{Y}_2$ → compare groups 1 to 2

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

Def A contrast is defined as.

$$\boxed{\underset{\sim}{L} = \underbrace{\tilde{C}}_{\text{contrast coeff matrix}}\ \tilde{\mu}}$$

vector of means

. Sum of coeffs in any contrast must be 0.

Ex
$$\boxed{L_j = C_j' Y_i = \sum_{j=1}^{n} c_{jj'} \bar{Y}_j}$$

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

$$\tilde{C} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & & \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \begin{array}{l} 1 \text{ vs } 2 \\ 1 \text{ vs } 3 \\ 1 \text{ vs } 4 \end{array}$$

3 contrasts = ~~…~~

$$\tilde{\mu} = \begin{pmatrix} \bar{Y}_{\cdot 1} \\ \bar{Y}_{\cdot 2} \\ \bar{Y}_{\cdot 3} \\ \bar{Y}_{\cdot 4} \end{pmatrix}$$

~> time points $t = 1$

So
$$L_j = C_j' Y_i$$

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

~~Ex $C_i \cdot C_j = 0$~~

~~$C = [c_1\ c_2\ \dots\ c_n]$~~

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

Ex Contrast 1 & 2, Dot product is $1 = 1+0+0+0$

Ex

$$\tilde{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \bar{\mu} \begin{pmatrix} \bar{Y}_{\cdot 1} \\ \bar{Y}_{\cdot 2} \\ \bar{Y}_{\cdot 3} \end{pmatrix} \begin{matrix} \rightarrow 1 \text{ vs } 2 \\ \rightarrow 2 \text{ vs } 3 \\ \rightarrow 3 \text{ vs } 4 \end{matrix}$$

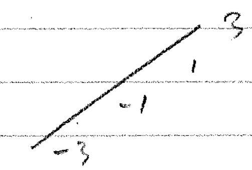
$\boxed{\text{Trend Analysis}}$ $n=4$

$n=4$

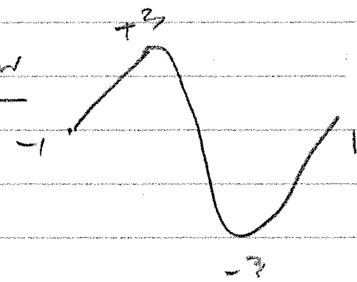
$$\tilde{C} = \begin{pmatrix} -3/\sqrt{20} & -1/\sqrt{20} & 1/\sqrt{20} & 3/\sqrt{20} \\ +1/\sqrt{4} & -1/\sqrt{4} & -1/\sqrt{4} & 1/\sqrt{4} \\ -1/\sqrt{20} & 3/\sqrt{20} & -3/\sqrt{20} & 1/\sqrt{20} \end{pmatrix} \begin{matrix} \rightarrow \text{linear trend} \\ \rightarrow \text{quadratic trend} \\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

Each row $\rightarrow$ 1 contrast

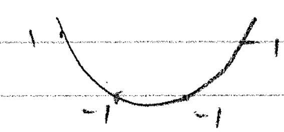
Note First row



Third row



Second row



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

$$t = \frac{\hat{L}_j - L_{j0}}{\sqrt{MSE\left(\sum_{j=1}^{n} c_{jj'}^2 / N\right)}} \qquad df = (N-1)(n-1)$$

☒ But we still have multiple comparisons issue

Summary $L_j = c_j' Y_i$ → basically a list of differences between some combination of mean

Feb 15, 2019

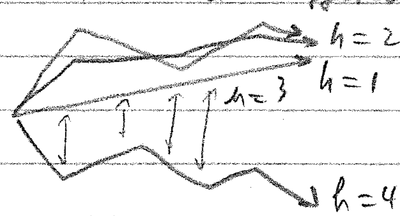
$\boxed{\text{Multisample Repeated measures ANOVA}}$

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

interaction term

- $h = 1, 2, \dots, s$ → treatment grp
- $i = 1, 2, \dots, N$ → subjects
- $j = 1, 2, \dots, n$ → time points.

Note $\gamma\tau$ → tests the difference in trajectories between (or among) 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.

☒ μ → grand mean

fixed → $\left\{ \begin{array}{l} \gamma_h \rightarrow \text{group effect} \\ \tau_j \rightarrow \text{time effect} \\ (\gamma\tau)_{hj} \rightarrow \text{interaction of time } j \text{ \& treatment / group } h \end{array} \right.$

random → $\pi_{i(h)}$ → individual different component for subject i in grp h

random → ε_{hij} → random error.

σ_π^2, σ_ε^2

☒ We have said that $\pi_{i(h)} \sim N(0, \sigma_\pi^2)$

$$e_{hij} \sim N(0, \sigma_\varepsilon^2)$$

average out to be zero $\left\{ \begin{array}{l} \sum^{s} \gamma_h = 0 \\ \sum^{n} \tau_j = 0 \end{array} \right.$ $\quad \sum^{s}\sum^{n} (\gamma\tau)_{hj} = 0$

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

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

t - comes from exclusion - 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/MSE
τ	Time	$n-1$	$N\sum_{j=1}^{n}(\bar{Y}_{..j} - \bar{\bar{Y}})^2$	$SST/n-1$	
$\gamma\tau$	Grp x Time	$(s-1)(n-1)$	$n\sum^{s}\sum^{n} N_h(\bar{Y}_{h.j} + \bar{\bar{Y}} - \bar{Y}_{h..} - \bar{Y}_{..j})^2$	$SSGT/(s-1)(n-1)$	$MS_{GT}/MS_{S(G)}$
π	Subject	$N-s$	$n\sum_h^{s}\sum_i^{N}(\bar{Y}_{hi.} - \bar{Y}_h)^2$	$SSS/N-s$	
ε	Error	$(N-s)(n-1)$	$\sum\sum\sum(Y_{hij} - \bar{Y}_{h.j} - \bar{Y}_{hi.} + \bar{Y}_{h..})^2$	$SSE/(N-s)(n-1)$	
	Total	$N_n - 1$	$\sum\sum\sum(Y_{hij} - \bar{\bar{Y}})^2$		

subjects are confounded by groups.

Subject → Rnd ↗ G1 → Tr1 ↘ / ↘ G2 → Tr2 ↗ compare

subject ~ Groups

Notes $\bar{Y}_{h \cdot \cdot}$ = averaging everybody @ all time j, in group h

$\bar{Y}_{hi\cdot}$ = averaging ~~one~~ one's response over time in group h

$\bar{\bar{Y}}_{\cdot\cdot\cdot}$ = grand mean → across groups, subjects & times

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

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

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

$\bar{Y}$ ...

→ interaction.

$\boxed{\text{Tests}}$ → the primary interest is Group × Time.

$H_0: (\gamma\tau)_{hj} = 0$

$H_a: (\gamma\tau)_{hj} \neq 0$

Conditions { sphericity conditions; Normal dist. within group

sphericity $V(Y_{ij} - Y_{ij'})$ const $\forall j \forall j'$

Test statistic $F = \dfrac{MSq_{(GT)}}{MSE}$ → error term

If the sphericity condition is violated, we can use MANOVA

MANOVA → multivariate ANOVA

Useful code

↳ dplyr ... group_by( ) → look at $\bar{\mu}$'s

ggplot ...

Assumed fixed

Again if $(\gamma\tau)$ sign, then other terms have no meaning

Feb 18, 2019 Recall multi-sample repeated measure ANOVA

$$Y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \varepsilon_{hij} + \pi_i$$

(Time effect) ☒ Test $H_0: \tau_1 = \tau_2 = \dots = \tau_h = 0$

Test statistics $F = \frac{MS_T}{MSE}$ → within-subject (effect, time effect?)

(Group Effect) ☒ test $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_j = 0$

between subject effects ← Test statistics $F = \frac{MSG}{MS_{S(G)}}$ → MS of subject

R lme(response ~ $\underbrace{\text{group + time + group * time}}_{\text{model}}$,

data = df,

rand effx → induces compound symmetry structure of cov. matrix → random = ~1| ~~Subject~~ Subject,

cor CompSymm(form ~ Month | Subject),

method = "REML")

Time Subject

↳ $\pi_i \sim N(0, \sigma_\pi^2)$ assumes subject fx is rnd → σ_π^2 and (assumes) σ_ε^2

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

↳ σ_π^2 while $Var(Y_{ij}, Y_{ij'}) = \sigma_\pi^2 + \sigma_\varepsilon^2$

Compound Symmetry

↳ $Cov(Y_{ij} \sim Y_{ij'}) = \sigma_\pi^2$

$Var(Y_{ij}, Y_{ij''}) = \sigma_\varepsilon^2$ ⟶

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

☒ Without random subject $\rightarrow \Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_\varepsilon^2 \end{pmatrix}$ WRONG

with random subject

$\hookrightarrow \Sigma = \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & & \sigma_\pi^2 \\ & \ddots & \\ \sigma_\pi^2 & & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix}$

☒ Once R runs $\Rightarrow$ Anova( ) $\rightarrow$ anova table.

If $\gamma\tau$ significant $\Rightarrow$ don't care abt γ & τ

can check residuals, ...

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

$\hookrightarrow$ We might want to test $Y_{hij} = \cancel{\mu} + \gamma_h + \tau_i + \pi_j + (\gamma\tau)_{hij} + \varepsilon_{hij}$

H_a : test $Y_{hij} = \cdots + \pi_j + \cdots \rightarrow$ with subject fx

$\cong H_0$: vs. $Y_{hij} = \cdots + 0 + \cdots \rightarrow$ no subject fx

This is $\equiv$ testing $\sigma_\pi^2 = 0$

If π_j not significant $\rightarrow$ gets absorbed into μ

$\hookrightarrow$ use gls( ) model WITHOUT the random subject term

$\rightarrow$ $\boxed{\text{anova (model 1, model 2)}}$

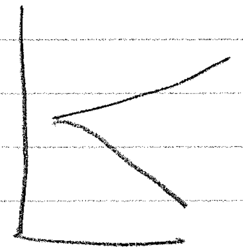
compare models ...

