NOTE 1

△ full rank factorization $A_{p\times q} = K_{p\times r} L_{r\times q}$ full column full rank. $\begin{bmatrix} x & x \\ x & x \\ x & x \end{bmatrix}\begin{bmatrix} x&x&x&x \\ x&x&x&x \end{bmatrix}$ rank = 2. trace: 对角线元素之和. $B'B$, $C'C$ symmetric

△ idempotent matrices $(A^2=A)$ ⇒ properties: ① All are singular (except I) ② $r(A)=tr(A)$ (rank, trace) $(tr(BC)=tr(CB))$ ③ eigenvalues are either 0/1 $A^2x=Ax=\lambda x$

④ symmetric A, if all eigenvalue 0/1, then idempotent. $P'AP=D$

△ Moore-Penrose Inverse $A_{m\times n}$ **Def.** A^+ satisfy ① AA^+ symmetric ② A^+A symmetric ③ $A^+AA^+=A^+$ ④ $AA^+A=A$. **Thm.** ① Each A has an A^+. $A^+=C'(CC')^{-1}(B'B)^{-1}B'$ $(A=A_LA_R=BC)$ **Properties:** ① unique. ② $(A')^+=(A^+)'$ ③ $(A^+)^+=A$ ④ $r(A^+)=r(A)$

⑤ if $A=A'$, then $A^+=(A^+)'$ ⑥ if A nonsingular, $A^{-1}=A^+$ ⑦ if A symmetric idempotent, $A^+=A$ ⑧ if $r(A_{m\times n})=m$, then $A^+=A'(AA')^{-1}$, $AA^+=I$.

⑨ if $r(A_{m\times n})=n$, then $A^+=(A'A)^{-1}A'$, $A^+A=I$. ⑩ matrices A^+A, AA^+, $I-A^+A$, $I-AA^+$ are symmetric idempotent.

△ Generalized Inverse **Def** $A_{m\times n}$, then A^- satisfies $AA^-A=A$ (M-P also is..) (not unique) **Property** let $X_{m\times n}$, $r(X)=k>0$, then

① $r(X^-)\geq k$, ② X^-X and XX^- idempotent. ③ $r(X^-X)=r(XX^-)=k$ ④ $X^-X=I$ iff $r(X)=n$. ⑤ $XX^-=I$ iff $r(X)=m$ ⑥ $tr(X^-X)=tr(XX^-)=k=r(X)$

⑦ if X^- is any g-inverse of X, then $(X^-)'$... (X'). ⇒ let $K=X(X'X)^-X'$, K is invariant for any g-inverse of $X'X$ (don't have to full rank, could singular, K unchanged) ① $K=XX^+$ ② $K=K'$, $K=K^2$ (symmetric idempotent) ③ $rank(K)=rank(X)=r$ ($tr(K)$) ④ $KX=X$, $X'K=X'$

⑤ $(X'X)^-X'$ is g-inverse of X for any g-inverse of $X'X$ ⑥ $X(X'X)^-$ is a g-inverse of X' for any g-inverse of $X'X$. (also $I-K$ is invariant)

Ex. $rank[A_{n\times p}]=r$. $rank(A_{11})=r$ $A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}$ ⇒ $A^-=\begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0\end{bmatrix}$ 另一个possible g-inverse hint: take $B=\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^- & I\end{pmatrix}$ $A_{22}=A_{21}A_{11}^-A_{12}$ $rank(BA)=rank(A)=r$

NOTE 2

Random Vector and Matrices Y, X be p×1 random vectors

$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p\end{pmatrix}$ is given by $E(Y)=\begin{pmatrix} E(Y_1) \\ \vdots \\ E(Y_p)\end{pmatrix}=\begin{pmatrix} u_1 \\ \vdots \\ u_p\end{pmatrix}=u$. and $\Sigma = Cov(Y)=E\{[Y-E(Y)][Y-E(Y)]'\}=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots \\ \sigma_{21} & \ddots & \\ \vdots & & \sigma_{pp}\end{pmatrix}$ positive definite ⇒ $E(X+Y)=E(X)+E(Y)$ ⇒ $\Sigma=E(YY')-E(Y)[E(Y)]'$

⇒ A, B const matrix. $Cov(AY)=A[Cov(Y)]A'$ $(A'=A^T)$ $Cov(AX,BY)=A\,Cov(X,Y)B'$

⇒ Generalized variance $|\Sigma|=|$ determinant, product of eigenvalues $\Sigma=|P'DP|=|P'||D||P|=|D|$ P: orthogonal $D=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_p\end{bmatrix}$ $\prod\lambda_i=|$

⇒ correlation matrices $\Omega=\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ & 1 & & \\ & & \ddots & \\ \rho_{p1} & & & 1\end{pmatrix}$ $-1\leq\rho_{ij}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\leq 1$ take $Var(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y})$ and $Var(\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y})$ $D_\sigma=(Diag\,\Sigma)^{\frac12}$ and $\Omega=D_\sigma^{-1}\Sigma D_\sigma^{-1}$

⇒ partitioned random vectors

$V=\begin{pmatrix} Y \\ X\end{pmatrix}=\begin{pmatrix} X_1 \\ X_2 \\ X_3\end{pmatrix}$ $u=E(V)=E\begin{pmatrix} Y \\ X\end{pmatrix}=\begin{pmatrix} E(Y) \\ E(X)\end{pmatrix}=\begin{pmatrix} u_Y \\ u_X\end{pmatrix}$ $\Sigma=Cov(V)=Cov\begin{pmatrix} Y \\ X\end{pmatrix}=\begin{pmatrix}\Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX}\end{pmatrix}$ $tr(\Sigma D)=tr(D\Sigma)$

⇒ let Y random vector $u=E(Y)$, $\Sigma=Cov(Y)$ p.d. then $E(Y'AY)=tr(A\Sigma)+u'Au$ where A symmetric $\Sigma=E(YY')-uu'$ ($Y'AY$ is quadratic form)

⇒ MGF of random vector Y $M_Y(t)=E(e^{t'Y})$ $t=\begin{pmatrix} t_1 \\ \vdots \\ t_n\end{pmatrix}$ $(t_i<M, \exists M)$

⇒ for $g_1(Y_1)\ldots g_m(Y_m)$ be functions of $Y_1\ldots Y_m$, if $Y_1\ldots Y_m$ mutually independent then $g_1\ldots g_m$ mutually indep.

