样本协方差: $Cov(x,y)=\sum_{i=1}^n(x_i-\bar{x})(y_i-\bar{y})/(n-1)$

大样本定理.: $Cov(x,y)=E((X-E(X))(Y-E(Y)))(n\to\infty)$

相关系数: $r=Cov(x,y)/\sqrt{(Var(x)Var(y))}={}^{\triangle}S_{xy}/\sqrt{(S_{xx}S_{yy})}$

In regression, one variable is considered independent variable(predictor); other dependent variable(outcome).

线性回归假设: I. The relationship between x and y is linear; II. y is distributed normally at each value x; III. The variance of y at every value of x is the same (homogeneity); IV. The observation is independent.

简单线性回归模型 假设: $E(\varepsilon_i)=0$; $Var(\varepsilon_i)=\sigma^2$; $Cov(\varepsilon_i,\varepsilon_j)=0$

最小二乘法: $\min\sum_{i=1}^n(y_i-y_i^{hat})^2$

$\beta_1^{hat}=\sum_{i=1}^n(x_i-\bar x)(y_i-\bar y)/\sum_{i=1}^n(x_i-\bar x)^2={}^{\triangle}S_{xy}/S_{xx}$

$\beta_0^{hat}=\bar y-\beta_1^{hat}\bar x$

$S^2=SSE/(n-2)=\sum_{i=1}^n(y_i-y^{hat})^2/(n-2)$ Unbiased for σ^2

极大似然法: $S_{MLE}^2=SSE/n=(n-2)S^2/n$

$\beta_1^{hat}\sim N(\beta_1,\sigma^2/S_{xx})$, $\beta_0^{hat}\sim N(\beta_0,\sigma^2(1/n+\bar x^2/\sum_{i=1}^n(x_i-\bar x)^2)$

$(n-2)S_{LSE}^2/\sigma^2\sim\chi^2(n-2)$.考虑到$\beta_0^{hat}$与 S^2 独立，所以

$(\beta_0^{hat}-\beta_0)/(S\sqrt{(1/n+\bar x^2/S_{xx})})\sim t(n-2)$

β_0 的$(1-\alpha)$CI: $[\beta_0^{hat}\pm S\,t(\alpha/2,n-2)/\sqrt{(1/n+\bar x^2/S_{xx})}]$

β_1 双边 t 检验: $H_0:\beta_1=0$ vs. $H_1:\beta_1\neq0$

$t=(\beta_1^{hat}-\beta_1)\sqrt{(S_{xx})}/S\sim t(n-2)$, reject H_0 if $|t|>t(\alpha/2,n-2)$

$(1-\alpha)$CI: $[\beta_1^{hat}-S\,t(\alpha/2,n-2)/\sqrt{S_{xx}},\ \beta_1^{hat}+S\,t(\alpha/2,n-2)/\sqrt{S_{xx}}]$

Est. of Mean of the Response Varibale for a Given Level of x:

$y_h^{hat}\sim N(\beta_0+\beta_1x_h,\sigma^2[1/n+(x_h-\bar x)^2/\sum_i(x_i-\bar x)^2])$ Est. of $E(y_h)$

CI of $E(y_h)$:$[y_h^{hat}\pm t(\alpha/2,n-2)S\sqrt{(1/n+(x_h-\bar x)^2/\sum_i(x_i-\bar x)^2)}]$

Prediction of a new observation

$y_h-y_h^{hat}\sim N(0,\sigma^2[1+1/n+(x_h-\bar x)^2/\sum_i(x_i-\bar x)^2])$

Predictive Itv.:$[y_h\pm t_{\alpha/2,n-2}S\sqrt{(1+1/n+(x_h-\bar x)^2/\sum_i(x_i-\bar x)^2)}]$

Analysis of Variance Approach to Regression Analysis

$SST=\sum_i(y_i-\bar y)^2=\sum_i(y_i-y_i^{hat})^2+\sum_i(y_i^{hat}-\bar y)^2=SSE+SSR$

SSE Sum of squares of Residual (Unexplained Variation)

SSR Sum of squares due to Regression (Explained Var.)

