page #1

1/31/17

Homework / Project 30%

Mid-term (take home) max (M.F) 40%

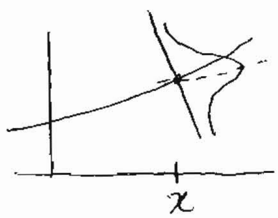
Final (in class) min (M.F) 30%

Scanned note 5% extra credit

GEE $\leftrightarrow$ GLMM

$y = a + bx$

population property



individuals

Review (Regression)

the basic technique in statistics is averaging.

Chi-square distribution w/ n degree of freedom (df)

$\chi^2(n)$

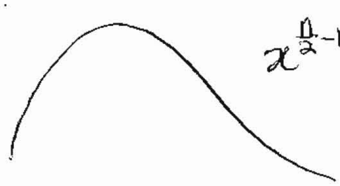
$$Z_1^2 + Z_2^2 + \cdots Z_n^2$$

with the Z_i iid $N(0,1^2)$

$N(\mu, \sigma)$

$\boxed{N(\mu, \sigma^2)}$

MGF $\quad M(t) = 1/(1-2t)^{1/2 \cdot n} \quad (t < \frac{1}{2})$



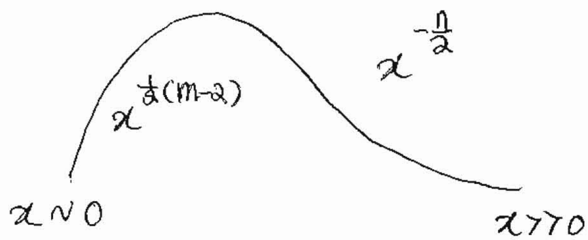
$x^{\frac{n}{2}-1} \exp(-\frac{x}{2})$

F-distribution with degrees of freedom (m, n)

$F(m, n) \qquad F = \dfrac{U/m}{V/n}$

two independent random variable U & V

Chi-square distribution w/ df m & n



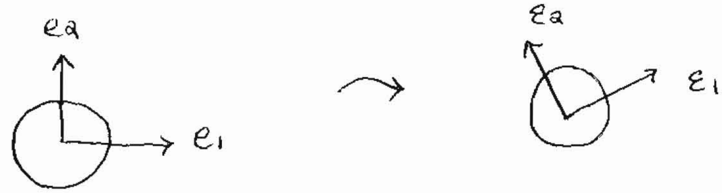
matrix A is orthogonal if $A^{-1} = A^T$

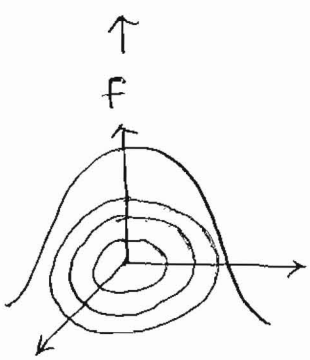
<u>Orthogonality Thm</u>

$\vec{x} = (x_1 \cdots x_p)^T$ $\quad x_i$ independent $N(0, \sigma^2)$

then $Y = AX$ where A is orthogonal.

Y_i independent $N(0, \sigma^2)$





$|A|^{-1}$

proof. $f(x)dx = f(x|y) = \left|\frac{\partial x}{\partial y}\right| dy$

$f(x) \propto \exp\left(\frac{1}{2} x^T \Sigma^{-1} x\right)$ $\qquad Y = AX$

$-\frac{1}{2\sigma^2} X^T X$ $\qquad \frac{\partial Y}{\partial X} = A$

$\uparrow$

$A^{-1}Y$

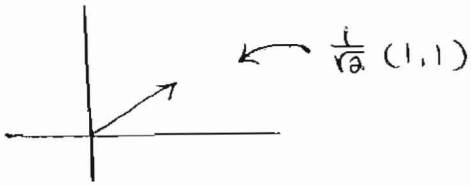
$$
\begin{aligned}
X^T X &= (A^{-1}Y)^T (A^{-1}Y) \\
&= Y^T (A^{-1})^T A^{-1} Y \\
&= Y^T A A^T Y \\
&= Y^T Y
\end{aligned}
$$

$f(y) \propto \exp\left(-\frac{1}{2\sigma^2} Y^T Y\right)$

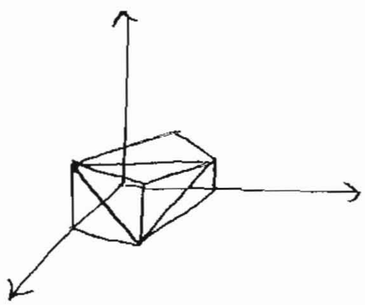
Helmert's Transformation

There exists an orthogonal $n \times n$ matrix P with first row $\frac{1}{\sqrt{n}}(1, \cdots 1)$

$\mathbb{R}^2$



Gram - Schmidt orthogonalization process



assumption:
if $(X_1 \cdots X_n)$ i.i.d $N(\mu, \sigma^2)$

<u>Thm</u> sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(1) $\bar{X}$ and S^2 are independent

(2) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

MGF

$$\frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$$

$$Z_i = \frac{X_i - \mu}{\sigma}$$

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{\bar{X} - \mu}{\sigma}$$

$$\Sigma(Z_i - \bar{Z})^2 = \Sigma\frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{nS^2}{\sigma^2}$$

$$\sum_i = w_i^2 = w^T w = z^T P^T P z = z^T z$$

$$\sum z_i^2 = z^T z = z^T A z + z^T B z \qquad \underline{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

$W = PZ$ with P is a Helmert's transformation.

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \qquad w_1 = \frac{1}{\sqrt{n}} \sum z_i = \frac{1}{\sqrt{n}} \mathbb{1}^T z \qquad \mathbb{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$w_1^2 = \frac{1}{n} \left(\sum_{i=1}^{n} z_i\right)^2 = z^T B z$$

$$(z_1 \cdots z_n)\, B \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

$$w_1^2 = w_1 \cdot w_1 = w_1^T \cdot w_1 = z^T \underbrace{\mathbb{1} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} \mathbb{1}^T}_{\frac{1}{2}\mathbb{1}\mathbb{1}^T} z$$

$$\sum_{i=2}^{n} w_i^2 = \sum_{i=1}^{n} w_i^2 - w_1^2$$

$z^T A z$

$$= \sum_{i=1}^{n} z_i^2 - z^T B z$$

w_i's are independent by the orthogonality of P.

w_1 is independent of $\sum_{i=2}^{n} w_i^2$

$$\sum_{i=1}^{n} (z_i - z)^2 = \sum z_i^2 - n\bar{z}^2$$

$$\sum_{i=1}^{n} z_i^2 + \sum_{i=1}^{n} \bar{z}^2 - 2\sum_{i=1}^{n} \bar{z} \cdot z_i$$

$$n\bar{z}^2 - 2n\bar{z}^2$$

$$\sum z_i^2 = \sum_{i=1}^{n} (z_i - \bar{z})^2 + n\bar{z}^2$$

$$\sum_{i=1}^{n} w_i^2 \qquad w_1^2$$

$$\sum_{i=2}^{n} w_i^2$$

$\bar{z}$ and $\sum (z_i - \bar{z})^2$ are independent

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 \sim \chi^2 (n-1)$$

Geary's Thm (1936)

If the sample mean $\bar{X} = \frac{1}{n} \Sigma X_i$ and sample variance $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ are independent.

then the population distribution is normal.

Fisher's Lemma

$X_1 \cdots X_n$ i.i.d $N(0, \sigma^2)$

$Y_i = \sum_{j=1}^{n} C_{ij} X_j \qquad i = 1, \cdots p < n$

row vector $(C_{i1} \cdots C_{in})$ are orthogonal

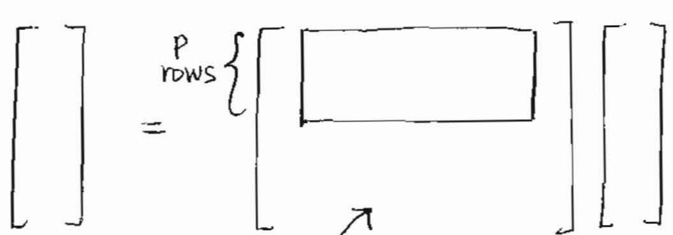
If $S^2 = \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{p} Y_i^2$

then S^2 is independent of $Y_1, Y_2, \cdots Y_p$

$S^2/a^2 \sim \chi^2 (n-p)$

$Y = CX$

$Y_1, \cdots Y_p, Y_{p+1} \cdots Y_n$



$Y \quad = \quad C \quad X$

Gram-schmidt orthogonalization

$$S^2 = \sum_{i=1}^{n} Y_i^2 * - \sum_{i=1}^{p} Y_i^2 = \sum_{i=p+1}^{n} Y_i^2$$

One-way ANOVA

$$X_{ij} \sim N(\mu_i, \sigma^2)$$

$$i = 1 \cdots r$$

$$j = 1 \cdots n_i$$

$$n = \sum_{i=1}^{n} n_i \qquad \text{total sample size}$$

$$X_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \qquad \text{group mean}$$

$$X_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij} \qquad \text{grand mean}$$

$$S_i^2 = \frac{1}{n_i} \sum^{n_i} (X_{ij} - X_{i\cdot})^2 \qquad i\text{th sample variance}$$

the total sum of square

$$SS = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - X_{\cdot\cdot})^2$$

$$= \sum_i \sum_j (X_{ij} - X_{i\cdot} + X_{i\cdot} - X_{\cdot\cdot})^2$$

$$\sum_j (X_{ij} - X_{i\cdot}) = 0 \longrightarrow \sum_j X_{ij} - n_i X_{i\cdot}$$

$$SS = \sum\sum (X_{ij} - X_{i\cdot})^2$$
$$+ \sum\sum (X_{i\cdot} - X_{\cdot\cdot})^2$$
$$+ \sum_i \sum_j (X_{ij} - X_{i\cdot})(X_{i\cdot} - X_{\cdot\cdot})$$

$$\underbrace{\sum_i (X_{i\cdot} - X_{\cdot\cdot}) \underbrace{\sum_j (X_{ij} - X_{i\cdot})}_{0}}_{0}$$

$$= \underbrace{\sum_i n_i S_i^2}_{\substack{\text{within group} \\ \text{residual} \\ SSE}} + \underbrace{\sum_i n_i (X_{i\cdot} - X_{\cdot\cdot})^2}_{\substack{\text{between group} \\ \text{treatment} \\ SST}}$$

$$SS = SSE + SST$$

$H_0 \cdot \mu_i = \mu \qquad i = 1 \cdots r$

$SS/\sigma^2 \sim \chi^2(n-1)$ under H_0

whether or not H_0 is true $\quad n_i S_i^2/\sigma^2 \sim \chi^2(n_i - 1)$

by chi-square addition property

$X_1 = \chi^2(n_1)$, $X_2 = \chi^2(n_2)$, X_1 & X_2 are indep

then $X_1 + X_2 \sim \chi^2(n_1 + n_2)$

$$SSE/\sigma^2 = \frac{\Sigma n_i S_i^2}{\sigma^2} \sim \chi^2\Big(\underbrace{\textstyle\sum_i (n_i - 1)}_{n-r}\Big)$$

$SST = \Sigma n_i (X_{i\cdot} - X_{\cdot\cdot})^2$

S_i^2 ~~are~~ and $X_{i\cdot}$ are independent

$$X_{\cdot\cdot} = \frac{1}{n} \Sigma n_i X_{i\cdot}$$

SST and SSE are independent

chi-square subtraction property

$X = X_1 + X_2$ and X_1 and X_2 are independent

if $X_1 \sim \chi^2(n_1)$

$X \sim \chi^2(n_1 + n_2)$

then $X_2 \sim \chi^2(n_2)$

$$\boxed{\begin{array}{l} X \sim \chi^2(n) \\ EX = n \\ \mathrm{var}\, X = 2n \end{array}}$$

$$SST \sim \chi^2(\underbrace{n-1-(n-r)}_{r-1})$$

$SS = SSE +$ ind SST

$$MS = \frac{SS}{n-1}$$

(mean square)

$$MSE = \frac{SSE}{n-r}$$

$\frac{\chi^2(r-1)}{(r-1)}$ when H_0 is true

$$MST = \frac{SST}{r-1}$$

$$F = \frac{MST}{MSE}$$

$\chi^2(n-r)/n-r$

Under H_0, $F \sim F(r-1, n-r)$

If treatment differ, this will tend to inflate SST.

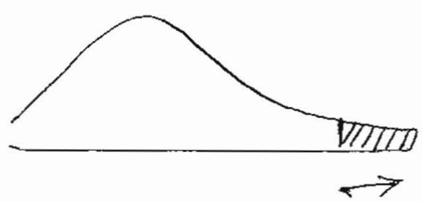
$$SST = \sum n_i (X_{i\cdot} - X_{\cdot\cdot})^2 \qquad \mathbb{E}X^2 = \mathrm{Var}X + (\mathbb{E}X)^2$$

$$= \sum_{i=1}^{r} n_i X_{i\cdot}^2 - n X_{\cdot\cdot}^2$$

$$\mathbb{E}(SST) = \sum_{i=1}^{r} n_i \underbrace{\mathbb{E}X_{i\cdot}^2}_{\frac{\sigma^2}{n} + \mu_i^2} - n \underbrace{\mathbb{E}X_{\cdot\cdot}^2}_{\frac{\sigma^2}{n} + \bar{\mu}^2} \qquad \bar{\mu} = \frac{1}{n}\sum n_i \mu_i$$

$$= \sum_{i=1}^{r} \sigma^2 + n_i \mu_i^2 - \sigma^2 - n\bar{\mu}^2$$

$$= \underbrace{(r-1)\sigma^2}_{} + \sum_{i=1}^{r} n_i (\mu_i - \bar{\mu}^2)$$



Model equation of one-way ANOVA

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1 \cdots r$$
$$j = 1 \cdots n_r$$
$$\epsilon_{ij} \text{ i.i.d } N(0, \sigma^2)$$

Source	df	SS	MS	F	p-value
Treatment	$r-1$	SST	$SST/r-1$	$\frac{MST}{MSE}$	
Residual	$n-r$	SSE	$SSE/n-r$		p
Total	$n-1$	SS	$SS/n-1$		

R. aov

```
model = aov(yield ~ treatment)
        aovp(     "           )
```

1/31/17
page #5

yeild	treatment
100	1
105	1
112	1
⋮	⋮
121	1
95	2
98	2
⋮	

TWO-way ANOVA

NO replications

$$X_{ij} = \mu + \underset{\text{treatment}}{\alpha_i} + \underset{\text{Block}}{\beta_j} + \epsilon_{ij}$$

$\Sigma \alpha_i = 0$ $\quad i = 1 \cdots r$

$\Sigma \beta_j = 0$ $\quad j = 1 \cdots n$

$$X_{ij} - X_{\cdot\cdot} = (X_{ij} - X_{i\cdot} - X_{\cdot j} + X_{\cdot\cdot}) + (X_{i\cdot} - X_{\cdot\cdot}) + (X_{\cdot j} - X_{\cdot\cdot})$$

$$\sum_i \sum_j (X_{ij} - X_{\cdot\cdot})^2 = \sum_i \sum_j (X_{ij} - X_{i\cdot} - X_{\cdot j} + X_{\cdot\cdot})^2 + n\left(\sum_{i=1}^{r} (X_{i\cdot} - X_{\cdot\cdot})^2\right) + r \sum_{j=1}^{n} (X_{\cdot j} - X_{\cdot\cdot})^2$$

$$\underset{nr-1}{SS} = \underset{(r-1)(n-1)}{SSE} + \underset{r-1}{SST} + \underset{n-1}{SSB}$$

Source	df	SS	F	p-value
Treatment	$r-1$	SST	$\frac{MST}{MSE}$	P_T
Block	$n-1$		$\frac{MSB}{MSE}$	P_B
Residual	$(r-1)(n-1)$	SSE		
Total	$nr-1$	SS		

two null hypothesis

```
model = aov ( yield ~ treatment + block)
        aovp (          "                 )
```

yetld	treatment	block
⋮	⋮	⋮

2/17/10

#1

Two-Way ANOVA Replications & Interactions

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \qquad i = 1 \cdots r$$

$$\Sigma \alpha_i = 0 \qquad \sum_j r_{ij} = 0, \ \forall i \qquad j = 1 \cdots n$$

$$\Sigma \beta_j = 0 \qquad \sum_i r_{ij} = 0, \ \forall j \qquad k = 1 \cdots m$$

Gender \ Treatment	High dose	Low dose	control
Female	F_1, F_2, F_3	F_4, F_5, F_6	
male			

$$X_{ijk} - X_{\cdots} = (x_{ijk} - x_{ij\cdot}) + (X_{i\cdot\cdot} - X_{\cdots}) + (X_{\cdot j\cdot} - X_{\cdots}) + (X_{ij\cdot} - X_{i\cdot\cdot} - X_{\cdot j\cdot} + X_{\cdots})$$

$$\sum_i \sum_j \sum_k (X_{ijk} - X_{\cdots})^2 = \sum_i \sum_j \sum_k (X_{ijk} - X_{ij\cdot})^2 + \underbrace{nm \sum_i (X_{i\cdot\cdot} - X_{\cdots})^2}_{= \sum_i \sum_j \sum_k}$$

$$SS = SSE + SST$$

$$rnm - 1 \qquad rn(m-1) \qquad r-1$$

$$+ rm \sum_j (X_{\cdot j\cdot} - X_{\cdots})^2 + m \sum_i \sum_j (X_{ij\cdot} - X_{i\cdot\cdot} - X_{\cdot j\cdot} + X_{\cdots})^2$$

$$+ SSB \qquad + SSI$$

$$n-1 \qquad (r-1)(n-1)$$

F	Source	d.f	sum of square	MS ~~M/S~~
	Source	d.f	sum of square	MS
MST/MSE	Treatment	$r-1$	SST	$SST/(r-1)$
MSB/MSE	Blocking Factor	$n-1$	SSB	$SSB/(n-1)$
MSI/MSE	Interaction	$(r-1)(n-1)$	SSI	$SSI/(r-1)(n-1)$
	Error	$rn(m-1)$	SSE	$SSE/rn(m-1)$
	Total	$rnm-1$ *	SS	$SS/(rnm-1)$

In R,

Yield ~ treatment * block

Multiple Regression

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

$i = 1 \cdots n$

$\rightarrow$ (1) $\epsilon_i \sim N(0, \sigma^2)$ i.i.d

$\rightarrow$ (2) $\mathbb{E}\epsilon_i = 0$, $\operatorname{var}\epsilon_i = \sigma^2$, independent

IF satisfy (1) then (2) is also satisfied.

Model Equation

$$\underset{n\times 1}{\vec{y}} = \underset{n\times p}{X}\,\underset{p\times 1}{\vec{\beta}} + \underset{n\times 1}{\vec{\epsilon}}$$

$n \gg p$

constant

$$\vec{y} = A\vec{\beta} + \vec{\epsilon}$$

Normal Equation

$$A^T A \beta = A^T y \nearrow \hat{\beta} = (A^T A)^{-1} A^T y$$

Lemma. IF $A_{n\times p}$ has full rank p then $A^T A$ is positive definite.

$$x^T A^T A x = 0 = (Ax)^T (Ax) = 0$$

$$\Leftrightarrow Ax = 0 \Leftrightarrow x = 0$$

Thm

The solutions $\hat{\beta} = (A^T A)^{-1} A^T y$ to the normal equation are both the maximum likelihood estimators and the least-square estimators of the parameter β.