If Significant, then Ha, not Ho, $\rightarrow$ keep π_j in the model

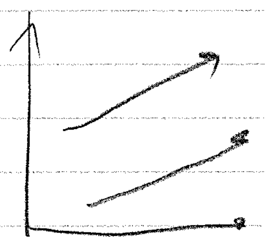
Post hoc comparison of means... dd() → multiple comparison

→ pairwise comparison adjustment

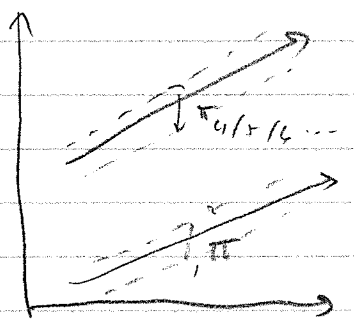
Ex



→ $\gamma\tau$ significant

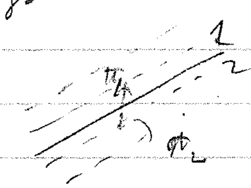


$\gamma\tau$ not significant, even though there is τ effect & γ effect



Test $\sigma^2_\pi = 0$ $\quad \pi \sim N(0, \sigma^2_\pi)$

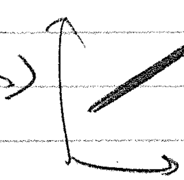
If $\sigma^2_\pi = 0$, then student/school effects don't vary

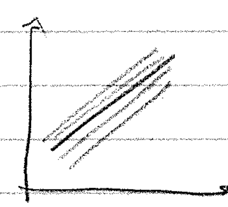


→ if $\sigma^2_\pi \neq 0$

So $\sigma^2_\pi = 0$ $\qquad \sigma^2_\pi \neq 0$

if they start out @ the same level →





If $\sigma^2_\pi = 0$ → good → more degrees of freedom → more power

↳ more df better

Feb 19 2019

Multivariate ANOVA – MANOVA

↳ general class of models for correlated data.

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

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

In general, this is a multivariate response.

- For us ... $Y = \begin{pmatrix} \text{response } t_1 \\ \text{response } t_2 \\ \vdots \\ \text{response } t_n \end{pmatrix}$ → our "responses" are commensurate
  - ↳ Expect strong correlated
  - ↳ measuring the same thing on the same measurement scale.

- 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 simple case of 2 treatments in which subjects are measured at n times each (balanced – same-time)

↳ 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?

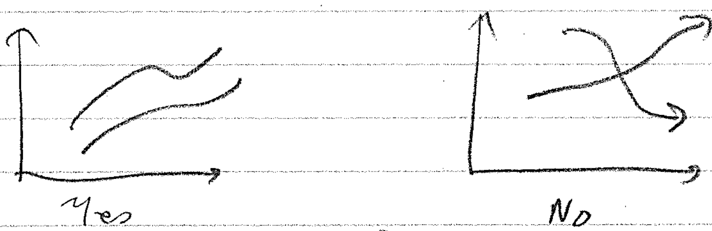
profile analysis

1 → group × time interaction (primary interest)

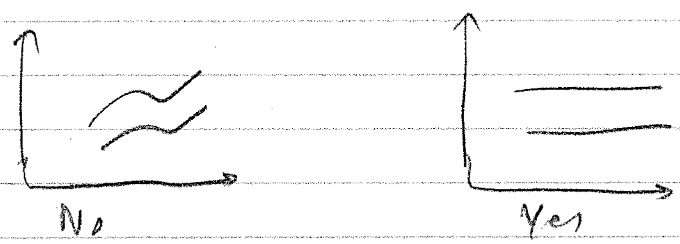
2 → time effect ⁓↗ ?

3 → group effect : ?

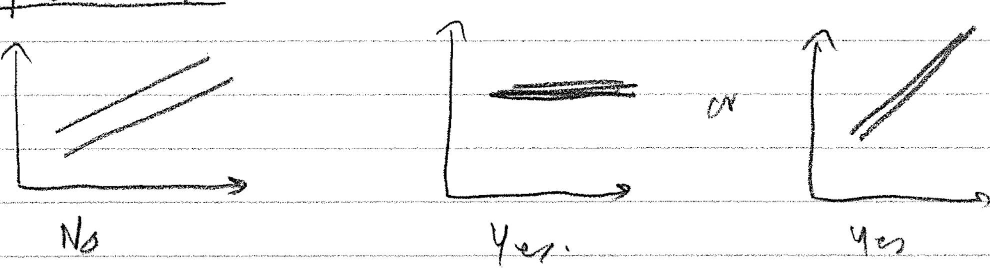
① Are the groups parallel?



② Are the groups flat?



③ Do they overlap?



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

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

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

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

Let

$$V_{i1} = Y_{i1} + Y_{i2} + Y_{i3}$$ → summing response over time

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

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

Note

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

- V_{i1} → tells me abt mean response thru time. nothing about time trends

- V_{i2}, V_{i3} → tells us something abt within-subject time trends.

- $T[\ ]$ → transformation matrix → original response → new vars.

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

the subsequent rows can be different.

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

- The remaining rows address change over time. (time effect), These are many ways to construct the remaining $n-1$ rows.

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 <u>invariant</u> to how we choose to look at change over time & how we characterize it.

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

q → 1 ↳ In our example, finding <u>no</u> group effect in V_{i2} or V_{i3} tells us there is <u>no group × time interaction</u>.

q → 2 → Whether ~~or~~ not there is a time trend of the mean of both V_{i2} & V_{i3} are 0.

q → 3 → 3^{rd} q can be addressed by looking at V_{i1} to see if means differ across groups.

- Repeated measures ANOVA requires

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

- In the book → $\boxed{\underline{Y}_i = \underline{\mu} + \underline{e}_i}$ → error vector

$\underline{\mu} = \mu + \underline{\tau}$ — $n\times1$ vect of means for time points

↳ $n\times1$ vect of time effects.

$$\Sigma_i = \sigma_\pi^2 1_n 1_n' + \sigma_e^2 I_n = \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\pi^2 \\ \sigma_\pi^2 & & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

covariance matrix

Same compound symmetry...

Consider (MANOVA in book) transformation matrix

$$\boxed{\underline{P}\,\underline{Y_i} = \underline{P}_{\mu} + \underline{P}\,\underline{e_i}}$$

$\underline{P}$ 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

$\underline{P} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ & & & \\ & & & \end{pmatrix}$ ← linear trend, ← quadratic trend } both G×T & Time effect

$\underline{PY} = \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} & y_{i4} \\ & & & \\ & & & \end{pmatrix}$

Q_1 are trajectories parallel for each group (G×T)

Q_2 are trajectories flat? (T)

Q_3 ~~are trajectories~~ ~~are trajectories quadratic~~?

So Q_1 →

Q_2 → can be formulated differently?

Q_3 → Group effect are the means the same

Note Polynomial contrast transformations are standard (get from R)

Feb 25, 2019

Hint lm(matrix(time 1, time 2, time 3) ~ Group + 1)

Disadvantage → if pple switch groups → oops

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
≡ Multilevel 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.

$$\boxed{Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + \varepsilon_{ij}}$$

$i \to$ subject $1 \to N$

$j \to$ time point $1 \to n_i$ → different number of time points for each subject.

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

advantage

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

Note ε_{ij} are conditionally independent, on the random effect.
(each subject has their own rnd fx)

2-level modeling strategy

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

Rnd fx } **Level 2**: $b_{0i} = \beta_0 + v_{0i}$ (between subjects)

$b_{1i} = \beta_1 \rightsquigarrow$ constant

Variance of Y_i

$$\boxed{Var(Y_{ij}) = Var(V_{0i}) + Var(\varepsilon_{ij})}$$

independent

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

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

Subjects are independent $\boxed{Cov(Y_{ij}, Y_{i'j}) = 0}$

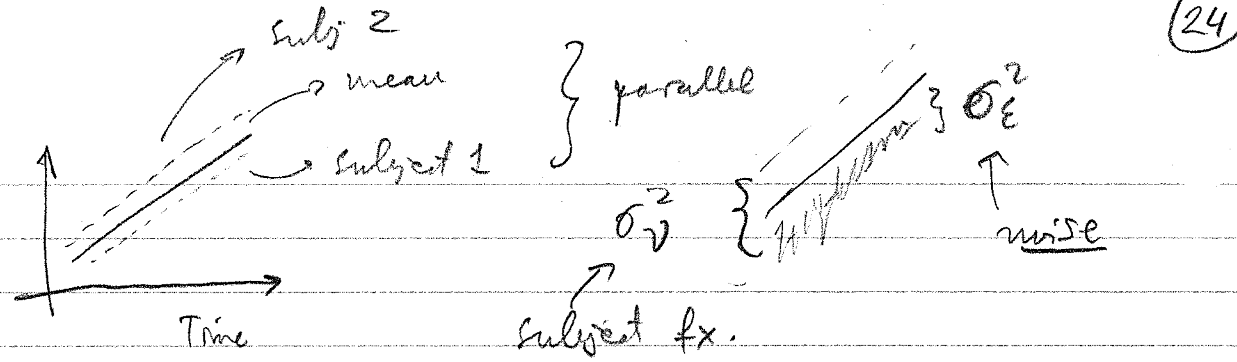
Get
→ Compound Symmetry (requires balance)

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

← intraclass corr.
↳ % of var due to ~~between subject fx~~
observed
↳ differences in subject

ideally, ICC large.
→ model good (noise small)

Ex



Hypothesis Testing for Model Parameters

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

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

Caution if $SE(\hat{\theta})$ is close to 0, theses tests are not reliable

(2) Likelihood ratio tests (LRT) (works better)

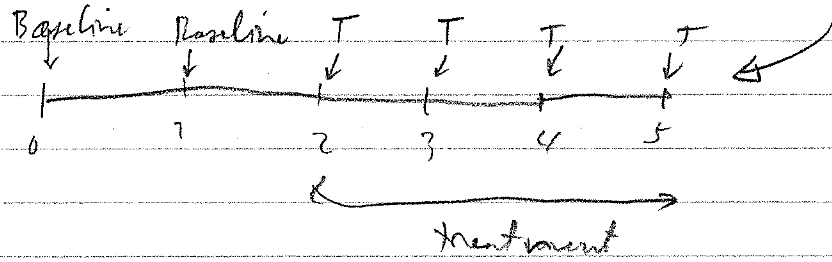
↳ particularly good at testing nested model.