NOTE 3

Multivariate Normal Distribution density function $\frac{N(0,1)}{\sqrt{\chi^2_v/v}}\sim t_v$. $\frac{\chi^2_v/v}{\chi^2_n/n}\sim F_{v,n}$

⇒ Density $(Y_{p\times1}\sim N(u,\Sigma))$ $f_Y(y)=|\Sigma|^{-\frac12}(2\pi)^{-\frac p2}e^{-\frac12\{(y-u)'\Sigma^{-1}(y-u)\}}$ MGF: $M_Y(t)=e^{t'u+\frac12t'\Sigma t}$

⇒ for B a constant matrix and C constant vector $BY+C\sim N(Bu+C, B\Sigma B')$ → var-cov matrix

⇒ marginal distribution, condition distribution and independence

$Y=\begin{bmatrix} Y_1 \\ Y_2\end{bmatrix}\sim N\left[\begin{pmatrix} u_1 \\ u_2\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right]$ (u,Σ) $\left[\begin{array}{c|cc} x & x & x \\ \hline x & x & x \\ x & x & x\end{array}\right]$ then ① $Y_1\sim N(u_1,\Sigma_{11})$ ③ Y_1 and Y_2 are independent iff $\Sigma_{12}=0$ ② $Y_1|Y_2=y_2\sim N(u_1+\Sigma_{12}\Sigma_{22}^{-1}(y_2-u_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ ★

⇒ Partial Correlation $V\sim N_q(u,\Sigma)$ and $V=\begin{pmatrix} y \\ x\end{pmatrix}$; $u=\begin{pmatrix} u_y \\ u_x\end{pmatrix}$; $\Sigma=\begin{pmatrix}\Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx}\end{pmatrix}$ let $\rho_{ij.r\ldots q}$ be the partial correlation between y_i and y_j $\rho_{ij.r\ldots q}=\sigma_{ij.r\ldots q}/\sqrt{\sigma_{ii.r\ldots q}\sigma_{jj.r\ldots q}}$

$\Omega_{y.x}=D_{y.x}^{-1}\Sigma_{y.x}D_{y.x}^{-1}$ $\Sigma_{y.x}=\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$ $D_{y.x}=[diag(\Sigma_{y.x})]^{\frac12}$

⇒ non-central χ^2 property ⇒ let $x\sim N(0,I_n)$, then $x'x\sim\chi^2_{(n)}$ ⇒ $x\sim N(u,I_n)$, $u=X'X\sim\chi^2_{(n,\lambda)}$, $\lambda=\frac12u'u$

Density $f(u)=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\frac{u^{\frac12n+k-1}e^{-\frac12u}}{2^{\frac12n+k}\Gamma(\frac12n+k)}$ $u>0$, $\lambda\geq0$ $\lambda^k=1$ if $\lambda=k=0$ ⇒ $u\sim\chi^2_n$ $f(u)=\frac{u^{\frac12n-1}e^{-\frac12u}}{2^{\frac n2}\Gamma(\frac n2)}$ $\Gamma(\frac12)=\sqrt{\pi}$. $\Gamma(1)=1$ $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$ sum of χ^2. $\chi^2_{r_1}+\cdots+\chi^2_{r_n}=\chi^2(r_1+\cdots+r_n)$

- mgf of $u\sim\chi^2(n,\lambda)$ is $(1-2t)^{-\frac n2}e^{-\lambda[1-(1-2t)^{-1}]}$ and $u\sim\chi^2_n$'s mgf is $(1-2t)^{-\frac n2}$
- $u_1\sim\chi^2(p_1,\lambda)$ $u_2\sim\chi^2(p_2,0)$ $W=(u_1/p_1)/(u_2/p_2)\sim F(p_1,p_2,\lambda)$ - $z\sim N(u,1)$ $u\sim\chi^2_n$ $t=\frac{z}{\sqrt{u/n}}\sim t_{n,\lambda}$

⇒ Additional results partition matrix

① $|\Sigma|=|\Sigma_{22}||\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|$ ② $\Sigma^{-1}=\begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I\end{bmatrix}\begin{bmatrix}(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0' & \Sigma_{22}^{-1}\end{bmatrix}\begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0' & I\end{bmatrix}$ partition matrix inverse

then $(X-u)'\Sigma^{-1}(X-u)=[X_1-u_1-\Sigma_{12}\Sigma_{22}^{-1}(X_2-u_2)]'x[\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]^{-1}[X_1-u_1-\Sigma_{12}\Sigma_{22}^{-1}(X_2-u_2)]+(X_2-u_2)'\Sigma_{22}^{-1}(X_2-u_2)$

$f(X_1,X_2)=\underbrace{MN(u_2,\Sigma_{22})}_{f(X_2)}\cdot\underbrace{MN(u_1+\Sigma_{12}\Sigma_{22}^{-1}(X_2-u_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}_{f(X_1|X_2)}$

NOTE 4 Quadratic Forms $X'AX$

⇒ let $x \sim N(\mu, \Sigma)$, assume A symmetric **THM** mgf of $X'AX$ is $M_{x'Ax}(t) = |I - 2tA\Sigma|^{-\frac{1}{2}} \cdot e^{\{-\frac{1}{2}\mu'[I-(I-2tA\Sigma)^{-1}]\Sigma^{-1}\mu\}}$

(proof equal) $M_{x'Ax}(t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{tX'AX} \left(\frac{1}{\sqrt{2\pi}^p |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}\right) dx$ $-\frac{1}{2}[x'(I-2tA\Sigma)\Sigma^{-1}x - 2\mu'\Sigma^{-1}x + \mu'\Sigma^{-1}\mu]$

THM $X_{p\times 1} \sim N(\mu, V)$, then $q = X'AX \sim \chi^2_{(r,\lambda)}$ (non-central parameter) (assume V positive definite), $\text{rank}(A) = r$ iff AV is idempotent (A symmetric)

Lemma if A and V are symmetric and V positive-definite ($P'P$ (var-cov matrix)), then AV (don't need to symmetric) have eigenvalue 0 and 1 implies AV is idempotent.