Linearity: $\hat{\beta}$ is linear in the data y

Unbiasedness: $\hat{\beta}$ is an unbiased estimator

* $\operatorname{var}(Ax) = A \operatorname{var}(X) A^T$

$$\mathbb{E}\hat{\beta} = (A^T A)^{-1} A^{*T} \mathbb{E}y$$

$$= \underline{(A^T A)^{-1} A^T A} \cdot \beta = \beta$$

Covariance: $\operatorname{Var}(\hat{\beta})$

$$= (A^T A)^{-1} A^T \underbrace{\operatorname{var}(y)}_{\sigma^2 I} \cdot A (A^T A)^{-1}$$

$$= (A^T A)^{-1} A^T A (A^T A)^{-1} \cdot \sigma^2 I$$

$$= \sigma^2 (A^T A)^{-1}$$

2/7/17

#2

$y = \alpha + \beta x + \epsilon$

$H_0: \beta = 0$

$H_a: \beta \neq 0$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \begin{matrix} \leftarrow A \\ \leftarrow X \end{matrix}$$

$$S_\beta = \frac{S_e}{\sqrt{\Sigma x_i^2}}$$

$$t = \frac{\hat{\beta} - \beta}{S_\beta}$$

<u>Gauss - Markov Thm</u>

Among all unbiased linear estimators $\tilde{\beta} = By$ of β. the least-squares estimator $\hat{\beta} = (A^TA)^{-1}A^Ty$ has the minimum variance in each component.

$\hat{\beta}$ is the BLUE or MVULE.

$SS = (y - A\beta)^T(y - A\beta)$

$SSE = (y - A\hat{\beta})^T(y - A\hat{\beta}) = (y - \hat{y})^T(y - \hat{y}) \qquad \hat{y} = A\hat{\beta}$

$SSR = (A\hat{\beta} - A\beta)^T(A\hat{\beta} - A\beta) = (\hat{\beta} - \beta)^TA^TA(\hat{\beta} - \beta)$

<u>Thm</u> (Sum of square Decomposition)

$SS = SSE + SSR$

$y - A\beta = y - A\hat{\beta} + A\hat{\beta} - A\beta$

multiple the vector on each side by its transpose

$(y - A\hat{\beta})^T \cdot (A\hat{\beta} - A\beta) = 0$

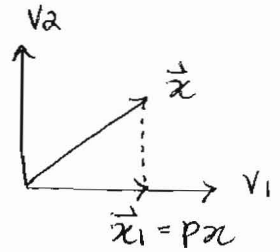
$(y - A\hat{\beta})^T A(\hat{\beta} - \beta) = 0$

$A^T(y - A\hat{\beta}) = A^Ty - \cancel{A^TA \cdot (A^TA)^{-1}} A^Ty = 0$

$$SSE = \min_{\beta} SS$$

Call a linear transformation $P: V \to V$ a projection onto V_1 along V_2 if V is the direct sum
$V = V_1 \oplus V_2$, and if $\vec{x} = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix}$ with $P\vec{x} = \vec{x}_1$

$$V_1 \oplus V_2 = \{ w : w = \lambda_1 v_1 + \lambda_2 v_2 ,\ v_1 \in V_1 ,\ v_2 \in V_2)$$



$P : A(A^TA)^{-1}A^T$ $\qquad\qquad$ $P \cdot y = A\underline{(A^TA)^{-1}A^Ty} = A\hat{\beta} = \hat{y}$

Lemma : P and $I-P$ are ~~independent~~ idempotent 연산을 여러번 적용하더라도 결과가 달라지지 않는 성질

Matrix M is idempotent if it is its own square

$$M^2 = M$$

<u>Thm</u>

$$SSE = Y^T(I-P)Y = (Y - A\beta)^T(I-P)(Y-A\beta)$$

$$SSR = (Y - A\beta)^T P (y - A\beta)$$

proof : $SSE = (Y - A\hat{\beta})^T(Y - A\hat{\beta}) = (Y - Py)^T(y - Py)$

$PY = A\hat{\beta}$ $\qquad\qquad$ $= Y^T(I-P)^T(I-P)y$

$= Y^T(I-P)y$

$$SSR = (\hat{\beta} - \beta)^T A^TA (\hat{\beta} - \beta)$$

$$\hat{\beta} - \beta = (A^TA)^{-1}A^Ty - (A^TA)^{-1}A^TA\beta$$

$$= (A^TA)^{-1}A^T(y - A\beta)$$

2/17/17

#3

$$SSR = (y - A\beta)^T A (A^T A)^{-1} A^T A (A^T A)^{-1} A^T (y - A\beta)$$
$$= (y - A\beta)^T P (y - A\beta)$$

$$SS = (y - A\beta)^T \cdot I \cdot (y - A\beta)$$
$$SSE = SS - SSR$$

Thm (trace Formula)

$$\mathbb{E}\, x^T A x = \text{trace}\,(A \cdot \text{var}(X)) + (\mathbb{E}x)^T A\, \mathbb{E}x$$

$$\mathbb{E}x^2 = \text{var}(x) + (\mathbb{E}x)^2$$

$A = [a_{ij}]$

$$\mathbb{E}\, a \cdot x^2 = \text{trace}\,(a \cdot \text{var}(X)) + a \cdot (\mathbb{E}x)^2$$

$$x^T A x = \sum_{ij} a_{ij} x_i x_j \qquad [x \cdots x_n] \begin{bmatrix} a_{11} & \\ & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

$$\mathbb{E}\, x^T A x = \sum_{ij} a_{ij}\, \mathbb{E}(x_i x_j)$$

$$= \sum_{ij} a_{ij} \left[\text{cov}(x_i, x_j) + \mathbb{E}x_i \cdot \mathbb{E}x_j \right]$$

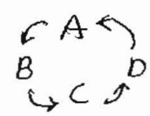
trace is invariant under cyclic permutation

$$\text{trace}\,(ABCD) = \text{trace}\,(DABC)$$
$$\text{trace}\,(ABC) \neq \text{trace}\,(ACB)$$
$$\text{trace}\,(AB) = \text{trace}\,(BA)$$
$$= \sum_i (AB)_{ii} = \sum_{ij} a_{ij} b_{ji}$$

$$\downarrow$$

$$\text{trace}\,(A \cdot B)$$

note $B = \text{var}(x)$ which is symmetric



$A_{n\times p}$
$\text{rank}(A) = p$
$\uparrow$
rk $\quad \epsilon_i \sim N(0, \sigma^2)$

<u>Thm</u> $\quad \text{trace } P = p$
$\text{trace}(I-P) = n-p$
$\mathbb{E}(SSE) = (n-p)\sigma^2 \longrightarrow \hat{\sigma}^2 = \dfrac{SSE}{n-p}$ unbiased estimator of σ^2

$$\text{trace } P = \text{trace } (A(A^TA)^{-1}A^T)$$
$$= \text{trace } (A^TA(A^TA)^{-1})$$
$$= \text{trace } (I_p) = p$$

$$\text{trace } I_n = n$$
$$\text{trace } (I-P) = n-p$$

$$SSE = (y - A\beta)^T (I-P)(y - A\beta)$$
$$\mathbb{E}y = A\beta \qquad var(y) = \sigma^2 I_n$$
$$\mathbb{E}\,SSE = \text{trace}(I-P \cdot var(y-A\beta)) + 0^T(I-P)0$$
$$= \text{trace}(\sigma^2 \cdot (I-P) \cdot I) = \sigma^2(n-p)$$

<u>Thm</u> If B is idempotent
(i) its eigenvalues λ's are 0 or 1.
(ii) its trace is its rank

$$B^2x = B(Bx) = B(\lambda x) = \lambda(Bx) = \lambda^2 x$$
$Bx = \lambda x$
$$\lambda = \lambda^2$$
$$\lambda = 0 \text{ or } 1$$

trace B = sum of eigen values

2/7/17

#4

$B = Q \Lambda Q^T$ $\quad$ $\text{trace}(B) = \text{trace}(Q \Lambda Q^T)$

$= \text{trace}(\Lambda Q^T Q)$

$= \text{trace}(\Lambda)$

Thm If P is a symmetric projector of rank r and x_i are independent $N(0, \sigma^2)$ the quadratic form $x^T P x \sim \sigma^2 \chi^2(r)$ orthogonality Thm

$x^T P x = y_1^2 + \cdots + y_r^2$, where $y_i \sim N(0, \sigma^2)$ i.i.d

plug in

$\boxed{P = Q \Lambda Q^T}$ $\quad$ $x^T Q \Lambda Q^T x$

$Q^T P Q = \Lambda$

$$\begin{array}{c} 1 \\ \vdots \\ r \\ r+1 \\ \vdots \\ n \end{array} \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

$x^T Q \Lambda Q^T x = y^T \Lambda y$

$= y_1^2 + \cdots + y_r^2$

* Thm Chi-square Decomposition Theorem

If $I = P_1 + \cdots + P_K$

with each P_i a symmetric projection matrix with rank n_i

(i) the rank sum $n = n_1 + \cdots + n_K$

(ii) each quadratic form $Q_i = x^T P_i x$ is chi-square $Q_i \sim \sigma^2 \chi^2(n_i)$

(iii) Q_i are mutually independent

(iv) $P_i P_j = 0 \quad (i \neq j)$

orthogonality of the projection P_i

This fundamental theorem gives all the distribution theory in the standard regression analysis.

$$I^2 = (\Sigma P_i)^2 = \sum_i P_i^2 + \sum\sum_{i<j} P_i P_j$$

$$= \sum_i P_i + \sum\sum_{i<j} P_i P_j$$

trace $P_i P_j = 0$

trace $(I) = n$ trace $(\sum_i P_i) = n$

$$\Rightarrow \sum\sum_{i<j} \text{trace } (P_i P_j) = 0$$

$$\text{trace } (P_i P_j) = \text{trace } (P_i^2 P_j^2) = \text{trace } (P_i P_i P_j P_j)$$

$$= \text{trace } (P_i P_j P_j P_i)$$

$$= \text{trace } (P_j P_i)^T (P_j P_i) = 0$$

trace $P_i P_j \geq 0$

$$\text{trace } (M^T M) = \sum_i \sum_j m_{ij}^2 = 0$$

$$m_{ij} = 0$$

$$\underline{P_i P_j = 0} \quad \forall i \neq j$$

$P_1 x$, $P_2 x \cdots P_K x$ are independent

$\text{cov}(P_1 x, P_2 x)$ $x^T P_1 P_2 x$

$$\underbrace{x^T x}_{\sigma^2 \chi^2(n)} = x^T I x = x_1^T P x + \underbrace{x^T P_2 x}_{\sigma^2 \chi^2(n_i)} + \ldots + x^T P_K x$$

$$Q_i = x^T P_i x = x^T P_i P_i x$$

$$= x^T P_i^T P_i x$$

$$= (P_i x)^T P_i x$$

Source	SS	F
1	Q_1	$Q_1/n_1 / Q_K/n_K$
2	Q_2	
3	Q_3	
⋮	⋮	
K	Q_K	

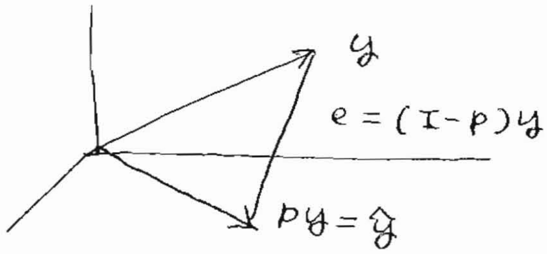
STAT 707
2/14/17
#1

Chi-square decomposition thm

Thm (i) $\hat{y} = Py \sim N(A\beta, \sigma^2 P)$

(ii) $e = y - \hat{y} = (I-P)y \sim N(0, \sigma^2(I-P))$

(iii) $e, \hat{y}$ are independent



$y = \underbrace{A}_{X}\beta$

Proof $\mathbb{E}\hat{y} = y = A\beta$ $\quad \hat{y} = Py$

$P: A(A^TA)^{-1}A^T$

$var(\hat{y}) = P\,var(y)P^T = P\sigma^2 I P = \sigma^2 P$

$cov(e, \hat{y}) = 0$

Thm (i) $\hat{\beta} \sim N(\beta, \sigma^2(A^TA)^{-1})$

(ii) $\hat{\beta}$ and SSE (or $\hat{\beta}$ and $\hat{\sigma}^2$) are independent.

proof $\hat{\beta} - \beta = (A^TA)^{-1}A^T(y - A\beta)$

$SSE = (y - A\beta)^T(I-P)(y - A\beta)$.

$(I-P)(y - A\beta)$ and $(A^TA)^{-1}A^T(y - A\beta)$ are independent.

$(\hat{\beta}-\beta)^T(I-P)(y - A\beta) \qquad cov = 0$

$(A^TA)^{-1}A^T(I-P)$

$(A^TA)^{-1}A^T - (A^TA)^{-1}A^TA(A^TA)^{-1}A^T = 0$

Corollary. β_i denote the ith element of β

$(A^TA)^{-1}_{ii}$ is the ith diagonal element of $(A^TA)^{-1}$

then

$$\frac{\beta_i - \hat{\beta}_i}{\hat{\sigma}\sqrt{(A^TA)^{-1}_{ii}}} \sim t_{n-p}$$

R output   est se t

Proof $\quad t_r \sim \dfrac{N(0, 1^2)}{\sqrt{\chi^2_r / r}}$

$\beta_i - \hat{\beta}_i \sim N(0, \sigma^2 (A^T A)^{-1}_{ii})$

$\hat{\sigma}^2 \sim \dfrac{\sigma^2}{n-p} \chi^2_{n-p}$

$\dfrac{MST}{MSE} \sim F$

* AIC

Akaike, Hirotogu (1973)

Information theory and the maximum likelihood principle

ARMA

$F(t) - b_1 F(t-1) - b_2 F(t-2) - \cdots$

$= a_0 \varepsilon(t) + a_1 \varepsilon(t) \cdots$

$\beta_1, \beta_2 \cdots \beta_m \cdots$

The model $f(\cdot | \theta)$ is a family of parametrized probability densities with $\theta \in \Theta$

We want to compare a general model with the true model θ_0

Likelihood ratio $\quad \lambda(\cdot) = \dfrac{f(\cdot | \theta)}{f(\cdot | \theta_0)}$

discrimination between θ and θ_0 at x

$\xi(\lambda(x))$ for some function ξ

mean discrimination, if θ_0 is true

$$\mathcal{D}(\theta, \theta_0 ; \xi) = \int_{-\infty}^{+\infty} \xi(\lambda(x)) f(x | \underline{\underline{\theta_0}}) dx$$

$$= \mathbb{E}_X \, \xi(\lambda(X))$$

STAT 707
2/14/17
#2

$$\theta = (\theta_1, \theta_2, \cdots \theta_i, \cdots)$$

$$\frac{\partial}{\partial \theta_i} \mathcal{D}(\theta, \theta_0 ; g) \Big|_{\theta=\theta_0} = \int_{-\infty}^{+\infty} \left(\frac{dg}{d\lambda} \frac{\partial \lambda}{\partial \theta_i} \right) \Big|_{\theta=\theta_0} f(x|\theta_0) dx$$

$$= \frac{dg}{d\lambda} \Big|_{\lambda=1} \int_{-\infty}^{+\infty} \frac{\partial f(x|\theta)}{\partial \theta_i} \Big|_{\theta=\theta_0} \frac{1}{\cancel{f(x|\theta_0)}} \cdot \cancel{f(x|\theta_0)} dx$$

$$\cancel{= g'(1) \cdot \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta_i} f(x|\theta) \Big|_{\theta=\theta_0} dx}$$

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{D} \Big|_{\theta=\theta_0} = \int_{-\infty}^{+\infty} \frac{d^2 g}{d^2 \lambda} \frac{\partial \lambda}{\partial \theta_i} \cdot \frac{\partial \lambda}{\partial \theta_j} \Big|_{\theta=\theta_0} f(x|\theta_0) dx$$

$$\cancel{+ \int_{-\infty}^{+\infty} \frac{dg}{d\lambda} \frac{\partial^2 \lambda}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta_0} f(x|\theta_0) dx}$$

$$\frac{\partial^2 \lambda}{\partial \theta_i \partial \theta_j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x|\theta) \Big|_{\theta=\theta_0} \cdot \frac{1}{f(x|\theta_0)}$$

$$\int f(x|\theta) dx = 1 \Rightarrow \int \frac{\partial f}{\partial \theta_i} dx = 0$$

$$\int \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} dx = 0$$

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{D} = g''(1) \int_{-\infty}^{+\infty} \left[\frac{\partial f(x|\theta)}{\partial \theta_i} \cdot \frac{1}{f(x|\theta_0)} \right] \Big|_{\theta=\theta_0} \cdot \left[\frac{\partial f(x|\theta)}{\partial \theta_j} \cdot \frac{1}{f(x|\theta_0)} \right] \Big|_{\theta=\theta_0} f(x|\theta_0) dx$$

$$\mathcal{D}(\theta, \theta_0 ; g) = \mathcal{D}(\theta_0, \theta_0 ; g)$$

$$\cancel{+ \sum \frac{\partial \mathcal{D}}{\partial \theta_i} \Big|_{\theta=\theta_0} (\theta_i - \theta_{i0})}$$

$$\frac{\partial \log f(x)}{\partial x} = \frac{1}{f(x)} \cdot \frac{\partial f}{\partial x}$$

$$+ \frac{1}{2} \sum\sum \frac{\partial^2 \mathcal{D}}{\partial \theta_i \partial \theta_j}\Big|_{\theta=\theta_0} (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) + o(\|\theta - \theta_0\|^2)$$

$$\mathcal{D}(\theta_0, \theta_0, \xi) = \xi(1)$$

$$\frac{\partial}{\partial \theta_i} \mathcal{D}\Big|_{\theta=\theta_0} = 0$$

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathcal{D}\Big|_{\theta=\theta_0} = \xi''(1) \int \frac{\partial \log f(x|\theta)}{\partial \theta_i}\Big|_{\theta=\theta_0} \cdot \frac{\partial \log f(x|\theta)}{\partial \theta_j}\Big|_{\theta=\theta_0} f(x|\theta_0)\,dx$$

Fisher Information

$$[I(\theta)]_{ij} = \mathbb{E}\left[\frac{\partial}{\partial \theta_i} \log f(|\theta) \cdot \frac{\partial}{\partial \theta_j} \log f(x|\theta)\right.$$
$$= -\mathbb{E}\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta)\right]$$

$$\mathcal{D}(\theta, \theta_0, \xi)$$
$$= \xi(1) + \frac{1}{2}\xi''(1)(\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|^2)$$

Make $\mathcal{D}$ behave like a distance

$\mathcal{D}(\theta_0, \theta_0, \xi) = 0$

$\Rightarrow \xi(1) = 0$

$\mathcal{D}(\theta, \theta_0, \xi) \geq 0$

$\Rightarrow \xi''(1) > 0$

$$\frac{1}{2}\xi''(1) = \frac{1}{2} \cdot \frac{2}{1} = 1$$

$\xi(t) = -2\log(t)$

$\xi'(t) = -\frac{2}{t}$

$\xi''(t) = \frac{2}{t^2}$

$$\mathcal{D}(\theta, \theta_0, \xi = -2\log(t))$$
$$= (\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0)$$
$$+ o(\|\theta - \theta_0\|^2)$$

$$-2\log(\lambda) = 2\log\left(\frac{1}{\lambda}\right)$$

Mean discrimination

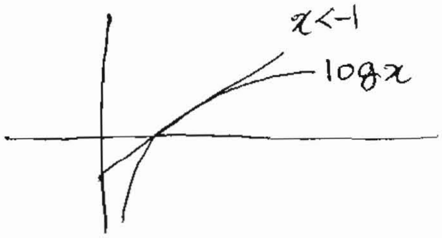
$$\mathcal{D}(\theta, \theta_0; \xi) = \int_{-\infty}^{+\infty} \xi(\lambda(x)) f(x|\theta_0)\,dx$$
$$= \mathbb{E}_x\, \xi(\lambda(x))$$

STAT707
2/14/17
#3

$$\hookrightarrow \int_{-\infty}^{+\infty} 2 \cdot \log \frac{f(x \mid \theta_0)}{f(x \mid \theta)} \cdot f(x \mid \theta_0)\, dx$$

$$= 2\mathbb{E}_X \log f(X \mid \theta) - 2\mathbb{E}_X \log f(X \mid \theta)$$

$\log(1+x) < x$, for $x > -1$



$- D(\theta, \theta_0) \leq 0$

$$-D = \int 2 \log \frac{f(x \mid \theta)}{f(x \mid \theta_0)}\, f(x \mid \theta_0)\, dx$$

$$\leq \int 2 \cdot \left[\frac{f(x \mid \theta)}{f(x \mid \theta_0)} - 1 \right] f(x \mid \theta_0)\, dx$$

$$= \int 2 \cdot f(x \mid \theta)\, dx - \int 2 \cdot f(x \mid \theta_0)\, dx$$

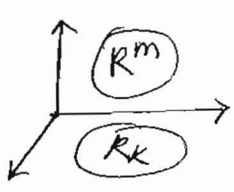
$$= 0$$

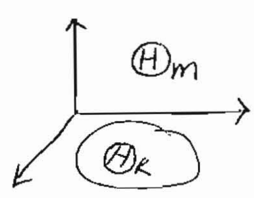
$$\hat{D}_n(\theta, \theta_0) = \frac{2}{n} \Sigma \log \frac{f(X_i \mid \theta_0)}{f(X_i \mid \theta)}$$

$\Theta_1 \subset \Theta_2 \subset \cdots \Theta_m$

$\dim(\Theta_K) = K$

Θ_K is the subspace of $\mathbb{R}^m$ which has the last $m-K$ elements equal to zero.





