matrix 相关 (vector: $x=\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}$)

$(X^T)^T=X$

$(XY)^T=Y^TX^T\neq X^TY^T$

$(A-B)^T=A^T-B^T$

Singular if $|X|=0$

对于 n×n non-singular:

① $XX^{-1}=X^{-1}X=I$

② $(X^{-1})^{-1}=X$

③ $(XY)^{-1}=Y^{-1}X^{-1}\neq X^{-1}Y^{-1}$

④ $(X^T)^{-1}=(X^{-1})^T$

证明 symmetric

$X^T=X$ 如果x是n×1, 则 $||x||=1$

Orthogonal:

iff $x^Ty=0=\sum_i x_iy_i=x\cdot y$ <vector>

norm of x: $x^Tx=||x||^2=\sum_i x_i^2$

iff $X^TX=I$ <matrix>

iff col of X form orthogonal set <matrix>

$\boxed{X^T=X^{-1}}$

Eigenvalues & Eigenvector (x) (λ_i)

$Ax=\lambda x$ $(A-\lambda I)x=0$

rank: # linearly independent col vectors

X: n×k matrix, ($r(X)\le k$ always)

full rank if $n\ge k$, $r(X)=k$

$r(XY)\le(r(X),r(Y))$

$r(X)=r(X^T)=r(X^TX)$

$r(X)=r(X^cX)=r(XX^c)$

for invertible P&Q,

$r(X)=r(PX)=r(XQ)$

Diagonalisable (P是A的 orthogonal matrix)

$P^TAP=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_i\end{bmatrix}=D$

$PP^T=I$

$r(A)=r(D)=$ # non-zero λ_i

Idempotent

iff $X^2=X$

是idempotent; 如果 matrix { $\lambda=0$ or $\lambda=1$

$H=X(X^TX)^{-1}X^T=H^2$

Trace: $\sum_i x_{ii}$ M1

$tr(cX)=c\,tr(X)$

$tr(X\pm Y)=tr(X)\pm tr(Y)$

$tr(XY)=tr(YX)$

如果是 symmetric + idempotent X, $r(X)=r(P^TXP)=tr(D)$

$tr(X)=r(X)$ $=tr(P^TPX)=tr(X)$

I−X 是 symmetric + idempotent

一些 theory

$A_1\cdots A_m$: symmetric matrix.

then,

a) they share same orthogonal matrix P that

{ P^TA_iP diagonalise all i=1..m

$A_iA_j=A_jA_i$

b) 已知其2可推3

① $A_1\cdots A_m$ 都 idempotent

② $\sum_i^m A_i$ idempotent

③ $A_iA_j=0$ for $i\neq j$

Hat matrix H

<symmetric & idempotent>

I−H 也是 $r(H)=tr(H)=r(X)$

$r(H)=r(X)$ $HX=X$, $X^TH=X^T$

H_{ii} ranges from [0,1],

被称为 point leverage.

注: 在 linear model 里,

$y^TXb=y^THy=SSreg$

$y^Ty-y^Tby=y^T(I-H)y=SSres$

positive-definite

$y^TAy=\sum_{i=1}^k\sum_{j=1}^k a_{ij}y_iy_j$

$y^TAy>0$ for $y\neq0$,

symm A with $\lambda_i>0$,

A is pos-def.

positive-semidefinite

$y^TAy\ge0$ for all y,

sym A with $\lambda_i\ge0$,

A is pos-semi def

求导

$\frac{\partial}{\partial b}b^TAb=Ab+A^Tb$

$\frac{\partial}{\partial b}a^Tb=a$

$\frac{\partial}{\partial b}b^Tb=2b$

General Linear Model

$y=X\beta+\varepsilon$, $\varepsilon\sim MVN(0,\sigma^2I)$

y: n×1 response vector, r.v.

X: n×p design matrix, NOT r.v.

p−1个 variable + 1个 intercept

ε: n×1 error term, r.v.

$E(y)=X\beta$, $var(y)=V=\sigma^2I$

$y\sim MVN(X\beta,\sigma^2I)$

$E(\hat y)=Xb$, $e_i=y_i-\hat y_i$

$\varepsilon_i=e_i+E(\hat y_i)-E(y_i)=y_i-E(y_i)$

Full rank model

$\beta=[\beta_0\cdots\beta_k]^T$

one-way classification.