(3) Can also be used to test covariance Structure

↳ divide p-value in half to avoid Type 2.

$\boxed{\text{Ex}}$ → Dataset in Book (1977)

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



Response : Score on Beck Depression Inventory.

<u>Note</u> MANOVA doesn't handle missing data

Compound Symmetry → subj corr constant.

<u>Feb 27, 2011</u>

<u>Recall</u> Random Intercept Model:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + V_{0i} + \varepsilon_{ij}$$

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

V_{0i}: subject effect $\sim N(0, \sigma_v^2)$

<u>Note</u>

V_{0i} + ε_{ij} are independent → $Var = Var(V_{0i}) + Var(\varepsilon_{ij})$

$E[Y_{ij} | V_{0i}] = \beta_0 + \beta_1 t_{ij} + V_{0i} + 0$ ~ $\underbrace{E[\varepsilon_{ij} | V_{0i}]}_{\text{just } 0}$

conditional expectation of Y_{ij} given V_{0i}

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

conditional expectation of $Y_{ij} | V_{0i}$

expectation averaged over population

<u>In fact</u> $E[Y_{ij}] = E\left[E[Y_{ij} | V_{0i}]\right]$

$\boxed{\text{Random Intercept models can handle unbalanced and missing data (some types of) — unlike MANOVA...}}$

Random Intercept Models have an <u>induced compound symmetry structure</u>.

<u>Recall</u> $Var(Y_{ij}) = Var[\underbrace{\beta_0 + \beta_1 t_{ij}}_{\text{fixed}} + V_{0i} + \varepsilon_{ij}] = Var(V_{0i} + \varepsilon_{ij})$

<u>Now</u> $Var(Y_{ij}) = Var(V_{0i} + \varepsilon_{ij})$ by independence

$= Var(V_{0i}) + Var(\varepsilon_{ij})$

$$\boxed{Var(Y_{ij}) = \sigma_V^2 + \sigma_\varepsilon^2}$$

variance is the same for every time point

- What $Cov(Y_{ij}, Y_{ij'})$ → same subj @ different time pts

$= Cov([\beta_0 + \beta_1 t_{ij} + V_{0i} + \varepsilon_{ij}], [\beta_0 + \beta_1 t_{ij'} + V_{0i} + \varepsilon_{ij'}])$

$= Cov([V_{0i} + \varepsilon_{ij}], [V_{0i} + \varepsilon_{ij'}])$

$= \underbrace{Cov(V_{0i}, V_{0i})} + \underbrace{Cov(\varepsilon_{ij'}, V_{0i})}_{0} + \underbrace{Cov(\varepsilon_{ij'}, V_{0i})}_{0} + \underbrace{Cov(\varepsilon_{ij}, \varepsilon_{ij'})}$ ← $Var(\varepsilon_{ij}, \varepsilon_{ij'})$

$= Var(V_{0i}) + \underbrace{Var(\varepsilon_{ij}, \varepsilon_{ij'})}$

$= Var(V_{0i}) + 0$

$= Var(V_{0i}) = \sigma_V^2$

So $Cov(Y_{ij}, Y_{ij'}) = \sigma_V^2$ ← constant

⇒ Compound symmetry in covariance matrix. $\begin{pmatrix} \sigma_V^2 + \sigma_\varepsilon^2 & & \sigma_V^2 \\ & \ddots & \\ \sigma_V^2 & & \sigma_V^2 + \sigma_\varepsilon^2 \end{pmatrix}$

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

↳ then our model (comp sym) might not work

☒ Side note

↳ In R, when we say "random = ~1|id"

id → is a var that uniquely identifies subjects.

id
1
1 } same subject
1
⋮
2
2 } same subject
2
⋮

☒ ~~How~~ How we code time matters... How we code time can affect the intercept.

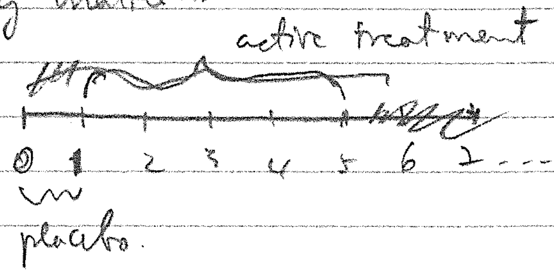
☒ VarCorr(model) → gives σ_v^2 (intercept) (due to subj)
σ_ε^2 (residual) (error)

☒ A "wrong" model that doesn't take into account will give a single stdv σ_ε^2 that is

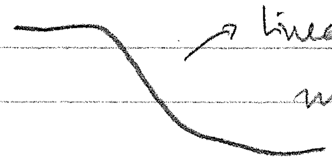
$$\underset{(\text{wrg})}{\sigma_\varepsilon^2} = \sigma_\varepsilon^2 + \sigma_v^2$$

☒ The slope & intercept of ME model & wrong model might not be large, but they matter...

☒ Look at how we coded time...

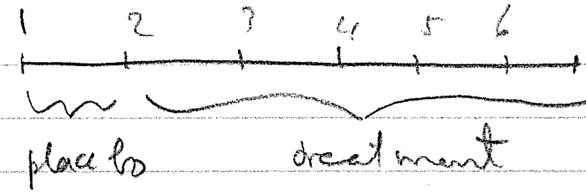


Expect



→ linear might not work...

What if we code time as



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

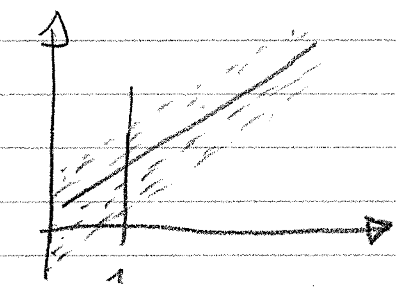
→ So the intercept ($t=0$) is before the study began

⇒ Extrapolating.

→ BAD.

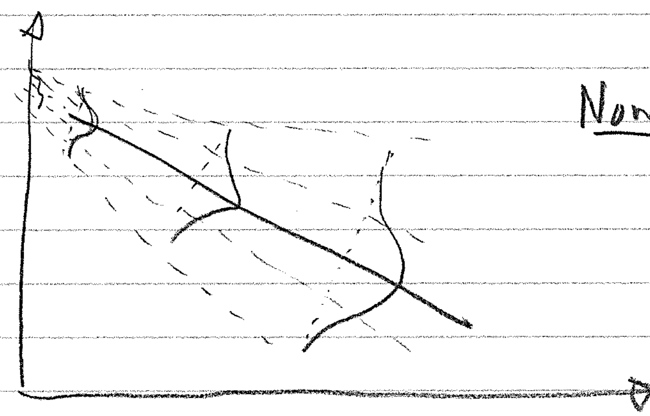
⇒ Change the subject-specific random effects.

Ex



→ 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

Mar 1, 2019

Last time ~ random intercept alone doesn't fit data well.

Recall $\boxed{\text{2-level model specification.}}$

note, we're not really fitting layer by layer

Level 1 $\quad Y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$

Level 2 $\quad b_{0i} = \beta_0 + v_{0i}$ and $\underbrace{b_{1i} = \beta_1 + v_{1i}}_{\text{slope}}$

intercept

just sth constant

random subject effect

So

$$\begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_v\right]$$

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

$Cov(v_0, v_1)$

$\boxed{\text{1-level formulation}}$

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

How many parameters do we estimate?

$\{\beta_0, \beta_1, \sigma^2_{v_0}, \sigma^2_{v_1}, \sigma^2_\varepsilon, \sigma_{v_0 v_1}\}$ $\rightarrow Cov(v_0, v_1)$

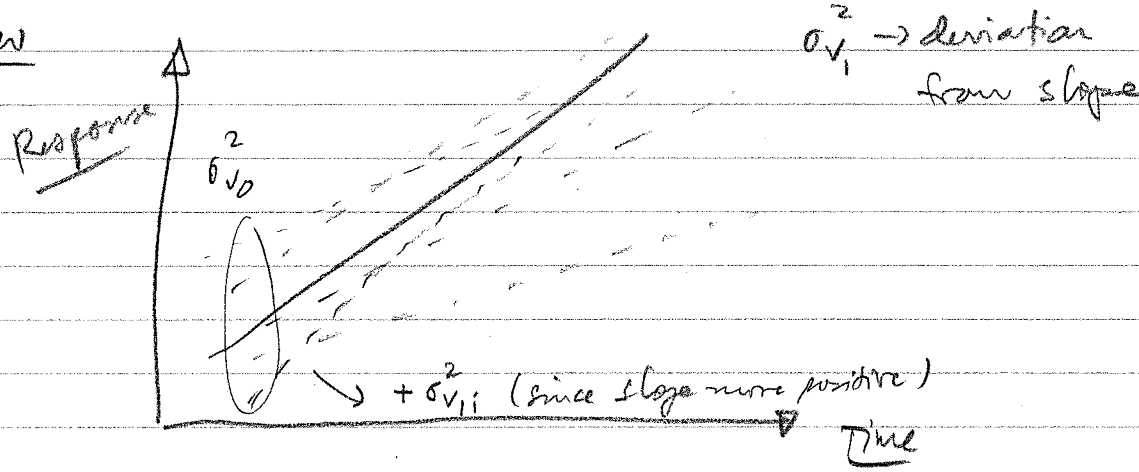
random error (var of rnd error)

var of intercept (subj-specific offset at $t=0$)

var of slope (subj-specific slope effect)

But if $\sigma_{v_0 v_1}$ is <u>positive</u> it tells us that those with an <u>high</u> intercept tend to have <u>steep</u> slope
<u>positive</u>

If $\sigma_{v_0 v_1}$ negative then the opposite occurs.



<u>Note</u> We don't have compound symmetry anymore.

