Vector space: addition 封(0 & inverse) ⊕ scalar multiplication. ($1x = x$)　projection matrix to M is not unique. but orthogonal projection matrix P_M is unique

linearly dependent: (if 0在其中，则是)，且有等价条件 $\Longleftrightarrow \exists\, \alpha_1 \cdots \alpha_n$ & $k \le n$ s.t. $x_k = \sum_{i \ne k} \alpha_i x_i$

basis: linearly independent ⊕ 任加入1个会变成 dependent.

dimension: basis 中 vector 的个数 (维数唯一定理: basis 不是唯一的，但 dimension 是)

coordinate: 对给定 basis 来说是唯一的

completion: 可把一组 independent vectors 补成 basis (不唯一的)　$M \cap M^c = \{\phi\}$　$V = M + M^c$

isomorphism: vector space U & V. ⊕ $x \in U, y \in V$ ⊕ $T(\alpha_1 x_1 + \alpha_2 y) = \alpha_1 T(x) + \alpha_2 T(y)$ ⊕ T is 1-1 (T^{-1} exist)

↳ indicates same dimension. $\longrightarrow$ $\forall$ n-dim vector space 与 R^n 同构.

linear subspace: 自含线性组合 $\longrightarrow$ indicate including 0.　sum & intersection are linear subspace.

flat: linear subspace 的平移 [平移 vector 不唯一]

Decomposition thm: $H \cap K = \{0\}$. ⊕ $z \in H+K \Longrightarrow$ unique decomposition $z = x + y$, $x \in H$, $y \in K$.

$\dim(H+K) = \dim(H) + \dim(K) - \dim(H \cap K)$.

linear transformation: $A: V \mapsto V_1 \subset V$.　$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$.

range $R(A) = \{y \mid y = Ax, x \in V\}$. $\longleftarrow$ dimension: Rank ($\rho(A)$)　$\Longrightarrow$ null space $N(A) = \{x \mid x \in V, A(x) = 0\}$ ↳ $\nu(A)$.

A nonsingular: A^{-1}存在 $\Longleftrightarrow$ $Ax = y$ has unique solution. $\Longleftrightarrow \nu(A) = 0$　$\rho(A) + \nu(A) = n$

$\Longleftrightarrow$ $Ax = 0$ 只有解 $x = 0$

basis $\{x_i\}_1^n$, $\{y_i\}_{i=1}^n \Longrightarrow$ unique nonsingular A, $Ax_i = y_i$

B nonsingular $\Longrightarrow \rho(AB) = \rho(BA) = \rho(A)$　$\rho(AB) \le \min(\rho(A), \rho(B))$　$\rho(A+B) \le \rho(A) + \rho(B)$

$P_{M|N}$: $M \cap N = \{0\}$ ⊕ $M + N = V$ ⊕ $z \in V$. $z = x + y \Longrightarrow P_{M|N} z = x$.

T is projection $\Longleftrightarrow$ T idenpotent. $T^2 = T$.

inner product: 对称⊕线性⊕~~…~~非负($\langle x, x\rangle$) $\xRightarrow{\text{induce}}$ Hilbert space: complete w.r.t. distance defined by inner product

$x \perp y$ (orthogonal) $\langle x, y\rangle = 0$ $\longrightarrow$ orthogonal subspace $\longrightarrow$ orthogonal basis $\longrightarrow$ orthogonal complement C

$\longrightarrow$ orthonormal basis $\|x\| = 1$

找 orthonormal basis: (Gram-Schmidt) $y_1 = x_1$,　$y_k = x_k - \sum_{i=1}^{k-1} \frac{\langle x_k, y_i\rangle}{\|y_i\|^2} y_i$　$k = 2, \cdots n$

$\{x_i\}$ ONB $\Longrightarrow y = \sum_i \langle y, x_i\rangle x_i = \sum_i \lambda_i x_i \Rightarrow \|y\|^2 = \sum_{i=1}^n \lambda_i^2$

$\min_{w \in M} \|z - w\|^2 \Longrightarrow \hat{w} = P_M z$　transpose of linear transformation: $\langle Ax, y\rangle = \langle x, A'y\rangle$

$A \ge 0 \Longleftrightarrow \forall x \ne 0, \langle x, Ax\rangle \ge 0$.　$A > 0 \Longleftrightarrow \forall x \ne 0, \langle x, Ax\rangle > 0$

orthogonal projection: $P_{M|M^\perp} = P_M \Longleftrightarrow P = P^2 = P^T$　$\|Px\|^2 \le \|x\|^2$

orthogonal orthogonal projections: $A \perp B \Longleftrightarrow R(A) \perp R(B) \Longleftrightarrow AB = \cancel{AB'} = A'B = 0$　$A = \sum_{i=1}^k A_i$ (sum of OP): A OP $\Longleftrightarrow A_iA_j = 0\ \forall i, j$

$M = \sum_{i=1}^k M_i$ sum of orthogonal subspace $\Longrightarrow \|Px\|^2 = \sum_{i=1}^k \|P_{M_i} x\|^2$.