$\beta=[\mu,\tau_1\cdots\tau_k]^T$

two-way classification

$\beta=[\mu,\beta_1\cdots\beta_a,\tau_1\cdots\tau_b]^T$

Parameter estimation b

LSE for β: minimise $\sum_i e_i=\sum_i(y_i-\hat y_i)$ 结果同LSE.

$=y-\hat y=y-Xb=(I-H)y$, $H=X(X^TX)^cX^T$

$SS_{Res}=e^Te=(y-Xb)^T(y-Xb)=y^Ty-y^TXb$

(与b有关的函数, 求导出极值)

full rank 时 UMVUE

$b=(X^TX)^{-1}X^Ty$, 因为(X^TX) invertible.

less than full rank,

$b=(X^TX)^cX^Ty$

$E(b)=\beta$, (LSE) $var(b)=\begin{cases}\sigma^2(X^TX)^{-1}, & \text{full}\\ \sigma^2(X^TX)^c, & \text{less than full}\end{cases}$

$b\sim MVN(\beta,\sigma^2(X^TX)^c)$

可通过 t^Tb 来 estimate $t^T\beta$,

$t^Tb\sim MVN(t^T\beta,\sigma^2t^T(X^TX)^ct)$

Estimability ($t^T\beta$): there exists matrix c s.t. $E(c)=t^T\beta$ (i.e. $t^T\beta$ has an unbiased estimator)

full rank: 都可以

less than full: iff $t^T(X^TX)^c(X^TX)=t^T$

常见题: ① $\beta_i=\mu+\tau_i$, t是[ ]

② $\tau_i-\tau_j$, treatment contrast, $\sum_i d_i\tau_i$ when $\sum_i d_i=0$

MLE

用MVN 的 pdf, 求 log likelihood.

⇒ minimise $(y-X\beta)^T(y-X\beta)$

Variance Estimation

MLE $\hat\sigma^2=\frac{SS_{Res}}{N}$

缺: biased, (n→∞, unbiased)

— 然而, $s^2=\frac{SS_{Res}}{n-r(X)}$, unbiased,

b is independent of SS_{Res},

t^Tb independent of s^2

因为 $b=(X^TX)^cX^Ty$

$SS_{Res}=\frac{1}{n-r}y^T(I-H)y$

$y\sim MVN(X\beta,\sigma^2I)$

By与y^TAy indep,

$(X^TX)^cX^T\cdot V\cdot(I-H)=0$,

Interval estimation

$\frac{SS_{Res}}{\sigma^2}=\frac{y^T(I-H)y}{n-r}\times\frac{n-r}{\sigma^2}=\frac{(n-r)s^2}{\sigma^2}\sim\chi^2_{n-r}$

$t^Tb\sim MVN(t^T\beta,\sigma^2t^T(X^TX)^ct)$

⇒ $z=\frac{t^Tb-t^T\beta}{\sqrt{\sigma^2t^T(X^TX)^ct}}\sim N(0,1)$ n = length(y), r = #param

⇒ $\frac{t^Tb-t^T\beta}{\sqrt{s^2t^T(X^TX)^ct}}\sim t(n-r)$

(1−α)% CI of $t^T\beta$ is

① $t^Tb\pm t_{1-\frac{\alpha}{2}}\sqrt{s^2t^T(X^TX)^ct}$

full rank model 时,

(1−α)% CI of β is → $se(b_i)$

② $b_i\pm t_{1-\frac{\alpha}{2}}\sqrt{s^2(X^TX)^{-1}_{ii}}$ (index 从0开始)

(1−α)% CI for $t^T\beta$ is

③ $t^Tb\pm t_{1-\frac{\alpha}{2}}se(t^Tb)$

100(1−α)% PI for y^* is

$(t^*)^Tb\pm t_{1-\frac{\alpha}{2}}\sqrt{s^2(1+t^T(X^TX)^ct)}$

predict 新 response 对于新的x来多少. #response's linear combination

是 interval for single response,

CI 是 interval for expected response

PI is wider as single observation is more variable than expected.