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

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

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

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

<u>Consider</u> $Cov(Y_{ij}, Y_{ij'}) = Cov[(\beta_0 + \dots \varepsilon_{ij}), (\beta_0 + \dots + \varepsilon_{ij'})]$

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

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

$$= \cancel{\sigma_{v_0} + t_{ij'} \sigma_{v_0 v_1}}$$

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

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

\# parameters = 4

\# of elements in Σ_i = covariance matrix of Y_{ij}

$$\boxed{\frac{t(t+1)}{2}}$$

instead of ~1|id, in R I say "~time|id"

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

Rnd Int + slope: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

nested

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

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 $-2(\ln L_0 - \ln L_A) \sim_{H_0} \chi^2$, df = # of parameters being introduced

df = difference between # of parameters being estimated

Rusby ex test statistic $= -2(\ln(-1144.590) - (-1109.579))$

$= 66.75 \sim \chi^2_{df = 6-4 = 2}$

Again,

H_0: estimate 4 params

H_a: estimate 6 params

Mar 4, 2019 Recall model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

- Matrix formulation

$$\boxed{\underset{[n_i \times 1]}{\underline{Y_i}} = \underset{[n_i \times p][p \times 1]}{\underline{X_i}\,\underline{\beta}} + \underset{[n_i \times r][r \times 1]}{\underline{Z_i}\,\underline{V_i}} + \underset{[n_i \times 1]}{\underline{\varepsilon_i}}}$$

vector of response for subject i

(0) $[X_i] \rightarrow$ design matrix for fixed effects. (FIXED EFFECTS,

(1) $[X_i][\beta] \equiv \beta_0 + \beta_1 t_{ij}$

(2) $[\beta] \rightarrow$ vector of population level coefs that relate fixed effects to response

(3) $[Z_i] \rightarrow$ design matrix for RANDOM effects

(4) $[V_i] \rightarrow$ vector of Random effects coefficients

(5) $[\varepsilon_{ij}] \rightarrow$ Random ~~effect~~ errors vector

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

- In Riesby example

$[Z_i] = [X_i]$

$$[X_i] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \qquad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad [V_i] = \begin{bmatrix} V_{0i} \\ V_{1i} \end{bmatrix}$$

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{i\,?} \end{pmatrix}$$

☒ Note $\underline{X}, \underline{Z}, \underline{\varepsilon}$ can have differents number of rows across individuals

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

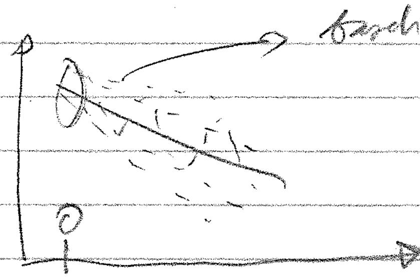
☒ Note $\beta_0, \beta_1 \to$ fixed effects

$v_{0i}, v_{1i} \to$ random effects $\sim MVN\left(\begin{pmatrix}0\\0\end{pmatrix}, \underline{\Sigma}_v\right)$

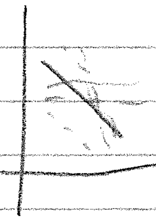
☒ what happens if we code time differently? (Recentered time)

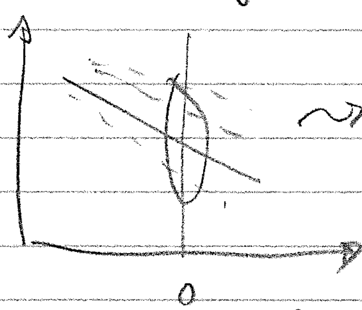
Fixed effects

	Est.	SE
Int	23.58	0.545
Slope (wk)	-2.38	0.209



☒ If we define $t=0$ at middle of study, then
$t_{ij} \to t_{ij} - 2.5$





$\leadsto$ $\boxed{\text{different var, lower intercept.}}$

Fixed Effects $\to$ as expected

	Est	SE
Int	17.63	0.562
Slope	-2.38	0.209

interaction

$+ \beta_3 Dx_i \cdot t_{ij})$

Now, talk about DIAGNOSIS

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 Dx_i}_{} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$$

$$\text{fixed}, \ [X_i] = \begin{pmatrix} 1 & 0 & Dx_i \\ 1 & 1 & Dx_i \\ \vdots & 2 & \vdots \\ \vdots & \vdots & \\ & (t) & (diag) \end{pmatrix}$$

But Dx_i alone assumes fixed → add interaction

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

$\boxed{\beta_3 \ Dx_i \times t_{ij}}$ → tests whether the trajectories through time is different by diagnosis.

So $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 Dx_i + \beta_3 Dx_i \times t_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

Level 1 $\quad Y_{ij} = b_0 + b_1 t_{ij} + b_{2i} Dx_i + b_{3i} Dx_i \times t_{ij} + \varepsilon_{ij}$

Level 2

$b_{0i} = \beta_0 + v_{0i}$ (fixed + random)

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

$b_{2i} = \beta_2$

$b_{3i} = \beta_3$

Note Dx_i is constant within individual (between subj)

t_{ij} not constant within an individual (within-subj)

Mar 6, 2019

Last time, we were looking at diagnosis Dx

Fixed effects: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 Dx_i + \beta_3 D_{ij} \cdot t_{ij}$
Rand effects $+ v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

In matrix $\underline{Y_i} = \underset{\text{fixed}}{\underline{X_i}} \vec{\beta} + \underset{\text{rnd}}{\underline{z_i}} \underline{v_i} + \varepsilon_{ij}$

where

$$\underline{X_i} = \begin{pmatrix} 1 & 0 & Dx_i & Dx_i \cdot (0) \\ 1 & 1 & Dx_i & Dx_i (1) \\ \vdots & 2 & \vdots & \vdots \\ & 3 & & \\ 1 & 4 & Dx_i & Dx_i \cdot (n) \\ & \vdots & & \end{pmatrix} \qquad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

$$z_i = \underline{\text{Random}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & \vdots \end{pmatrix} \qquad \underline{v_i} = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \end{pmatrix}$$

Compare (H_a) versus H_0: the model without diagnosis
$Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

Note smaller model is
NESTED in the bigger model → likelihood ratio test

Riesby data ☆ Use tapply( ) for summary statistics...

random ~ week | id → slope + intercept random
~ 1 | id → intercept random

Run lme → $\boxed{+\ln L_A = -1107.45}$

trajectory thru time → insignificant of $Dx_i \cdot t_{ij}$ coef
is insignificant...

Difference between models → degrees of freedom.

H_a: $\beta_0, \beta_1, \underbrace{\beta_2, \beta_3}$, $\sigma^2_{v_0}, \sigma^2_{v_1}, \sigma_{v_0 v_1}, \sigma^2_\varepsilon$ ‖ H_0: $\beta_0, \beta_1, \sigma_{v_0}, \sigma^2_{v_1}, \sigma^2_{v_0 v_1}$

$2 = 8-6$

Anyways, $\ln(L_0) = -1109.519$

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

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

Simpler model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$

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

$\rightarrow$ 4: $\sigma^2_{v_0}$, $\sigma^2_{v_1}$, $\sigma_{v_0 v_1}$, σ^2_{ε}

Need to estimate Cov Matrix $\rightarrow$ $\boxed{\hat{\Sigma}_i = \underline{Z}_i \cdot \Sigma_v + \underline{Z}_i^T + \sigma^2_{\varepsilon} \cdot I_{n_i}}$

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

Σ_i

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

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

$\rightarrow$ Before we derived $\rightarrow$ $\boxed{Var(Y_{ij}) = \sigma^2_{v_0} + t^2_{ij}\, \sigma^2_{v_1} + 2t_{ij}\, \sigma_{v_0 v_1} + \sigma^2_{\varepsilon}}$

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

How to test? $\rightarrow$ generate $\hat{\Sigma}_i$ from σ^2_{ε}, $\underline{Z}_i$

get Cov matrix get VarCov $\rightarrow$ $\underline{\Sigma_v}$

$\hat{\Sigma}_i = Z$ %*% Σ_r %*% $Z^T + \sigma^2_\varepsilon \cdot I_{n_i}$

To get actual $\Sigma_i \rightarrow$ cor( ) (In R code)

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

Note: $Z_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & 2 \\ \vdots & \vdots \end{pmatrix}$ only true for balanced.

If time different $\Rightarrow$ $Z_i = \begin{pmatrix} \end{pmatrix}$ unique to individual

→ just do it again. ~~But~~ and again for different i.

→ So we can generalize, so long as # of measurements same

Mar 8, 2019

$\boxed{\text{TIME - VARYING COVARIATES} \rightarrow \text{PREDICTORS OR EXP. VARS}}$

- Generally speaking, some common covariates don't change thru time: sex, biomarkers
- but some do → risk factors (blood pressure) (treatment in crossover study)
- We can easily incorporate time-varying covariates into model → we can place them into level 2 or we can think of putting them into the design matrix

Ex

$$\begin{array}{ccc} \beta_0 & \beta_1 & \beta_2 \end{array}$$
$$\begin{bmatrix} 1 & 0 & x_{i1} \\ 1 & 1 & x_{i2} \\ 1 & 2 & \vdots \\ 1 & \vdots & \vdots \\ \vdots & & \end{bmatrix}$$

intercept → time, covariate, time varying

Back to Riesby data

The original data contains 2 varying time covariate

(1) Blood plasma concentration of imipramine

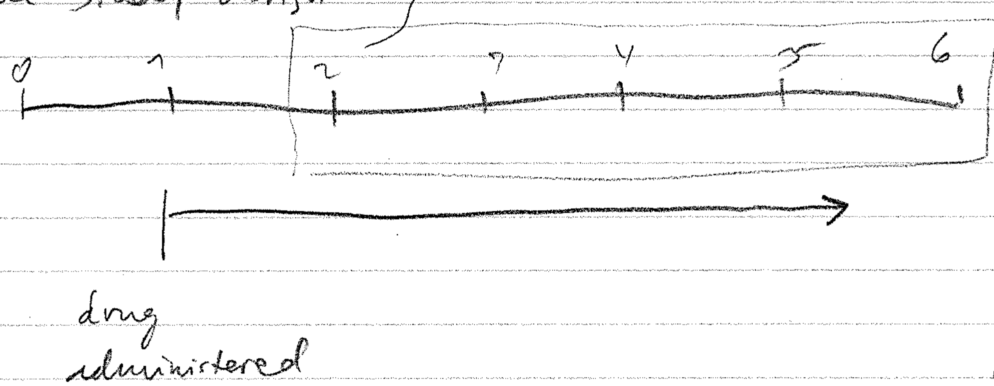
(2) Plasma concentration of desipramine ← transform

Note imipramine biotransform into desipramine.