$W(\theta, \theta_0)$
$= (\theta, \theta_0)^T I(\theta_0)(\theta - \theta_0)$
$= \|\theta - \theta_0\|^2$

Define an inner product

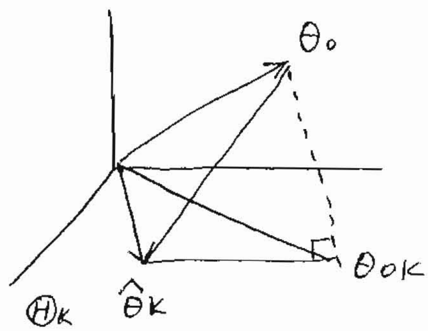
Σ positive definite $\quad \langle \alpha , \beta \rangle = \alpha^T \Sigma \beta$

~~$\langle \alpha + \beta \rangle$~~

$$W(\hat{\theta}_K, \theta_0)$$
$$= \| \theta_{0K} - \theta_0 \|^2 + \| \hat{\theta}_K - \theta_{0K} \|^2$$

θ_{0K} is the projection of θ_0 on Θ_K

$\hat{\theta}_K$ is the estimation of θ_0 on Θ_K



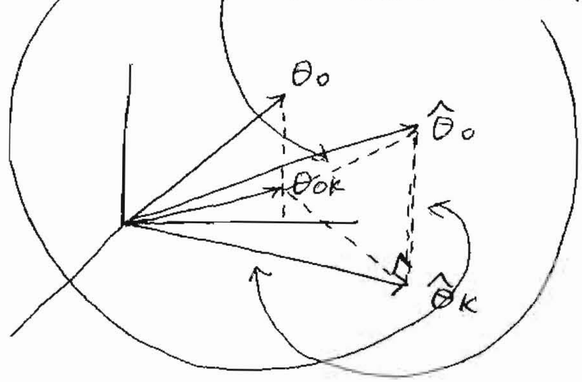
$$n \hat{D}_n (\hat{\theta}_0, \theta_{0K}) \approx n \cdot \| \hat{\theta}_0 - \theta_{0K} \|^2$$
$$\| \quad (\hat{\theta}_0 - \theta_{0K})^T I(\theta_0) (\hat{\theta}_0 - \theta_{0K})^T$$

Similarly,

$$n \hat{D}_n (\hat{\theta}_0, \theta_{0K}) \approx n \cdot \| \hat{\theta}_K - \theta_{0K} \|^2$$

$$\Rightarrow n \hat{D} (\hat{\theta}_K, \hat{\theta}_0) \approx n \cdot \| \hat{\theta}_0 - \theta_{0K} \|^2 - n \cdot \| \hat{\theta}_K - \theta_{0K} \|^2$$

$(\hat{\theta}_0 - \theta_0) - (\theta_{0K} - \theta_0)$



STAT 707
2/14/17
#4

$E\chi^2(K)=K$

$$n \cdot \hat{\mathcal{D}}_n(\hat\theta_K, \theta_0)$$
$$\approx n \cdot \|\cancel{\theta_{0K}} - \theta_0\|^2 + n\|\hat\theta_0 - \theta_0\|^2 - 2n\langle \hat\theta_0 - \theta_0, \theta_{0K} - \theta_0\rangle \approx 0$$
$$- n\|\hat\theta_K - \theta_{0K}\|^2$$

$$n \cdot W(\hat\theta_K, \theta_0) - n\hat{\mathcal{D}}_n(\hat\theta_K, \hat\theta_0)$$
$$\approx \underbrace{\underbrace{n\|\hat\theta_K - \theta_{0K}\|^2}_{d.f=K} + \underbrace{n\|\hat\theta_K - \theta_{0K}\|^2}_{d.f=K} - \underbrace{n\|\hat\theta_0 - \theta_0\|^2}_{d.f=m}}$$

$$n E W(\hat\theta_K, \theta_0) = n\hat{\mathcal{D}}_n(\hat\theta_K, \hat\theta_0) + E[\;\sim\sim\sim\;]$$
$$+ 2K - m$$

$$2\sum \log f(X_i|\hat\theta_0) - 2\sum \log f(X_i|\hat\theta_K)$$

$$\Theta_1 \subset \Theta_2 \subset \cdots \subset \Theta_m \cdots \subset \Theta$$

$$AIC = -2\sum \log f(X_i|\hat\theta_K) + 2K$$

AICC = AIC correction for small sample size n

$$= AIC + \frac{2K(K+1)}{n-K-1}$$

STAT 107
2/21/17
#1

MANOVA & ANCOVA

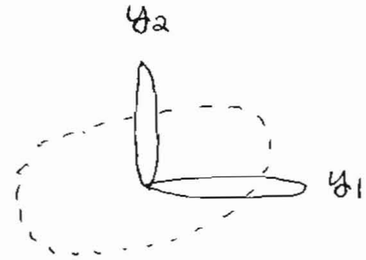
Beck's Depressing Index (self-rated)
Hamilton Rating Scale (doctor-rated)

	BDI	HRS	anti-depresant	therapy
patient 1	y_{11}	y_{21}	control	clinic
			low-dose	cognitive

$y_1 \sim X$
$y_2 \sim X$

$$t = \frac{Z}{\sqrt{\chi^2/r}}$$

$$t^2 = \frac{Z^2}{\chi^2/r} \sim F_{1r}$$



Data : $p \times 1$ random vector
$\vec{x}_1, \ldots, \vec{x}_n$

$$\begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}$$

1) common mean vector $\vec{\mu}$
$\mathbb{E}\vec{x}_i = \vec{\mu}$, $i = 1 \cdots n$

2) Homoskedasticity
the data for all subjects have common covariance matrix Σ

3) The subjects are independently sampled.

4) The subjects are sampled from a multivariate normal distribution

Hypothesis $H_0 : \vec{\mu} = \vec{\mu}_0$
$H_a : \vec{\mu} \neq \vec{\mu}_0$
$\Downarrow$
at least one $\mu_i \neq \mu_{0j}$

$$t_i = \frac{\bar{x}_j - \mu_{0j}}{\sqrt{s_i^2/n}}$$

$$t^2 = \frac{(\bar{x}-\mu_0)^2}{s^2/n} = n(\bar{x}-\mu_0)\cdot\frac{1}{s^2}(\bar{x}-\mu_0) \sim F_{1,n-1}$$

$$s^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$$

Hotelling's T-square

(A31)

$$T^2 = n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) = (\bar{x}-\mu_0)^T S_{\bar{x}}^{-1}(\bar{x}-\mu_0)$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\vec{x}_i-\bar{x})(x_i-\bar{x})^T \qquad S_{\bar{x}} = \frac{1}{n}S$$

$$n(\bar{x}-\mu_0)^T\Sigma^{-1}(\bar{x}-\mu_0) \sim \chi_p^2$$

(proof. MGF)

$$T^2_{p,n-1} = \frac{p(n-1)}{n-p}F_{p,n-p}$$

$$\frac{(n-p)}{p(n-1)}T^2_{p,n-1} = F_{p,n-p}$$

Paired Hotelling's T-square

$$\begin{matrix} \vec{x}_{11} & \vec{x}_{21} \\ \vdots & \vdots \\ \vec{x}_{1n} & \vec{x}_{2n} \end{matrix}$$

$\vec{x}_{1i}$ subject i measured at $t=0$

$\vec{x}_{2i}$ i $t=3$

$\vec{Y}_i = \vec{x}_{1i} - \vec{x}_{2i}$ $\qquad H_0: \vec{\mu}_Y = 0$ $\qquad T^2 = n\bar{\vec{y}}^T S_y^{-1}\bar{\vec{y}}$

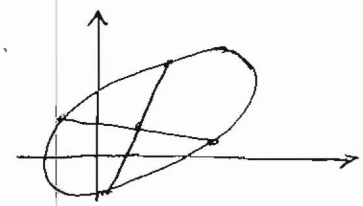
$\vec{Y}_1, \cdots \vec{Y}_n$ $\qquad H_a: \vec{\mu}_Y \neq 0$

STAT707
2/21/17
#2

$$F_{p,n-p} = \frac{n-p}{p(n-1)} T^2$$

ex) $\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$ $\quad \mu \neq 0$?

$$\frac{\mu_1 + \mu_2}{2} - \mu_3 = 0 \,?$$



Simultaneous $(1-\alpha) \times 100\%$ confidence interval

$c^T Y$ $\quad \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix}$ $\quad c_1 = \frac{1}{2}$, $c_2 = \frac{1}{2}$, $c_3 = -1$, $c_4 = 0$

$$c^T \vec{\mu} \in c^T \vec{x} \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,\alpha}} \sqrt{\frac{1}{n} c^T S c} \qquad S = \begin{bmatrix} s_1^2 & & \\ & \ddots & \\ & & s_p^2 \end{bmatrix}$$

$$c = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \mu_1 \in \bar{x}_1 \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,\alpha}} \sqrt{\frac{s_1^2}{n}}$$

$$c^T \vec{\mu} = \mu_1 \qquad \mu_1 \in \bar{x}_1 \pm t_{n-1,\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n}}$$

Two-sample test

$\vec{x}_{11} \cdots \vec{x}_{1n_1}$ $\qquad H_0: \vec{\mu}_1 = \vec{\mu}_2$

$\vec{x}_{21} \cdots \vec{x}_{2n_2}$ $\qquad H_a: \vec{\mu}_1 \neq \vec{\mu}_2$

$\mu_{1k} \neq \mu_{2k}$ at least one pair of k

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, 2$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$$

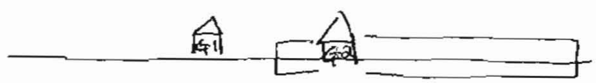
$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

$$T^2 = (\bar{x}_1 - \bar{x}_2)^T \left(S_p\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)^{-1} (\bar{x}_1 - \bar{x}_2)$$

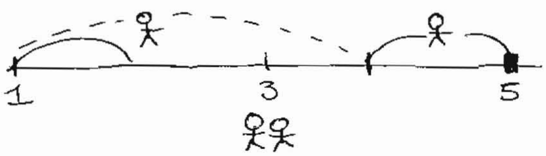
$$F_{p,\, n_1+n_2-p-1} = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}\, T^2$$

Hotelling's law

$ED = \frac{L}{8}$



$ED = \frac{L}{4}$



Multivariate ANOVA (MANOVA)
one-way

$$\vec{Y}_{ij} = \begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijp} \end{pmatrix} \quad \text{subject } j \text{ in group } i$$

Group 1	Group 2	Group t
Y_{11}	Y_{21}	Y_{t1}
Y_{12}	Y_{22}	Y_{t2}
$\vdots$	$\vdots$	$\vdots$
Y_{1n_1}	Y_{2n_2}	Y_{tn_t}

STAT 707
2/21/17
#3

n_i number of subjects in Group i

$N = n_1 + n_2 + n_3 \ldots + n_t$

1) common mean $\vec{\mu}_i$ (data from group i)
2) common covariance matrix $\Sigma \leftarrow$ Bartlett's test χ^2
3) Independence
4) Normality

Null : $\vec{\mu}_1 = \vec{\mu}_2 = \cdots = \vec{\mu}_t$

Ha : $\mu_{ik} \neq \mu_{jk}$ at least one $i \neq j$ and at least one variable k

$k = 1 \cdots p$

$$\vec{Y}_{i\cdot} = \frac{1}{n_i}\sum_{i=1}^{n_i} \vec{Y}_{ij} \qquad \vec{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{t}\sum_{j=1}^{n_i} \vec{Y}_{ij}$$

$$\underset{p \times p}{T} = \sum_{i=1}^{t}\sum_{j=1}^{n_i} \underbrace{(Y_{ij} - Y_{\cdot\cdot})}_{Y_{ij} - Y_{i\cdot} + Y_{i\cdot} - Y_{\cdot\cdot}}(Y_{ij} - Y_{\cdot\cdot})^T$$

$$T = \sum_{i=1}^{t}\sum_{j=1}^{n_i} (Y_{ij} - Y_{i\cdot})(Y_{ij} - Y_{i\cdot})^T \qquad E_{p \times p}$$

$$+ \sum_{i=1}^{t} n_i (Y_{i\cdot} - Y_{\cdot\cdot})(Y_{i\cdot} - Y_{\cdot\cdot})^T \qquad H_{p \times p}$$

$SS = SSE + SST$

$$F = \frac{SST / K-1}{SSE / N-K} = \frac{SST}{SSE} \cdot \frac{N-K}{K-1}$$

$HE^{-1} \cdot c \sim ?$

$\frac{p(p-1)}{2}$ $p \times p$

λ_i eigenvalues of HE^{-1} $i=1,\dots,p$

Wilks's Lambda (1932)

$$\frac{|E|}{|H+E|} = \prod_{i=1}^{p} \frac{1}{1+\lambda_i}$$

Hotelling - Lawley's trace (1951)

$$\text{trace}\,(HE^{-1}) = \sum_{i=1}^{p} \lambda_i$$

most robust Pillai's trace (1954)

$$\text{trace}\,(H(H+E)^{-1}) = \sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i}$$

Roy's maximum root (1953)

largest eigenvalue of $HE^{-1} = \max_{i=1}^{p} \{\lambda_i\}$

Wilk Λ

$$c \cdot \frac{1-\Lambda^d}{\Lambda^d} \sim F \text{ (approx)}$$

Hotelling U

$$c \cdot U \sim F \text{ (approx)}$$

Pillai V

$$c \cdot \frac{V}{d-V} \sim F \text{ (approx)}$$

Roy Θ

$$c \cdot \Theta \sim F \text{ (approx)}$$

STAT707
2/21/17
#4

two-way MANOVA without replication

Y_{ijk} observation for variable k
from block j in treatment i

$$\vec{Y}_{ij} = \vec{\mu} + \vec{\alpha}_i + \vec{\beta}_j + \vec{\epsilon}_{ij}$$

$$T = b\sum_{i=1}^{t}(y_{i\cdot}-y_{\cdot\cdot})(y_{i\cdot}-y_{\cdot\cdot})^T \leftarrow H_{p\times p}$$

$$+ \, t\sum_{j=1}^{b}(y_{\cdot j}-y_{\cdot\cdot})(y_{\cdot j}-y_{\cdot\cdot})^T \leftarrow B_{p\times p}$$

$$+ \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij}-y_{i\cdot}-y_{\cdot j}+y_{\cdot\cdot})(y_{ij}-y_{i\cdot}-y_{\cdot j}+y_{\cdot\cdot})^T \leftarrow E_{p\times p}$$

MANOVA

$\mathrm{trace}\,(H(H+E)^{-1})$
$\mathrm{trace}\,(B(B+E)^{-1})$

$\sim$

ANOVA

Trt	$F = MST/MSE$
Blk	$F = MSB/MSE$
Err	

Generally,

$$\underset{n\times p}{Y} = \underset{n\times r}{X}\ \underset{r\times p}{B}$$

$$P = X(X^TX)^{-1}X^T$$

$$\begin{array}{c|c} & 1\ 2 \cdots\ 5 \qquad\qquad r \\ \hline \begin{matrix}1\\2\\\vdots\\n\end{matrix} & \left[\begin{array}{c|c|c} X & X & X \end{array}\right] \end{array}$$

$$I = P_1 + P_2 + \cdots + P_5 + \underbrace{(I - P_1 - \cdots - P_5)}$$

$$E = X^T(I-P_1-\cdots-P_5)X$$

ANCOVA

a general linear model that includes both ANOVA (categorical) predictors and regression (continuous) predictors

Introducing further explanatory variables

regressions $y = X\beta + \epsilon$

augmented regression model

$$
\begin{aligned}
y &= X\beta + Z\gamma + \epsilon \\
&= (X, Z)\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \epsilon \\
&= W\delta + \epsilon
\end{aligned}
$$

→ column of Z are linearly independent of the column X

X : $n \times p$ rank p

Z : $n \times q$ rank q

$W = (X, Z)$

$\delta = \begin{pmatrix}\beta \\ \gamma\end{pmatrix}$

Lemma: $R = I - P = I - X(X^T X)^{-1} X^T$

then $Z^T R Z$ is positive definate.

proof $R = I - P$ is independent

$$x^T Z^T R Z x = 0$$

$$x^T Z^T R R Z x = 0$$

$$R Z x = 0$$

$$(1 - P) Z \vec{x} = 0$$

$$P Z \vec{x} = P \vec{x}$$

$$\underbrace{X(X^T X)^{-1} X^T Z \vec{x}}_{\vec{y}} = Z\vec{x}$$

$$\rightarrow \underbrace{X\vec{y} = Z\vec{x}}_{\vec{x} = 0}$$

STAT 707
2/21/17
#5

Thm $\quad R_A = I - W(W^TW)^{-1}W^T$

$L = (X^TX)^{-1}X^TZ$

$$\hat{\delta} = \begin{pmatrix} \hat{\beta}_A \\ \hat{\gamma}_A \end{pmatrix}$$

(i) $\hat{\gamma}_A = (Z^TRZ)^{-1}Z^TR\vec{y}$

(ii) $\vec{\beta}_A = (X^TX)^{-1}X^T(y - Z\hat{\gamma}_A) = \hat{\beta} - L\hat{\gamma}_A$