Hypothesis testing

① 模型整体意义, β=0.

$SS_{Res}=e^Te=y^Ty-y^TXb$ variation explained by error

$SS_{Total}=y^Ty$ total variation

$SS_{Reg}=y^TXb=(Xb)^T(Xb)=y^T\hat y$ variation explained by the model

$\frac{SS_{Reg}}{\sigma^2}=\frac{(Xb)^T(Xb)}{\sigma^2}=\frac{y^THy}{\sigma^2}\sim\chi^2(k=r,\lambda=\frac{\beta^TX^TX\beta}{2\sigma^2})$

indep

$\frac{SS_{Res}}{\sigma^2}\sim\chi^2(n-r)$

如果 under H0,

$\frac{SS_{Reg}}{\sigma^2}\sim\chi^2(r)$

M3

random vector y, $[y_1\cdots y_n]^T$

向量 $E(y)=[E(y_1)\cdots E(y_n)]^T$

$E(a)=a$ $E(a^Ty)=a^TE(y)$

$E(ay)=aE(y)$

var(y) ≡ $E[(y-\mu)(y-\mu)^T]$ <n×n>

$=V$ (covariance matrix)

symmetric & pos-semi-def (但不一定 invertible, 不一定 full-rank)

$V_{ii}=var(y_i)$

$V_{ij}=cov(y_i,y_j)$ ($i\neq j$)

$=E[(y_i-\mu_i)(y_j-\mu_j)]$

if y_i,y_j indep. $=E(y_iy_j)-\mu_i\mu_j$

$var(a^Ty)=a^TVa$ <a为vector>

$var(Ay)=AVA^T$ <A为matrix>

Quadratic form

symme matrix 是 pos-def (⇔)

iff 它的λ_i都大于0.

Symme matrix 是 pos-semidef ⇔

iff 它的λ_i都 ≥0

$E(y^TAy)=tr(AV)+\mu^TA\mu$,

MVN

$x_i\sim N(0,1)$ i=1..n

$x=[x_1..x_n]^T\sim MVN(0,I_n)$

$y=Ax+b$, ~ $MVN(\mu=b,V=AA^T)$

if $y\sim MVN(\mu,V)$

$Ay+b\sim MVN(A\mu,AVA^T)$

如果 V 是 invertible,

$z=V^{-\frac12}(y-\mu)\sim MVN(0,I)$

if $y\sim MVN(\mu,I_k)$

$x=y^Ty=\sum_i^k y_i^2\sim\chi^2(k,\lambda=\frac12\mu^T\mu)$

$E(x)=k+2\lambda$

$var(x)=2k+8\lambda$

一些 theorem

① $x_i\sim\chi^2(k_i,\lambda_i)$, $\sum_i x_i\sim\chi^2(\sum_i k_i,\sum_i\lambda_i)$

② $y\sim MVN(\mu,V)$, A: n×n symmetric, AV: idempotent, $r(AV)=k$

$y^TAy\sim\chi^2(k,\lambda=\frac12\mu^TA\mu)$

③ $y\sim MVN(\mu,V)$, y: n×1 r.v.

V is invertible,

$y^TV^{-1}y\sim\chi^2(n,\lambda=\frac12\mu^TV^{-1}\mu)$

④ $y\sim MVN(\mu,I_n)$: y: n×1 r.v.

A: n×n symmetric.

$y^TAy\sim\chi^2(k,\lambda=\frac12\mu^TA\mu)$

⑤ { $y\sim MVN(0,V)$ / $y\sim MVN(0,I_n)$ }, A: n×n symmetric { idempotent, $r(A)=k$ }

$y^TAy\sim\chi^2_k$

⑥ $y\sim MVN(\mu,\sigma^2I_n)$, A: n×n symm

$\frac{1}{\sigma^2}y^TAy\sim\chi^2(k,\lambda=\frac{1}{2\sigma^2}\mu^TA\mu)$ Idem + $r(A)=k$