Cor ① if $x \sim N(0, I)$, then $X'AX$ is χ^2_r iff A is idempotent of rank r. ② $x \sim N(0, V)$, $X'AX \sim \chi^2_r$ iff AV idempotent rank r.

③ $X \sim N(\mu, \sigma^2 I)$, $\frac{X'X}{\sigma^2} \sim \chi^2_{(n, \frac{\mu'\mu}{2\sigma^2})}$ ④ $X \sim N(\mu, I)$, $X'AX \sim \chi^2_{(r, \frac{1}{2}\mu'A\mu)}$ iff A idempotent of rank r. → 不需要 idempotent

做题思路：先证明 AV is idempotent，然后有 χ^2_n，再去求 $\text{rank}(A) = n$ 是多少 $E(\chi^2_n) = n$, $Var(\chi^2_n) = 2n$

Independence **THM** when $X \sim N_n(\mu, \Sigma)$, then $X'AX$ (χ^2) and BX (normal) are distributed independently iff $B\Sigma A = 0$

THM $X \sim N(\mu, \Sigma)$, $X'AX$ (χ^2) and $X'BX$ (χ^2) are distributed independently iff $A\Sigma B = 0$ (or $B\Sigma A = 0$)

Additional results let $Y_{n\times 1} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim N(\mu, \Sigma)$, $q_1 = Y'A_1Y$, $q_2 = Y'A_2Y$, $T = BY$ where $B_{r\times n}$ and A_1, A_2 symmetric

⇒ ① $E(q_1) = \text{trace}(A_1\Sigma) + \mu'A_1\mu$ ② $Var(q_1) = 2\,\text{trace}(A_1\Sigma A_1\Sigma) + 4\mu'A_1\Sigma A_1\mu$

③ $Cov(q_1, q_2) = 2\,\text{trace}(A_1\Sigma A_2\Sigma) + 4\mu'A_1\Sigma A_2\mu$ ④ $Cov(Y, q_1) = 2\Sigma A_1\mu$ ⑤ $Cov(T, q_1) = 2B\Sigma A_1\mu$

有些一眼就要认出来的：① $\sum_{i=1}^n (Y_i - \bar{Y})^2 = Y'(I_n - \frac{J_n}{n})Y$ ② $\bar{z}^2 = \frac{z'Jz}{n^2}$ ($\bar{z} = \frac{z'1}{n}$, $\bar{z}^2 = \bar{z}\bar{z} = \frac{z'1}{n}\frac{1'z}{n} = \frac{z'Jz}{n^2}$)

③ $\sum_{i<j}(Y_i - Y_j)^2 = Y'(nI - J)Y$ ④ $(I - \frac{J}{n})J = 0$ ⑤ $\sum Y_i^2 = Y'IY$

idempotent: $\frac{J}{n}$ (rank = 1), $(I_n - \frac{J}{n})$ (rank = n−1), $(nI - J)$ (trace = n(n−1)) not idempotent

$Q = X_1X_2 - X_3X_4$ construct $A = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 0 \end{pmatrix}$

⇒ $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, $(n-1)S^2 \sim \chi^2_{(n-1)}$, $\bar{X}$ iid of S^2

⇒ $Cov(☆, □) = E[(☆ - E(☆))(□ - E(□))]$

⇒ $MSE(Y'AY) = Bias(Y'AY) + Var(Y'AY) = [E(Y'AY - \sigma^2)]^2 + Var(Y'AY) = E(\hat{\theta} - \theta)^2$

⇒ 求矩阵的逆：① $[A|I] \to [I|A^{-1}]$ ② $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ ⇒ 求 rank：初等变换后不能约的行数

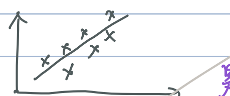
NOTE 5 Simple Linear Regression

$r = \frac{\hat{Cov}(X,Y)}{\sqrt{\hat{Var}(X)}\sqrt{\hat{Var}(Y)}} = \frac{\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}}{\sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}\sqrt{\frac{\sum(Y_i - \bar{Y})^2}{n-1}}} = \frac{S_{XY}}{S_{XX}S_{YY}}$ $T: \frac{\hat{\beta}_i}{S\{\hat{\beta}_i\}}$

⇒ $\sum(☆_i - \bar{☆})(□_i - \bar{□}) = S_{☆□}$

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ① $E(\varepsilon_i) = 0$ ② $Var(\varepsilon_i) = \sigma^2$ const

dependent (random) intercept slope independent (nonrandom) ③ $Cov(\varepsilon_i, \varepsilon_j) = 0$.



y-intercept slope (are called regression equation) 是 Y_i 的线性组合

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ (estimate) s.t. $y = 3 + 5x$. when x increase by 1 unit, y increase by 5 unit. • also $\hat{\beta}_1 = \sum k_i Y_i$

区分！ regression model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ least squares regression line: $\hat{Y} = 3 + 5x$ for $k_i = \frac{(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}$

Method of Least Squares (smallest of sum of squares) does not require normality • $\sum k_i = 0$