Orthogonal Transformation: $\|Ax\| = \|x\| \Longleftrightarrow \langle Ax, Ay\rangle = \langle x, y\rangle\ \forall x, y \Longleftrightarrow A^T = A^{-1} \Longleftrightarrow \forall$ ONB $\{x_i\}$, $\{Ax_i\}$ is also ONB.

A matrix is a linear function. A ~~symmetric~~ $n \times n$ matrix is a linear transformation. (square matrix)

matrix A: $R(A)$ A 的 column space.　$tr(A+B) = tr(A) + tr(B)$　$tr(ABC) = tr(BCA) = tr(CAB)$.

$\det(A) = \sum (-1)^{f(i_1 \cdots i_n)} a_{1i_1} \cdots a_{ni_n} = \prod$ eigenvalues.　$tr(A) = \sum$ eigenvalues.

Orthogonal matrix: $\Gamma^T \Gamma = \Gamma \Gamma^T = I$　区别 orthogonal project: 是描述投影到的空间的

Spectral Decomposition: symmetric $A \Longrightarrow A = \Gamma D \Gamma'$　Γ orthogonal matrix $= (\gamma_1 \cdots \gamma_n)$　$D = \mathrm{diag}(\lambda_1 \cdots \lambda_n) \Leftarrow$ all eigen.

$= \sum_{i=1}^n \lambda_i \gamma_i \gamma_i^T$　$R(A)$, $N(A)$都可由Γ的列组成

~~$A > 0 \Longleftrightarrow$~~

Symmetric $A > 0 \Longleftrightarrow x'Ax > 0$

Singular value Decomposition: $A = U_{n \times n} D_{n \times p} V_{p \times p}^T$　D: eigenvalues of AA^T或A^TA.　U: eigenvectors of AA^T.　V: eigen- of A^TA

QR Decomp: $A = QR = (Q_1, Q_2)\begin{pmatrix} R_1 & * \\ 0 & 0 \end{pmatrix}$　$= U_1 D_1 V_1^T$ (只取非0的部分)　$U_1^T U_1 = I_r$　$V_1^T V_1 = I_r$　$R(A) = R(U_1)$　$R(A^T) = R(V_1)$　$N(A) = R(V_2)$

rank(A) $= r$,　Q_1 $n \times r$ 维.　R_1 上三角阵.

Orthogonal projection matrix $\Longrightarrow$ 所有 eigenvalue 非1即0. $\Longrightarrow \rho(P) = tr(P) \Longrightarrow P \ge 0$

投 $X(X^TX)^{-1}X^T$ 唯一给出. $P = Q_1 Q_1^T = U_1 U_1^T = XX^+ \longrightarrow$ MPGI [因此 P 是 r 维的, 与 X 同维]

generalized inverse: $AA^-y = y$ for $\forall y \in R(A) \Longleftrightarrow AA^-A = A \Longrightarrow$ unique MPGI: $V_1 D_1^{-1} U_1^T = A^+$ $\Longleftarrow$ $(AA^+)^T = AA^+$, $(A^+A)^T = A^+A$, $AA^+A = A$, $A^+AA^+ = A^+$

$\rho(A_{n \times p}) = p \Longrightarrow A^+ = (A^TA)^{-1}A^T$

OP sum. $\sum_{i=1}^m P_i = I \Longrightarrow \|y\|^2 = \sum_{i=1}^m \|P_i y\|^2 = \sum_{i=1}^m y^T P_i y$

P is OP ⊕ $EY = 0$ ⊕ $Var(Y) = \sigma^2 I \Longrightarrow E(Y^TPY) = \sigma^2 \mathrm{rank}(P)$

OLS problem: $\min_{\mu \in \mathcal{V}} \|y-\mu\|^2$ assume $EY=\mu$ ⊕ $Var(y)=\sigma^2 I$ ⊕ $\mathcal{V}$ is linear subspace (r维). $\Rightarrow \hat{\mu} = P_{\mathcal{V}} y$ orthogonal projection.