Independence of quadratic form

① $y\sim MVN(\mu,V)$, V invertible,

A&B 是 symmetric n×n

iff $AVB=0$ ⇒ y^TAy indep y^TBy

② $y\sim MVN(\mu,\sigma^2I)$, A&B sym n×n

y^TAy indep y^TBy iff $AB=0$

③ $y\sim MVN(\mu,V)$, A&B sym n×n

y^TAy indep By iff $BVA=0$.

t-distribution

indep. $z\sim N(0,1)$, $z^2\sim\chi^2_1$

$\chi^2(df=r,\lambda=0)$

$T=\frac{z}{\sqrt{\chi^2_r/r}}\sim t_r$

F-distribution

indep. χ^2_m, χ^2_n

$F=\frac{\chi^2_m/m}{\chi^2_n/n}\sim F(m,n)$

M2

Estimability 补充

$E(c^Ty)=t^T\beta$, 则 $t^T\beta$ estimable 有解.

$E(z_0^TX^Ty)=t^T\beta$ ⇐ $X^TXz=t$

存在 z_0:

$t^T(X^TX)^c(X^TX)=t^T$ OR

$(X^TX)(X^TX)^ct=t$

⇒ $z_0=(X^TX)^ct$

$c^T=z_0^TX^T$

$c=Xz_0=X(X^TX)^ct$

$t^T=z^TX^TX$

判断 estimable,

$c_1^T\beta$, $c_2^T\beta$ estimable & c full rank

$c(X^TX)^c(X^TX)=c$ & c full rank

注: deviance (model) → SS_{Res}

可构建:

$F=\frac{\frac{SS_{Reg}}{\sigma^2}/r}{\frac{SS_{Res}}{\sigma^2}/n-r}=\frac{MS_{Reg}}{MS_{Res}}=\frac{SS_{Reg}/r}{s^2}\sim\chi^2(m,\lambda)$, $s^2=SS_{Res}/n-r$

r = # params in X $=F_{r,n-r}$

是个 one-tail test,

F越大, p-value越小, 越reject H0.

被r个 parameter 解释的 variation 多.

General Linear Hypothesis

full rank model

① H0: $C\beta=\delta^*$, C: m×p, $r(C)=m$.

C: m rows, 每个row 都 linearly independent.

检查 C 是否 estimable,

$C(X^TX)^c(X^TX)=C$,

H0: $C\beta=0$ testable.

如果 less than full rank,

H0: $C\beta=\delta^*$ testable (✓)

因为 $Cb-\delta^*$ 是 estimable.

大多数情况下 $\delta^*=0$,

$b\sim MVN(\beta,\sigma^2(X^TX)^{-1})$

$Cb\sim MVN(C\beta,C\sigma^2(X^TX)^{-1}C^T)$

$Cb-\delta^*\sim MVN(C\beta-\delta^*,C\sigma^2(X^TX)^{-1}C^T)$

⇒ $\frac{(Cb-\delta^*)^T[C(X^TX)^{-1}C^T]^{-1}(Cb-\delta^*)}{\sigma^2}\sim\chi^2(m,\lambda)$, $m=r(C)$

$\lambda=\frac{(C\beta-\delta^*)^T[C(X^TX)^{-1}C^T]^{-1}(C\beta-\delta^*)}{2\sigma^2}$

under H0, λ=0,

$F=\frac{(Cb-\delta^*)^T[C(X^TX)^{-1}C^T]^{-1}(Cb-\delta^*)/m}{s^2}\sim F_{m,n-r}$, $m=r(C)$, $r=r(X)$

for less than full rank model,

$F=\frac{Cb[C(X^TX)^cC^T]^{-1}Cb/m}{s^2}\sim F_{m,n-r}$.

③ one-way classification

$\bar y_i-\bar y_j\pm t_{\alpha/2}s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$

H0: $\tau_1=\cdots=\tau_k$, { $\tau_1-\tau_2=0$ … $\tau_{k-1}-\tau_k=0$ }

$m=r(C)=$ # hypothesis $=k-1$

④ $\beta=\begin{bmatrix}\gamma_1\\ \gamma_2\end{bmatrix}$ γ_1: m×1, γ_2: (p−m)×1, (部分params=0) H0: $\gamma_1=0$

解释: significance of γ_1 in presence of γ_2, m = # β in γ_1

$F=\frac{[R(\beta)-R(\gamma_2)]/m}{s^2}\sim F_{m,n-r}$

大, reject H0.