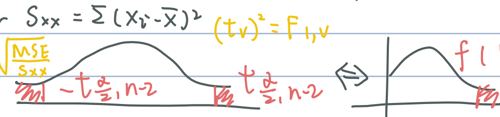
⇒ $S_{YY} = \sum(Y_i - \hat{Y}_i) = \sum[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]^2$; for $\frac{\partial S}{\partial \beta_1} = 0$, solve that $\hat{\beta}_1 = \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}} = \frac{r\sqrt{S_{XX}S_{YY}}}{S_{XX}} = r\sqrt{\frac{S_{YY}}{S_{XX}}}$

for $\frac{\partial S}{\partial \beta_0} = 0 \to \boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}}$ does not require normality, too

⇒ the unbiased estimator of σ^2 $E = \frac{SSE}{n - r(x)}$; $S^2 = MSE = \frac{SSE}{n-2} = \frac{\sum(Y_i - \hat{Y})^2}{n-2} = \frac{\sum e_i^2}{n-2}$ ⇒ the MLE of σ^2 is $\frac{SSE}{n} = \frac{n-2}{n}S^2$ (invariant)

and as $\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$ $Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2)$ and $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{S_{XX}})$ for $S_{XX} = \sum(X_i - \bar{X})^2$ $(t_v)^2 = F_{1,v}$

$100(1-\alpha)\%$ confidence interval for β_1 is $\hat{\beta}_1 \pm t_{n-2, \frac{\alpha}{2}} S\{\hat{\beta}_1\}$ $S\{\hat{\beta}_1\} = \sqrt{\frac{MSE}{S_{XX}}}$



$-t_{\frac{\alpha}{2}, n-2}$ $t_{\frac{\alpha}{2}, n-2}$ ⇔ $f(1, n-2, \alpha)$

$\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{S_{XX}}} \sim N(0,1)$ $\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$ $\hat{\sigma}^2 \sim \frac{SSE}{n-2} \sim \frac{\chi^2_{n-2}}{n-2}\sigma^2$

so $\frac{\hat{\beta}_1}{\sigma/\sqrt{S_{XX}}} \Big/ \sqrt{\frac{\chi^2_{n-2}}{n-2}} = \frac{\hat{\beta}_1 / (\sigma/\sqrt{S_{XX}})}{\sqrt{\hat{\sigma}^2/\sigma^2}} = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{XX}}} \sim t_{n-2}$ $\hat{\beta}_0 \sim N(\beta_0, \sigma^2[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}])$ $\hat{\beta}_0 \pm t_{n-2, \frac{\alpha}{2}} S\{\hat{\beta}_0\}$ for $S\{\hat{\beta}_0\} = \sqrt{MSE(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}})}$

{ 计算总的 (mean of) $E(Y_n) - \hat{Y}_n \sim N(0, \sigma^2[\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}])$ → confidence

{ 计算 one new observation $Y_{h(new)} - \hat{Y}_h \sim N(0, \sigma^2[1 + \frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}])$ prediction

two sided $100(1-\alpha)\%$

$\left(\hat{Y}_n - t_{\frac{\alpha}{2}, n-2} \cdot S\sqrt{\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}},\ \hat{Y}_n + t_{\frac{\alpha}{2}, n-2} \cdot S\sqrt{\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}}\right)$ ($S = \sqrt{MSE}$) lwr upr

$\left(\hat{Y}_n - t_{\frac{\alpha}{2}, n-2} \cdot S\sqrt{1 + \frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}},\ \hat{Y}_n + t_{\frac{\alpha}{2}, n-2} \cdot S\sqrt{1 + \frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum(X_i - \bar{X})^2}}\right)$

fit / $\hat{\beta}_0 + \hat{\beta}_1 X_n$ $S_{XX} = \frac{MSE}{(S\{\hat{\beta}_1\})^2}$

① $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$ ② $0 \le R^2 \le 1$ ③ $r = \pm\sqrt{R^2}$ (±看斜率)

State H_0, H_1, test F etc. In $\alpha = 0.05$, we have enough evidence to indicate model is useful ⇒

ANOVA Table

Some RMKS ① $\sum e_i = 0$ ② $\sum Y_i = \sum \hat{Y}_i$ ③ $\sum X_i(Y_i - \hat{Y}_i) = \sum X_i e_i = 0$

④ $\sum \hat{Y}_i(Y_i - \hat{Y}_i) = 0$ ⑤ line pass through $(\bar{X}, \bar{Y})$ ⑥ $SST = SSR + SSE$

	Sum of Squares (SS)	Degrees of freedom (df)	Mean squares (MS)	F
Regression	SSR $\sum(\hat{Y}_i - \bar{Y})^2$	1	MSR = SSR/1	MSR/MSE $\sim T^2$
Error	SSE $\sum(Y_i - \hat{Y}_i)^2$	n-2 ($Y'(I-H)Y$, rank of B is n-2)	MSE = SSE/(n-2) $\hat{\sigma}^2$	$F_{1, n-2}$
Total	SST $\sum(Y_i - \bar{Y})^2$ (S_{YY})	n-1		

⇒ P-value 小：null hypothesis is rejected, indicating that x contributed information for X.

⇒ residual standard error 是 $\hat{\sigma} = \sqrt{MSE}$ ⇒ Std. error 是 $S\{\hat{\beta}_i\}$

⇒ coefficient of determination 是 multiple (R^2) correlation ($r = \pm\sqrt{R^2}$) R-squared (ie. 0.374, 37.4% of the proportion of variations of x is explained by the model)

⇒ 4 methods ① F to $F_{(1,18, 0.05)}$ ② p-value to α. ③ t to $t_{(18, 0.025)}$. ④ if give you confidence interval contains 0.