Analysis of Variation (ANOVA) Table

	SS	Df	MS	F
Regression	SSR	1	MSR=SSR	MSR/MSE
Error	SSE	n-2	MSE=SSE/(n-2)	
Total	SST	n-1		

F 检验: $H_0:\beta_1=0$(No linear relationship) vs. $H_1:\beta_1\neq0$

Reject H_0 if $F>F(\alpha,1,n-2)$

Coefficient of Determination (R-squared)

$R^2=SSR/SST=1-(SSE/SST)$

$r=\pm\sqrt{(R^2)}$ True for Simple Linear Regression only

满秩分解 Factorization: $A_{p*q}=K_{p*r}L_{r*q}$, r=r(A), K, L full rank

Let $B=A_{22}-A_{21}A_{11}^{-1}A_{12}$; $D=A_{11}-A_{12}A_{22}^{-1}A_{21}$

Remark 3.1 A^{-1}=

$A_{11}^{-1}+A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1}$	$-A_{11}^{-1}A_{12}B^{-1}$
$-B^{-1}A_{21}A_{11}^{-1}$	B^{-1}

=

D^{-1}	$-D^{-1}A_{12}A_{22}^{-1}$
$-A_{22}^{-1}A_{21}D^{-1}$	$A_{22}^{-1}+A_{22}^{-1}A_{21}D^{-1}A_{12}A_{22}^{-1}$

$(A+cc')^{-1}=A^{-1}-A^{-1}cc'A^{-1}/(1+c'A^{-1}c)$

幂等矩阵 Idempotent matrix $A^2=A$

对于幂等矩阵 A，有 r(A) = tr(A)

证明：考虑 A 的满秩分解 A=BC, $A^2=BCBC=BC$, $CB=I_r$

所以 tr(A)=tr(BC)=tr(CB)=r=rank(A).

幂等矩阵的特征值均为 0 或 1.

Theorem 秩 r 矩阵 A_{n*p}假如可划分为$[[A_{11}\ A_{12}][A_{21}\ A_{22}]]$，且 A_{11} 是 r 阶满秩矩阵，则 $A^-=[[A_{11}^{-1}\ 0][0\ 0]]$

类似地，若 $A_{12}/A_{21}/A_{22}$ 分别是 r 阶满秩矩阵，则 A^-分别可以是$[[0\ 0][A_{12}^{-1}\ 0]]/[[0\ A_{21}^{-1}][0\ 0]]/[[0\ 0][0\ A_{22}^{-1}]]$

广义逆 Generalized Inverse A^-: $AA^-A=A$; rank(A)=k, A_{m*n}

$rank(A^-)\ge rank(A^-A)=rank(AA^-)=k$; A^-是 A的广义逆

A^-A 和 AA^- 是幂等的; $A^-A=I\leftrightarrow rank(A)=n$

$AA^-=I\leftrightarrow rank(A)=m$; $tr(A^-A)=tr(AA^-)=k$

$K=A(A'A)^-A'$对任意$(A'A)^-$均不变

$K=K'$; $K=K^2$ (Symmetric, Idempotent)

rank(K) = rank(A) = tr(K); KA = A; A'K = A

$(X'X)^-X'$ is a g-inverse of X for any g-inverse of X'X

$X(X'X)^-$ is a g-inverse of X' for any g-inverse of X'X

Vector and Matrix Calculus:

x 是向量： $u=f(\mathbf{x})$

$f(\mathbf x)=\mathbf a'\mathbf x=\mathbf x'\mathbf a\to\partial u/\partial\mathbf x=\mathbf a$

$f(\mathbf x)=\mathbf x'A\mathbf x\to\partial u/\partial\mathbf x=2A\mathbf x$

X 是矩阵： $u=f(X_{p*p})$, $\partial u/\partial X=[\partial u/\partial X_{ij}]$

$f(X)=tr(XA)\to\partial u/\partial X=A+A'-diag(A)$

$f(X)=\ln|A|\to\partial u/\partial X=2X^{-1}-diag(X^{-1})$

x 是常数： $A_{n*n}=[a_{ij}]$, a_{ij} 是 x 的函数，

$f(x)=A^{-1}\to\partial u/\partial x=-A^{-1}\,\partial A/\partial x\,A^{-1}$

$f(x)=\ln|A|\to\partial u/\partial x=tr(A^{-1}\partial A/\partial x)$

Random Vector and Multivariate Normal Distribution

$Cov(\mathbf Y)=\Sigma=E((\mathbf Y-E(\mathbf Y))(\mathbf Y-E(\mathbf Y)')=[\sigma_{ij}]=[Cov(Y_i,Y_j)]\ge\mathbf 0$