$R(\beta)$: SS_{Reg} of full model (β)

$R(\gamma_2)$: SS_{Reg} of model with γ_2

⑤ model relevance using corrected sum of squares. 其余 param 为0

默认 β_0 有意义, 则 $\gamma_2=\beta_0$.

$R(\beta_0)=y^T1(1^T1)^{-1}1^Ty=n\bar y^2$ correction factor

$R(\gamma_0)=y^Ty-R(\beta_0)$

$F=\frac{(SS_{Reg}-\frac{(\sum_i y_i)^2}{N})/(r-1)}{SS_{Res}/(n-r)}\sim F(r-1,n-r)$

理解为: response variable

$y-\bar y$, no intercept,

$SS_{Total}=(y-\bar y)^T(y-\bar y)=\sum_i^N(y_i-\bar y)^2$

$=y^Ty-R(\beta_0)$

M3

补充 Goodness-of-fit

$BIC=N\ln(\frac{SS_{Res}}{N})+p\ln(n)$

$C_p=\frac{SS_{Res}(model)}{s^2(full)}+2p-N$

Normal Equation

$X^TXb=X^Ty$

④ (1−α)% CI for σ^2

$\left(\frac{(n-r)s^2}{\chi^2_{n-r}(1-\frac{\alpha}{2})},\frac{(n-r)s^2}{\chi^2_{n-r}(\frac{\alpha}{2})}\right)$

Conditional Inverse

① $XX^cX=X$ $(X^TX)(X^TX)^cX^T=X^T$

$(X^T)^c=(X^c)^T$ $X(X^TX)^c(X^TX)=X$

$HX=X$ $X^TH=X^T$

② X^cX, XX^c, $I-X^cX$, $I-XX^c$ are idempotent.

③ $A(A^TA)^cA^T$ unique, sym, idemp.

$r(A(A^TA)^cA^T)=r(A)=r$

$I-A(A^TA)^cA^T$ unique, sym, idemp.

$r(I-A(A^TA)^cA^T)=n-r$.

Steps.

① 找 M: r(A)×r(A) non-singular.

② replace M in A with $(M^{-1})^T$

③ step 2 的 matrix transpose $=A^c$

one way classification

categorical.

$$X = \begin{bmatrix} 1_{n_1} & 1_{n_1} & & 0 \\ 1_{n_2} & 0 & 1_{n_2} & 0 \\ \vdots \\ 1_{n_k} & 0 & 0 & 1_{n_k} \end{bmatrix} (n \times p)$$

$$\beta = \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_k \end{bmatrix}$$ (less than full rank) ($\mu+\tau_i$ estimable)

$$(X^TX)^c = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \frac{1}{n_1} & & \\ 0 & & \ddots & \frac{1}{n_k} \end{bmatrix}$$

$$b = \begin{bmatrix} 0 \\ \bar{y}_1 \\ \vdots \\ \bar{y}_k \end{bmatrix}$$ 只需要每个 treatment 的 mean. $\hat{y}_i = \widehat{\mu+\tau_i}$

保证 $\mu+\tau_i$ & $\tau_j-\tau_i$ constant，改变任何 μ/τ_i，可获 valid b

$$s_i^2 = \frac{1}{n_i-1}\sum_j^{n_i}(y_{ij}-\bar{y}_i)^2$$

$$s^2 = \frac{(n_1-1)s_1^2+\cdots+(n_k-1)s_k^2}{(n_1-1)+(n_2-1)+\cdots+(n_k-1)} = \frac{SS_{Res}}{n-r}$$

$$F = \frac{(Cb)^T[C(X^TX)^cC^T]^{-1}(Cb)/m}{s^2} \sim F_{m,n-r}$$ m = # hypothesis

Summary last line & ANOVA 可以验证 $\tau_1=\tau_2=\cdots=\tau_k$，F 大 p 小，reject H_0，var 有效

two-way classification

更多的 categorical factor 为 X.