Multiple Regression

$Y_{n\times1} = X_{n\times r}\beta_{r\times1} + \varepsilon_{n\times1}$

Assumption: ① $E(\varepsilon)=0$, ② $Cov(\varepsilon)=\sigma^2 I$. ③ X full column rank ($rank(X'X)=rank(X)=k+1$)

$\Rightarrow$ sum of squares of deviations. $S=(Y-X\hat\beta)'(Y-X\hat\beta)=\sum_{i=1}^n (Y_i-\hat Y_i)^2$. normal equation: $X'Y=X'X\hat\beta$ $\hat\beta=(X'X)^{-1}X'Y$. when $\beta=\hat\beta$, minimum

$\hat\varepsilon = Y-X\hat\beta = Y-X(X'X)^{-1}X'Y=[I-H]Y$. when $H=X(X'X)^{-1}X'$. so $\hat Y = X\hat\beta = HY$.

Notes: ① $X'\hat\varepsilon=0$ ($X'H=X'$, $HX=X$, $X'(I-H)=0$, $(I-H)X=0$) ② H, $(I-H)$ symmetric idempotent. ③ $\hat Y'\hat\varepsilon=0$. ④ $Var(\hat\beta)=(X'X)^{-1}\sigma^2$ when $Var(\hat\beta_i)=C_{ii}\sigma^2$

⑤ $E(\hat\beta)=\beta$. (unbiased) ⑥ $tr(I-H)=n-r$ ⑦ $\hat\varepsilon'\hat\varepsilon = tr(YY'(I-H))$ ⑧ $E(YY')=\sigma^2 I + X\beta\beta'X'$ ⑨ $E(\frac{\hat\varepsilon'\hat\varepsilon}{n-r})=\sigma^2$

Generalized Least Square Estimator: $Cov(\varepsilon)=V$的情况下, $\min (Y-X\beta)'V^{-1}(Y-X\beta)$ $\Rightarrow$ $\tilde\beta=(X'V^{-1}X)^{-1}X'V^{-1}Y$.

Gauss-Markov Thm (best linear unbiased estimator): t is a vector, find blue of $t'\beta$. $\Rightarrow$ let $E(\lambda'Y)=t'\beta$. $=\lambda'E(Y)=\lambda'X\beta$. so $\lambda'X=t$.

$Var(\lambda'Y)=\lambda'V\lambda$. $\lambda'=t'(X'V^{-1}X)^{-1}X'V^{-1}$ then blue of $t'\beta=\lambda'Y=t'(X'V^{-1}X)^{-1}X'V^{-1}Y$. $Var(\lambda'Y)=t'(X'V^{-1}X)^{-1}t$

$MLE(\beta)=\tilde\beta=(X'X)^{-1}X'Y$ $MLE(\sigma^2)=\tilde\sigma^2=\frac{1}{n}(Y'Y-\tilde\beta'X'Y)$ 比MSE偏估一点点, $SSE=Y'(I-H)Y$. $Y\sim N(X\beta,\sigma^2 I)$ $\hat\beta$和SSE, $\hat\beta$和$\hat\sigma^2$ independent

Distribution $\frac{SSE}{\sigma^2}=\frac{1}{\sigma^2}Y'(I-H)Y\sim\chi^2(rank(\frac{I-H}{\sigma^2}),\frac12(X\beta)'(\frac{I-H}{\sigma^2})(X\beta))\sim\chi^2(n-r(X),0)\sim\frac{[n-r(X)]\hat\sigma^2}{\sigma^2}$ when $Y\sim N(X\beta, I\sigma^2)$

Ex. $Y=X\beta+\varepsilon$ $Y\to X'Y\to(X'X)^{-1}$. $\to$ $\hat\beta=(X'X)^{-1}X'Y\to\hat\sigma^2=\frac{1}{n-2}(Y'Y-\hat\beta'X'Y)$

Deviations from Means: $Y_{n\times1}=X\beta_{(k+1)\times1}+\varepsilon_{n\times1}$, $\hat\beta=(X'X)^{-1}X'Y$, $Var(\hat\beta)=(X'X)^{-1}\sigma^2$, let $X=(1\ X_1)$, $\beta'(\beta_0\ b')$ 因为不想把β_0算进去

Let $\bar X'=(\bar x_{.1}, \bar x_{.2}\cdots\bar x_{.k})$, $1'1=n$, $1'Y=n\bar y$, $1'X_1=n\bar X'$. so $\hat\beta=\begin{bmatrix}\frac1n+\bar X'S^{-1}\bar X & -\bar X'S^{-1}\\ -S^{-1}\bar X & S^{-1}\end{bmatrix}\begin{bmatrix}n\bar y\\ X_1'Y\end{bmatrix}$ when $S=X_1'X_1-n\bar X\bar X'=Z'Z$. ($Z=X_1-1\bar X'$)

$\Rightarrow\hat\beta_0=\bar y-\bar X'\hat b=\frac1n 1'Y-\bar X'\hat b$, $\hat b=S^{-1}(X_1'Y-n\bar y\bar X)=(Z'Z)^{-1}Z'Y$, $Var(\hat b)=(Z'Z)^{-1}\sigma^2$. $Var(\hat\beta_0)=\frac{\sigma^2}{n}+\bar X'var(\hat b)\bar X$, $cov(\hat\beta_0,\hat b')=-\bar X'var(\hat b)$

Partitioning Total Sum of Squares $\Rightarrow SST=Y'Y-\frac1n Y'11'Y=Y'(I-\frac1n J)Y$, $\frac{SST}{\sigma^2}\sim\chi^2(n-1,\frac{\beta'X'X\beta-\frac1n(1'X\beta)^2}{2\sigma^2})$ $=Y'Y-\tilde\beta'X'Y$

$\Rightarrow SSR=SST-SSE=Y'(I-\frac1n 11')Y-Y'(I-H)Y=Y'(H-\frac1n 11')Y=\hat b'Z'Y=Y'Z(Z'Z)^{-1}Z'Y$, $\frac{SSR}{\sigma^2}\sim\chi^2(k,\frac{b'(Z'Z)b}{2\sigma^2})$ $SSE=Y'(I-H)Y$.