$Cov(A\mathbf Y)=ACov(\mathbf Y)A'$ A 是常数矩阵

$Cov(A\mathbf X,B\mathbf Y)=ACov(\mathbf X,\mathbf Y)B'$

Mahalabolis Distance: $(\mathbf Y-\mu)'\Sigma^{-1}(\mathbf Y-\mu)$ (Standardized Dist.)

Generalized Variance = $|\Sigma|$

Remark 4.1 一系列观察 $\mathbf y_i=[y_{i1}\ldots y_{ip}]\sim$(i.i.d.)$(\mu,\Sigma)$

样本均值向量 $\mu^{hat}=\bar{\mathbf y}=[\bar y_1\ldots\bar y_p]$, $\bar y_j=\sum_{i=1}^n y_{ij}/n$

样本协方差阵 $\Sigma^{hat}_{p*p}=S=\sum_{i=1}^n(\mathbf y_i-\bar{\mathbf y})(\mathbf y_i-\bar{\mathbf y})'$

相关系数矩阵 $\Omega=[\rho_{ij}]=[\sigma_{ij}/\sqrt{(\sigma_{ii}\sigma_{jj})}]$

Remark 4.2 令 A 为对称矩阵, $E(\mathbf Y'A\mathbf Y)=tr(A\Sigma)+\mu'A\mu$

证明： $E(\mathbf Y'A\mathbf Y)=E((\mathbf Y-\mu)'A(\mathbf Y-\mu))+\mu'A\mu$

$E((\mathbf Y-\mu)'A(\mathbf Y-\mu))=E\,tr((\mathbf Y-\mu)'A(\mathbf Y-\mu))=E\,tr(A(\mathbf Y-\mu)(\mathbf Y-\mu)')$

$=tr(A\,E((\mathbf Y-\mu)(\mathbf Y-\mu)'))=tr(A\Sigma)$

Theorem: $Y_1,\ldots,Y_m$ 独立→$g_1(Y_1),\ldots,g_m(Y_m)$独立.

MGF: $M_Y(\mathbf t)=E(e^{\mathbf t'\mathbf Y})$

Let $\mathbf Y_{p*1}\sim N(\mu,\Sigma)$, $M_Y(\mathbf t)=e^{\mathbf t'\mu+\mathbf t'\Sigma\mathbf t/2}$

$f_Y(\mathbf y)=|\Sigma|^{-1/2}(2\pi)^{-p/2}\exp(-(\mathbf y-\mu)'\Sigma^{-1}(\mathbf y-\mu)/2)$

Remark 4.4 令 B 为常数矩阵，**C** 为常数向量，则

$\mathbf X=B\mathbf Y+\mathbf C\sim N(B\mu+\mathbf C,B\Sigma B')$

证明： $M_X(\mathbf t)=E(e^{\mathbf t'(B\mathbf Y+\mathbf C)})=E(e^{\mathbf t'B\mathbf Y})e^{\mathbf t'\mathbf C}$. 记 $\mathbf t^*=B'\mathbf t$，则

$M_X(\mathbf t)=E(e^{\mathbf t^{*\prime}\mathbf Y})e^{\mathbf t'\mathbf C}=M_Y(\mathbf t^*)e^{\mathbf t'\mathbf C}=\exp(\mathbf t^{*\prime}\mu+\mathbf t^{*\prime}\Sigma\mathbf t^*/2+\mathbf t'\mathbf C)$

$=\exp(\mathbf t'(B\mu+\mathbf C)+\mathbf t'B\Sigma B'\mathbf t/2)$. 即证.