ANCOVA: continous + categorical

rank:

$$\vec{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \varepsilon_{11} \\ \vdots \end{bmatrix}$$

不可测任意一个 $\varepsilon_{ij}=0$，但可测 interaction 整体

n 需 > r，否则 $s^2 = \frac{SS_{Res}}{n-r}$ 不存在.

one-factor model $r(X) = p-1$

two-factor additive model:

$r(X) = 1+(a-1)+(b-1) = a+b-1$

$X: n \times (1+a+b)$

two-factor interaction model:

$X: n \times (1+a+b+a^*b)$

$r(X) = 1+a-1+b-1+(a-1)(b-1)$

$= ab.$

① continous variable 和 intercept 占用 1 个 parameter.

② catagorical var A with a levels & continous & cata var interaction 占用 a-1 个 ~~that~~ parameters

③ cate var 和 cate var 的 interaction 占用 (a-1)(b-1) 个 parameters.

Diagnostic

判断 model 好坏.

① residual $e = (I-H)y \sim MVN(0, \sigma^2(I-H))$

Standarised residual $z_i = \frac{e_i}{\sqrt{s^2(1-H_{ii})}}$

② leverage. (extremeness of a point in the x space)

$H_{ii} = (X(X^TX)^cX^T)_{ii}$

high potential effect to the fit.

③ Cook's Distance.

$$D_i = \frac{1}{k+1} z_i^2 \left(\frac{H_{ii}}{1-H_{ii}}\right)$$

检查一个点对 model fit influence.

$D_i < 0.5$，安全；$D_i > 1$，显著.

General Linear model Assumption

① linear relationship holds

② $\varepsilon \sim MVN(0, \sigma^2 I)$

③ 符合 model 的模型:

a) residual 在 0 左右的红线，分布均匀

b) constant var (which 3)

若上扬，则 heteroskedasticity.

c) normally distributed (which 2)

Model Selection

调整 $\Delta SS_{Reg} = SS_{Reg}(full) - SS_{Reg}(reduced)$

Δ 大, significant, F-stat 也大.

① forward selection.

每次添加 f stat 最大的 variable，直到没有 significant var.

对应 R-code 看 Sum of Sq 代表 Δ

$$F\text{-value}_i = \frac{\text{Sum of Sq}_i/df_i}{\text{当前 model RSS}/n-r} \sim F(df_i, n-r)$$ (包含 var i 的 model 的 s^2)

r = 加上 var i 模型的 # param (intercept 也算) $s_i^2 = RSS_i/n-r$

② Backward elimination.

full model 开始，依次删 least sig.

每轮的 full: remove var 之前

reduced: remove 当前 var 之后

① 和 ② 可能无法做到 optimal.

③ stepwise selection.

measure: $AIC = n\ln\left(\frac{SS_{Res}}{n}\right)+2p$

p: # param in the model (reduce 后)

AIC 越小越好.

缺点：有可能有不 significant vari 因为未比对 Fstat.

优点：易找 optimal.

相比于 prefer 复杂模型的 s^2 (因为 s^2 & R^2 只会降)

$$R^2: 1-\frac{SS_{Res}}{SS_{total}-\sum_i y_i^2/n}$$

Adjusted R^2.

$$1-\frac{n-1}{n-k-1}\cdot(1-R^2)$$

增加 # param penalty. 但仍不如 AIC

Experimental Design

方法

① control group: 建立 base level.

placebo, as treatment

② Blocking. 将 subjects 分组. 根据 confounding factor, each subpopulation is homogeneous in the confounding variables.

③ Randomisation: 随机给 subject 分 treatment (within each block)

avoid impact of lurking factors

④ Blind & Double-blind testing

avoid response bias.

⑤ replication: 越多，减少偶然性，precision 越高.