SSR independent SSE, $F\sim F(r(X)-1,n-r(X),\frac{b'(Z'Z)b}{2\sigma^2})$ 这里的b是没有β_0的.

Source	df	SS	MS	F-statistics
Regression	$r(\mathbf{X})-1$ k	$\hat{\mathbf{b}}'\mathbf{Z}'\mathbf{Y}$	$\frac{SSR}{r(\mathbf{X})-1}$	$F=\frac{MSR/\sigma^2}{MSE/\sigma^2}$
Error	$n-r(\mathbf{X})$ n-k-1	$\mathbf{Y}'\mathbf{Y}-\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{Y}$	$\frac{SSE}{n-r(\mathbf{X})}$ MSR MSE	
Total	$n-1$	$\mathbf{Y}'(\mathbf{I}-\frac1n\mathbf{1}\mathbf{1}')\mathbf{Y}$		

Adjust R^2 $R_a^2=1-[(1-R^2)\frac{n-1}{n-(k+1)}]<R^2$. to penalize excessive use of var.

as $R^2=\frac{SSR}{SST}$. $F=\frac{MSR}{MSE}=\frac{SSR/k}{SSE/n-k-1}=\frac{R^2(n-k-1)}{(1-R^2)k}$ 所以有F就有R^2.

Hypothesis Testing

χ^2 and F test under H_0 then all centraled.

$H_0: K'\beta=m$. Let $Y\sim N(X\beta,\sigma^2 I)$ Then $\hat\beta=(X'X)^{-1}X'Y\sim N(\beta,(X'X)^{-1}\sigma^2)$, $K'\hat\beta-m\sim N[K'\beta-m,\underbrace{K'(X'X)^{-1}K\sigma^2}_{Var(K'\hat\beta)}]$

$\Rightarrow$ if we let SSH. $Q=(K'\hat\beta-m)'[\underbrace{K'(X'X)^{-1}K}_{\text{rank is } q}]^{-1}(K'\hat\beta-m)$ then $\frac{Q}{\sigma^2}\sim\chi^2\{q,\frac{1}{2\sigma^2}\underbrace{(K'\beta-m)'[K'(X'X)^{-1}K]^{-1}(K'\beta-m)}_{\text{is 0 when } H_0}\}$ Q and SSE independent.

test statistics $F(H)=\frac{Q/q}{SSE/[N-r(X)]}=\frac{Q}{q\hat\sigma^2}\sim F[q,N-r(X),\frac{1}{2\sigma^2}(K'\beta-m)'[K'(X'X)^{-1}K]^{-1}(K'\beta-m)]$ when $\hat\sigma^2=\frac{SSE}{N-r(X)}$

under null hypothesis. $\tilde\beta$. $\Rightarrow\hat\beta-\tilde\beta=(X'X)^{-1}K(K'(X'X)^{-1}K)^{-1}(K'\hat\beta-m)$

△ without H_0. $SSE=(Y-X\hat\beta)'(Y-X\hat\beta)$ △ under H_0. $SSE_{H_0}=(Y-X\hat\beta)'(Y-X\hat\beta)+(\hat\beta-\tilde\beta)'X'X(\hat\beta-\tilde\beta)=SSE+Q$

Likelihood Ratio Test $F=\frac{\hat\beta'X'Y/(k+1)}{Y'Y-\hat\beta'X'Y/(n-k-1)}$ H_0 is rejected if $F>F_{\alpha,k+1,n-k-1}$ for $\hat\sigma^2=\frac{(y-X\hat\beta)'(y-X\hat\beta)}{n}$

$\Rightarrow SSR(X_3|X_1,X_2)=SSR(X_1,X_2,X_3)-SSR(X_1,X_2)=SSE(X_1,X_2)-SSE(X_1,X_2,X_3)$ (要找对自由度参数!)

¡Important! 在算partial F test的时候有两种方法 $\frac{(SSR_{full}-SSR_{red})/q}{MSE_{full}}$ ① 正常算SSE. 但有可能$\hat\beta'X'Y=[(X'X)^{-1}X'Y]'X'Y$很繁. F test一定是正的!

② $Q=SSE_{red}-SSE_{full}$, $Q=(K'\hat\beta-m)'[K'(X'X)^{-1}K]^{-1}(K'\hat\beta-m)$, 挑好算的算. (是为了证明是否reduced和full有一样utility)

More Complex Model

qualitative / categorical Predictor. indicator / dummy variable (split to n-1)

Coefficients of partial determination $R^2_{Y2,13}=\frac{SSR(X_2|X_1,X_3)}{SSE(X_1,X_3)}=\frac{SSR(X_1,X_2,X_3)-SSR(X_1,X_3)}{SSE(X_1,X_3)}$

Interaction Model. $b_3x_1x_2$ (Slope are different if the effect of X_1 on Y depends on X_2 value)

Polynomial Regression Model often use centered predictor ($x_i=X_i-\bar X$) as otherwise x and x^2 are highly correlated. multicollinearity↓ concave up or downward.

Confidence/Prediction intervals

Confidence region for β: $P[\frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{(k+1)\hat\sigma^2}\le F_{\alpha,k+1,n-k-1}]=1-\alpha$ $\left(\frac{\chi^2/k+1}{\chi^2/n-k-1}\right)$ $P[(\hat\beta-\beta)'X'X(\hat\beta-\beta)\le(k+1)\hat\sigma^2F_{\alpha,k+1,n-k-1}]=1-\alpha$

Confidence interval for β_j: $P[-t_{\frac\alpha2,n-k-1}\le\frac{\hat\beta_j-\beta_j}{S\{\hat\beta_j\}}\le t_{\frac\alpha2,n-k-1}]=1-\alpha$ so CI is $\hat\beta_j\pm t_{\frac\alpha2,n-k-1}S\{\hat\beta_j\}$ we estimate α ↑↓ if x is increase!