$E\|y-\mu\|^2 = n\sigma^2 = E\|\hat{\mu}-\mu\|^2 + E\|y-\hat{\mu}\|^2 = r\sigma^2 + (n-r)\sigma^2 \Rightarrow \hat{\sigma}^2 = \frac{\|y-\hat{\mu}\|^2}{n-r}$

full rank case: $\mathcal{V} = R(X)$. rank p. $\Rightarrow P_X = X(X^TX)^{-1}X^T \Rightarrow \hat{\mu} = P_X y \Rightarrow \hat{\sigma}^2 = \frac{\|y-\hat{\mu}\|^2}{n-p}$

BLUE: $\widehat{\langle b, \mu \rangle} = \langle c, y \rangle$ 无偏 ⊕ variance 最小. 无偏 $\Leftrightarrow Pc = Pb \Leftrightarrow c = b + Q a$. some $a \in \mathbb{R}^n$.

Gauss-Markov: unique BLUE of $\langle b, \mu \rangle$ is $\langle Pb, y \rangle$ [即使X不满秩, P也唯一] if estimate $\langle a, \beta \rangle$. $a \in R(X^T)$. $a = X^T b$. $= \langle b, Py \rangle = \langle b, \hat{\beta} \rangle$

normal equation (for $\hat{\beta}$) $(X^TX)\hat{\beta} = X^TY$. (解不唯一) $\Leftrightarrow \hat{\beta}$ 是 OLS.

① 满秩: $\hat{\beta} = (X^TX)^{-1}X^TY$. $E\hat{\beta} = \beta$. $Var(\hat{\beta}) = \sigma^2 (X^TX)^{-1} \Rightarrow \langle a, \hat{\beta} \rangle$ is unique BLUE for $\langle a, \beta \rangle$.

② 非满秩: $\hat{\beta}^* = (X^TX)^+ X^TY$. $\hat{\beta} = \hat{\beta}^* + z$. $z \in N(X^TX) = N(X) = R(X^T)^\perp$ 这个等式可以直推到 SVD.

$\hat{\beta}$ 一般都有偏. but if $a \in R(X^T)$. $\langle a, \beta \rangle$ is estimable.

对任何的 $\hat{\beta}$, $\langle a, \hat{\beta} \rangle$ 都是 $\langle a, \beta \rangle$ 的 unique BLUE.

} 实践中直接把多出来那 $p-r$ 个 β 设成 0.

3 ways dealing with $\rho(X) < p$: ① estimates for estimable $c^T\beta$ always unique.

② set some of β to 0

③ additional linear constraints. [not in $R(X^T)$, $\underbrace{t \text{ num of vectors}}_{\Delta}$. $t = p-r$] $\hat{\beta} = (X^TX + \Delta^T\Delta)^{-1}X^TY$.

GLS: $\varepsilon \sim (0_n, \sigma^2 \Sigma_{n\times n})$ multiply $\Sigma^{-\frac{1}{2}}$

consistency: $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \delta) = 0 \quad \forall \delta > 0$ $\quad [\, P(|\hat{\theta}_n-\theta|>\delta) \le \delta^{-2} MSE(\hat{\theta}_n) = \delta^{-2}[Bias(\hat{\theta}_n)^2 + Var(\hat{\theta}_n)] \,]$

① $P = P_X$: $P_{ii} \to 0 \Rightarrow \hat{\mu}_i$ is consistent. ② $\max_i P_{ii} \to 0$. $\Rightarrow \forall a \in \mathbb{R}^n$. $a^T\hat{\mu}$ is asymptotically normal.

MGF of multivariate normal: $M_X(t) = E(e^{t^TX}) = \exp(t^T\mu + t^T\Sigma t/2) \Leftrightarrow \forall a \in \mathbb{R}^n$. $a^TX \sim N(a^T\mu, a^T\Sigma a)$

prove normal $\Leftrightarrow \exists\, r \ge 0$ iid normal $z_1 \cdots z_r$. $A_{n\times r} \Rightarrow \Sigma = AA^T$, $X = \mu + Az$

Σ singular $\Rightarrow$ no joint distribution for X.

conditional distribution: $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $\quad X_1 | X_2 = a \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a-\mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

$X \sim N(\mu, \Sigma)$. $AX \perp BX \Leftrightarrow A\Sigma B^T = 0$.