Type of Design

① CRD

没有 confounding，类 one-way classi

$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i=1..a$, $j=1..n_i$

$= \mu_i + \varepsilon_{ij}$

$\hat{\mu}_i = \hat{y}_i$, $var(\hat{\mu}_i) = \frac{\sigma^2}{n_i}$

常见分配资源问题，minimize var，用 Lagrange multiplier,

$$f(n_1..n_k, \lambda) = \sigma^2\sum_1^k\frac{1}{n_i} + \lambda(\text{constraints})$$ $n_1..n_k$

$\Rightarrow n_i^2 = \frac{\sigma^2}{\lambda}$, $n_i = \frac{n}{k}$

② CBD

1 个 confounding，适用 two-way model

$$y_{ijk} = \mu + \underset{\substack{\downarrow \\ \text{size of block}}}{\beta_i} + \underset{\substack{\downarrow \\ \text{\# treatments}}}{\tau_j} + \varepsilon_{ijk}$$, $i=1..a$, $j=1..b$, $k=1..n_{ij}$

block 是 confounding 的划分.

$$X = \Big[\underbrace{\mu\ \beta_1 \cdots \beta_a}_{X_1}\ \underbrace{\tau_1 \cdots \tau_b}_{X_2}\Big] \quad \beta = \begin{bmatrix} \mu \\ \beta_i \\ \tau_j \end{bmatrix} \begin{matrix} 1 \\ a \\ b \end{matrix}$$

$$y = [X_1|X_2]\begin{bmatrix}\beta_1 \\ \hline \beta_2\end{bmatrix} + \varepsilon$$，对应 $\begin{bmatrix} b_1 \\ \hline b_2 \end{bmatrix}$

存在 iff $H_1 = X_1(X_1^TX)^cX_1^T$ 存在、

b_2 对应 model (仅 treatment)

$X_{2|1} = (I-H_1)X_2$ <design matrix>

normal equation:

$(X_{2|1}^TX_{2|1})b_2 = X_{2|1}^Ty$

$y = [I-H_1]X_2\beta_2 + \varepsilon$, $\hat{\beta}_2 = b_2$

Estimability

If $t^T\tau$ is estimable

$t^T[X_{2|1}^TX_{2|1}]^cX_{2|1}^TX_{2|1} = t^T$

H_1 和 $X_{2|1}$ 的简便计算

block = b   size of block = k

每个 block size 一样时，

$$H_1 = \frac{1}{k}\begin{bmatrix} J_k & & \\ & J_k & \\ & & \ddots\ J_k \end{bmatrix} (k\times b)\times(k\times b)$$

如果 # treatment = k 时，

1 sample per treatment in each B,

$$X_2 = \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \quad X_{2|1} = \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} - \frac{1}{k}\begin{bmatrix} J_k \\ \vdots \\ J_k \end{bmatrix}$$

$$X_{2|1}^TX_{2|1} = b\left[I_k - \frac{1}{k}J_k\right]$$

$$(X_{2|1}^TX_{2|1})^c = \frac{1}{b}I_k$$

CBD & CRD 对比

① estimator 一样 ($t^T\tau$)，CBD 可以 reduce variance.

② 没有 blocking factor (B 无效) CRD 优秀 (# var 少，df of s^2 高)

③ Latin Squares

2 个 confounding，因为 CBD dimension 过多.

LS. 使用前提：# blocks = # treatment levels

有 n 个 treatment，

each confounding fator 也需分 n 个 treatment.

构造：each treatment appears only once in each row & col

estimator：同 CRD 与 CBD

var(Latin) ≤ var(CRD)

④ BIBD

treatments levels > blocksize

block = b，size of block = k，

treatment level = t，t > k

需要：a) 每个 treatment 在 block 内最多出现 1 次

b) 每个 treatment 出现 $r = \frac{bk}{t}$ 次 (总)

c) 每组 treatment 在 B 中出现次数相同，$\lambda = r\left(\frac{k-1}{t-1}\right)$ times.

estimator：和 CRD 的不同.

$$b_2 = \frac{k}{\lambda t}\vec{q}$$

model:

$$y = X_1\begin{bmatrix}\mu \\ \beta\end{bmatrix} + X_2\tau + \varepsilon$$

Orthogonally Block Design

不同 design with same treatment

If $H_1^AX_2 = H_1^BX_2$，则

$X_{2|1}^A = X_{2|1}^B$ (equavalent Blocking)

CRD, CBD, Latin 都是 ↑

BIBD does not have treatment orthogonal to the block.