(iii) $y^TR_Ay = (y - Z\hat{\gamma}_A)^TR(y - Z\hat{\gamma}_A) = \underbrace{y^TRy}_{\text{original error term}} - \hat{\gamma}_A^TZ^TR\vec{y}$

STAT 707
2/28/17
#1

ANCONA

$$y = X\beta + \epsilon \quad \hat\beta$$

augumented $y = (X, Z)\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \epsilon \quad \begin{pmatrix}\hat\beta_A \\ \hat\gamma_A\end{pmatrix}$

$$= W\cdot\delta + \epsilon \qquad L = (X^TX)^{-1}X^TZ$$

$$R = I - P = I - X(X^TX)^{-1}X^T$$

$$R_A = I - P_A = I - W(W^TW)^{-1}W^T$$

Thm (i) $\hat\gamma_A = (Z^TRZ)^{-1}Z^TR\vec{y}$

(ii) $\hat\beta_A = (X^TX)^{-1}X^T(y - Z\hat\gamma_A) = \hat\beta - L\hat\gamma_A$

(iii) $Y^TR_AY = (Y - Z\hat\gamma_A)^TR(Y - Z\hat\gamma_A)$

$$= Y^TRY - \hat\gamma_A^TZ^TRY$$

Proof: $X\beta + Z\gamma$

$$= X\beta + PZ\gamma + (I-P)Z\gamma$$

$$= X(\beta + (X^TX)^{-1}X^TZ\gamma) + RZ\gamma$$

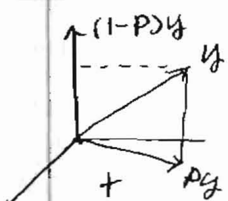
$$= (X, RZ)\begin{pmatrix}\alpha \\ \gamma\end{pmatrix} \qquad \alpha = \beta + (X^TX)^{-1}X^TZ\gamma$$

$$= V\lambda$$

$V\lambda = 0 \Leftrightarrow X\beta + Z\gamma = 0$

X, Z independent $\vec\beta = 0$, $\vec\gamma = 0$

$\lambda = 0$



V is Full-rank matrix $p+q$

$X^TR = 0$, $RX = 0$

$$(I - X(X^TX)^{-1}X^T)X = 0$$

$$X - X(X^TX)^{-1}X^TX = 0$$

Normal equation

$$V\lambda = Y$$

$$\hat{\lambda} = (V^TV)^{-1}V^TY \qquad (V^TV)\lambda = V^TY$$

$$V^T = (X, RZ)^T$$
$$= \begin{pmatrix} X^T \\ (RZ)^T \end{pmatrix}$$
$$= \begin{pmatrix} X^T \\ Z^TR \end{pmatrix}$$

$$V^TV = \begin{pmatrix} X^T \\ Z^TR \end{pmatrix} (X, RZ)$$

$$= \begin{pmatrix} X^TX & \overset{0}{\underset{\sim}{X^TRZ}} \\ \underset{0}{Z^TRX} & Z^TRZ \end{pmatrix} = \begin{pmatrix} X^TX & 0 \\ 0 & Z^TRZ \end{pmatrix}$$

$$(V^TV)^{-1} = \begin{pmatrix} (X^TX)^{-1} & \\ & (Z^TRZ)^{-1} \end{pmatrix}$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}+D^{-1}CMBD^{-1} \end{bmatrix}$$

where $M = (A - BD^{-1}C)^{-1}$

when $B * C = 0 \qquad M = A^{-1}$

$$\hat{\lambda} = \begin{pmatrix} (X^TX)^{-1} & \\ & (Z^TRZ)^{-1} \end{pmatrix} \begin{pmatrix} X^T \\ Z^TR \end{pmatrix} y = \begin{pmatrix} (X^TX)^{-1}X^TY \\ (Z^TRZ)^{-1}Z^TRY \end{pmatrix}$$

$$\hat{\gamma}_A = (Z^TRZ)^{-1}Z^TRY$$

(ii) $\hat{\alpha} = (X^TX)^{-1}X^Ty \leftarrow \hat{\beta}$

$\alpha = \beta + (X^TX)^{-1}X^TZ\gamma$

$\hat{\alpha} = \hat{\beta}_A + (X^TX)^{-1}X^TZ\hat{\gamma}_A$

$\hat{\beta}_A = \hat{\beta} - \underset{\sim}{(X^TX)^{-1}X^TZ}\hat{\gamma}_A$

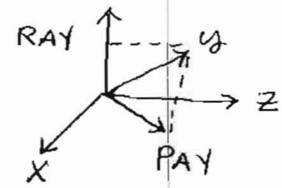
$= (X^TX)^{-1}X^TY - (X^TX)^{-1}X^TZ\hat{\gamma}$

$= (X^TX)^{-1}X^T(y - Z\hat{\gamma})$

$Y = X\beta + \epsilon \quad \leftarrow \hat{\beta}$

$Y - Z\gamma = X\beta + \epsilon \quad \leftarrow \hat{\beta}_A$

STAT 707
2/28/17
#2



(iii) $RAY = Y - X\hat\beta_A - Z\hat\gamma_A$

$= Y - X(X^TX)^{-1}X^T(Y - Z\hat\gamma_A) - Z\hat\gamma_A$

$= Y - P(Y - Z\hat\gamma_A) - Z\hat\gamma_A$

$= (I - P)(Y - Z\hat\gamma_A)$

$= R(Y - Z\hat\gamma_A)$

$Y^TRAY = Y^TR_ARAY = (Y - Z\hat\gamma_A)^T R \cdot R(Y - Z\gamma_A)$

$RAY = RY - RZ(Z^TRZ)^{-1}Z^TR\vec{Y}$

$Y^TRAY = Y^TRY - \hat\gamma_A^T Z^TRY$

Thm. (Sum of square decomposition)

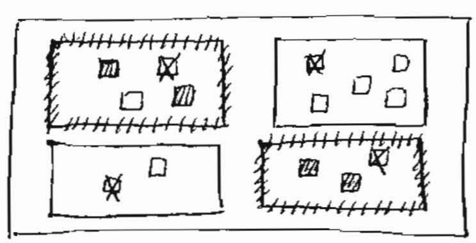
$SSE = SSE_A + \hat\gamma_A^T Z^TRY$

the sum of squares attributable to the new explanatory variables Z is $Y^TRZ\hat\gamma_A$.

Perform tests in a sequential manner.

HLM : applications

Multi-Stage sampling

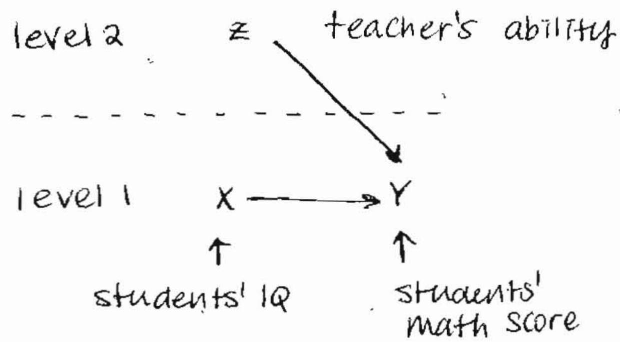


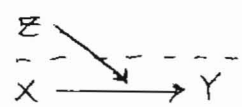
A common mistake in research is to ignore the fact that the sampling scheme was a two-stage one, and to pretend that the secondary unit were selected independently.

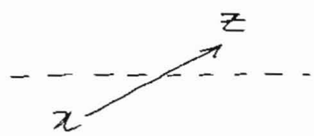
Multi-stage samples are preferred in practice because the cost of interviewing or testing persons are reduced enormously. If these persons are geographically or organizationally grouped.

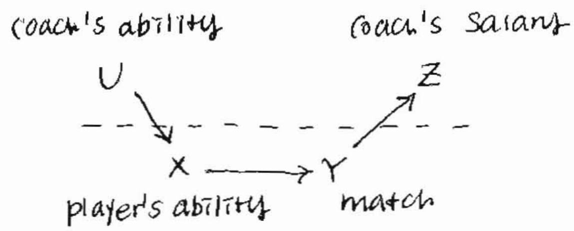
primary unit   Secondary unit

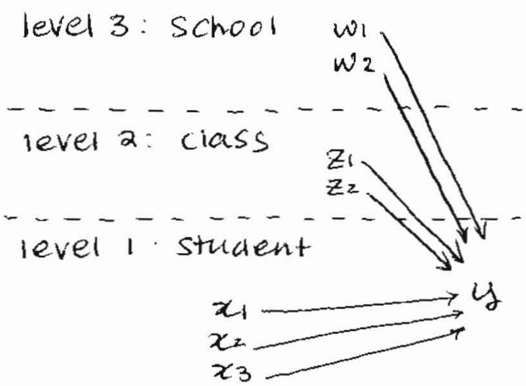
level-2 unit   level-1 unit





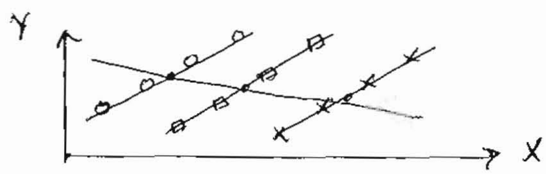






Statistical treatment of clustered data

→ Aggregation (simplest way)



the degree of resemblance between level-1 units belonging to the same level-2 unit can be expressed by the intraclass correlation coefficient (ICC)

random effect ANOVA model

$$Y_{ij} = \mu + U_j + R_{ij}$$

↑

observed for level-1 unit i within level-2 unit j

$U_j \sim N(0, \tau^2)$ population between group variance

$R_{ij} \sim N(0, \sigma^2)$ population within group variance

$$Var(Y_{ij}) = Var(U_j) + Var(R_{ij})$$
$$= \tau^2 + \sigma^2$$

STAT 707
2/28/17
#3

$$\text{cov}(Y_{ij}, Y_{i'j}) = \text{cov}(U_j + R_{ij}, U_j + R_{i'j})$$
$$= \text{Var}(U_j)$$
$$= \tau^2$$

$$\rho_I^2 = \frac{\text{population btw group variance}}{\text{total variance}} = \frac{\tau^2}{\tau^2+\sigma^2}, \quad 0 \le \rho_I < 1$$

$$F = \frac{MST}{MSE} \quad \begin{matrix} \leftarrow \text{between} & N-1 \\ \leftarrow \text{within} & M-N \end{matrix}$$

n_j : number of level-1 unit within j^{th} level-2 unit

N: number of level-2 units

total sample size is $M = \sum_{j=1}^{N} n_j$

$$\hat{\rho} = \frac{F-1}{F+\tilde{n}-1} \qquad \tilde{n} = \bar{n} - \frac{s^2(n_j)}{N\bar{n}}$$

$$s^2(n_j) = \frac{1}{N-1}\sum_{j=1}^{N}(n_j - \bar{n})^2$$

$$S^2_{\text{between}} = \frac{1}{\tilde{n}(N-1)}\sum_{j=1}^{N} n_j (Y_{\cdot j} - Y_{\cdot\cdot})^2$$

$$S^2_{\text{within}} = \frac{1}{M-N}\sum_{j=1}^{N}\sum_{i=1}^{n_j}(Y_{ij} - Y_{\cdot j})^2$$

$$\mathbb{E}S^2_{\text{between}} = \tau^2 + \frac{\sigma^2}{\tilde{n}}$$

$$\mathbb{E}S^2_{\text{within}} = \sigma^2$$

$$\hat{\sigma}^2 = S^2_{\text{within}}$$

$$\hat{\tau}^2 = \left\{ S^2_{\text{between}} - \frac{S^2_{\text{within}}}{\tilde{n}} \right\} \vee 0$$

$$\hat{\rho}_I = \frac{\hat{\tau}^2}{\hat{\sigma}^2 + \hat{\tau}^2}$$

In the simple random sample

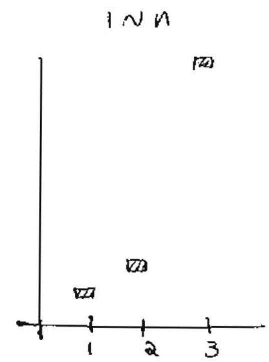
$$\text{Standard Error} = \frac{\text{Standard deviation}}{\sqrt{\text{sample size}}}$$

$n_j = n$

design effect : $1 + (n-1)\rho_I$

indicate how much the sample size is to be adjusted because of the sampling design used

$$M_{\text{Effective}} = \frac{Nn}{\text{design effect}}$$



$M = Nn$

$$N_{ES} = \text{design effect} \cdot N_{srs}$$
$$= N_{srs} + (n-1)\rho_I \cdot N_{srs}$$

STAT 707
3/7/17
#1

HLM · application

The Random intercept Model

j = the index for group ($j = 1, 2 \cdots N$)

i = the index for the individual within the group ($i = 1, 2 \cdots n_j$)

For individual i in group j

Y_{ij} dependent variable

x_{ij} the explatory variable at individual level

For group j

z_j the explatory variable at the group level

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + R_{ij}$$ ← error

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \beta_2 z_j + R_{ij}$$

OLS.

nesting structure has no effect.

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij}$$

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij}$$

group dependent intercept

be split into an average intercept and the group deviation

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_1 = \gamma_{00}$$

$$Y_{ij} = \gamma_{00} + U_{0j} + \gamma_{10} x_{ij} + R_{ij}$$

$$= \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij}$$

U_{0j} as fixed parameter, $\sum_{j=1}^{N} U_{0j} = 0$

$N-1$ parameters

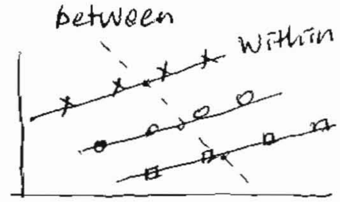
It's impossible to use a group-level variable z_j as an explanatory variable because it would be redundant given the fixed group effect.

r level-1 explanatory variable $\qquad q$ level -

Include group size as an explanatory variable group average

within-group regression

between-group regression



$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij}$$

$$\frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j} = \frac{\sum_{i=1}^{n_j} \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij}}{n_j}$$

$$\gamma_{00} + \underline{\underline{\gamma_{10}}}\, \bar{x}_{\cdot j} + \underline{U_{0j} + R_{ij}} = Y_{\cdot j}$$

The average in group j

$$z_j = \bar{x}_{\cdot j}$$

$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \underline{\gamma_{01} \bar{x}_{\cdot j}} + U_{0j} + R_{ij}$$

within group regression coefficient is allowed to differ ~~than~~ from the between-group regression coefficients.

$$Y_{ij} = (\gamma_{00} + \gamma_{01} \bar{x}_{\cdot j} + U_{0j}) + \gamma_{10} x_{ij} + R_{ij}$$

within group line $\quad Y = (\gamma_{00} + \gamma_{01} \bar{x}_{\cdot j}) + \gamma_{10} x_{ij}$

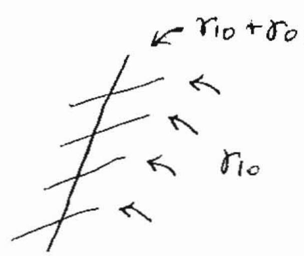
Take the group average

$$\bar{Y}_{\cdot j} = \gamma_{00} + (\gamma_{10} + \gamma_{01})\, \bar{x}_{\cdot j} + U_{0j} + \bar{R}_{\cdot j}$$

$$Y = \gamma_{00} + \underline{(\gamma_{10} + \gamma_{01})}\, x$$

test if between-group & within-group coefficient are the same or not.

$H_0 \cdot \ \gamma_{01} = 0 \qquad \gamma_{01} > 0$



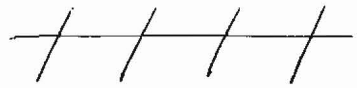
replace x_{ij} by the within-group deviation score

$$x_{ij} - \bar{x}_{\cdot j}$$

3/7/17
#3

31

$$Y_{ij} = \tilde{\gamma}_{00} + \tilde{\gamma}_{10}(X_{ij} - \bar{x}_{\cdot j}) + \underbrace{\tilde{\gamma}_{01}\bar{x}_{\cdot j}} + U_{0j} + R_{ij}$$



$$Y_{ij} = \gamma_{00} + \tilde{\gamma}_{10} X_{ij} + (\tilde{\gamma}_{01} - \tilde{\gamma}_{01})\bar{X}_{\cdot j} + U_{0j} + R_{ij}$$

$\tilde{\gamma}_{10} = \gamma_{10}$

$\tilde{\gamma}_{01} - \tilde{\gamma}_{10} = \gamma_{01}$ $\Rightarrow$ $\begin{cases} \tilde{\gamma}_{10} = \gamma_{10} & \text{slop of within-group reg} \\ \tilde{\gamma}_{01} = \gamma_{01} + \gamma_{10} & \text{slop of between-group re} \end{cases}$

parameter estimation

ML (maximum likelihood)

REML (restricted / residual ML)

differ little w.r.t estimating the regression coefficient

differ variance componant

Three-level Random Intercept Model

$Y_{ijk} = \beta_{0jk} + \beta_1 x_{ijk} + R_{ijk}$ $\quad Var(R_{ijk}) = \sigma^2$

$\beta_{0jk} = \delta_{00k} + U_{0jk}$ $\quad var(U_{0jk}) = \tau^2$ indep.

$\delta_{00k} = \gamma_{000} + V_{00k}$ $\quad var(V_{00k}) = \varphi^2$

We can define 1st deviation

$(X_{ijk} - \bar{x}_{\cdot jk})$

within class

$(\bar{X}_{\cdot jk} - \bar{x}_{\cdot\cdot k})$

between class

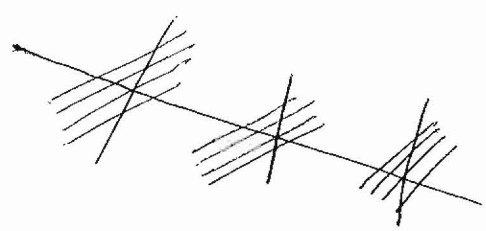
$\bar{X}_{\cdot\cdot k}$

between school

school

class

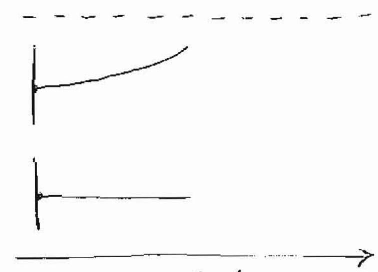
student



STAT 707
3/21/17
#1

Apr. 20 5:35-7:25 Thursday (Make-up class) 921HE

HLM application
Some subjects progress faster than others heterogeneity of regressions across groups it's modeled by random slope.