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

- Recall study design

1st measure at which we expect fx.



data set with plasma concentration only contains these 4 time points.

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

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

Main Effects Model

imi-concentration

desip concentration

change in depression score

$$Y_{ij} - Y_{i0} = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln I_{ij} + \beta_3 \ln D_{ij} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}$$

response

baseline

- Need to recode time

$$\begin{array}{cccc} 2 & 3 & 4 & 5 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 0 & 1 & 2 & 3 \end{array} \rightarrow \text{to get meaningful intercept}$$

$\boxed{\text{How do we want to code the plasma concentration?}}$

Think about β_0 → represents the expected dosage score at $t=0$ and when plasma concentrations are effectively zero..

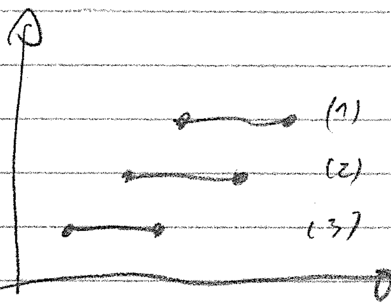
↳ makes more sense to center ln(plasma concentration)

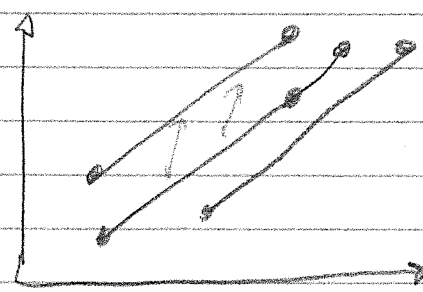
→ change β_0. β_0 now describes the expected dosage score for someone at $t=0$ with average concentration ~~[illegible]~~ levels.

$$\boxed{\begin{array}{l} \ln(I_{ij}) \rightarrow \ln I_{c_{ij}} = \ln I_{ij} - \ln \bar{I}_j \\ \ln(D_{ij}) \rightarrow \ln D_{c_{ij}} = \ln D_{ij} - \ln \bar{D}_j \end{array}}$$

Note when we center a time-varying covariate, we are assuming ~~fx~~ between subject & within-subject are the same.

(no within subj) →





between-subj fx
no within-subj fx

within-subj fx
no between-subj fx

mean across all subjects

$$\ln I_{ij} = \overline{\ln I_{ij}} + (\ln I_{ij} - \overline{\ln I_{ij}})$$

$$\ln D_{ij} = \underbrace{\overline{\ln D_{ij}}}_{\text{between}} + \underbrace{(\ln D_{ij} - \overline{\ln D_{ij}})}_{\text{within}}$$

- We could separate the between & within pieces in the model. → add terms

$$Y_{ij} - Y_{i0} = \beta_0 + \beta_1 t_{ij} + \beta_2 (\ln I_{ij} - \overline{\ln I}) + \beta_3 (\ln D_{ij} - \overline{\ln D}) + \beta_4 \overline{\ln I} + \beta_5 \overline{\ln D} + V_{0i} + V_{1i} t_{ij} + \varepsilon_{ij}$$

Nov 11, 2019 — Recall Time-varying Covariates

change from baseline

$$(Y_{ij} - Y_{i0}) = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln I_{ij} + \beta_3 \ln D_{ij} + V_{0i} + V_{1j} t_{ij} + \varepsilon_{ij}$$

Now, $\ln I_{ij}$, $\ln D_{ij}$ } both have been "centered" by subtracting the overall means of $\ln I$ and $\ln D$ across all subjects / observations.

$\boxed{\text{Between vs Within Effects of Time-varying Covariates}}$

We can separate each chemical's effect into between and within subject components.

varies BETWEEN SUBJECT

$$\ln I_{ij} = \ln \overline{I}_{i\cdot} + (\ln I_{ij} - \ln \overline{I}_i)$$

deviation at time of subj i from $\overline{\ln I}$ ↳ WITHIN PIECE

averaged across time, for each individual (i) → varies. This is the BETWEEN SUBJECT averaging over j

If we include the between & within subject preds separately,

(2)
$$\boxed{\begin{aligned}(Y_{ij} - Y_{i0}) = \beta_0 + \beta_1 t_{ij} &+ \beta_2 (\ln I_{ij} - \overline{\ln I_{i\cdot}}) \\ &+ \beta_3 (\ln D_{ij} - \overline{\ln D_{i\cdot}}) + \beta_4 \overline{\ln I_{i\cdot}} \\ &+ \beta_5 \overline{\ln D_{i\cdot}} + v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij}\end{aligned}}$$

overall effect

If $\beta_2 = \beta_4$, then $\beta_2 (\ln I_{ij} - \overline{\ln I_{i\cdot}}) + \beta_4 \overline{\ln I_{i\cdot}}$

$$= \beta_2 (\ln I_{ij} - \overline{\ln I_{i\cdot}} + \overline{\ln I_{i\cdot}})$$

$$= \beta_2 \ln I_{ij}$$

This assumes that the between and within effects are equal.

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

library(dplyr)

Then merge into original df → id ... ... mean

id		mean
1	replicated	$\bar{I}$
1		$\bar{I}$
1		$\bar{I}$

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

Note $\beta_1 \ldots \beta_5$ → fixed effects → no "i" index
$v_{0i}, v_{1i} \ldots$ → random fx → "i" index.

<u>Compare models</u> without splitting (W-B) and with.

<u>Model 1</u> $\ln L_0 = -749$	<u>Model 2</u> $\ln L_1 = -747$
$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$
$\sigma^2_{v_0}, \sigma^2_{v_1}, \sigma_{v_0 v_1}, \sigma^2_{\varepsilon}$	$\sigma^2_{v_0}, \sigma^2_{v_1}, \sigma_{v_0 v_1}, \sigma^2_{\varepsilon}$

<u>Nested F-test</u> $\longrightarrow$ $\chi^2 = -2(\ln L_0 - \ln L_a) = 3.08 \overset{H_0}{\sim}$ 1X

p-value = 0.275 → null model is adequate

↳ seperating out within - between subjects is probably not something we want to do...

Interaction model → between time-varying covariates and time itself

$$(Y_{ij} - Y_{i0}) = \beta_0 + \beta_1 t_{ij} + \beta_2 \ln I_{ij} + \beta_3 \ln D_{ij} + \beta_4 \ln I_{ij} t_{ij}$$ (int)

$$+ \beta_5 \ln D_{ij} t_{ij} + \underbrace{V_{0i} + V_{1i} t_{ij} + \varepsilon_{ij}}_{\text{random}}$$

(int)

☒ (β_0) is the mean week 2 change in depression score for patients with average log chemical levels

☒ (β_1) is the average weekly change in change scores for patients with average drug levels.

☒ (β_2) → change score difference for a one unit ↑ in ln IMI at week 2

☒ (β_3) ⇒ ..... in ln DMI at week 2

☒ (β_4) & (β_5) → indicate per week <u>change</u> in <u>drug effect</u> on depression change scores

↳ compare models → interaction model is better

<u>Note</u> df = 2 because Δ # parameters = 2 -

Nov 17, 2018

$\boxed{\text{Estimation}}$ Just get a general feel about … how we estimate β variance parameters, and the individual → specific random effects.

Two types of estimation used

1. Likelihood-based $\xrightarrow{est}$ β, Σ_σ, σ_ε^2
2. Empirical Bayes $\xrightarrow{est}$ v_i (random fx intercept/slope)

Bayesian estimation → 1. make a guess about distribution of a parameter.
prior distribution

2. Look at the data
3. Update the guess abt distribution (posterior distribution)

→ Involves iterating, Goes from estimating β, Σ_v, σ_ε^2 and Estimating v_i

→ $\Sigma_{v_i|y_i}$ → covariance matrix of random effects.

Iterate until converge on a solution.

Two potential problems → (I) Exceed maximum of allowed iterations and/or ~~?~~

And/or (II) Can't get Δ in iterations to be small enough → don't converge.

Estimation method

$\boxed{\text{ML versus REML}}$

ML: Maximum likelihood

REML: Restricted Maximum likelihood.

ML: produces estimates for variance parameters (Σ_v, σ_ε^2) that are biased. How biased depends on # of independent and # of parameters estimating.

- Consider a standard multi regression model.
  - $\boxed{\hat{\sigma}^2_{ML} = \frac{SSE}{N}}$
  - $\boxed{\hat{\sigma}^2_{REML} = \frac{SSE}{N-(p+1)}}$

the intercept. Type 1 ↑↑↑

p # of predictors

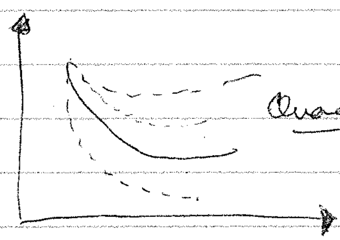
ML → gives estimates of the variance that too small.
→ Increase in type I error rate
→ Confidence intervals aren't as "confident" as they should be.

- Why don't we use REML all the time?
  ↳ Because we can't compare Nested models using REML
  ↳ Need ML for this...

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

$\boxed{\text{Chapter 5: CURVILINEAR TREND}}$

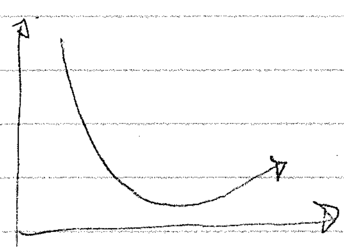
Ex



Quadratic

Fixed effects Linear + Quadratic.
Random effects Linear + Quadratic.

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

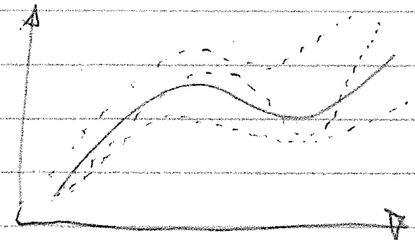


change in direction of trend when 1^{st} derivative is zero.

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

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

[R] Can also have cubic trend



$\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + v_{0i}$

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

Nov 18, 2019

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

Consider model → $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + v_{0i} + v_{1i} t_{ij} + v_{2i} t_{ij}^2 + \varepsilon_{ij}$

How many parameters do we need to estimate?