$\chi^2_m \sim Gamma(\frac{m}{2}, \frac{1}{2})$. $E(e^{tX}) = (\frac{1}{1-2t})^{\frac{m}{2}}$ $\quad \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = f(x)$ $\quad$ Non Central χ^2_m: $z_1 \cdots z_m \sim N(\mu_i, 1)$

$P(y-\mu) \sim N(0, \sigma^2 P)$. $\sigma^{-2}\|P(y-\mu)\|^2 \sim \chi^2_r$ $\quad \sigma^{-2}\|y-\mu\|^2 \sim \chi^2_n$ $\quad X = \sum_{i=1}^m z_i^2 \sim \chi^2_m(\delta^2)$. $\delta^2 = \sum_{i=1}^m \mu_i^2$

A symmetric. $z \sim N(0_n, I_n)$: $z^TAz \sim \chi^2_p \Leftrightarrow A$ is orth proj on a p-dim space

Cochran's Thm: $Q(x) = \sum_{i=1}^k Q_i(x) = \sum_{i=1}^k x^T T_i x$ ~~(T_i symmetric)~~ $Y \sim N(\mu, I)$ $Q(Y) \sim \chi^2$, then.

$Q_i(Y) \sim \chi^2 \Leftrightarrow T_i^2 = T_i \Leftrightarrow T_iT_j = 0 \ i \ne j \Leftrightarrow \sum_{i=1}^k rank(T_i) = rank(T)$. $T = \sum T_i$.

$\sim \chi_{rank(T_i)}(Q_i(\mu))$ $\quad \cancel{Q_i(Y) \sim \chi^2_{rank(T_i)}(Q_i(\mu))}$, ind

independent between i.

F distribution: $X \sim \chi^2_m$. $Y \sim \chi^2_n$. $\frac{X/m}{Y/n} \sim F_{m,n}$. $\quad$ t distr: $X \sim N(0,1)$. $Y \sim \chi^2_m$. $\frac{X}{\sqrt{Y/m}} \sim t_m$.

$Y \sim N(\mu, \sigma^2 I_n)$. $\mu = X\beta \in \mathcal{V} = R(X)$. $\rho(X) = r$. $\Rightarrow \hat{\mu} \sim N(\mu, \sigma^2 P)$. ; $\hat{\mu}(\hat{\beta}) \perp (I-P)Y$. ; $\sigma^{-2}\|P(y-\mu)\|^2 \sim \chi^2_r$; $\sigma^{-2}\|y-Py\|^2 \sim \chi^2_{n-r}$

$a \in R(X^T)$. $\theta = a^T\beta$. estimable. $\hat{\theta} = a^T\hat{\beta}$ BLUE of θ. $\Rightarrow E(\hat{\theta}) = \theta$. $\sigma^2_{\hat{\theta}} = \sigma^2 a^T(X^TX)^-(X^TX)(X^TX)^- a$; $se(\hat{\theta}) = \hat{\sigma}\sqrt{a^T(X^TX)^- X^TX(X^TX)^- a}$

$\Rightarrow T = \frac{\hat{\theta}-\theta}{se(\hat{\theta})} \sim t_{n-r}$

nested model: $\begin{cases} Y = X_1\beta + \varepsilon \\ Y = X_2\beta + \varepsilon \\ Y = X\beta + \varepsilon \end{cases}$ $\quad \mathcal{V}_2 \subset \mathcal{V}_1 \subset \mathcal{V}$. $R(X_2) \subset R(X_1) \subset R(X)$. $\quad \mathbb{R}^n = \mathcal{V}_2 + \mathcal{V}_2^\perp(\mathcal{V}_1) + \mathcal{V}_1^\perp(\mathcal{V}) + \mathcal{V}^\perp \Rightarrow \|y\|^2 = \|P_{\mathcal{V}_2}Y\|^2 + \|P_{\mathcal{V}_2^\perp(\mathcal{V}_1)}Y\|^2 + \|P_{\mathcal{V}_1^\perp(\mathcal{V})}Y\|^2 + \|P_{\mathcal{V}^\perp}Y\|^2$

$RSS_2 = \|Q_{\mathcal{V}_2}Y\|^2 = \|y\|^2 - \|P_{\mathcal{V}_2}Y\|^2$

$RSS_1 = \|Q_{\mathcal{V}_1}Y\|^2 = \|y\|^2 - \|P_{\mathcal{V}_2}Y\|^2 - \|P_{\mathcal{V}_2^\perp(\mathcal{V}_1)}Y\|^2$ $\quad \frac{(RSS_1 - RSS_{true})/(rank(X) - rank(X_1))}{RSS_{true}/(n - rank(X))} \sim F_{rank(X)-rank(X_1),\, n-rank(X)}(\delta^2)$