$\rightarrow Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij}$ β_{0j}, β_{1j} group-dependent

$\rightarrow \beta_{0j} = \gamma_{00} + U_{0j}$

$\beta_{1j} = \gamma_{10} + U_{1j}$ $\leftarrow$ random interaction between group and X

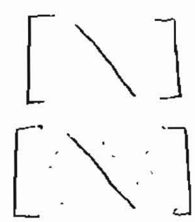
$Y_{ij} = \gamma_{00} + U_{0j} + \gamma_{10} x_{ij} + \underline{U_{1j} x_{ij}} + R_{ij}$

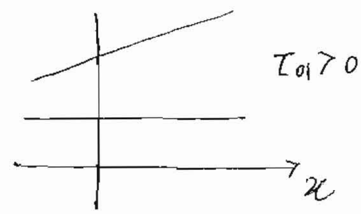
$Var(R_{ij}) = \sigma^2$

$Var(U_{0j}) = \tau_0^2 = \tau_{00}$

$Var(U_{1j}) = \tau_1^2 = \tau_{11}$

$Cov(U_{0j}, U_{1j}) = \tau_{01}$ $\leftarrow = 0$

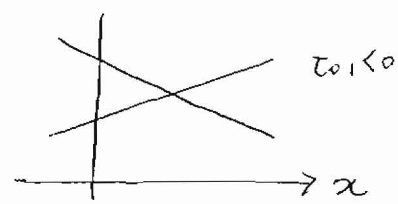




R_{ij} are i.i.d

$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix}$ are i.i.d

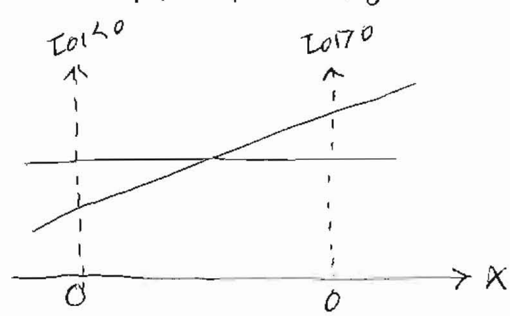
R_{ij} and $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix}$ are independent.



$Y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2j} + \cdots + \beta_{pj} x_{pj} + R_{ij}$

$\begin{pmatrix} U_{0j} \\ U_{1j} \\ \vdots \\ U_{pj} \end{pmatrix}$ $p+1$

$\frac{(p+1)p}{2}$



The intercept variance τ_{00} and intercept-by-slope covariance τ_{01} depend on the origine for the x-variable.

The Scale for X is defined so that $x=0$ has a well-interpretable meaning.

e.g. when X refers to time, let $x=0$ correspond to the start/end

In nesting structures of individuals within group

let $x=0$ correspond to the overall mean of population/sample

$\underbrace{IQ - 100}$ $\cancel{IQ_{ij} - IQ_{\cdot j}}$

grand-mean centered

In a random slope model, the within-group coherence cannot be simply expressed by the intraclass correlation coefficient or its residual version.

If we wish to reduce the unexplained variability associated with U_{0j} & U_{1j}.

we expand equations by predicting the group-dependent regression coefficients β_{0j} & β_{1j} from level-2 variables

① $\beta_{0j} = \gamma_{00} + U_{0j}$

$\beta_{1j} = \gamma_{10} + U_{1j}$

② $\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + U_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} z_j + U_{1j}$

$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \beta_{ij}$

$$Y_{ij} = \gamma_{00} + \gamma_{01} z_j + U_{0j} + \gamma_{10} x_{ij} + \underbrace{\gamma_{11} z_j x_{ij}}_{\text{cross-level interaction}} + U_{1j} x_{ij} + R_{ij}$$

It's not contradictory to look for a specific cross-level interaction even if no significant random slope was found.

STAT 707
3/21/17
#2

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \cdots + \beta_{pj} x_{ij} + R_{ij}$$

$$\beta_{hj} = \gamma_{h0} + \gamma_{h1} Z_{1j} + \cdots + \gamma_{hq} Z_{qj} + U_{hj}$$

$h = 0, 1, \cdots p$

$$Y_{ij} = \gamma_{00} + \sum_{h=1}^{p} \gamma_{h0} x_{hij} + \sum_{k=1}^{q} \gamma_{0k} Z_{kj} + \sum_{k=1}^{q} \sum_{h=1}^{p} \gamma_{hk} Z_{kj} x_{hij}$$

$$+ U_{0j} + \sum_{h=1}^{p} U_{hj} x_{hij} + R_{ij}$$

$var(R_{ij}) = \sigma^2$

$var(U_{hj}) = \tau_{hh} = \tau_h^2$

$cov(U_{hj}, U_{kj}) = \tau_{hk}$

$$\begin{bmatrix} \tau_{00} & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & \tau_{pp} \end{bmatrix}_{(p+1)\times(p+1)}$$

i	j	Y	$x_1 \cdots x_p$	$Z_1 \cdots Z_q$	$Z_1 x_1 \cdots Z_q x_q$
		1 columns	p columns	q columns	pq columns

Whether a level-one variable was obtained as a cross-level interaction or not is immaterial to the computer program.

Difference between level-1 and level-2 variables is not of any importance for the parameter estimation.

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \gamma_{01} \bar{x}_{\cdot j} + R_{ij}$$

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_{1j} = \gamma_{00} + \cancel{U_{hj}}$$

group-mean centered random slope model

$$Y_{ij} = \tilde{\beta}_{0j} + \boxed{\tilde{\beta}_{1j}(x_{ij} - \bar{x}_{\cdot j})} + \tilde{\gamma}_{01} \bar{x}_{\cdot j} + R_{ij}$$

$$\tilde{\beta}_{0j} = \tilde{\gamma}_{00} + U_{0j}$$

$$\tilde{\beta}_{1j} = \tilde{\gamma}_{10} + \cancel{U_{hj}}$$

$$Y_{ij} = \underline{\gamma_{00}} + U_{0j} + \underline{\gamma_{00} x_{ij}} + U_{1j} x_{ij} + \underline{\gamma_{01} \bar{x}_{\cdot j}} + R_{ij}$$

$$Y_{ij} = \underline{\tilde{\gamma}_{00}} + \tilde{U}_{0j} + \underline{\tilde{\gamma}_{10} x_{ij}} + U_{1j} x_{ij} + \underline{(\tilde{\gamma}_{01} - \tilde{\gamma}_{10}) \bar{x}_{\cdot j}} - U_{1j} \bar{x}_{\cdot j} + R_{ij}$$

One should be reluctant to use group-mean centered random slope models unless there is a clear theory that not in the first place that the absolute score x_{ij} but rather the relative score $x_{ij} - \bar{x}_{\cdot j}$ is related to Y_{ij}

* teacher's rating of student performance

$Y_{ijk} = \beta_{0jk} + \beta_{1jk} x_{ijk} + R_{ijk}$ level-1

add $+ \delta_{01k} z_{jk}$

$\beta_{0jk} = \delta_{00k} + U_{0jk}$

$\beta_{1jk} = \delta_{10k} + U_{1jk}$ $+ \delta_{11k} z_{jk}$] → level-2

$\delta_{00k} = \gamma_{000} + V_{00k}$

$\delta_{10k} = \gamma_{100} + V_{10k}$] → level-3

$\delta_{01k} = \gamma_{010} + V_{01k}$

$\delta_{11k} = \gamma_{110} + V_{11k}$

In R

nlme

lme4

SES → SES+5

SES → SES−5

x_{ij} SES, Y_{ij} Math Achieve, Z_j MEANSES

$\to \sigma^2$

$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij}$

$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + U_{0j}$ $\to \tau_0^2$

$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + U_{1j}$ $\to \tau_1^2$

$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + U_{0j} + \gamma_{10} x_{ij} + \gamma_{11} x_{ij} Z_j + U_{1j} x_{ij} + R_{ij}$

τ_0^2 8.83

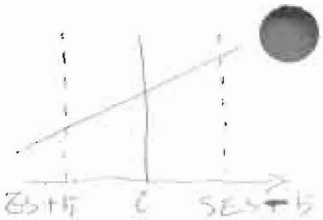
τ_1^2 0.4129

σ^2 36.83

−0.11

γ_{00} 12.665

γ_{10} 2.3938



STAT707
3/21/17
#3

Test for fixed parameters

$H_0 \cdot \quad \gamma_h = 0$

$$T(\gamma_h) = \frac{\hat{\gamma}_h}{SE(\hat{\gamma}_h)}$$

the approximation by the t-distribution is not exact
if the normality assumption for the random coefficient holds.

Coefficient of a level-1 variable

$\square\ x_{ij} \qquad \square\ x_{ij} z_j$

$d.f. = M - r - 1$

total number of level-1 unit

total number of explanatory variables

$7185 - 3 - 1$

Coefficient of a level-2 variable

$\square\ z_j$

$d.f = N - q - 1$

number of level-2 unit

number of exp var at level 2

$100 - 1 - 1$

STAT 707
3/28/17
#1

HLM: application
tests for the fixed parameters

1. test a single fixed parameter

$H_0 : \gamma_h = 0$

$T : \dfrac{\hat{\gamma}_h}{S.E(\hat{\gamma}_h)} \xrightarrow{\text{normal dist}}$ ✓ p-value

d.f level-1 coeff ← # of explanatory variables

$d.f = M - r - 1$

level-1 x_{hij}, $x_{hij} z_{kj}$

level-2 z_{kj}

↑ total # of level-1 unit

$\sum_{j=1}^{N} n_j$

level-2 coeff

$d.f = N - q - 1$

↑ total # of level-2 unit

2. Multi-parameter tests for fixed effect

$\boxed{H_0 : \vec{\gamma} = 0}$

↑ q regression parameters

$(\gamma_1, \gamma_2, \cdots \gamma_q)'$

→ Wald χ^2 test

$\hat{\gamma}^T \hat{\Sigma}^{-1} \hat{\gamma} \xrightarrow{\text{approximated}} \chi^2(q)$

↑

$\hat{\gamma}^T \Sigma^{-1} \hat{\gamma} \sim \chi^2(q)$

γ $\boxed{\tau^2_h \quad \sigma^2}$

deviance $= -2 \cdot \log L$

a measure of lack of fit between model and data

One interpret only differences in deviance values for several models fitted to the same data set

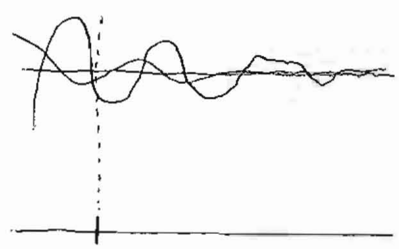
M_0 with m_0 parameters deviance D_0
M_1 m_1 D_1

add $m_1 - m_0$ parameters
$D_0 - D_1 \sim \chi^2(m_1 - m_0)$
$\uparrow \quad \uparrow$
ML estimation

~~Wald χ^2 test~~
deviance test

The deviance produced by the REML method can be used in deviance tests only if the two models compared how the same fixed parts and only differ in their random parts.

Wald and deviance tests are very close to each other for intermediate / large sample size



Halved p-value for variance parameters variance are by definition non-negative.

$$H_0 \cdot \tau_p^2 = 0 \Rightarrow \tau_{0p} = \tau_{1p} = \cdots = \tau_{p-1,p} = 0$$

(P)

random part $\quad U_{0j} + x_{1ij} U_{1j} + \cdots + x_{pij} U_{pj} + R_{ij}$

$(p+1)$ parameters $\begin{bmatrix} U_{0j} \\ \vdots \\ U_{pj} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma_U \right)$

3/28/17
#2

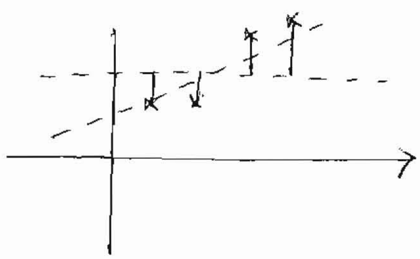
$D_0 - D_1 \sim \chi^2(p+1)$
halt this tail probability

① Including a random slope implies inclusion of the fixed effect
② Be reluctant to include random slopes for interactions.
$x_{hij} \cdot z_{kj}$
$x_{hij} \cdot x_{pij}$

$$R^2 = \cancel{\frac{SSR}{SSE}} \quad \frac{SSR}{SS_{Tot}} = 1 - \frac{SSE}{SS_{Tot}}$$

proportional reduction of prediction error

$$R^2 = \frac{Var(Y_i) - Var(Y_i - \sum_h \beta_h x_{hi})}{Var(Y_i)}$$



Consider a two-level random effects model w/ a random intercept and some predictor variables w/ fixed effect but no other random effects.

$$Y_{ij} = \gamma_0 + \sum_{h=1}^{g} \gamma_h x_{hij} + U_{0j} + R_{ij}$$

level-1: the proportional reduction of error ~~of~~ for predicting an individual outcome.

$$R_1^2 = 1 - \frac{Var(Y_{ij} - \sum_h \gamma_h x_{hij})}{Var(Y_{ij})}$$

empty model
$Y_{ij} = \gamma_0 + U_{0j} + R_{ij}$
$\uparrow$
$\hat{\sigma}^2 + \hat{\tau}_0^2$

$Var(Y_{ij}) = Var(U_{0j}) + Var(R_{ij})$

level - 2 : the proportional reduction of error for predicting a group mean

$$R_2^2 = 1 - \frac{var(\bar{Y}_{\cdot j} - \Sigma_h \gamma_h \bar{x}_{h\cdot j})}{var(\bar{Y}_{\cdot j})} \qquad \sigma^2 + \frac{\tau_0^2}{n}$$

Assumptions of the Hirerarechial Linear Model

$$Y_{ij} = \gamma_0 + \sum_{h=1}^{r} \gamma_h x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} x_{hij} + R_{ij} \qquad p \leq r$$

① Does the Fixed part contain the right variables, X_1 to X_r ?

② Does random part , X_1 to X_p ?

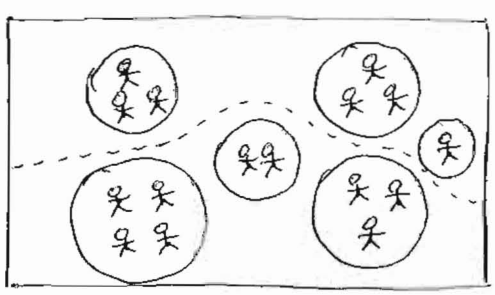
③ Are the level-one residuals normally distributed ?

④ Do the level - one have constant variance ?

⑤ Are the level - two random coefficients normally distributed ?

⑥ Do have a constant covariance matrix ?

Cross - Validation



For a two-level model, two independent halves are obtained by randomly distributing the level-2 units into two subsets.

3/28/17
#3

Model Checks

1 Include Contextual effect (group mean)

2 Check whether variables have random effects

3 Explained variance R_1^2, R_2^2

IF the fraction of explained variance decreases when a fixed effect is added to the model, this can be a sign of misspecification

a small decrease can be a result of chance fluctuations

a decrease by a magnitude of ≥ 0.05 is a warning of possible misspecification

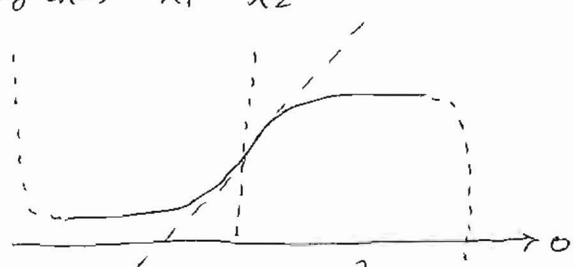
Ans Q1 Specification of the fixed part

what is the right set of variables in the fixed part depends ~~on~~ in the first place on the domain in which the explanation is being considered.

Transformation

1. aggregation to group mean
2. calculating product or other interaction variables
3. non-linear transformation

$Y \quad x_1 \quad x_2 \quad \bar{x}_1 \quad x_1x_2 \quad 1/x_1 \quad \log(x_2) \quad x_1^2 \quad x_2^2$



$$x_{1+}^2 = \begin{cases} x_1^2, & x_1 \geq 0 \\ 0, & x_1 < 0 \end{cases}$$

$$x_{1-}^2 = \begin{cases} x_1^2, & x_1 \leq 0 \\ 0, & x_1 > 0 \end{cases}$$

Ans Q2 Specification of the random

① If certain variables are mistakenly omitted from the random part, the test of their fixed coefficients may also be unreliable

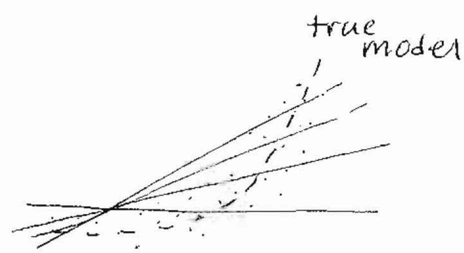
② A misspecification of the fixed part of the model can lead to a misspecification of the random part.

Eg. an ~~incorect~~ incorrect fixed part shows up in an unnecessarily complex random part.

Fixed	random
x_1	x_1
x_2	x_2
x_3	x_3
x_4	$\vdots$
x_5	
$\vdots$	$\vdots$
x_r	x_r

Use this to genera[...] detec[...]

$$
\begin{cases}
Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + R_{ij} \\
\beta_{0j} = \gamma_{00} + U_{0j} \\
\beta_1 = \gamma_{10} \\
\beta_2 = \gamma_{20}
\end{cases}
$$



Wrong model to use

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij}$$
$$\beta_{0j} = \gamma_{00} + U_{0j}$$
$$\beta_{1j} = \gamma_{10} + U_{1j}$$

Fixed		random
$1+x$	$\longrightarrow$	$1+x$
Then choose $1+x+x^2$	$\longleftarrow$	
	$\longrightarrow$	$\cancel{1+x}$
		$\downarrow$
		1

3/28/17

#4

Testing for heteroscedasticity 이분산성

the residuals R_{ij} and U_{hj} are assumed to be homoscedastic

1. variances depends on an explanatory variable

$$\sigma^2(x) = \sigma_0^2 + 2\sigma_{01}\, x$$

2. without a specific connection to some explanatory variable based on the estimated least-squares residuals within each group.

   only level-1 explanatory variables are considerd,
   level-2 variables are disregarded.

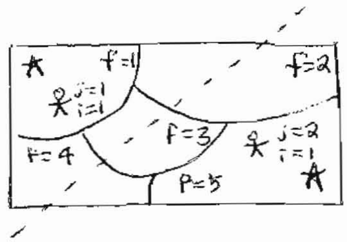
$$Y_{ij} = \gamma_0 + \gamma_1 x_{1ij} + \gamma_2 x_{2ij} + \gamma_3 z_j + \gamma_4 U_{1j} + U_{0j}$$

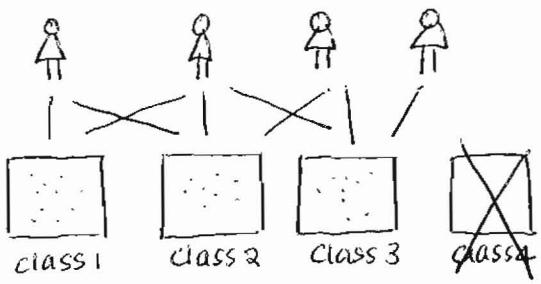
$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} \quad \text{for any } j$$

HLM: application

Crossed Random Effect

nested Random Effect





In a study of Students (i) within schools (j), neighborhood also have effects.

In a study of students within classes, classes are taught by several teachers.

Student i, nested within school j

with a crossed random effect, neighborhood f, $f = 1 \dots F$

$$Y_{ij} = \gamma_0 + \sum_{h=1}^{r} \gamma_h x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} x_{hij} + R_{ij} + W_{f(i,j)}$$

$W_1, W_2 \cdots W_F$

the usual assumption is made that W_1 to W_F are mutually independent random variables.

also independent of the other random effects (U and R)

all $W_f \sim N(0, \tau_W^2)$

Interpretation: part of the variability in the dependent variable is accounted for by neighborhood.