Confidence interval for $a'\beta$: (a is a row factor) 如 $(0,1,-1)\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}=\beta_1-\beta_2$. $t=\frac{a'\hat\beta-a'\beta}{\hat\sigma\sqrt{a'(X'X)^{-1}a}}\sim t_{n-k-1}$ so CI is $a'\hat\beta\pm t_{\frac\alpha2,n-k-1}\hat\sigma\sqrt{a'(X'X)^{-1}a}$

Confidence interval for $E(y_0)$ given x: 求平均气温 (average). $Var(\hat y_0-E(y_0))=Var(\hat y_0)=x_0'(X'X)^{-1}x_0\sigma^2$. CI: $x_0'\hat\beta\pm t_{\frac\alpha2,n-k-1}\hat\sigma\sqrt{x_0'(X'X)^{-1}x_0}$ (mean) $x=\begin{pmatrix}1\\x_1\\x_2\\ \vdots\end{pmatrix}$

Prediction interval for a future observation: 预测一天气温 (one particular). $\hat y_0=x_0'\hat\beta$, $Var(y_0-\hat y_0)=[1+x_0'(X'X)^{-1}x_0]\sigma^2$ ($Var(\hat y_0)+Var(y_0)$) CI: $\hat y_0\pm t_{\frac\alpha2,n-k-1}\hat\sigma\sqrt{1+x_0'(X'X)^{-1}x_0}$

Confidence interval for σ^2: as $\frac{(n-k-1)\hat\sigma^2}{\sigma^2}\sim\chi^2_{(n-k-1)}$, so $P[\chi^2_{1-\frac\alpha2,n-k-1}\le\frac{(n-k-1)\hat\sigma^2}{\sigma^2}\le\chi^2_{\frac\alpha2,n-k-1}]=1-\alpha$ CI: $\left(\frac{(n-k-1)\hat\sigma^2}{\chi^2_{\frac\alpha2,n-k-1}},\frac{(n-k-1)\hat\sigma^2}{\chi^2_{1-\frac\alpha2,n-k-1}}\right)$

Simultaneous intervals: ① Familywise CL: $100(1-\alpha)\%$ every interval ② $\boxed{\text{Bonferroni CI}}$: choose $\alpha_c=\frac{\alpha_f}{m}$. so is $a_i'\hat\beta\pm t_{\frac{\alpha}{2m},n-k-1}\hat\sigma\sqrt{a_i'(X'X)^{-1}a_i}$

③ $\boxed{\text{Scheffé CI}}$: $a'\beta\pm\hat\sigma\sqrt{(k+1)F_{\alpha,k+1,n-k-1}a'(X'X)^{-1}a}$ (m任意,如自助餐)

Ex. 把 X. β. Y 写出来 $\to X'X\to(X'X)^{-1}\to X'Y\to\hat\beta=(X'X)^{-1}X'Y$ (least square estimator) $SSE=\sum_{i=1}^4(Y_i-\hat Y)^2$ $(\hat Y=X\hat\beta)$

Confidence Ellipse 是考虑到了 correlation $(X'X)^{-1}$

Model valiation and diagnostics

1. Residuals: $\hat{\varepsilon} = Y - X\hat{\beta} = (I-H)Y = (I-H)\varepsilon$. for $H = X(X'X)^{-1}X'$, $\hat{\varepsilon}_i = \varepsilon_i - \sum_{j=1}^{n} h_{ij}\varepsilon_j$

Prop: ① $E(\hat{\varepsilon}) = 0$. ② $Var(\hat{\varepsilon}) = \sigma^2(I-H)$ ③ $Cov(\hat{\varepsilon}, Y) = \sigma^2(I-H)$ NOT independent ④ $Cov(\hat{\varepsilon}, \hat{Y}) = 0$. ⑤ $\bar{\hat{\varepsilon}} = \sum^{n} \frac{\hat{\varepsilon}_i}{n} = \frac{\hat{\varepsilon}'1}{n} = 0$. ⑥ $\hat{\varepsilon}'Y = SSE$. ⑦ $\hat{\varepsilon}'\hat{Y} = 0$. ⑧ $\hat{\varepsilon}'X = 0'$

①Residuals: 由data知道的,与model无关. $e_i = y_i - \hat{y}_i$ ② Standardized residual (rstandard) $z_i = \frac{e_i}{\hat{\sigma}}$ ← $\sqrt{MSE}$ (internally) $Var(\hat{\varepsilon}) = \sigma^2(I-H) = \sigma^2\begin{bmatrix}1-h_{11} & & \\ & \ddots & \\ & & 1-h_{kk}\end{bmatrix}$

③ studentized residual $r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$ ④ Deleted residual $\hat{\varepsilon}_{(i)} = e_{(i)} = y_i - X_i'\hat{\beta}_{(i)}$ ⑤ studentized deleted residuals (rstudent) (externally): $t_i = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}}$

2. model in centered form $y_i = \alpha + \beta_1(x_{i1} - \bar{x}_1) + \cdots + \beta_k(x_{ik} - \bar{x}_k) + \varepsilon_i$, $y = (1\ X_c)\binom{\alpha}{\beta_1} + \varepsilon$. $\beta_1 = (\beta_1, \ldots, \beta_k)'$, $X_c = (I - \frac{1}{n}J)X_1$ then

$\hat{\alpha} = \bar{y}$, $\hat{\beta}_1 = (X_c'X_c)^{-1}X_c y$. $\hat{y} = \bar{y}1 + X_c(X_c'X_c)^{-1}X_c'y = (\frac{1}{n}1'y)1 + H_c y = (\frac{1}{n}J + H_c)y$, as $\hat{y} = Hy$. so $H = \frac{1}{n}J + H_c = \frac{1}{n}J + X_c(X_c'X_c)^{-1}X_c'$

so $h_{ii} > \frac{1}{n}$. also. $H = H^2$. then $h_{ii} = h_{ii}^2 + \sum_{j\neq i} h_{ij}^2$ so $h_{ii} \le 1$. ⇒ 3. Prop: ① $\frac{1}{n} \le h_{ii} \le 1$ ② $-0.5 \le h_{ij} \le 0.5$ ③ $tr(H) = \sum_{i=1}^{n} h_{ii} = k+1$