$RSS_{true} = \|Q_{\mathcal{V}}Y\|^2 = \|P_{\mathcal{V}^\perp}Y\|^2$

$\mathcal{V} \setminus \mathcal{V}_1 = \mathcal{V}_1^\perp(\mathcal{V})$ $\quad \delta^2 = \|P_{\mathcal{V}_1^\perp(\mathcal{V})}\mu\|^2/\sigma^2$

non-central χ^2: $y_i \sim N(\mu_i, 1)$ indep. $\Rightarrow y^T y \sim \chi_n^2(\delta^2)$ $\delta^2 = \|\vec{\mu}\|^2$, $Ey^Ty = n+\delta^2$.

BLUE: $\varphi = c^T\beta$ OLS: $(X^TX)\hat{\beta} = X^TY$. $\Rightarrow \hat{\varphi} = c^T\hat{\beta}$ $P_S = X(X^TX)^{-1}X^T$. $\Rightarrow Var(y) = \sigma^2 I_n$. $Var(\hat{\mu}) = \sigma^2 P_S$.

factorization thm: $f(y;\theta) = h_\theta(t(y))g(y)$. $\Rightarrow t(y)$ sufficient.

Rao-Blackwell thm: $\tilde{g}$ unbiased of θ, t sufficient. $\hat{g} = E[\tilde{g}|t]$ MVUE $\xrightarrow{\text{density complete}}$ UMVUE.
$E_\theta f(y)=0 \Rightarrow f(y)=0$ a.s.

MLE: $\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\, L(\theta)$. $\Rightarrow$ ① $\hat{\theta}_{MLE}$ consistent; ② $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{D} N(0, I(\theta)^{-1})$.

argue for VMUE: ① 计算 likelihood. ② sufficient stat. ③ Rao-Blackwell.

F-test: $R^n = \underbrace{\mathfrak{S}_0}_{q} + \underbrace{(\mathfrak{S}-\mathfrak{S}_0)}_{p-q} + \underbrace{\mathfrak{S}^\perp}_{n-p}$ $\|y\|^2 = \|P_{\mathfrak{S}_0}y\|^2 + \underbrace{\|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2 + \|Q_{\mathfrak{S}}y\|^2}_{\|Q_{\mathfrak{S}_0}y\|^2}$ $\mu\in\mathfrak{S}_0 \Leftrightarrow \mu\in\mathfrak{S}$. $\cancel{F = \frac{\|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2/(p-q)}{\|Q_{\mathfrak{S}}y\|^2/(n-p)}}$

$$F = \frac{\|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2/(p-q)}{\|Q_{\mathfrak{S}}y\|^2/(n-p)} \sim F_{p-q,n-p}\left(\frac{\|P_{\mathfrak{S}-\mathfrak{S}_0}\mu\|^2}{\sigma^2}\right) = \frac{(\|Q_{\mathfrak{S}_0}y\|^2 - \|Q_{\mathfrak{S}}y\|^2)/(p-q)}{\|Q_{\mathfrak{S}}y\|^2/(n-p)}$$

原因: $\|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2/\sigma^2 \sim \chi^2_{p-q}\left(\frac{\|P_{\mathfrak{S}-\mathfrak{S}_0}\mu\|^2}{\sigma^2}\right)$

Calculate $\|Q_{\mathfrak{S}}y\|^2$ with QR decomposition: $X = Q^T_{X_{xp}} R_{p\times p}$ full rk upper triangular. $QX = R$.

let $Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$ Q_1 $p\times p$. then $P_S = Q_1^TQ_1$. $\Rightarrow \|Q_S y\|^2 = \langle y, y\rangle - \langle Q_1 y, Q_1 y\rangle$.

Connection between LRT & F test. LRT: $\Lambda(Y) = \dfrac{\sup_{\mu\in\mathfrak{S}_0} L(\mu,\sigma^2)}{\sup_{\mu\in\mathfrak{S}} L(\mu,\sigma^2)} \Rightarrow -2\log\Lambda(Y) \sim \chi^2_{p-q}$. n large. rely on an approximation.

$$F = \frac{n-p}{p-q}\left[\Lambda(Y)^{-\frac{2}{n}} - 1\right]$$ doesn't rely on approximation.