Model crossed Random Effects

Dummy variables b_f ($f=1,\dots,F$) are defined

$$b_{fij} = \begin{cases} 1 & \text{if student } i \text{ in school } j \text{ lives in neighborhood } f. \\ 0 & \text{lives in a different neighborhood} \end{cases}$$

$$W_{f(i,j)} = \sum_{f=1}^{F} W_f b_{fij}$$

the slope variance satisfy the restrictio

$Var(W_1) = Var(W_2) \cdots = Var(W_F)$

$W_{f(1,1)} = W_1$

$b_{111} = 1$

$b_{211} = 0$

$b_{311} = 0$

$b_{411} = 0$

$b_{511} = 0$

1. Random Slops in a crossed design

$W_{0f} + W_{1f}x_{ij}$

$W_{0f(ij)} + W_{1f(ij)} \cdot x_{ij}$

can be written as

$$\sum_{f=1}^{F}\left(W_{0f} b_{fij} + W_{1f} \cdot b_{fij} \cdot x_{ij}\right)$$

$W_{0f} \sim N(0, \tau^2_{W_0})$

$W_{1f} \sim N(0, \tau^2_{W_1})$

For different value of f, these random slopes are uncorrelated.

The slop of b_f and $b_f X$ for each single, f, is correlated.

$Var(W_{01}) = \dots = Var(W_{0F})$

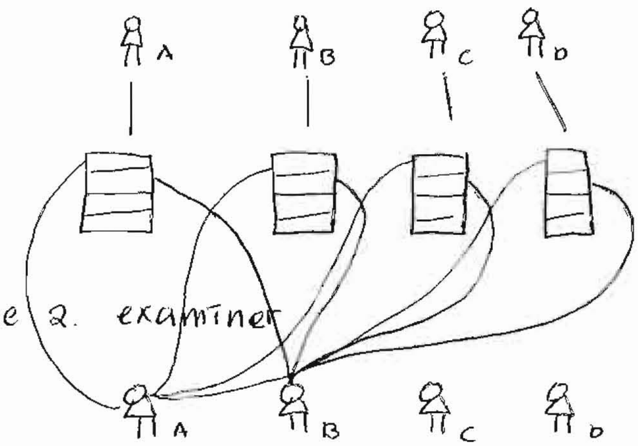
$Var(W_{11}) = \dots = Var(W_{1F})$

$Cov(W_{01}, W_{11}) = \dots = Cov(W_{0F}, {}_{1F})$

STAT 707
4/4/17
p2

2. Multiple Roles

role 1. teacher



role 2. examiner

student ID	teacher ID	examiner
1	1	1
2	1	2
3	1	3
4	2	1
5	2	
6		

$b_{fi} = \begin{cases} \cancel{\text{If}}\ 1, & \text{if student } i \text{ } \cancel{\text{uses}} \text{ was taught by person } f \\ 0, & \text{another person} \end{cases}$

$c_{fi} = \begin{cases} 1, & \text{if student } i \text{ was } \cancel{\text{taught}} \text{ examined by person } f \\ 0, & \text{another person} \end{cases}$

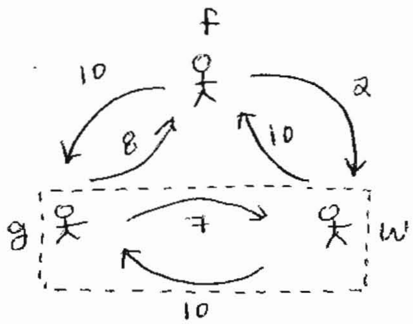
$\underbrace{W_{0,f(i)} + W_{1,g(i)}}$

$\sum_{f=1}^{F} (W_{0f}\, b_{fi} + W_{1f}\, c_{fi})$

$\underbrace{W_{01}} + W_{13}$ For $i=3$

$f(3)=1$
$g(3)=3$

3. Social Network



reciprocity : if f likes g, than chances are higher that g also likes f.

popularity : indicated by a person receiving high scores from the other.

outgoingness / activity : indicated by a person giving high scores to others

$$Y_{fg} = \mu + \underbrace{A_f}_{\text{outgoingness}} + \underbrace{B_g}_{\text{popularity}} + \underbrace{U_{fg}}_{\text{reciprocity}} + \underbrace{R_{fg}}_{\text{random error}}$$

$U_{fg} = U_{gf}$

pair (Y_{fg}, Y_{gf}) pair j ; 1 to $\frac{F(F-1)}{2}$ $\binom{F}{2}$

two relations $i = 1, 2$

$S_{fij} = 1$ if person f is the sender

$r_{fij} = 1$ if person f is the receiver

$S_{f1j} = 1$ $\quad F_{f2j} = 1$

$$\left[\, F \overset{i=1}{\underset{i=2}{\rightleftharpoons}} g \,\right] \quad j$$

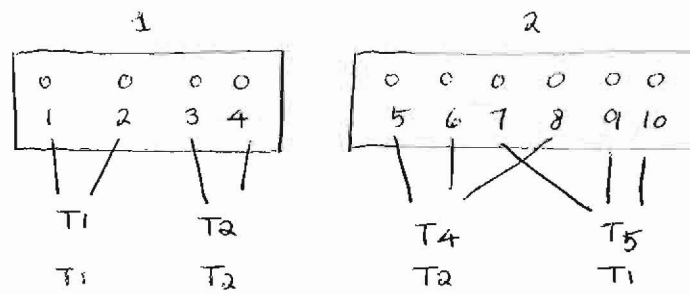
$S_{f2j} = 1 \Rightarrow r_{f1j} = 1$

STAT 707
4/4/17
P3

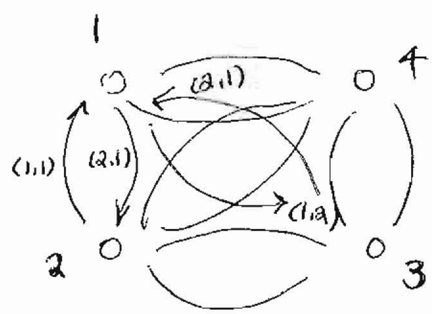
$$Z_{ij} = \mu + \sum_{f=1}^{F} (A_f s_{fij} + B_f r_{fij}) + U_j + R_{ij}$$

$$Var(A_1) = \cdots = Var(A_F)$$
$$Var(B_1) = \cdots = Var(B_F)$$
$$Cov(A_1 B_1) = \cdots = Cov(A_F B_F)$$

In R



* Results from Model 4, Model 5 and Model 1 are equivalent.



i	j	s	k
1	1	1	2
2	1	2	1
1	2	1	3
2	2	3	1
1	3	1	4
2	3	4	1
1	4	2	3
2	4	3	2
1	5	2	4
2	5	4	2
1	6	3	4
2	6	4	3

Model

$$Z_{ij} = \mu + \sum_{f=1}^{F} (A_f s_{fij} + B_f r_{fij}) + U_j + R_{ij}$$

$$Var(A_1) = \cdots = Var(A_F)$$
$$Var(B_1) = \cdots = Var(B_F)$$
$$Cov(A_1 B_1) = \cdots = Cov(A_1 B_1) = 0$$

4/20/17
STAT 707
#1

HLM : applrcation

Ordinary linear model $Y = f(x) + \epsilon$

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

Testing for heteroscedasticity

residuals

Y_{ij} = Fixed part $+ \sum U_{hj} x_{hij} + R_{ij}$

R_{ij} ↑ level 1 $\quad$ U_j ↑ level 2 residual

$R_{ij} \sim N(0, \sigma^2)$ $\quad$ $U_j \sim N(0, \tau_0^2)$

$n_j - r - 1$ is not too small $(\geq 10, \geq 5)$

S_j^2 : the resulting estimated residual variance for group j

$j=1$ · $Y_{11}, Y_{21}, Y_{31}, \cdots Y_{n,1}$ $\quad S_1^2$

j $\quad Y_{1j}, Y_{2j} \cdot \quad Y_{n_j j}$ $\quad S_j^2$

Y_{i1} = Fixed part $+ R_{i1}$

Weighted average of the logorithm

$$\frac{\sum_j df_j \cdot \log(S_j^2)}{\sum_j df_j} \sim \log(\sigma^2)_{ML}$$

a standardized residual dispersion

$$d_j = \sqrt{\frac{df_j}{2}} \cdot \left(\log(S_j^2) - \frac{\sum_j df_j \log(S_j^2)}{\sum_j df_j}\right)$$

$\sim N(0, 1^2)$

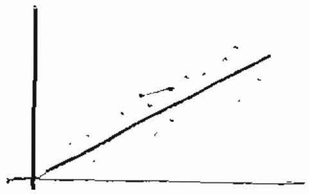
omit the residuals for group with low with-in group d.f

$$\sum_{j=1}^{n} d_j^2 \sim \chi^2(N-1)$$

to check if σ^2 is a constant

Inspection of level-1 residuals

$Y = f(x) + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$



Q-Q plot

observed

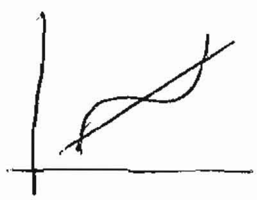
expected

$(z_i, \tilde{\gamma}_i)$

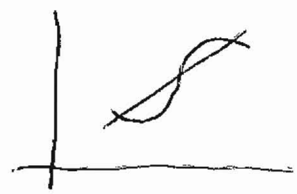
raw OLS residuals

standardized OLS residual

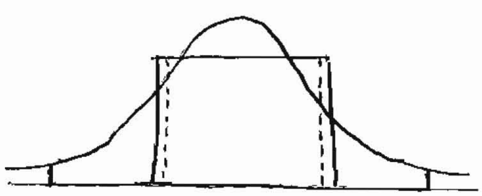
divided by their standard deviation



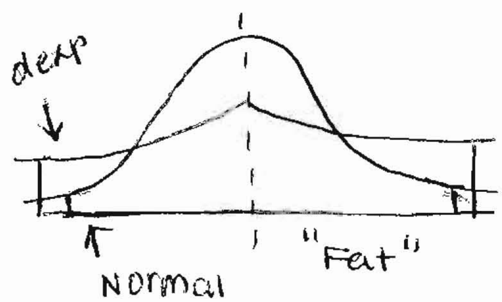
fat tail



thin tail



"thin"



longitudinal data

teacher / | \ student

individual / | \ observation

GLMM

GEE

4/20/17
STAT 707
#2

repeated measurements

$t = 1, 2 \cdots m$

Fixed occasion design — week 0, week 4, week 24

Variable occasion design

person 1	0, 1, 2	missing value	3
person 2	0, 2, 3	"	1
	0, 1, 2 3		

person 1 — 0, 50, 120

person 2 — 2, 48, 101

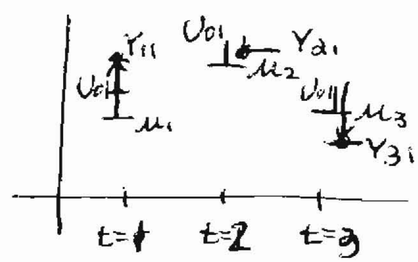
) too many missing value
NOT a good idea to
use fixed occasion design

$t = 1 \cdots m$ — level-1 unit

i — level-2 unit

We assume that the absent data are missing at random the fact that they are missing does not itself provide relevant information about the studied phenomena.

Compound symetry model



$$Y_{ti} = \mu_t + U_{0i} + R_{ti}$$

$U_{0i} \sim N(0, \tau_0^2)$

$R_{ti} \sim N(0, \sigma^2)$

U_{0i} & R_{ti} are independent

$$d_{hti} = \begin{cases} 1, & t = h \\ 0, & t \neq h \end{cases}$$

$$\mu_t = \sum_{h=1}^{m} \mu_h d_{hti}$$

$$Y_{ti} = \underbrace{\sum \mu_h d_{hti}}_{+ \sum_{k=1}^{8} \alpha_k Z_{ki} + \sum_{k=1}^{8} \gamma_k Z_{ki} S(t)} + U_{0j} + R_{ti}$$

	$t=1$	$t=2$	$t=3$
p1	2 mg/mL	4	3
p2	0.5	3	1.2

Y	ID	d_1	d_2	d_3
2	1	1	0	0
4	1	0	1	0
3	1	0	0	1
0.5	2	1	0	0
3	2	0	1	0
1.2	2	0	0	1

$$\vec{Y}_i = \begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix}$$

$$\mathrm{cov}(Y_i) = \begin{bmatrix} \tau_0^2+\sigma^2 & \tau_0^2 & \cdots & \tau_0^2 \\ \tau_0^2 & \tau_0^2+\sigma^2 & \cdots & \tau_0^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_0^2 & \tau_0^2 & \cdots & \tau_0^2+\sigma^2 \end{bmatrix}$$

$F(Y_1 \cdots Y_n)$
$= F(Y_{n_1}, \cdots Y_{n_n})$
exchangable

IF measurements are ordered in time.
the correlation is often larger between nearby measurements than between measurements that are for apart.

4/20/17
STAT 707
#3

Random Slope

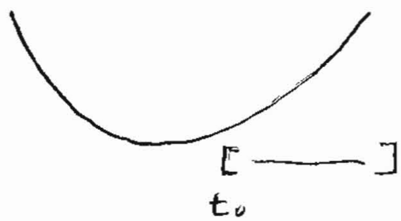
$$Y_{ti} = \text{Fixed part} + U_{0i} + U_{1i}(t-t_0) + R_{ti}$$

the value t_0 is subtracted from t in order to let the intercept variance refer to the reference point, t_0.

$$var(Y_{ti}) = \tau_0^2 + 2\tau_{01}(t-t_0) + \tau_1^2(t-t_0)^2 + \sigma^2$$
$$cov(Y_{ti}, Y_{si}) = \tau_0^2 + \tau_{01}(t-t_0+s-t_0) + \tau_1^2(t-t_0)(s-t_0)$$

$$t - t_0 = -\frac{2\tau_{01}}{2\tau_1^2}$$

$$t = t_0 - \frac{\tau_{01}}{\tau_1^2}$$



$$U_{0j} + U_{1j}(t-t_0) + U_{2j}(t-t_0)^2$$

U_{0j}	2	σ^2, τ_0^2
$U_{0j} + U_{1j}$	4	
$U_{0j} + U_{1j} + U_{2j}$	7	

q

$$\frac{(q+1)(q+2)}{2} + 1$$

m $\quad t = 1, 2 \cdots m$

$$\boxed{\frac{(m+1)\cdot m}{2}}$$

Fully multivariate model

$$Y_{ti} = \text{Fixed part} + U_{ti}$$

$$d_{hti} = \begin{cases} 1 & h = t \\ 0 & h \neq t \end{cases}$$

$$Y_{ti} = \text{Fixed part} + \sum_{h=1}^{m} U_{hi} d_{hti}$$

$$Y_{ti} = \text{Fixed part} + \sum_{h=1}^{m} U_{hi} d_{hti}$$

$$\begin{pmatrix} U_{1i} \\ U_{2i} \\ \vdots \\ U_{mi} \end{pmatrix} \sim N(0, \Sigma_U)$$

$$\downarrow$$

$$\frac{(m+1)m}{2}$$

U_{hi} for $h = 1 \ldots m$ are random at level 2 with mean 0 and unconstrained covariance matrix.

Incompleted paired data

	$t=1$	$t=2$
P_1	80	90
P_2	65	70
P_3	95	92
P_4	x	70
⋮		
P_{10}	86	x

$$t = \frac{d}{S(d)}$$

$$Y_{ti} = \gamma_0 + \gamma_1 d_{1ti} + \sum_{h=0}^{1} U_{hi} d_{hti}$$

$t = 0$ & 1

$$Y_{0i} = \gamma_0 + U_{0i}$$
$$Y_{1i} = \gamma_0 + \gamma_1 + U_{1i}$$

$$\begin{pmatrix} U_{0i} \\ U_{1i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix}\right)$$

$$\to Y_{1i} - Y_{0i} = \gamma_1 + \underline{U_{1i} - U_{0i}}$$

$U_{1i} - U_{0i}$ and $U_{1i'} - U_{0i'}$ are independent

teacher
/ | \
Student
math | ....
Writing

Multivariate Multilevel Model

Y_{hij} ~~is~~ the measurement on ~~its~~ h's variable for included i in group j

4/20/17
STAT707
#4

$$\vec{Y}_{ij} = \begin{pmatrix} Y_{1ij} \\ \vdots \\ Y_{mij} \end{pmatrix}$$

$K+1$ (univariate multilevel)

K

1. Conclusion can be drawn about the correlations between the dependent variables.

2. The test of specific effects for single dependent variable are more powerful.
   (If strongly correlated dependent variable)

3. test whether the effect of an explanatory variable on Y_1 is larger than its effect on Y_2.