令 $\mathbf Y=[[\mathbf Y_1]\ [\mathbf Y_2]]\sim N([[\mu_1][\mu_2]],[[\Sigma_{11}\ \Sigma_{12}]\ [\Sigma_{21}\ \Sigma_{22}]]$,

$\mathbf Y_1\sim N(\mu_1,\Sigma_{11})$

$\mathbf Y_1|(\mathbf Y_2=\mathbf y_2)\sim N(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf y_2-\mu_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

$\mathbf Y_1$ 和 $\mathbf Y_2$ 独立当且仅当$\Sigma_{12}=0$

偏相关 Partial Correlation:

$\mathbf v=[[\mathbf y][\mathbf x]]\sim N_q([[\mu_1][\mu_2]],[[\Sigma_{yy}\ \Sigma_{yx}]\ [\Sigma_{xy}\ \Sigma_{xx}]])$

$\Sigma_{y.x}=\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}=[\sigma_{ij.r\ldots q}]$, $D_{y.x}=[diag(\Sigma_{y.x})]^{1/2}$

$\rho_{ij.r\ldots q}=\sigma_{ij.r\ldots q}/\sqrt{(\sigma_{ii.r\ldots q}\sigma_{jj.r\ldots q})}$, $\Omega_{y.x}=D_{y.x}^{-1}\Sigma_{y.x}D_{y.x}^{-1}$

二次型 Quadratic Forms: 令 $\mathbf x\sim N(\mu,\Sigma)$，A 是对称矩阵，则

$M_{\mathbf x'A\mathbf x}(\mathbf t)=|I-2\mathbf tA\Sigma|^{-1/2}\exp(-\mu'(I-(I-2\mathbf tA\Sigma)^{-1})^{-1}\Sigma^{-1}\mu/2)$

$E(\mathbf x'A\mathbf x)=tr(A\Sigma)+\mu'A\mu$, $Var(\mathbf x'A\mathbf x)=2tr[(A\Sigma)^2]+4\mu'A\Sigma A\mu$

Non-central χ^2 distribution

$\mathbf x\sim N(\mathbf 0,I_n)$, $\mathbf x'\mathbf x\sim\chi^2(n)$; $\mathbf x\sim N(\mu,I_n)$, $u=\mathbf x'\mathbf x\sim\chi^2(n,\lambda)$, $\lambda=\mu'\mu/2$

$f(u)=e^{-\lambda}\sum_{k=0}^\infty\lambda^ku^{n/2+k-1}e^{-u/2}/(k!2^{n/2+k}\Gamma(n/2+k))$, $\mu>0$, $\lambda\ge0$

$M_u(t)=(1-2t)^{-n/2}\exp(-\lambda[1-(1-2t)^{-1}])$; $E(u)=n+2\lambda$,

$Var(u)=2n+8\lambda$; $u_i\sim\chi^2(n_i,\lambda_i)$ 独立, $\sum u_i\sim\chi^2(\sum n_i,\sum\lambda_i)$

Non-central F distribution

$u_1\sim\chi^2(p_1,\lambda)$, $u_2\sim\chi^2(p_2,0)$, $w=u_1p_2/u_2p_1\sim F(p_1,p_2,\lambda)$

$E(w)=(1+2\lambda/p_1)p_2/(p_2-2)$

Non-central t distribution

$z\sim N(\mu,1)$, $u\sim\chi^2(n)$, z is independent of u, then

$t=z/\sqrt{(u/n)}\sim$non-centered t distribution

Theorem 5.1 $\mathbf x_{p*1}\sim N(\mu,\Sigma)$, A 对称, r=rank(A), $\lambda=\mu'A\mu/2$ 则

$q=\mathbf x'A\mathbf x\sim\chi^2(r,\lambda)\leftrightarrow A\Sigma$幂等.

推论

$\mathbf x\sim N(\mathbf 0,I)$，则 $\mathbf x'A\mathbf x\sim\chi^2(r)\leftrightarrow A$ 幂等

$\mathbf x\sim N(\mathbf 0,\Sigma)$，则 $\mathbf x'A\mathbf x\sim\chi^2(r)\leftrightarrow A\Sigma$幂等

$\mathbf x\sim N(\mu,\sigma^2I)$, then $\mathbf x'\mathbf x/\sigma^2\sim\chi^2(n,\mu'\mu/(2\sigma^2))$

$\mathbf x\sim N(\mu,I)$，则 $\mathbf x'A\mathbf x\sim\chi^2(r,\mu'A\mu/2)\leftrightarrow A$ 幂等

Theorem 5.2

$\mathbf x\sim N(\mu,\Sigma)$, A 对称，则 $\mathbf x'A\mathbf x$ 与 $B\mathbf x$ 独立 $\leftrightarrow B\Sigma A=0$.