Kronecker product: $A_{n\times m}$. $B_{r\times s} \Rightarrow A\otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots \\ \vdots & & \end{pmatrix}_{nr\times ms}$

① $(A\otimes B)\otimes C = A\otimes(B\otimes C)$ ② $(A\otimes B)(C\otimes D) = (AC\otimes BD)$. ③ $(A\otimes B)^T = B^T\otimes A^T$. ④ A.B. invertible $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$

1-way ANOVA Kronecker: $\mathfrak{S}_0 = R(J_n)$. $\mathfrak{S} = R(I_p\otimes J_m)$. p categories, m indv in each.

$\|Q_{\mathfrak{S}_0}y\|^2 = \sum_{i,j}(y_{ij} - \bar{y}_{..})^2$ $\|Q_{\mathfrak{S}}y\|^2 = \sum_{i,j}(y_{ij} - \bar{y}_{i.})^2 \Rightarrow \|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2 = \|Q_{\mathfrak{S}_0}y\|^2 - \|Q_{\mathfrak{S}}y\|^2$

H_0: $F \sim F_{p-1,n-p}$, $\Leftrightarrow$ H_1: $F \sim F_{p-1,n-p,\delta^2}$ $\delta^2 = \|P_{\mathfrak{S}-\mathfrak{S}_0}\mu\|^2 = \|Q_{\mathfrak{S}_0}\mu\|^2 - \|Q_{\mathfrak{S}}\mu\|^2$

power = P(reject H_0 | H_1 true)

general table $\longrightarrow$ 1 way ANOVA Table

Source	df	SS	EMS		SS	df
$\mathfrak{S}_0$	q	$\|P_{\mathfrak{S}_0}y\|^2$			$mp\bar{y}_{..}^2$	1
$\mathfrak{S}$	p	$\|P_{\mathfrak{S}}y\|^2$		$\Rightarrow$	$m\sum_{i=1}^{p}\bar{y}_{i.}^2$	p
$\mathfrak{S}-\mathfrak{S}_0$	$p-q$	$\|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2$			$m\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2$	$p-1$
$\mathfrak{S}^\perp$	$n-p$	$\|Q_{\mathfrak{S}}y\|^2$	σ^2		$\sum_{i,j}(y_{ij} - \bar{y}_{i.})^2$	$n-p$

Another example of testing. NH: $B\mu = 0$ $B_{r\times n}$, $\dim(B) = r$.

$\mathfrak{S} = R(P_S)$. $\mathfrak{S}-\mathfrak{S}_0 = R(P_SB^T)$

parametric version: NH: $\psi_1 = A_1\beta = 0$ $(A_1)_{r\times p}$ full rk. $(y = X\beta + \varepsilon)$. $A_1\beta$ estimable $\Leftrightarrow R(A_1^T)\subset R(X^T)$ $\Leftrightarrow A_1^T = X^TB^T$.

NH $\Leftrightarrow$ $BX\beta = 0 = B\mu = 0$ Back to old case $\Rightarrow \mathfrak{S}-\mathfrak{S}_0 = R(P_SB^T)$

$\hat{\varphi} = A_1\hat{\beta}$ $\hat{Var}(\hat{\varphi}) = \hat{\sigma}^2 A_1(X^TX)^{-1}A_1^T \Rightarrow \|P_{\mathfrak{S}-\mathfrak{S}_0}y\|^2 = \hat{\sigma}^2\hat{\varphi}^T[\hat{Var}(\hat{\varphi})]^{-1}\hat{\varphi} \Rightarrow F = \frac{1}{r}\hat{\varphi}^T[\hat{Var}(\hat{\varphi})]^{-1}\hat{\varphi} \sim F_{r,n-p,\delta^2}$.

$$\delta^2 = \frac{\|P_{\mathfrak{S}-\mathfrak{S}_0}\mu\|^2}{\sigma^2} = \frac{\beta^TA_1^T[A_1(X^TX)^{-1}A_1^T]^{-1}A_1\beta}{\sigma^2}$$

special case: $r=1$. NH: $c^T\beta = 0$. $\mathfrak{S}-\mathfrak{S}_0 = R(\underbrace{P_S c}_{z}) = R(z)$ $\hat{\mu} = P_S y \Rightarrow F = \dfrac{\langle c, \hat{\mu}\rangle^2}{\|P_S c\|^2\hat{\sigma}^2} \sim F_{1,n-p}$.

if NH: $c^T\mu = k$ $c^T\beta = k$.