4. Carry out a single test of the joint effect of an explanatory variable on ~~served~~ several Y's.

$$Y_{hij} = \gamma_{0h} + \gamma_{1h} x_{1ij} + \cdots + \gamma_{ph} x_{pij} + U_{hj} + R_{hij}$$

dummy variable

$$d_{shij} = \begin{cases} 1, & h = s \\ 0, & h \neq s \end{cases}$$

$$Y_{hij} = \sum_{K=0}^{p} \sum_{s=1}^{m} \gamma_{ks} d_{shij} x_{kij} + \sum_{s=1}^{m} U_{sj} d_{shij} + \sum_{s=1}^{m} R_{sij} d_{shij}$$

$h = 1 \cdots m$ level-1

$i = 1 \cdots n_j$ level-2

$j = 1 \cdots N$ level-3

<u>NO</u> level-1 random effect

STAT 707
4/25/17
#1

$N = 1 \cdot \cdot j$

n_j $\qquad R_{ij} \sim N(0, \sigma^2)$

$H: = \Sigma d_j^2 \sim \chi^2 \; (N-1)$

$$d_j = \sqrt{\frac{df_j}{2}} \cdot \left(\log\left(s_j^2\right) - \frac{\Sigma df_j \log(S_j^2)}{\Sigma df_j} \right)$$

within sd

$$= \sqrt{2df_j} \cdot \left(\log(S_j) - \frac{\Sigma df_j \log(S_j)}{\Sigma df_j} \right) \sim N(0, 1^2)$$

$$Y_{ij} = \underline{\mu_{ij} + U_{0j}} + R_{ij} \qquad \sigma$$

$$Y_{ij} = \mu_{ij} + R^*_{ij} \qquad \sigma^*$$

GEE

Generalized Estimating Equation

" Longitudinal data analysis using generalized linear models "

by Kung-Yee Liang & Scott Zegar (1986)

Random ~~effect~~ effort model (GLMM, HLM)

$$g(\mathbb{E}(y_{ij} \mid b_j)) = x_{ij}^T \beta + z_{ij}^T b_j$$

$$\mathbb{E}[y_{ij} \mid U_{0j}] = \mu_{ij}$$

Marginal Models

$$\mathbb{E}(y_{ij}) = \mu_{ij}$$

$$g(\mu_{ij}) = x_{ij}^T \vec{\beta}$$

link function $\qquad$ covariantes

Marginal Variance

$$var(y_{ij}) = \phi\, V(\mu_{ij})$$

Covariance $y_{ij}, y_{i'j}$ is known function of $\mu_{ij}, \mu_{i'j}$ and $\vec{\alpha}$ (parameters)

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

$$l = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)$$

$\mathbb{E}y - b'(\theta) = 0$

$\mu = b'(\theta)$

Score: $$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \equiv U$$

$\mathbb{E}U = 0$

$\mathbb{E}(U^2) = -\mathbb{E}(U')$

$$U' = -\frac{b''(\theta)}{a(\phi)}$$

$$\mathbb{E}\left(\left(\frac{\partial l}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta^2}\right)$$

$$\mathbb{E}U^2 = \mathbb{E}\left[\left[\frac{y - b'(\theta)}{a(\phi)}\right]^2\right] = \frac{var(y)}{a(\phi)^2} = \frac{b''(\theta)}{a(\phi)} \rightarrow var(y) = a(\phi)\cdot \underbrace{b''(\theta)}_{V(\theta)}$$

$g(\mu) = x\beta \quad j = 1 \cdots p$

$$\frac{\partial l}{\partial \beta_j} = 0 \text{ (MLE)}$$

$i = 1 \cdots n$

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a(\phi)} \cdot \frac{1}{V(\theta)} \cdot \frac{\partial \mu_i}{\partial \beta_j}$$

$\mu = b'(\theta)$

$\mu' = b''(\theta)$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\underbrace{a(\phi) V(\mu_i)}_{var(y_i)}} \cdot \frac{\partial \mu_i}{\partial \beta_j} = 0$$

TTL $(y_i; \theta_i \phi)$

$\Sigma\, l(y_i; \theta_i \phi)$

STAT 701
4/25/17
#2

Quasilikeihood

$$U_i = \frac{Y_i - \mu_i}{\phi\, V(\mu_i)}$$

$Var(\mu_i) \propto V(\mu_i)$
proportional

$$\sum_{i=1}^{n} \frac{Y_i - \mu_i}{\phi\, V(\mu_i)} \cdot \left(\frac{\partial \mu_i}{\partial \beta_j}\right) = 0 \rightsquigarrow U(\beta) = \underset{p\times n}{D^T} \underset{n\times n}{V^{-1}} \underset{n\times 1}{\frac{Y-\mu}{\phi}} = \vec{0}_{p\times 1}$$

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$V_{n\times n} = \begin{bmatrix} V(\mu_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V(\mu_n) \end{bmatrix} \qquad D_{n\times p} = \frac{\partial \mu}{\partial \beta}$$

$$\mathbb{E}\left[-\frac{\partial U}{\partial \beta}\right] = \mathcal{I} = D^T V^{-1} D/\phi$$

$$\mathbb{E}[U^2] \equiv \mathcal{I} \qquad Var(\hat{\beta}) = \mathcal{I}^{-1} = \phi \cdot (D^T V^{-1} D)^{-1}$$

Fisher's scoring method

$$b^{(m)} = b^{(m-1)} + [\mathcal{I}^{(m-1)}]^{-1} U^{(m-1)}$$

Y_{ij} = Fixed part $+ U_{0j}\beta_0 + U_{1j}\beta_1^{x_i} + R_{is}$

n subjects

ith subject $1, 2 \cdots t_i$ observation

ith subject

A_i is a $t_i \times t_i$ diagonal matrix with $V(\mu_{ti})$ as the jth diagonal matrix.

marginal variance

$$A_i = \begin{bmatrix} V(\mu_{i1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V(\mu_{it_i}) \end{bmatrix}$$

$R_i(\alpha)$ is a $t_i \times t_i$ working correlation matrix for ith subject.

independence $R = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$ 0 $(x_1 \cdots x_p) \overset{\alpha}{=} (x_{i1} \cdots x_{ip})$

exchangeable $R = \begin{bmatrix} 1 & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & & & \alpha \\ \alpha & \alpha & \alpha & \ddots & \alpha \\ \alpha & \alpha & \alpha & \alpha & 1 \end{bmatrix}$ 1

AR-1 $R = \begin{bmatrix} 1 & \alpha & \alpha^2 & & \\ \alpha & 1 & \alpha & \alpha^2 & \\ \alpha^2 & \alpha & 1 & \ddots & \\ & & \ddots & \ddots & \alpha \\ & & & \alpha & 1 \end{bmatrix}$ 1

unconstraint $R = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$ $\frac{t_i(t_i-1)}{2}$

$V_i(\alpha) = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}$

covariance matrix

$AB = BA$

$A^{1/2} B A^{1/2} = (A^{1/2}R)A^{1/2}$

$= A^{1/2}(A^{1/2}R)$

$= AR = RA$

RA

For $Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{it_i} \end{pmatrix}$

$\mu = g^{-1}(X\beta)$

GEE estimate of β is

$$U(\beta) = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T [V_i(\alpha)]^{-1} (\vec{Y}_i - \vec{\mu}_i) = \vec{0}_p$$

$R_i(\alpha)$ and $\phi \longleftrightarrow \beta$

$$r_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{[V_i]_{ij}}}$$

$\underline{var(\hat{\beta}) = M_0^{-1}}$

$$M_0 = \sum_{i=1}^{m} \left(\frac{\partial \hat{\mu}_i}{\partial \beta} \right)^T V_i^{-1} \left(\frac{\partial \hat{\mu}_i}{\partial \beta} \right)$$

$\uparrow$ $V_i(\hat{\alpha})$

not consistant

STAT 707
4/25/17
#3

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

$$M_1 = \sum_{i=1}^{n} \underbrace{\left(\frac{\partial \hat{\mu}_i}{\partial \beta}\right)^T V_i^{-1}}_{\text{consistant}} \underbrace{(y_i - \hat{\mu}_i)(y_i - \mu_i)^T V_i^{-1}}_{\text{information sandwich estimate}} \underline{\left(\frac{\partial \hat{\mu}_i}{\partial \beta}\right)}$$

$H_0: \beta_j = 0$

Testing

$H_0: C\beta = d$ — C is $c \times p$ matrix

c linearly independent constrains

Wald's test

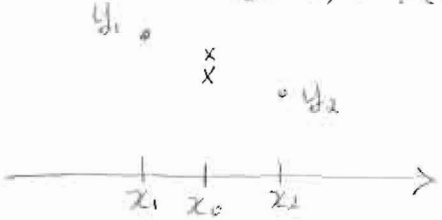
$$Q_c = (C\hat{\beta} - d)^T [C \, Var(\hat{\beta}) C^T]^{-1} (C\hat{\beta} - d) \sim \chi^2_c$$

STAT707
5/2/17
#1

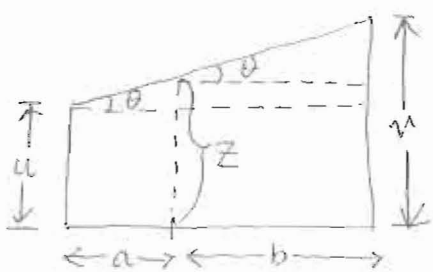
Marginal Model specify the covariance (GEE)

Random effects Model specify the random effect (GLMM)

$$\hat{y}(x_0) = f(x_0, (x_1, y_1)(x_2, y_2))$$



linear interpolation



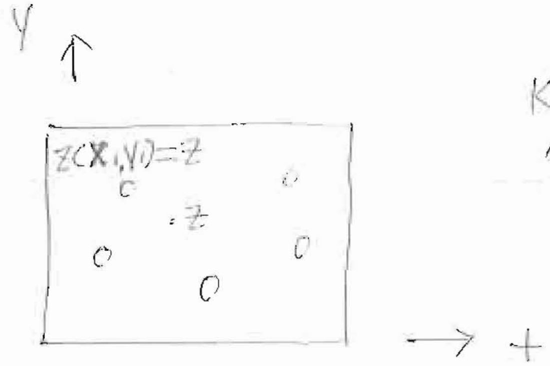
$$\tan\theta = \frac{v - z}{b} = \frac{v - u}{a+b}$$

$$z = v - \frac{b}{a+b}(v - u)$$

$$= \frac{a}{a+b} v + \frac{b}{a+b} u$$

$$= \frac{ab}{a+b} \cdot \frac{1}{b} v + \frac{ab}{a+b} \cdot \frac{1}{a} u$$

Tobler's First Law of Geography
All attribute values on a geographic surface are related to each other, but closer values are more strongly related than more ~~distinct~~ distance ones



$z_1 = z(x_1, y_1)$

$\vdots$

z_n

Krige estimation

$$\hat{z} = \sum_i \frac{c}{d^{\alpha}} z_i$$

$\alpha = 1 \sim 2$

by D.G. Krige

$$\hat{z}_0 = \sum_i \underbrace{\lambda_i}_{} z_i$$

$$Z = \underset{\substack{\uparrow \\ \text{constant term}}}{C} + \underset{\substack{\uparrow \\ \text{random effect}}}{R}$$

intercept $\Leftarrow$ constant term

$= \mu(x)$

$\to$ For general case

$$\left.\begin{aligned} &E(\hat{z}_0 - z_0) = 0 \\ &\min_{\lambda_i} Var(\hat{z}_0) \end{aligned}\right] \to \text{min MSE}$$

BLUE

unbiased estimator

$\to E(Z(x,y)) = C$

$Var(Z(x,y)) = \sigma^2$

Based on unbiased property,

$$E\hat{z}_0 = E\sum_i \lambda_i z_i = \sum_i \lambda_i E z_i = C \cdot \sum_i \lambda_i$$

$$= E z_0 = C$$

$$\sum_i \lambda_i = 1$$

$$Var(\hat{z}_0 - z_0) = Var\left(\sum_i \lambda_i z_i - z_0\right)$$

$$= Var\left(\sum_i \lambda_i z_i\right) - 2 \cdot cov\left(\sum_i \lambda_i z_i, z_0\right) + Var(z_0)$$

$$= \sum_i \sum_j \lambda_i \lambda_j cov(z_i, z_j) - 2\sum_i \lambda_i cov(z_i, z_0) + cov(z_0, z_0)$$

$Z(x,y) = C + R(x,y)$

$cov(z_i, z_j) = cov(R_i, R_j)$

$= C_{ij}$

$$= \sum_i \sum_j \lambda_i \lambda_j C_{ij} - 2\sum_i \lambda_i C_{0i} + C_{00}$$

STAT707
5/2/17
#2

$r_{ij} = \sigma^2 - C_{ij}$ semivariogram

$$Var(\hat{Z}_0 - Z_0) = \sum_i \sum_j \lambda_i \lambda_j (\sigma^2 - r_{ij}) - 2 \sum_i \lambda_i (\sigma^2 - r_{0i}) + \sigma^2 - r_{00}$$

$$= \sigma^2 - \sum_i \sum_j \lambda_i \lambda_j r_{ij} - 2\sigma^2 + 2 \sum_i \lambda_i r_{0i} + \sigma^2 - r_{00}$$

$$= 2 \sum_i \lambda_i r_{0i} - \sum_i \sum_j \lambda_i \lambda_j r_{ij} - r_{00}$$

$$\frac{\partial}{\partial \lambda_k} = \sum_j \lambda_j r_{kj} + \sum_i \lambda_i r_{ik}$$

$$= \sum_j \lambda_j r_{kj} + \sum_j \lambda_j r_{jk}$$

$$= \sum_j \lambda_j (r_{kj} + r_{jk})$$

$$L = Var(\hat{Z}_0 - Z_0) + \phi \left(\sum_i \lambda_i - 1 \right)$$

$$\frac{\partial L}{\partial \lambda_i} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \phi} = 0$$

$$\frac{\partial L}{\partial \lambda_i} = 2 r_{0i} - \sum_j (r_{ij} + r_{ji}) \lambda_j \iff r_{0i} - \sum_j r_{ij} \lambda_j = 0$$

$$\sum_i \lambda_i = 1$$

$$2\lambda_1 r_{01} + 2\lambda_2 r_{02} + \cdots + 2\lambda_n r_{0n}$$

$$2\lambda_2 r_{02} \downarrow 2 r_{02}$$

$$\frac{\partial L}{\partial \phi} = \sum \lambda_i - 1$$

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} & 1 \\ r_{21} & r_{22} & \cdots & r_{2n} & 1 \\ \vdots & \vdots & & \vdots & \\ r_{n1} & r_{n2} & & r_{nn} & 1 \\ 1 & 1 & & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ 0 \end{bmatrix} = \begin{bmatrix} r_{01} \\ r_{02} \\ \vdots \\ r_{0n} \\ 1 \end{bmatrix}$$

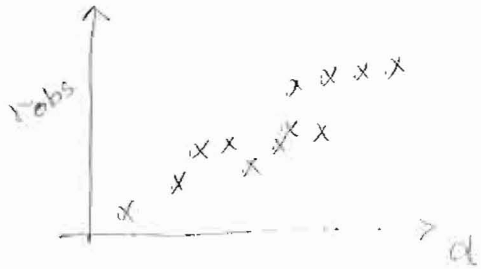
$r_{ij} = \sigma^2 - C_{ij}$

$$\frac{n(n+3)}{2}$$

$$\frac{1}{2} \mathbb{E} (Z_i - Z_j)^2$$

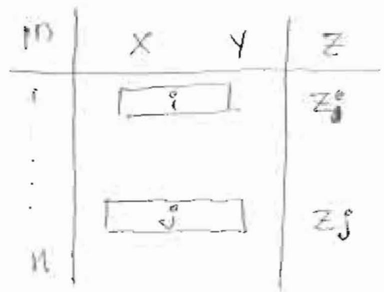
$$= \frac{1}{2} \mathbb{E} (R_i - R_j)^2 = \frac{1}{2} \mathbb{E} R_i^2 + \frac{1}{2} \mathbb{E} R_j^2 - \mathbb{E} R_i R_j = \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 - C_{ij} = \sigma^2 - C_{ij}$$

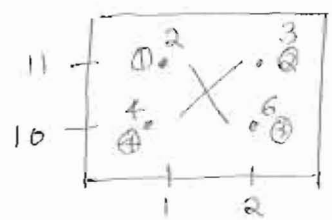
$\hat{\gamma}_{ij} = \frac{1}{2}(z_i - z_j)^2$

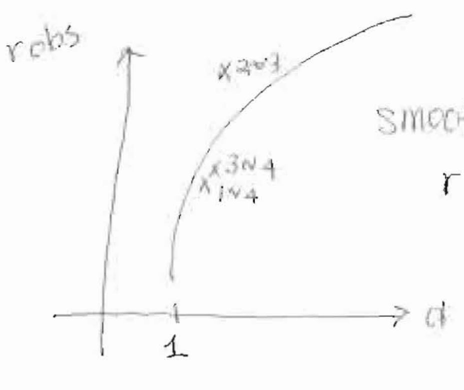


$\frac{(1+n)n}{2}$

$= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$

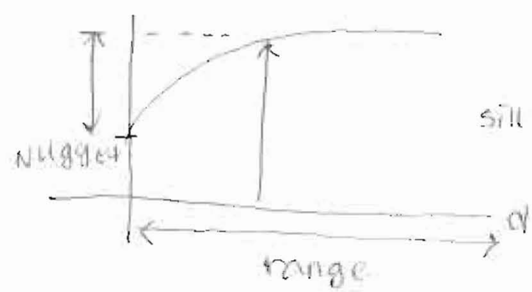






smooth curve

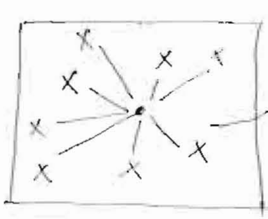
$r = r(d)$



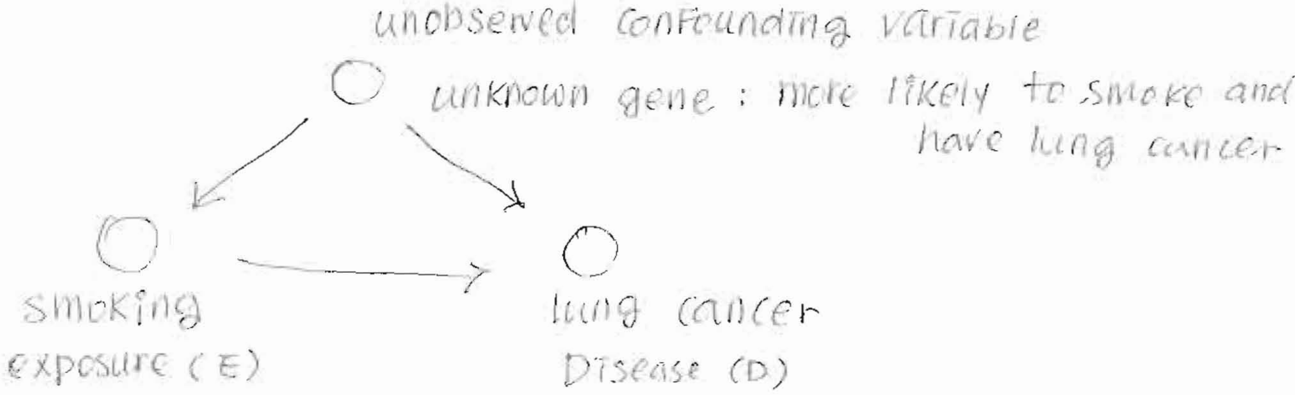
$r \sim d$

$[\gamma_{ij}]\,[\lambda] = [\gamma_{0j}]$

$\sum_i \lambda_i z_i = \hat{z}_0$



STAT707
5/2/17
#3



$$\frac{P(D=1 \mid E=1)}{P(D=1 \mid E=0)} = RR_{ED} = \cancel{8}9$$ ← Relative Risk

Cornfield (1959)'s inequality
If E and D are conditionally independent, ~~RR_{EU}~~ $(E \perp D \mid U)$
and then $RR_{EU} \geq RR_{ED}$

By the definition,

$$\frac{P(U=1 \mid E=1)}{P(U=1 \mid E=0)} \geq 9$$ ← # quite large   NOT realistic

Assume that
$f_1 = P(U=1 \mid E=1)$
$f_0 = P(U=1 \mid E=0)$

$$RR_{EU} = \frac{P(U=1 \mid E=1)}{P(U=1 \mid E=0)}$$

$$RR_{UD} = \frac{P(D=1 \mid U=1)}{P(D=1 \mid U=0)}$$

$RR_{ED} \geq 1$ $\quad RR_{EU} \geq 1$

Under the assumption, $E \perp D \mid U$

$$RR_{ED} = \frac{P(D=1 \mid E=1)}{P(D=1 \mid E=0)} = \frac{\sum_{u=0,1} P(D=1, U=u \mid E=1)}{\sum_{u=0,1} P(D=1, U=u \mid E=0)}$$ joint prob.