↳ $\beta_0, \beta_1, \boxed{\beta_2}$, ~~$\beta$~~ → fixed effects

$\sigma_\varepsilon^2, \sigma_{v_0}^2, \sigma_{v_1}^2, \boxed{\sigma_{v_2}^2}, \sigma_{v_0 v_1}, \boxed{\sigma_{v_1 v_2}}, \boxed{\sigma_{v_2 v_0}}$ ← rand ~~efx~~ } 10

$$\Sigma_v = \begin{bmatrix} \sigma_{v_0}^2 & & \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 & \\ \sigma_{v_0 v_2} & \sigma_{v_1 v_2} & \sigma_{v_2}^2 \end{bmatrix}$$

If we consider only linear model → have $\boxed{6}$ parameters

↳ χ^2 comparison has $df = 4 = 10 - 6$.

[R] y ~ week + I(week^2)

- The quadratic term for time is just like time interacting with itself → $t_{ij} \times t_{ij}$

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

$\boxed{R}$ → merge (df 1, df 2, by = 0) → row-by-row merge

Mar 18, 2019

COVARIANCE PATTERN MODEL

→ Extension of multiple regression

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i$$

$$e_i \sim N(0, \sigma_\varepsilon^2)$$

Problem σ_ε^2 is fixed

$$\rightarrow \Sigma_\varepsilon = \sigma^2 \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{pmatrix} = \sigma^2 I$$

which assumes independence across individuals and within individuals.

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

→ In the mixed model setting → $\underbrace{\sigma_{v_0}^2 \cdots \text{ and } \sigma_{v_k}^2}_{\text{within-subj}}$ σ_ε^2 ← between subj

$\boxed{\text{Covariance Pattern Models}}$ → 2 steps

1. Modeling covariance
2. Modeling the mean response

Potential Problem → we might not have enough df to estimate all param

Think about the most general covariance pattern model

$$\boxed{\underline{Y}_i = \underline{X}_i\underline{\beta} + \underline{e}_i} \qquad \underline{e}_i \sim N\left[\underline{0}, \Sigma_i\right]$$

if $\underline{Y}_i$ is $n_i \times 1$, $i = 1, 2, \ldots, N$ subj

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

Ex

Before

$$\begin{array}{cc} & \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & 2 \\ \vdots & \vdots \\ 1 & 5 \end{bmatrix} \\ \underline{X}_i = & \begin{array}{cc} \uparrow & \uparrow \\ \text{int} & \text{time} \end{array} \end{array}$$

→ assumes a linear relationship between time ~~and response~~ points and response

indicators

Now,

$$\begin{array}{cc} & \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \underline{X}_i = & \begin{array}{cccc} \uparrow & & & \\ \text{int} & I(t_{ij}=0) & I(t_{ij}=1) \cdots & I(t_{ij}=4) \end{array} \end{array}$$

→ design matrix

⇒ imposes ~~that~~ no structure at all on how times of measurement relate to Y.

- Before > Now → less params to be estimated
- Now > Before → doesn't impose time structure

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

☐ what about covariance of $\underline{e}_i$?

If we have p measurements taken on each individual, how many variance / covariance are there in $\underline{\Sigma}_i$?

$$\boxed{q = \underbrace{p}_{\text{variances}} + \underbrace{\binom{p}{2}}_{\text{covariance}} = \frac{p(p+1)}{2}}$$

This is what MANOVA does...

when $p = 2 \Rightarrow q = 3$
$p = 4 \Rightarrow q = 10$
$p = 10 \Rightarrow q = 55$

☐ If we place no structure on $\underline{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{p1} \\ \sigma_{12}^2 & \ddots & \vdots \\ \sigma_{13} & & \\ \vdots & & \sigma_p^2 \\ \sigma_{1p} & & \end{pmatrix}$

$\boxed{\text{Compound Symmetry}}$ (?)

↳ $\underline{q = 2}$

$$\underline{\Sigma}_i = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \cdots & \sigma_1^2 \\ \sigma_1^2 & \ddots & \vdots \\ \sigma_1^2 & & \vdots \\ \vdots & & \\ \sigma_1^2 & \cdots \ \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}$$

↳ ~~df~~ save df,
but price → 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 on different extremes in terms of placing structure on $\underline{\Sigma}_i$.

↳ and even random-intercept and LMM.

First-order Autoregressive

↳ covariance between two time points j and j' is given by

$$\boxed{\sigma_{jj'} = \sigma^2 \underbrace{\rho^{|j-j'|}}_{\text{correlation}}}$$

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho & & \rho^{n-1} \\ \rho & 1 & & \\ \vdots & & \ddots & \\ & & & \rho \\ \rho^{n-1} & \cdots & \rho & 1 \end{pmatrix}$$

→ also assumes variances are all equal.

$q = 2$ (σ and ρ)

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

Toeplitz structure

$$\Sigma_i = \begin{pmatrix} \theta_1 & & & \theta_n \\ \theta_2 & \theta_1 & & \\ \vdots & & \theta_2 & \\ \theta_n & \cdots & \theta_2 & \theta_1 \end{pmatrix}$$

$$\boxed{q = n}$$

\# measurements

covariance between 2 time points:

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

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

$q < n$

Exponential Structure

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & & & \\ \rho^{\cdots} & 1 & & \\ \rho^{\cdots} & & \ddots & \\ \rho^{\cdots} & & & 1 \\ \rho & & & \end{pmatrix}$$

$q = 2$

$\rho^{\cdots} = \rho^{|j'-j|}$ ← exp

time pts don't have to be equally spaced

$\boxed{\text{Review}}$ Can bring 1 sided sheet of notes...

- Fundamentals of Longitudinal Data (within-subject, datasets...
- Visualize longitudinal data ("spaghetti plot...)
- Repeated Measures ANOVA (time effect + <u>compound symmetry</u> + no missing data + Balanced
- MANOVA → entire set of longitudinal observation at once, transformation matrices, questions that can be answered with profile analysis, no structure on with-subj' M_{cov}
- <u>Mixed Models</u> → Fixed and random effects, estimates (β_i, variance components σ_V^2, ...)
  → much more flexible in terms of structure of Σ_i
  (ex: random int, rnd int + slope, higher order...)

rnd int = compound sym

Vare

- Estimation: ① 2-step iteration: between estimates of β_i, Σ_V and V_i (need to know basics)
  ② ML vs REML. ML gives biased est of variance terms, especially $\hat{\sigma}_\varepsilon^2$
  REML is unbiased. <u>Catch</u>: We need ML to conduct likelihood ratio test.

ε_{ij} & V_{0i} independent
ε_{ij}, $\varepsilon_{ij'}$ independent

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

$$\begin{cases} E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} \\ E(Y_{ij} | V_{0i}) = \beta_0 + \beta_1 * t_{ij} + V_{0i} \end{cases}$$

getVarCov() → cov matrix
VarCorr → correlation

<u>April 1, 2019</u> <u>Recall</u> Covariance pattern models → $\boxed{\text{need balanced data}}$

need same # of time pts

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

most expens → ② Compound symmetry → # params = 2 on each subj'

↳ assumes constant variance, covariances...

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

$\rho_{ij'} = \rho^{|j'-j|}$ } → need equally spaced measurements

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

All nested within unstructured when compare...

④ Toeplitz (Banded) $q = k$

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

$1-2: \theta_{12}$
$1-3: \theta_{13} \ldots$ } can set threshold $\theta_{ij} = 0$ for some ij

↳ more flexible.

⑤ Exponential $q = 2$. Assumes constant variance. Assume the same correlation for each time separation

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

↳ does not require equally spaced - measurements.

- How do we specify these in R?

$\boxed{Ex}$ 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 trtmt 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 (WPSS) for Episodic Disorder (1-6) → lower = better

Grp 0 : TCA - No drug
Grp 1 : No drug - TCA

week ① ② ③ ④ ⑤ ⑥

−2.5 −1.5 −0.5 0.5 1.5 2.5 → linear

−0.5 0 0.5 0.5 0 −0.5 (quad)

Start

Model	$-\ln L$	q
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 (gls)

unstructured → cor Symm (form = ~ week/id), weights ...

needs to be a numeric count + measurement occ. 1, 2

To get unstructured matrix ⇒ cor Symm (form = ~ week / id), weights = varIdent (form ~ 1| week)

↳ allows variances to be unique at each time pts.

Comp Symm → Cor Comp Symm (form = ~ week / id), "

April 3, 2019 $\boxed{\text{Back data}}$ Recall

Model	$-\ln L$	q
Unstructured	−472.9676	21 ← 6 time pts
Comp Symm	−592.9134	2
Comp Symm w/ Heterogeneous Var	−586.569	7
Autoregressive (1)	−498.174	2
Toeplitz	−494.433	6

$$\text{ex} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_3\sigma_1 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \\ \rho^2\sigma_3\sigma_1 & \cdot & \sigma_3^2 \end{pmatrix} \leftarrow \text{Comp Symm w/ Heterogeneous Var}$$

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)}\sqrt{V(X_2)}} \rightarrow \text{different}$$

Comp Symm with Heterogeneous Var

Comp Symm

$$\boxed{\begin{array}{l}\text{corCompSymm (form = \textasciitilde week | id),} \\ \text{weights = varIdent (form=\textasciitilde 1 | week)}\end{array}}$$

Hetero. Var

Now, compare each structure to unstructured
All cov structures are nested in unstructured

Note abt Toeplitz in R → not part of the built-in R cov structures
But we can trick it into fitting one by using ARMA model

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

p → largest time gap

q = 0

Comp Symm to Unstructured
CS with Hetero var to Unstructured...

unstructured

Model	$-\ln L$	q	$-2(\ln L_0 - \ln L_A)$	q_A	p
CS		2	239.92	21	<<
CS - Hetero		7	227.02	21	<<
AR(1)		2	50.41	21	0.0001
Toeplitz		6	... [illegible]	21	0.0011

↳ when doing these tests, take p-value and divide by 2
→ avoid Type 1 error.

In each case → reject null hyp → unstructured model necessary
→ conclude need unstructured structure

↳ choose unstructured

Done with covariance modeling...

now back to model mean

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

Since high order term significant
→ keep all terms

↳ get model ⇒ run again with REML → get unbiased estimates...