4. outlier (残差很大的点) 看residual plot [±3]. studentized residual

异常点VS高杠杆点VS强影响点（Outlier，High Leverage Point，Influential Point）

一，定义

异常点（Outlier）：残差很大的点（即：因变量y的值是极端值的观测值）

高杠杆点（High Leverage Point）：远离样本空间中心的点（即：自变量x的值是极端值的观测值）

强影响点（Influential Point）：对模型有较大影响的点，如果删除该点能改变拟合回归方程

5. influential observations. ① Leverage $\boxed{h_{ii} > \frac{2(k+1)}{n}}$

② $Dffits_{(i)} = \frac{\hat{Y}_i - \hat{Y}_{(i)}}{\sqrt{\hat{\sigma}_{(i)} h_{ii}}} = e_i\sqrt{\frac{n-(k+1)-1}{SSE(1-h_{ii}) - e_i^2}}\left(\frac{h_{ii}}{1-h_{ii}}\right)$ $\boxed{> 2\sqrt{\frac{k+1}{n}}}$

③ Cook's distance $D_i = \frac{(\hat{Y}_{(i)} - \hat{Y}_i)'(\hat{Y}_{(i)} - \hat{Y}_i)}{(k+1)\ \hat{\sigma}^2} = \frac{r_i^2}{k+1} \times \frac{h_{ii}}{1-h_{ii}}$, $(r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}})$ $\left(\frac{e_i^2}{(k+1)MSE} \cdot \frac{h_{ii}}{(1-h_{ii})^2}\right)$ $P(F_{k+1,n-k-1} < D_i) \ge 0.5$. find $\boxed{F_{0.05,k+1,n-k-1} < D_i}$ then influential

Multicollinearity

如,总的F-test说sig.但单独每个T都不显著. $(VIF)_i = \frac{1}{1-R_i^2}$ ≥ 10 有共线性

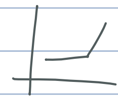
condition number $K = \sqrt{\frac{\lambda_1}{\lambda_p}}$ (min/max) condition indexes: $\sqrt{\frac{\lambda_1}{\lambda_i}}$ (all sorted) ≥ 30. 有共线性

Variable Selection

① find the highest adjust R^2. ② Stepwise regression see AIC: $n\log\frac{SSE_p}{n} + 2p$ BIC: $n\log\frac{SSE_p}{n} + p\log n$

Piecewise linear regression

分段线性回归. $E(y) = \beta_0 + \beta_1 X_1 + \beta_2(X_1 - 70)X_2$ ($X_2 = 1$ if $X_1 > 70$, 0 otherwise)



Ex. dummy variables d_1 and d_2. $d_1 = 1$ if $x > 320$. $d_2 = 1$ if $x > 500$, 0 otherwise. $E(y) = \beta_0 + \beta_1 x + \beta_2 d_1(x-320) + \beta_3 d_2(x-500)$

$H_0: \beta_2 = \beta_3 = 0$. H_1: not all $\beta_i = 0$, $i = 2, 3$. $F = \frac{SSR(d_1(x-320), d_2(x-500) \mid x)/2}{MSE_{full}}$ reject H_0 if $F > F_{0.05,2,n-4}$

看结论的时候: Estimate: $\hat{\beta}_1$, Std. Error: $S\{\hat{\beta}_1\}$, t-value: $\frac{\hat{\beta}_1}{S\{\hat{\beta}_1\}}$

Residual 那行后面写的 xx degrees of freedom 就是 n-r(x).

anova(fit)出来的表格. Mean Sq 是MSE, Sum Sq 是 extra sum of square、如 SSR(Ht | Age) 直接加和得SST. order matters.

Adj R (interpretation): after adjusting for sample size and the number of independent variables..

overall utility: ① 式子. ② H_0, H_1 ③ F-statistic / p-value ④ conclusion (与 $F_{Df, Res.Df}$ 比)

anova(reg2, reg1) 出来的F是 partial F-test, RSS 分别是 SSE_{red} 和 SSE_{full}, 如果F很小则说明 full model 更好.

⇒ regression model 如 $Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$ $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ VS least squares regression equation: $\hat{Y} = -4 + 3x_1$

MLE of σ: $\hat{\sigma}_M^2 = \frac{n-(k+1)}{n}\hat{\sigma}^2$ (有个原则是想证明什么就放在H_1里.如果reject null 则说明成立)

如要 test 单个也可以用 t-test ($\beta_i > 0$) $T = \frac{\bar{x} - u_0}{S/\sqrt{n}}$ $\left(\frac{estimates}{std.error}\right)$

Mis-Specified Model 如用 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 代替 $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$ 求 bias of $\hat{\beta}_0$ and $\hat{\beta}_1$

for this Ex. $\beta = \begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix}$ $\gamma = \begin{pmatrix}\beta_2 \\ \beta_3\end{pmatrix}$ $X = \begin{pmatrix}1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2\end{pmatrix}$ ← if x is this $Z = \begin{pmatrix}4 & 8 \\ 6 & 6 \\ 4 & 8\end{pmatrix}$

思路是: β^* 是 mis 的. $\hat{\beta}^* = (X'X)^{-1}X'Y$. $E(\hat{\beta}^*) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'[X\beta + Z\gamma] = \beta + \boxed{(X'X)^{-1}X'Z\gamma}$ 这个就是 bias

在诊断里.模型可以进行log. 用 Box-Cox 看一下 $y^* = \begin{cases}\frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0\end{cases}$