$\because$ independent If $A \perp B$, then $p(A|B) = P(A)$

$$= \frac{\sum_{u=0,1} P(D=1 \mid U=u, \cancel{E=1}) \cdot P(U=u \mid E=1)}{\sum_{u=0,1} P(D=1 \mid U=u, \cancel{E=0}) \cdot P(U=u \mid E=0)} \begin{matrix} \leftarrow f_1 \\ \leftarrow f_2 \end{matrix}$$

$$= \frac{P(D=1 \mid U=1) \cdot \overset{f_1}{P(U=1 \mid E=1)} + P(D=1 \mid V=0) \, P(U=0 \mid E=1)}{P(D=1 \mid U=1) \, \underset{f_0}{P(V=1 \mid E=0)} + P(D=1 \mid U=0) \cdot P(U=0 \mid E=0)}$$

$$= \frac{RRUD \cdot f_1 + (1-f_1)}{RRUD \cdot f_0 + (1-f_0)}$$

$F_1 > F_0$ based on $RRED \geq 1$ $RREU \geq 1$

rewrite this form,

$$RRED = \frac{RRUD \cdot F_0 \cdot \frac{F_1}{F_0} + (1-F_0) \cdot \frac{F_1}{F_0} - (1-F_0) \frac{F_1}{F_0} + (1-F_1)}{RRUD \cdot F_0 + (1-F_0)}$$

$$= \frac{F_1}{F_0} + \frac{1 - \frac{F_1}{F_0}}{RRUD \cdot f_0 + (1-f_0)}$$

$RRUD \uparrow \rightarrow RRED \uparrow$

$$RRED \leq \lim_{RRUD \to +\infty} \frac{f_1}{f_0} - \frac{\frac{f_1}{f_0} - 1}{RRUD f_0 + (1-f_0)}$$

$$= \frac{f_1}{f_0} = RREU$$

5/9/17
STAT 707
#1

Spatial · Kriging
Causal Inference : Cornfield's inequality
propensity score
↳ D. B. Rubin

Cochran & Rubin (1973)
Controlling Bias in Observational Studies : a review

Rosenbaum & Rubin (1983)
The central role of the propensity score in observational studies for causal effects

Difference between experiments and observational study
↳ unable to control any factors
ex) Smoking

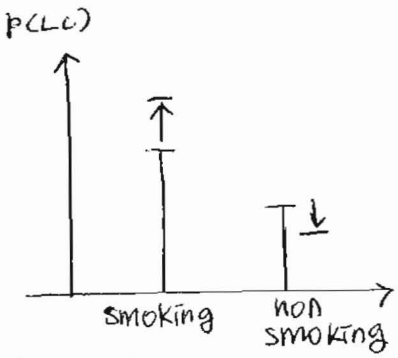
age group gender

Randomized Block Design

	(1,2,3) Treat 1	(4,5,6) Treat 2	(7,8,9) Treat 3
F	$\boxed{P1}$		
M		P2	

P1 (F)
P2 (M)

P21



Age	gender	BMI	Smoke	Lung Cancer
20	M	27	Y	N
35	M	25	Y	Y

Propensity Score Matching

$\mathbb{E}(r_1) - \mathbb{E}(r_0) > 0$

$z = 0, 1$ $\qquad \vec{x} = (\quad \ldots \quad)$ high dimension

(r_0, r_1)

	z	r_0	r_1
P1	1	X	1
P2	0	1	X

propensity score: $e(\vec{x}) = P(z=1 \mid \vec{x})$

Thm1 Treatment Assignment and observed covariates are conditionally independent given the propensity score.

constant $\rightarrow$ $\vec{x} \perp\!\!\!\perp z \mid e(\vec{x})$ $\qquad$ "$\perp$" = "$\perp\!\!\!\perp$" Independent

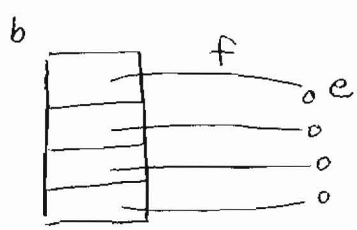
$\uparrow$ simplest way to summerize x

Thm2 Let $b(x)$ be a function of $\vec{x}$. Then $b(\vec{x})$ is a balancing score. i.e.

$$\vec{x} \perp\!\!\!\perp z \mid b(\vec{x})$$

iff $b(x)$ is finer than $e(x)$.

$$e(x) = f(b(x)) \text{ for some } f.$$



proof ① Assume $b(x)$ is finer than $e(x)$.

$$e(x) = \Pr(z=1 \mid \vec{x})$$

$$\Pr(z=1 \mid b(x)) = \mathbb{E}[I(z=1) \mid b(x)]$$
$$= \mathbb{E}_x[\mathbb{E}[\mathbb{E}[I(z=1) \mid b(x), x]]]$$
$$= \mathbb{E}[e(x) \mid b(x)] = e(x)$$

5/9/17
STAT 707
#2

② Suppose $b(x)$ is a balance score.
but $b(x)$ is not finer than $e(x)$.

$\exists\ x_1 \ \& \ x_2$

$b(x_1) = b(x_2)$

$e(x_1) \neq e(x_2)$

$\rightarrow Pr(Z=1 \mid x_1) \neq Pr(Z=1 \mid x_2) \rightarrow x \not\perp\!\!\!\perp Z \mid b(x_2)$

Def. we say treatment assignment is strongly ignorable, given $\vec{x}$
if $(r_0, r_1) \perp\!\!\!\perp Z \mid U$

Thm3 If treatment assignment is strongly ignorable, given $\vec{x}$, then it is strongly ignorable given any balancing score $b(x)$. i.e.

$(r_0, r_1) \perp\!\!\!\perp Z \mid \vec{x}$ for all $\vec{x}$

imply

$\underbrace{(r_0, r_1) \perp\!\!\!\perp Z \mid b(\vec{x})}$ for all $b(\vec{x})$

proof: $X \perp Y \Leftrightarrow P(X \mid Y) = P(X)$

$\rightarrow P(Z=1 \mid r_0, r_1, b(\vec{x})) = \underbrace{P(Z=1 \mid b(\vec{x}))} = e(\vec{x})$

$Pr(Z=1 \mid r_0, r_1, b(X))$
$= E[I(Z=1) \mid r_0, r_1, b(x)]$
$= E[E[I(Z=1) \mid r_0, r_1, x] \mid r_0, r_1, b(x)]$
$= E[e(x) \mid r_0, r_1, \underbrace{b(x)}]$
$= e(x)$

$H_0: \mathbb{E}[r_1] - \mathbb{E}[r_0] = 0$

$$\mathbb{E}[r_1 \mid z=1] - \mathbb{E}[r_0 \mid z=0]$$
$$= \mathbb{E}_x\left[\mathbb{E}[r_1 \mid z=1, \vec{x}] - \mathbb{E}[r_0 \mid z=0, \vec{x}]\right]$$

<u>strongly ignorable</u> $\quad \mathbb{E}_x\left[\mathbb{E}[r_1 \mid \vec{x}] - \mathbb{E}[r_0 \mid \vec{x}]\right] = \mathbb{E}[r_1] - \mathbb{E}[r_0]$

$\vec{x} = (x_1, x_2, \cdots)$

$\hat{e}(a) = Pr(z=1 \mid \vec{x} = 0)$

	T	C
(0,1)	5	2
(0,0)	8	2
(1,0)	10	16
(1,1)	2	1

$$e(x) = Pr(z=1 \mid x) = \frac{Pr(z=1)\,Pr(x \mid z=1)}{P(z=1)\,Pr(x \mid z=1) + P(z=0)\,Pr(x \mid z=0)}$$

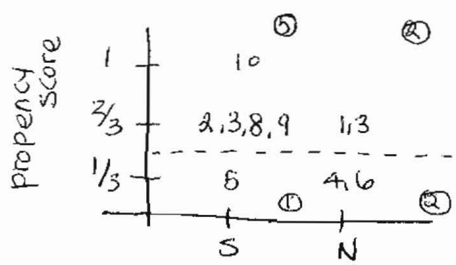
Age	Gender	Smoking	Lung Cancer
0	0	0	0
0	0	1	0
0	0	1	0
1	0	0	1
1	0	1	1
1	0	0	0
0	1	0	0
0	1	1	0
0	1	1	0
1	1	1	1

$Pr(z=1 \mid x(0,0)) = \frac{2}{3} = \hat{e}((0,0))$

$Pr(z=1 \mid x(1,0)) = \frac{1}{3} = \hat{e}((1,0))$

$Pr(z=1 \mid x(0,1)) = \frac{2}{3} = \hat{e}((0,1))$

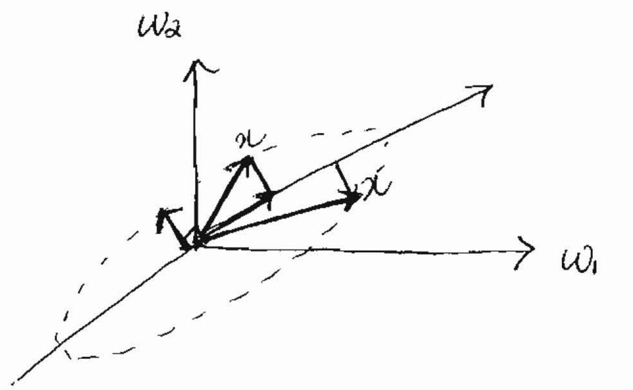
$Pr(z=1 \mid x(1,1)) = 1 = \hat{e}((1,1))$



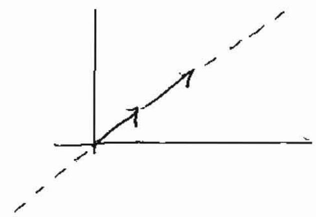
MatchIt ← R package

5/9/17
STAT 707
#3

Principal Component Analysis (PCA)



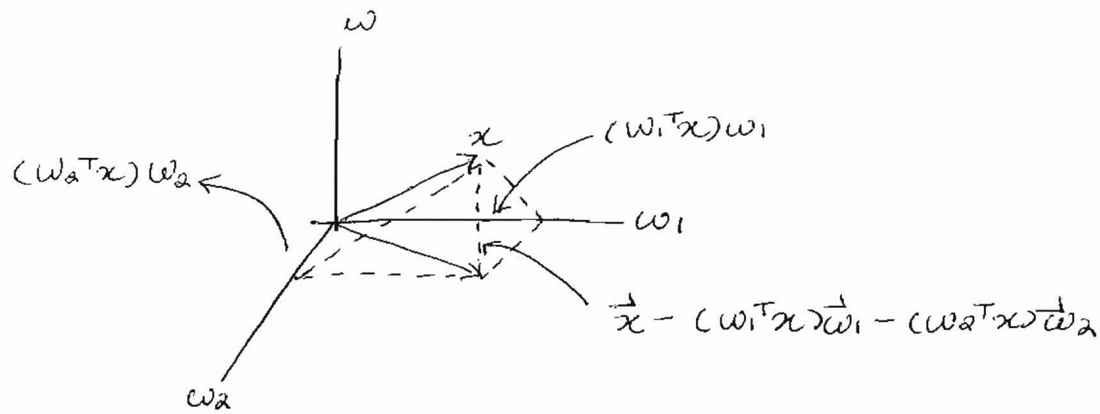
Find direction which express as muc



$\vec{x} \in \mathbb{R}^n$

data set: $(\vec{x}_1, \vec{x}_2, \cdots \vec{x}_n)$

$$\vec{w}_1 = \arg\max_{\|w\|=1} \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i)^2$$



$K \to K+1$

$$w_{k+1} = \arg\max \sum_{i=1}^{n} (w^T(\vec{x} - \sum_{j=1}^{k} w_j^T x)\vec{w}_j))^2$$

$$w^T \underbrace{x^T x}_{M} w$$

$$\arg\max_{\|w\|=1} w^T M w$$

$$F = w^T M w - \lambda (w^T w - 1)$$

$$\frac{\partial F}{\partial w} = 2M\vec{w} - 2\lambda\vec{w} = 0$$

$$\therefore M\vec{w} = \lambda\vec{w}$$

$$w^T \underbrace{x^T x}_{M} w$$

$$\frac{w^T M w}{w^T w} = \frac{w^T \lambda \vec{w}}{w^T w} = \lambda$$

Rayleigh Quotient

STAT 707
5/16/17
#1

Final Review

(HW#3)

① $Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} x_{ij} + \gamma_{11} Z_j x_{ij} + R_{ij}$,

$i = 0, 1, 2$
$j = 1, 2, 3$

Suppose $Z_j = Y_{0j}$, and $t_i = x_{ij}$

let

$$\beta = \begin{pmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{pmatrix} \leftarrow \text{true coefficient}$$

$\hat{\beta}$ = the MLE based on full data set $(i=0,1,2,\ j=1,2,3)$
$\tilde{\beta}$ = the MLE based on partial data set $(i=1,2,\ j=1,2,3)$

Show (1) $\mathbb{E}\hat{\beta} = \mathbb{E}\tilde{\beta} = \beta$

(2) ~~[illegible]~~ $var(\hat{\gamma}_{kl}) \le var(\tilde{\gamma}_{kl})$ $k=0,1,\ l=0,1$

change to $X^T X \ge X_*^T X_*$

proof:

$$Y = \begin{bmatrix} y_{01} \\ y_{11} \\ y_{21} \\ y_{02} \\ y_{12} \\ y_{22} \\ y_{03} \\ y_{13} \\ y_{23} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & y_{01} & 0 & 0 \\ 1 & y_{01} & t_1 & t_1 y_{01} \\ 1 & y_{01} & t_2 & t_2 y_{01} \\ 1 & y_{02} & 0 & 0 \\ 1 & y_{02} & t_1 & t_1 y_{02} \\ 1 & y_{02} & t_2 & t_2 y_{02} \\ 1 & y_{03} & 0 & 0 \\ 1 & y_{03} & t_1 & t_1 y_{03} \\ 1 & y_{03} & t_2 & t_2 y_{03} \end{bmatrix}$$

$\vec{\epsilon} = [R_{ij}]$ $\qquad Y = X\beta + \epsilon$

$\hat{\beta} = (X^T X)^{-1} X^T Y \longrightarrow \hat{\beta} = (X^T X)^{-1} X^T (X\beta + \epsilon)$

$= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon$

$Y = \underset{\sim}{X\beta} + \epsilon$ $\qquad = \beta + (X^T X)^{-1} X^T \epsilon$

$\mathbb{E}\hat{\beta} = \beta$ $\qquad \therefore \mathbb{E}\hat{\beta} = \beta$, $\mathbb{E}\tilde{\beta} = \beta$

(2) ~~E var~~

assume that σ^2 is given

$$var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$X = \begin{bmatrix} 1 & y_{01} & 0 & 0 \\ 1 & y_{02} & 0 & 0 \\ 1 & y_{03} & 0 & 0 \\ & & & \\ & X_* & & \\ & & & \end{bmatrix}$$

$$X = \begin{bmatrix} A \\ X_* \end{bmatrix}$$

$$X^T X = \begin{bmatrix} A^T \cdot X_*^T \end{bmatrix} \begin{bmatrix} A^T \\ X_* \end{bmatrix}$$

$$= A^T A + X_*^T X_*$$

Run simulation to check

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ y_{01} & y_{02} & y_{03} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & y_{01} & 0 & 0 \\ 1 & y_{02} & 0 & 0 \\ 1 & y_{03} & 0 & 0 \end{bmatrix} + X_*^T X_*$$

$$= \begin{bmatrix} 3 & \Sigma y_{0j} & 0 & 0 \\ \Sigma y_{0j} & \Sigma y_{0j}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + X_*^T X_*$$

$$X^T X \geq X_*^T X_*$$

(2)

Athlete	Age	Club	Time
1	38	0	91.500
1	40	0	93.633
1	43	1	92.717
2	53	1	96.017
2	55	1	98.117
2	56	1	91.383
2	58	1	93.167
3	37	1	83.183
3	40	1	81.500
3	42	0	82.017
4	41	0	91.167
4	45	1	94.217
4	46	0	99.100
5	54	1	105.400
5	58	1	111.133
6	57	1	90.250
6	60	1	89.450
6	62	1	94.620

Run two models (GLMM)

log(time) ~ club + log(age) , 1| athlete

log(time) ~ log(age), 1| athlete

test the effect of variable "club"

t-test or Deviance-test (χ^2 test)

Run two models (GEE) : marginal model

log(time) ~ club + log(age) corr mat = exchangeable

log(time) ~ log(age)

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

③ Kriging (1-D)

0 ——— 1

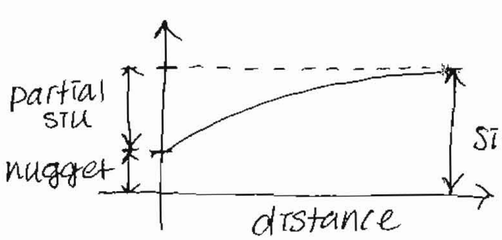
⓪ Z_0

① at 0 : 5% Z_1

② at 0.5 : 4% Z_2

③ at 1 : 3% Z_3

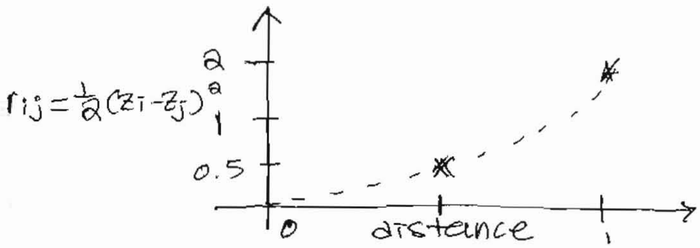
use a quadratic function with nugget = 0
in the semi-varigram vs. distance graph



Find the interpolation at 0.2 by using kriging data pair

difference
of coordinates

data pair	distance	$\frac{1}{2}(z_i - z_j)^2$
(1,2)	0.5	$\frac{1}{2}(5-4)^2 = \frac{1}{2}$
(2,3)	0.5	$\frac{1}{2}(4-3)^2 = \frac{1}{2}$
(3,1)	1	$\frac{1}{2}(3-5)^2 = 2$



$r = 2d^2$

$$z_0 = \sum_{i=1}^{3} \lambda_i z_i$$

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} & 1 \\ r_{21} & r_{22} & \cdots & r_{2n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ r_{n1} & r_{n2} & & r_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ 0 \end{bmatrix} = \begin{bmatrix} r_{01} \\ r_{02} \\ \vdots \\ r_{0n} \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & \frac{1}{2} & 2 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ 2 & \frac{1}{2} & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.18 \\ 1.28 \\ 1 \end{bmatrix}$$

pair	d	$\hat{r}$
(0,1)	0.2	0.08
(0,2)	0.3	0.18
(0,3)	0.8	1.28

$$z_0 = \hat{\lambda}_1 \cdot 5\% + \hat{\lambda}_2 \cdot 4\% + \hat{\lambda}_3 \cdot 3\%$$

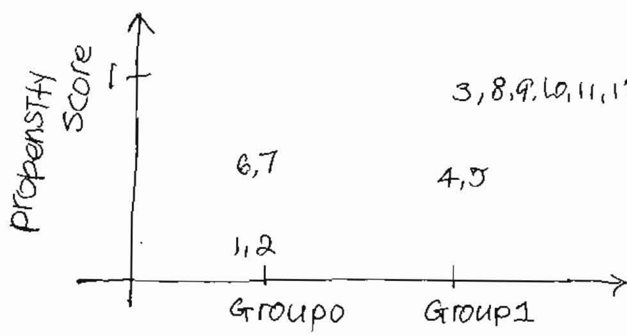
STAT 707
5/16/17
#3

(4) ~~property~~ propensity score matching

calculate propensity scores

$$\hat{e}(a) = \text{Prob}(Z=1 \mid x = a)$$

ID	Age	History of family member	Group
1	≤30	0	0
2	≤30	0	0
3	≤30	1	1
4	30~50	0	1
5	30~50	1	1
6	30~50	0	0
7	30~50	1	0
8	≥50	0	1
9	≥50	0	1
10	≥50	0	1
11	≥50	0	1
12	≥50	1	1



		Group0	Group1	e
≤30	0	1, 2		$0 = \frac{0}{2}$
≤30	1		3	$1 = \frac{1}{1}$
30~50	0	6	4	$0.5 = \frac{1}{2}$
30~50	1	7	5	$0.5 = \frac{1}{2}$
≥50	0		8, 9, 10, 11	$1 = \frac{4}{4}$
≥50	1		12	$1 = \frac{1}{1}$