April 5, 2019

think back to mixed models...

$$\underline{Y}_i = \underline{X}_i \underline{\beta} + \underline{Z}_i \underline{V} + \underline{\epsilon}_i \qquad \underline{v}_i \sim N[\underline{0}, \underline{\Sigma}_v]$$

fixed (for $\underline{X}_i \underline{\beta}$), random (for $\underline{Z}_i \underline{V}$)

$$e_i \sim N(0, \sigma_\epsilon^2)$$

$$\hat{\underline{\Sigma}}_i = \underbrace{\underline{Z}_i \underline{\Sigma}_v \underline{Z}_i^T}_{\text{the correlation structure comes from this term}} + \underbrace{\sigma_e^2 I}_{\text{contribute to variance on diagonal}}$$

$$\boxed{\text{Chap 7: MIXED MODELS WITH (AUTO)CORRELATED ERRORS}}$$

↳ Same model: $Y_i = \underline{X}_i \underline{\beta} + \underline{Z}_i \underline{V} + \underline{e}_i$

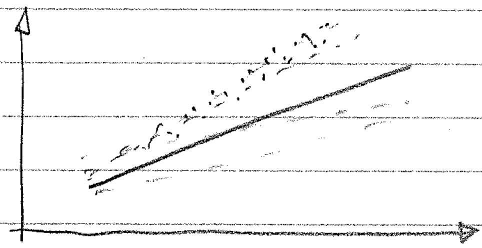
$$V_i \sim N[\underline{0}, \underline{\Sigma}_v]$$

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

Now, $\boxed{\hat{\underline{\Sigma}}_i = \underline{Z}_i \underline{\Sigma}_v \underline{Z}_i^T + \sigma_\epsilon^2 \underline{\Omega}_i}$

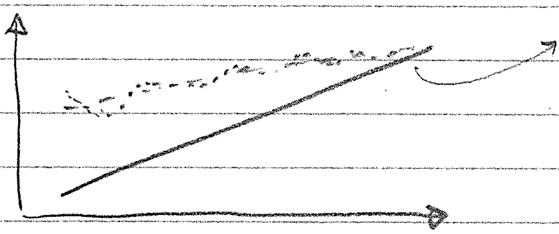
↑ no longer identity

Ex Random Intercept + Slopes...



$\varepsilon \sim N(0, \sigma_\varepsilon^2)$

Error scatters randomly around individual trajectories...



errors correlated with each other around individual trajectories

$\varepsilon_i \sim N(0, \sigma^2 \Omega_i)$

Consider AR(1) errors $\boxed{\varepsilon_j = \rho\, \varepsilon_{j-1} + \xi_j \text{ where } \xi_j \sim N(0, \sigma^2)}$

$\rho \rightarrow$ autocorrelation coefficient.

Note if we had AR(2), then $\boxed{\varepsilon_j = \rho_1 \varepsilon_{j-1} + \rho_2 \varepsilon_{j-2} + \xi_j}$

$\boxed{\text{STATIONARITY}} \rightarrow Var(\varepsilon_j) = Var(\varepsilon_{j'})$

Consider AR(1) model...

$$Var(\varepsilon_j) = Var(\rho\, \varepsilon_{j-1}) + Var(\xi_j)$$

$$= \rho^2 Var(\varepsilon_{j-1}) + Var(\xi_j)$$

$$= \rho^2 Var(\varepsilon_{j-1}) + \sigma^2$$

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

So
$$\boxed{\begin{array}{l} Var(\varepsilon_j) = \rho^2 Var(\varepsilon_j) + \sigma^2 = \dfrac{\sigma^2}{1-\rho^2} \\ Cov(\varepsilon_j, \varepsilon_{j-1}) = \dfrac{\rho\sigma^2}{1-\rho^2} \end{array}}$$

Theorem
$$\boxed{\Sigma_V = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & & \rho^{n-1} \\ \rho & \ddots & \\ \vdots & & \\ \rho^{n-1} & & 1 \end{pmatrix}}$$

Before in Ch. 6, we said

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

equivalent

→ scaling or variance is different

$$\sigma^2 = \frac{\sigma^{*2}}{1-\rho^2}$$

~~Ex Value~~

Another model → moving average ...

↳ we can assume $\boxed{\varepsilon_j = \xi_j - \theta\,\xi_j}$ where $\begin{cases} \xi_j \sim N(0,\sigma^2) \\ \theta = \text{autocorr. coef for moving average process} \end{cases}$ $\boxed{(MA)(1)}$

If we assume stationarity at each time point, then

$$\Sigma = \sigma^2\begin{bmatrix} 1+\theta & -\theta & & & 0 \\ -\theta & \ddots & \ddots & & \\ & -\theta & \ddots & \ddots & \\ & & \ddots & \ddots & -\theta \\ 0 & & & -\theta & 1+\theta \end{bmatrix}$$

- 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...

$$\Sigma = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} \gamma_0 & & & \rho^{n-2}\gamma_1 \\ \gamma_1 & \ddots & & \\ \rho\gamma_1 & & \ddots & \\ \vdots & & & \\ \rho^{n-2}\gamma_1 & \cdots & & \gamma_0 \end{bmatrix}$$

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 & remaining lags decrease exponentially

☒ Toeplitz $\Sigma = \sigma^2 \begin{bmatrix} 1 & & & \rho_{n-1} \\ \rho_1 & 1 & & \\ \vdots & & \ddots & \\ \rho_{n-1} & & \rho_1 & 1 \end{bmatrix}$

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

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

☒ $\boxed{\text{NON-STATIONARY}}$ what happens if Var not constant thru 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 $\sigma^2\Omega_i$, AR(1), MA(1)

	AR(1) ⇓	MA(1) ⇓
- ARMA(1,1) – allows for an arbitrary large correlation for a time gap of 1 that decays rapidly for large gaps	decay w time separation @ constant rate	allows correlation when time gap = 1, no corr for larger 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 ~~structure~~ regardless of the time measurements under consideration.

Ex Consider random int model with Toeplitz structure on Ω

$$\sigma^2 \begin{bmatrix} \theta & & & \\ \theta_1 & \theta & & \\ \vdots & & \ddots & \\ \theta_5 & & & \theta \end{bmatrix} \qquad \alpha^2 \begin{bmatrix} 1 & & \\ \rho & \ddots & \\ \rho & \rho & 1 \end{bmatrix}$$

Random Effect

compound symmetry

So we can only specify a Toeplitz structure with $6-1=5$ parameters.

$\boxed{\text{Non-stationarity}}$ → we can allow variances at each time point to differ

↳ $\boxed{V(\varepsilon_j) = \rho^2 V(\varepsilon_{j-1}) + \sigma^2}$

Mansour (1985) gave the following framework...

- Start with

$$
\begin{aligned}
V(\varepsilon_0) &= 0 \quad \text{(before study began)} \\
V(\varepsilon_1) &= \sigma^2 \\
V(\varepsilon_2) &= \rho^2\sigma^2 + \sigma^2 = \sigma^2(1+\rho^2) \\
V(\varepsilon_3) &= \;--- = \sigma^2(1+\rho^2+\rho^4) \\
&\;\vdots
\end{aligned}
$$

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

$$
\Rightarrow \sigma^2\Omega_i = \sigma^2 \begin{bmatrix} 1 & & & \\ \rho & 1+\rho^2 & & \\ \vdots & \rho & 1+\rho^2+\rho^4 & \\ \vdots & \vdots & \vdots & \ddots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \sum_0^n \rho^{2j} \end{bmatrix}
$$

$\boxed{\text{How do we compare models?}}$

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

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

⇒ use $\boxed{\text{AIC}}$ → takes log likelihood & penalizes it based on # of parameters that we've estimated.

$\boxed{AIC = -2\ln L + 2p}$ → lower is better

p → # params estimated

$\boxed{BIC = -2\ln L + p\ln N}$ → not sure what N is
observations ? # ind. individuals
→ unclear → don't use BIC

Ex ri.ind → $\sigma^2 \Omega_i = \sigma^2 I_i$ → AIC = 1201.8

find intercept?

ri. AR1 → AIC = 1014.4

ri. nar1 → nonstationary AIC = 1010.7 ...

April 10, 2019

April 12, 2019

A marginal model for longitudinal data has the following 3-part specification

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

$E(Y_{ij}/X_{ij}) = \mu_{ij}$ is assumed to depend on the covariates (predictors) through a known link function

$g(\mu_{ij}) = \eta_{ij} = \underline{X}_{ij}^T \underline{\beta}$ $\left(\text{mean}/_{1-\text{mean}}\right)$

Ex Identity, ln (odds), ln (mean) $= \beta_0 + \beta_1 X_{ij} \cdots$
(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 a 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 $\underline{\alpha}$ (and also the means, μ_{ij})

What do we need to estimate?

- β's
- Maybe ϕ
- α's ← nuisance parameters

→ MARGINAL MODELS

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

Note $E(Y_{ij} | \underline{X}_i) = E(Y_{ij} | X_{i1}, \dots X_{in}) = E(Y_{ij} | X_{ij})$

This implies that given X_{ij} → there's no dependence of Y_{ij} on X_{ik} for $j \neq k$. This is fine for time-invariant covords, 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}

$\boxed{\text{What abt GEE's?}}$

→ There's no easy way to ~~specify~~ specify the (joint) distribution of all the responses when responses are discrete.

→ we need special equations to provide us with the parameters 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...