Theorem 5.3

$\mathbf x\sim N(\mu,\Sigma)$, A 和 B 对称, r=rank(A)，则

$\mathbf x'A\mathbf x$ 和 $\mathbf x'B\mathbf x$ 均幂等$\leftrightarrow A\Sigma B=B\Sigma A=0$.

Remark 4.3 $\mathbf Y_{p*1}\sim N(\mu,\Sigma)$，则 $M_Y(\mathbf t)=e^{\mathbf t'\mu+\mathbf t'\Sigma\mathbf t/2}$.

例 5.1 令 $\mathbf y=[Y_1,\ldots,Y_n]'\sim N(\alpha\mathbf 1,\sigma^2I)$， $U=\sum(Y_i-\bar y)^2/\sigma^2$,

$V=n(\bar y-\alpha)^2/\sigma^2$,求 U 和 V 的分布，并证明 U 和 V 独立.

证明： $\bar y=\mathbf 1'\mathbf y/n=A\mathbf y$, $(\mathbf y-\bar y\mathbf 1)=(I-\mathbf{11}'/n)\mathbf y=(I-J/n)\mathbf y=B\mathbf y$

B 是幂等的，因为$(I-J/n)^2=I-2J/n+J^2/n^2=I-J/n$,

rank(B)=tr(B)=n-1.$U=(\mathbf y-\bar y\mathbf 1)'(\mathbf y-\bar y\mathbf 1)/\sigma^2=\mathbf y'B'B\mathbf y/\sigma^2=\mathbf y'(B/\sigma^2)\mathbf y\sim\chi^2(n-1,\lambda)=\chi^2(n-1)$.因为$(\bar y-\alpha)/(\sigma/\sqrt n)\sim N(0,1)$，所以 $V\sim\chi^2(1)$.因为 $A\Sigma(B/\sigma^2)=\mathbf 1'/n\,\sigma^2I\,(I-\mathbf{11}'/n)/\sigma^2=\mathbf 0$，所以 U 和 $\bar y$ 独立，进而 U 和 V 独立，因为 V 是 $\bar y$ 的函数.

例 5.2 $\mathbf Y=[Y_1,\ldots,Y_n]'\sim N(\mu\mathbf 1,\sigma^2I)$, $Q_1=n(\bar Y)^2$, $Q_2=\sum(Y_i-\bar Y)^2$.

1.证明 $\bar Y$ 与 Q_2 独立; 2.证明 Q_1 与 Q_2 独立; 3.求 Q_1,Q_2 的分布.

解 $\bar Y=\mathbf 1'\mathbf y/n=A\mathbf y$, $Q_2=\sim\sigma^2\chi^2(n-1)$. $A\Sigma B'=\mathbf 0$， $\bar Y\perp Q_2$， $Q_1\perp Q_2$.

$Q_1=\sim\sigma^2\chi^2(1,n\mu^2/2\sigma^2)$.

例 5.3 $\mathbf y\sim N_3(\mu,\sigma^2I)$, $\mu'=[3\ -2\ 1]'$, A=[[2 -1 -1][-1 2 -1][-1 -1 2]]/3 B=[[1 1 1][1 0 -1]]/3. 求 $\mathbf y'A\mathbf y/\sigma^2$ 的分布；探究 $\mathbf y'A\mathbf y$ 与 $B\mathbf y$、$\mathbf y'A\mathbf y$ 与 $y_1+y_2+y_3$ 是否独立.

解：易得σ^2AI 幂等，所以 $\mathbf y'A\mathbf y/\sigma^2\sim\chi^2(n-1,38/3)$.

∵$B\sigma^2IA\neq O$，所以 $\mathbf y'A\mathbf y$ 与 $B\mathbf y$ 不独立.

$y_1+y_2+y_3=[1\ 1\ 1]\mathbf y=\mathbf C'\mathbf y$, $\mathbf C'\sigma^2IA=O$，所以 $\mathbf y'A\mathbf y$ 与 $y_1+y_2+y_3$ 独立.