$$\frac{\langle c, \hat{\mu}\rangle - \langle c, \mu\rangle}{\sqrt{\sigma^2\|P_S c\|^2}} \sim N(0,1)$$

2 way ANOVA. $r\times c$, 每格 m indvc.

$$\theta = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1c} \\ \theta_{21} & \cdots & \theta_{2c} \\ \vdots & & \\ \theta_{r1} & \cdots & \theta_{rc} \end{pmatrix}_{r\times c} \Rightarrow vec(\theta^T) = \begin{pmatrix} \theta_{11} \\ \vdots \\ \theta_{r1} \\ \theta_{12} \\ \vdots \\ \theta_{r2} \\ \vdots \\ \theta_{rc} \end{pmatrix} \quad \mu = Vec(\theta^T) \otimes J_m. \quad X = (I_r \otimes I_c \otimes J_m) \Rightarrow \mu = X\, Vec(\theta^T)$$

$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \end{pmatrix}$ $\quad S = R(X).$

$S = R(I_r \otimes I_c \otimes J_m).$

① grand mean $\quad S_1 = R(I_r \otimes \overset{J_c}{\cancel{I_c}} \otimes J_m) \quad \dim(S_1) = 1.$

② main A. $\quad S_A = R(I_r \otimes J_c \otimes J_m) \overset{decorr.}{\Longrightarrow} S_A^* = R(G_r \otimes J_c \otimes J_m) \quad \dim(S_A^*) = r-1. \quad G_r = I_r - J_r(J_r^TJ_r)^{-1}J_r^T$

③ main B. $\quad S_B = R(J_r \otimes I_c \otimes J_m) \overset{decorr}{\Longrightarrow} S_B^* = R(J_r \otimes G_c \otimes J_m) \quad \dim(S_B^*) = c-1$

④ interaction AB. $\quad S_{AB}^* = R(G_r \otimes G_c \otimes J_m) \quad \dim(S_{AB}^*) = (r-1)(c-1).$

⑤ $S^\perp \quad \dim(S^\perp) = rc(m-1) = n - rc$

$$\hat\alpha = A_A^T y = (G_r \otimes \frac{J_c^T}{c} \otimes \frac{J_m^T}{m})y \qquad \hat\beta = A_B^T y = (\frac{J_r^T}{r} \otimes G_c \otimes \frac{J_m^T}{m})y \qquad \hat\gamma = (G_r \otimes G_c \otimes J_m^T/m)y = A_{AB}^T y.$$

$$Var(\hat\alpha) = \sigma^2 A_A^T A_A = \frac{\sigma^2}{cm} G_r. \qquad Var(\hat\beta) = \frac{\sigma^2}{rm} G_c \qquad Var(\hat\gamma) = \sigma^2 A_{AB}^T A_{AB} = \frac{\sigma^2}{m}(G_r \otimes G_c)$$

2 way ANOVA table.

Source	df	SS	E(MS)
mean	1	$rmc\bar{y}_{+++}^2$	$\sigma^2 + rcm\bar{\mu}_{+++}^2$
A	$r-1$	$mc\|\hat\alpha\|^2$	$\sigma^2 + mc\|\alpha\|^2/(r-1)$
B	$c-1$	$mr\|\hat\beta\|^2$	$\sigma^2 + mr\|\beta\|^2/(c-1)$
AB	$(r-1)(c-1)$	$m\|Vec(\hat\gamma)\|^2$	$\sigma^2 + \frac{m\|vec(\gamma)\|^2}{(c-1)(r-1)}$
Error	$n-rc$	$rc(m-1)\hat\sigma^2$	σ^2

$$\hat\sigma^2 = \frac{1}{n-rc}\|(I - P_S)y\|^2$$

f test. $\delta^2 = \frac{SS}{\sigma^2}$. replace y with μ.

$$\hat\alpha = \begin{pmatrix} \bar y_{1++} - \bar y_{+++} \\ \vdots \\ \bar y_{r++} - \bar y_{+++} \end{pmatrix} \quad \hat\beta = \begin{pmatrix} \bar y_{+1+} - \bar y_{+++} \\ \vdots \\ \bar y_{+c+} - \bar y_{+++} \end{pmatrix} \quad \hat\gamma = \begin{pmatrix} \bar y_{11+} - \bar y_{1++} - \bar y_{+1+} + \bar y_{+++} \\ \vdots \\ \bar y_{rc+} - \bar y_{r++} - \bar y_{+c+} + \bar y_{+++} \end{pmatrix}$$

mean model.