We still need models for
$$\begin{cases} \text{mean } (1) & \eta_{ij} = x_i \beta \\ \text{variance } (2) & V(Y_{ij} | X_{ij}) = \phi\, v(\mu) \\ \text{correlation structure } (3) & \underline{R}(\underline{\alpha}) \rightsquigarrow \text{corr. matrix} \end{cases}$$

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

$$\boxed{\underline{\Sigma}_i = \underline{V}_i = A_i^{1/2} \underline{R}(\underline{\alpha}) A^{1/2}}$$

where A_i is a diagonal matrix with $V(Y_{ij}|X_{ij}) = \phi\, v(\mu_{ij})$ on diagonal

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 $\rightsquigarrow$ $R_i(\alpha) = \rho$ (same as compSymm)
3. AR(1) $=$ $R_i(\alpha) = \rho^{|j-j'|}$ (same as exponential)

NOT autoregressive ...

like Toeplitz

4. M-dependent (or banded) : $R_i(\alpha) = \rho_{|j-j'|}$

The GEE estimator is the solution to

$$\boxed{\sum_i^N \left[\underline{D}_i^T \left[V(\hat{\alpha})\right]^{-1} (\underline{Y}_i - \underline{\mu}_i) \right] = 0} \rightsquigarrow \text{estimating eqn}$$

where

$$\underline{D}_i = \frac{\partial \mu_i}{\partial \beta}$$

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

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

in which case,

$$\boxed{\sum \left[x_i \, (R_i(\hat{\alpha}))^{-1} (Y_i - x_i\beta) \right] = 0}$$

$$\rightarrow \quad \hat{\beta} = \left[\sum x_i^T (R_i(\lambda))^{-1} x_i \right]^{-1} \left[\sum x_i^T (R_i(\lambda))^{-1} Y_i \right]$$

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

$\hookrightarrow$ 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's, obtain estimates of α and ϕ

Iterate until convergence (~ w/ some tolerance)

Result is a consistent estimate of β

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

{ model based ; Empirical }

April 15, 2019 GEEs provide us 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)

$$V(\hat{\beta}) = \left[\sum_{i=1}^{N} D_i^T \hat{V}^{-1} D_i\right]^{-1}$$

② Robust (empirical)

$$V(\hat{\beta}) = M_0^{-1} M_1 M_0^{-1}$$

where

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

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

☒ Important to remember → If variance of $\underline{Y}_i$ is misspecified then using model-based SEs will be wrong...

⇒ so most programs default to using robust estimators called the "sandwich" estimator ($M_0^{-1} M_1 M_0^{-1}$)

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

  → in finite samples, it may be biased and the variance of $V(\vec{\beta})$ ~~so~~ can also be large...

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

$\boxed{\text{Example}}$ Data obtained from a randomized study on smoking cessation published by Gruder 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)

<u>Timing</u>

enrolment — 0 (Baseline) — end of study — 6 — 12 — 24

<u>Response</u> Binary → $\begin{cases} 0: \text{smoking} \\ 1: \text{Not smoking} \end{cases}$

<u>Covariates</u>

Race $\begin{cases} 0 & \text{white} \\ 1 & \text{non white} \end{cases}$

☒ Coding for group and Helmert 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

H1 ↑ experimental

H2 ↑ observational

H3 ↖ observational

Note: makes no sense
to calc. correlation for binary data / count data → use tetrachoric

April 17, 2019 — GEEs & missing data

- MANOVA / RM ANOVA ⇝ no missing data is allowed
- Mixed modeling ⇝ used ML ~~REML~~ → can handle missing data that are either MCAR (missing compl @ rand) or MAR (mis @ rand)
- 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)

Types of missing data

① MCAR → probability of an observation being missing cannot depend on any characteristics of the individuals observed or un observed ... (just random)

② MAR ⇝ prob. of observation being missing can depend on observed characteristics, but not unobserved ...

③ Non-ignorable ⇝ prob. of obs being missing depends on unobserved characteristics

Running in R. waves = time → allows for intermittent missingness
family = binomial → Y/N response ...
Scale.fix = T → for binary response...

⇝ Wald statistics...

April 19, 2019

GEE → only models population...

$\boxed{\text{MIXED MODELING FOR BINARY DATA}}$

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

Our likelihood function is a function of the unknown parameter, p.
↳ p_i → probability that observation i is a "success" = $P(Y_i = 1)$
$1-p_i$ → ... "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 covariate $\underline{X}$. Need to relate $\underline{Y}$ to $\underline{X}$ via logit link.

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

We're trying to get $\hat{\beta}$ that maximizes the likelihood...

$$= \frac{1}{1+\exp[-x_i^T\beta]}$$

$$= \Psi(x_i^T\beta)$$

For a single Bernoulli trial,

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

$$\boxed{\text{Likelihood fn} = \prod_i^n \Psi_i^{Y_i}(1-\Psi_i)^{1-Y_i}}$$

Usually, we take log likelihood → $\boxed{l = \ln(L) = \sum_{i=1}^{N}\left[Y_i \ln \Psi_i + (1-Y_i)\ln(1-\Psi_i)\right]}$

then $\frac{\partial l}{\partial \beta} = \sum (y_i - \psi_i) x_i = 0$ (1)

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

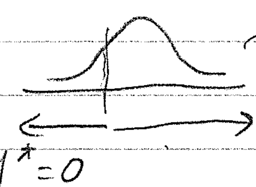
We can't solve this analytically $\Rightarrow$ need an iterative approach.
$\rightarrow$ Use Newton-Raphson algorithm to get estimates of β.

$$\boxed{\beta_{k+1} = \beta_k - \underbrace{\left[\frac{\partial^2 l}{\partial \beta \partial \beta^T}\right]^{-1}}_{\text{provides us with } Var(\hat{\beta})} \cdot \frac{\partial l}{\partial \beta}}$$

$\hookrightarrow$ Take the diagonal elements & they will represent the variances of β's. (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... (skip)
- Threshold concept $\rightsquigarrow$ (skip) $\rightsquigarrow$ dichotomizing
- Latent variable ... $\longrightarrow$



Mixed effects models 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 them, 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 Y_{ij} given the random fx is normal, with $Var(Y_{ij} | \underline{v}_i) = \sigma_v^2 \Rightarrow \phi = \sigma_v^2, V(\mu_{ij}) = 1)$

     Also, given RE, it is assumed that the Y_{ij} are independent.

  2. Conditional mean of Y_{ij} is assumed to depend on both fixed & random fx via
     $$\eta_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T v_i \quad . \text{ Also, } \underline{v}_i \sim N(\underline{0}, \underline{\Sigma}_v)$$

  3. $$E[Y_{ij} | \underline{v}_i] = \eta_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T v_i$$

     The link function is the identity, in the continuous response setting. ~~I, no relation~~

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

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

$$E[\underline{Y}_i | \underline{v}_i] = \underline{X}_i \underline{\beta} + \underline{Z}_i \underline{v}_i \ , \ E[\underline{Y}_i] = \underline{X}_i \underline{\beta}$$

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

$\Rightarrow \beta$ have the interpretation of being population-averaged, ... in how mean responses change over time and how these changes relate to the covariates...

The conditional covariance of Y_i given V_i is assumed to be a diagonal matrix with

$$Cov(Y_i \mid V_i) = Cov(\varepsilon_i) = \sigma^2 I_n \quad \text{(diagonal)}$$

What about the marginal covariance of Y_i?

$$Cov(Y_i) = \underline{Z_i}\, \underline{\Sigma_V}\, Z_i^T + \sigma^2 I_n$$

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

- How does this extend to binary longitudinal responses?

(1) Conditional on a single RE, $\underline{v_i}$, the $\underline{Y_{ij}}$ are independent and have a Bernoulli dist with

$$V(y_{ij} \mid \underline{v_i}) = \underbrace{E[\,y_{ij} \mid \underline{v_i}\,]}_{p} \cdot \underbrace{[\,1 - E(y_{ij} \mid v_i)\,]}_{q} \quad \text{with } \phi = 1$$

(2) Conditional mean of Y_{ij}:

$$\eta_{ij} = \underline{X_{ij}}^T \underline{\beta} + \underline{z_{ij}}^T \underline{v_i} = \underline{X_{ij}}^T \underline{\beta} + \underline{v_i} \quad \text{where } z_{ij} = 1 \ \forall\, i, j \ldots$$

(3) link function is the logit.

$$\ln\left(\frac{P(Y_{ij}=1 \mid V_i)}{P(Y_{ij}=0 \mid V_i)}\right) = \eta_{ij} = \underline{X_{ij}}^T \beta + V_i$$

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

conditional on subject random fx.

- Now, the regression parameter (β) interpretations are subject-specific
  - Consider a simple, mixed effects logistic regression...

$$\text{logit}(p_{ij}) = \underline{x}_{ij}^T \underline{\beta} + V_i \quad \text{(random intercept...)}$$

$$V_i \sim N(0, \sigma_V^2)$$

- 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}

$$\text{logit}(p_{ij} \mid x_{ijk} = x^*) = V_i + \beta_1 x_{ij1} + \dots + \beta_k \cdot x^* + \dots + \beta_p x_{ijp}$$

$$\text{logit}(p_{ij} \mid x_{ijk} = x^*+1) = V_{0i} + \beta_1 x_{ij1} + \dots + \beta_k (x^*+1) + \dots + \beta_p x_{ijp}$$

April 24, 2019

So

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

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.

$\Rightarrow$ $\boxed{\text{subject specific}}$. Note that this interpretation makes more sense for covariates that vary within an individual. If x_k is constant for all time points (within ~~an~~ 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 $V_{0i} - V_{0i'}$

$$\underset{x_k = 0}{\uparrow} \qquad \underset{x_k = 1}{\uparrow}$$

If the link function is not the identity,

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

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

$\Rightarrow$ Inferences for β in mixed-effects model are subject-specific.
Inferences for β in marginal model are at population level.

$\boxed{\text{Ex}}$ Consider simple example using 3 subjects A, B, C.

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

Subj	Baseline prop	Post-Baseline-prop	Difference	log odds rat
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 subj-specific of treatment. We can produce population effects two different ways.

identical if we treat response as continuous

① Average the subject-specific effects

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

② Compare pop-average at baseline to pop. avg post-baseline: $-0.37 - 0.50 = -0.13$.

Now, treating these as Bernoulli and applying logit link.

① Averaging log odds ratio we get

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

So, the effect of individuals probability of disease is $e^{-0.697} \simeq 0.5$

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

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

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

So, population effect is $-0.532 \neq 0.5$

$\boxed{Q}$ which of the ~~effect~~ estimate is better? -0.69 or -0.532?

↳ They're both reasonable:

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

There is $1-e^{-0.697} \simeq 0.5$ reduction in odds of disease

(subj - specific mixed-effects model) → of most interest to you and your doctor.

(marginal models)

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

→ different perspective. of most interest to public health professional.