例 5.4 $\mathbf y\sim N_n(\mu\mathbf 1,\Sigma)$, $\Sigma=\sigma^2[[1\ \rho\ \ldots][\rho\ 1\ \ldots]\ \ldots\ [\ldots\ \rho\ 1]]$, 求 $U=\sum_{i=1}^n(y_i-\bar y)^2/(\sigma^2(1-\rho))$的分布.

解： $\Sigma=\sigma^2((1-\rho)I+\rho J)$，记 $U=\mathbf y'A\mathbf y=\mathbf y'B\mathbf y/[\sigma^2(1-\rho)]$，则 $B=(I-J/n)$, $A=B/[\sigma^2(1-\rho)]$， 而 $A\Sigma=B$，B 幂等，所以 $A\Sigma$幂等.

rank(A)=rank(B)=tr(B)=n-1, 所以 $U\sim\chi^2(n-1,\lambda)=\chi^2(n-1)$

Remark 3.3 $X(X'X)^-X'$是不变的. X 是 m*n 的秩 k 矩阵.

Remark 4.5 $p(\mathbf Y_1|\mathbf Y_2=\mathbf y_2)=p(\mathbf y_1,\mathbf y_2)/p(\mathbf y_2)\propto p(\mathbf y_1,\mathbf y_2)$

$p(\mathbf y_1,\mathbf y_2)\propto\exp(-[[\mathbf y_1-\mu_1]\ [\mathbf y_2-\mu_2]]'\Sigma^{-1}[[\mathbf y_1-\mu_1]\ [\mathbf y_2-\mu_2]]/2)\propto\exp(((\mathbf y_1-\mu_1)'B^{-1}(\mathbf y_1-\mu_1)-2(\mathbf y_1-\mu_1)'B\Sigma_{12}\Sigma_{22}^{-1}(\mathbf y_2-\mu_2)+(\mathbf y_2-\mu_2)'A(\mathbf y_2-\mu_2))/2)$

$\propto\exp(-(\mathbf y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(\mathbf y_2-\mu_2))'B^{-1}(\mathbf y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(\mathbf y_2-\mu_2))/2)$

$\mathbf Y_1|\mathbf Y_2=\mathbf y_2\sim N(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf y_2-\mu_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

其中 $B=Cov(\mathbf Y_1|\mathbf Y_2)=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Remark 4.6 Correlation and Partial Correlation

y-血压; x_1-摄入热量; x_2-体重. $\rho_1=corr(y,x_1)$, $\rho_2=corr(y,x_2)$.

$\rho_3=corr(x_1,x_2)\neq0$. 如何将 x_2 加入线性模型 $y=a+bx_1+\varepsilon_1$?

假设 x_1 与 x_2 之间也有线性关系 $x_2=c+dx_1+\varepsilon_2$，

记残差 $r_y=y-(a^{hat}+b^{hat}x_1)$， $r_2=x_2-(c^{hat}+d^{hat}x_1)$，则

$\rho_(y,x_2).x_1=corr(r_y,r_2)$为给定 x_1 时， x_2 和 y 的偏相关系数.

偏相关为 0 时，意味着 x_2 中能解释 y 的信息均包含在 x_1 中，因此无需考虑将 x_2 加入模型；而偏相关不为 0 时，意味着 x_2 包含 x_1 以外的能够解释 y 的额外信息，此时应考虑将 x_2 加入模型.

Remark 5.1 二次型矩生成函数的表达式推导

$M_{\mathbf x'A\mathbf x}(t)=E(e^{t\mathbf x'A\mathbf x})=c_1\int\ldots\int\exp(t\mathbf x'A\mathbf x-(\mathbf x-\mu)'\Sigma^{-1}(\mathbf x-\mu)/2)d\mathbf x$

$=c_1\int\ldots\int\exp(*)d\mathbf x$. 其中， $c_1=(2\pi)^{-p/2}|\Sigma|^{-1/2}$.