$$vec(\bar y^T) \sim N(vec(\theta^T), \sigma^2 diag(\frac{1}{m_{ij}}))$$

$X = I_r \otimes I_c. \quad A_A = G_r \otimes \frac{J_c}{c}$

$A_B = \frac{J_r}{r} \otimes G_c$

$A_{AB} = G_r \otimes G_c$

weighted least square $\quad y = X\beta + \varepsilon \quad Var(\varepsilon) = \sigma^2\Sigma. \quad \Sigma = AA^T$

$\Downarrow$

$A^{-1}y = A^{-1}X\beta + A^{-1}\varepsilon \quad Var(A^{-1}\varepsilon) = \sigma^2 I$

$z = W\beta + \eta \Rightarrow \hat\beta_w = (W^TW)^{-1}W^Tz = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$

generalized linear models. (drop normality) $\quad g(E\,y_i) = x_i\beta$

$\mu_i = Ey_i. \Rightarrow \eta_i = g(\mu_i)$ link function.

$Var(\eta_i) = V(\mu_i)$ variance function exponential family: $f(y;\theta,\varphi) = \exp\left\{ \frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi) \right\}$. θ: canonical parameter

φ: scale parameter

- log likelihood: $l(\theta) = \frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi)$
- score function $u(\theta) = \frac{\partial l}{\partial \theta} \Rightarrow \begin{cases} E\,u(\theta) = 0 \Rightarrow EY = b'(\theta) \\ E\,u^2(\theta) = I(\theta) = -E\frac{\partial u}{\partial \theta} \end{cases}$
- $Var(Y) = -a(\varphi)b''(\theta) = V(\mu)$. $\mu = EY = b'(\theta)$
- link function choices: [canonical parameter]
  1. normal: $\theta = \mu$, $a(\varphi) = \sigma^2$. $g(\mu) = \mu$.
  2. binomial: $\cancel{\theta = \log\frac{\pi}{1-\pi}}$ $\binom{n}{y}\pi^y(1-\pi)^y$. $\theta = \log\frac{\pi}{1-\pi}$ $g(\mu) = \log\frac{\mu}{1-\mu}$
  3. poisson. $\frac{\lambda^y}{y!}e^{-\lambda}$ $\theta = \log\lambda$ $g(\mu) = \log\mu$.
- estimate β: iteratively weighted. least square

maximize $L(\beta) = \prod_{i=1}^{n} l_i(\beta)$. $l_i(\beta) = \log f(y_i;\theta,\varphi)$

$$\frac{\partial L}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \beta_j} = \sum_i \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}$$

$$= \sum_i \frac{y_i - b(\theta_i)}{a(\varphi)}\left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1}\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^{-1} x_{ij}$$

$$= \sum_i \frac{y_i - b(\theta_i)}{a(\varphi)}\frac{1}{b''(\theta_i)}\frac{g'(\mu_i)}{(g'(\mu_i))^2} x_{ij}$$

$$= \sum_i \frac{(y_i - b(\theta_i))g'(\mu_i)}{V(\mu_i)[g'(\mu_i)]^2} x_{ij}$$

let $W_i = \frac{1}{V(\mu_i)[g'(\mu_i)]^2}$ $\cancel{z_i = y_i + g'(\mu_i)(y_i - \mu_i)}$

$z_i = \eta_i + g'(\mu_i)(y_i - \mu_i)$

intuition. $g(y_i) = g(\mu_i) + g'(\mu_i)(y_i - \mu_i)$

$\approx \eta_i + g'(\mu_i)(y_i - b'(\theta_i)) \Rightarrow z_i - \eta_i = g'(\mu_i)(y_i - \mu_i)$

so $\frac{\partial l}{\partial \beta_j} = \sum_i (z_i - \eta_i) W_i x_{ij}$ $W = Diag(W_i)$ $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix}$

$X_{\cdot j} W(z - \eta) = \frac{\partial l}{\partial \beta_j}$

jth column

$$\begin{pmatrix} \frac{\partial l}{\partial \beta_1} \\ \vdots \\ \frac{\partial l}{\partial \beta_p} \end{pmatrix} = X^T W(z - \eta) = 0 \Rightarrow \hat{\beta} = (X^T W X)^{-1} X^T W z.$$

note z dependens on η & μ which depends on β.

so specify $\mu^{(0)} \Rightarrow \eta^{(0)}, z^{(0)} \Rightarrow \beta^{(0)}$

then iterate.