记 $V^{-1}=(I-2tA\Sigma)\Sigma^{-1}$， $\theta'=\mu'(I-2tA\Sigma)^{-1}$，则

$(*)=-[\mathbf x(I-2tA\Sigma)^{-1}\Sigma^{-1}\mathbf x-2\mu'\Sigma^{-1}\mathbf x+\mu'\Sigma^{-1}\mu]/2$

$=-(\mathbf x-\theta)'V^{-1}(\mathbf x-\theta)/2-(\mu'\Sigma^{-1}\mu-\theta'V^{-1}\theta)/2$，于是

$M_{\mathbf x'A\mathbf x}(t)=(|V|/|\Sigma|)^{-1/2}\exp(-(\mu'\Sigma^{-1}\mu-\theta'V^{-1}\theta)/2)$，其中

$\theta'V^{-1}\theta=\mu'\Sigma^{-1}(I-2t\Sigma A)^{-1}\mu=\mu'\Sigma^{-1}(\Sigma(I-2tA\Sigma)^{-1}\Sigma^{-1})^{-1}\mu$

$=\mu'(I-2tA\Sigma)^{-1}\Sigma^{-1}\mu$. 所以

$M_{\mathbf x'A\mathbf x}(t)=|I-2tA\Sigma|^{-1/2}\exp(-\mu'(I-(I-2tA\Sigma)^{-1})^{-1}\Sigma^{-1}\mu/2)$.

Remark 5.2 用矩生成函数推导二次型的期望和方差

$E(\mathbf y^n)=d^nM_y(t)/dt^n|_{t=0}$， $Var(\mathbf y)=\partial\ln M_y(t)/\partial t^2|_{t=0}$.

期望的证明： 见 Remark 4.2

方差的证明：

令 $c(t)=I-2tA\Sigma$，则 $c(0)=I$. 令 $k(t)=\ln M_{\mathbf x'A\mathbf x}(t)$，则

$k(t)=-(\ln|c(t)|)/2-\mu'(I-c^{-1}(t))\Sigma^{-1}\mu$.

$\partial k(t)/\partial t=-tr(c^{-1}(t)\,\partial c(t)/\partial t)/2+\mu'\,\partial c^{-1}(t)/\partial t\,\Sigma^{-1}\mu/2$

$=tr(c^{-1}(t)A\Sigma)+\mu'c^{-1}(t)A\Sigma c^{-1}(t)\Sigma^{-1}\mu$

$\partial k^2(t)/\partial t^2=tr(\partial c^{-1}(t)/\partial t\,A\Sigma)+\mu'[\partial c^{-1}(t)/\partial t]A\Sigma[c^{-1}(t)]\Sigma^{-1}\mu$
$\quad+\mu'[c^{-1}(t)]A\Sigma[\partial c^{-1}(t)/\partial t]\Sigma^{-1}\mu$

$=2tr(c^{-1}(t)A\Sigma c^{-1}(t)A\Sigma)+4\mu'c^{-1}(t)A\Sigma c^{-1}(t)A\Sigma c^{-1}(t)\Sigma^{-1}\mu$

$Var(\mathbf x'A\mathbf x)=\partial^2k(t)/\partial t^2|_{t=0}=2tr((A\Sigma)^2)+4\mu'A\Sigma A\mu$

多元回归 Multiple Regression 有 p 个独立变量的线性回归模型：

$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_px_{ip}+\varepsilon_i$; $\mathbf y_{n*1}=X_{n*r}\beta_{r*1}+\varepsilon_{n*1}$. r = p+1.

假设： $E(\varepsilon)=\mathbf 0$; $Cov(\varepsilon)=\sigma^2I$; X 列满秩。

LSE $\beta^{hat}=(X'X)^{-1}X'\mathbf y$.

$\varepsilon^{hat}=\mathbf y-X\beta^\wedge=[I-X(X'V^{-1}X)^{-1}X'V^{-1}]\mathbf y=[I-H]\mathbf y\sim N(\mathbf 0,\sigma^2(IV-X(X'V^{-1}X)^{-1}X'))$

$\sigma^{2\,hat}=\varepsilon^{hat\prime}\varepsilon^{hat}/(n-r)=\varepsilon^{hat\prime}\varepsilon^{hat}/(n-(p+1))=SSE/(n-p-1)$.

$X'\varepsilon^\wedge=\mathbf 0$; $H=X(X'X)^{-1}X'$对称幂等; $\mathbf y^{\wedge\prime}\varepsilon^\wedge=0$;

I-H 对称幂等; $E(\beta^\wedge)=\beta$; $Cov(\beta^\wedge)=(X'X)^{-1}\sigma^2$;

$tr(I-H)=n-r$; $\varepsilon^{\wedge\prime}\varepsilon^\wedge=tr(\mathbf y'\mathbf y(I-H))$;

$E(\mathbf{yy}')=\sigma^2I+X\beta\beta'X'$; $E(\varepsilon^{\wedge\prime}\varepsilon^\wedge/(n-r))=\sigma^2$.

Generalized Least Squares (GLS) Estimation

假设： $Cov(\varepsilon)=V$ 已知。$\beta^\sim=(X'V^{-1}X)^{-1}X'V^{-1}\mathbf y$. 方差比$\beta^\wedge$小。

B.L.U.E. 最佳线性无偏估计

令$\lambda'\mathbf y$ 是 $\mathbf t'\beta$的观测值的线性函数和估计量，即 $E(\lambda'\mathbf y)=\mathbf t'\beta$.

而 $E(\lambda'\mathbf y)=\lambda'E(\mathbf y)=\lambda'X\beta$,所以$\lambda'X=\mathbf t'$. 而 B.L.U.E.是 $\mathbf t'\beta$的具有最小方差的线性无偏估计量， $Var(\lambda'\mathbf y)=\sigma^2\lambda'V\lambda$

Gauss-Markov 定理：在$\lambda'X=\mathbf t'$时， $W=\lambda'V\lambda$在 $\lambda'=\mathbf t'(X'V^{-1}X)^-X'V^{-1}$ 时取到最小值，即 GLS 就是 B.L.U.E.

MLE $\mathbf y=X\beta+\varepsilon$, $\varepsilon\sim N(0,\sigma^2I)$, σ^2 未知

$L(\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp(-(\mathbf y-X\beta)'(\mathbf y-X\beta)/(2\sigma^2))$

$\partial\ln L/\partial\beta=(X'\mathbf y-X'X\beta)/\sigma^2=0$

$\partial\ln L/\partial\sigma^2=-n/(2\sigma^2)+(\mathbf y-X\beta)'(\mathbf y-X\beta)/(2\sigma^4)=0$

$MLE(\beta)=(X'X)^{-1}X\mathbf y=\beta^\wedge$

$MLE(\sigma^2)=(\mathbf y-X\beta^\wedge)'(\mathbf y-X\beta^\wedge)/n=\mathbf y'(I-H)\mathbf y/n=(\mathbf y'\mathbf y-\beta^{\wedge\prime}X'\mathbf y)/n$

$=(n-2)\sigma^{2\wedge}/n$ $((n-p-1)\sigma^{2\wedge}/\sigma^2\sim\chi^2(n-p-1))$

$Cov(\beta^\wedge)=(X'X)^{-1}\sigma^2$. $(\beta^\wedge\sim N(\beta,\sigma^2(X'V^{-1}X)^{-1}))$

$SSE=\mathbf y'(I-H)\mathbf y=n*MLE(\sigma^2)=(n-r(X))LSE(\sigma^2)$

$\beta^\wedge$和 SSE 是独立的。记 $LSE(\sigma^2)=\sigma^{2\sim}$.

$(n-r(X))\sigma^{2\sim}/\sigma^2\sim\chi^2(n-r(X))$.

中心化形式 Centered Form

$y_i=\beta_0+\beta_1x_{i1}+\ldots+\beta_px_{ip}+\varepsilon_i=\alpha+b_1(x_{i1}-\bar x_1)+\ldots+b_p(x_{ip}-\bar x_p)+\varepsilon_i$.

$\mathbf y_{n*1}=[\mathbf 1\ Z][[\alpha]\ [\mathbf b]]+\varepsilon_{n*1}$, $Z=X_{1\sim p}-\mathbf 1\bar X'$, $\bar X'=[\bar x_1\ \ldots\ \bar x_p]$

$\alpha^\wedge=\beta_0^\wedge+\beta_1^\wedge\bar x_1+\ldots+\beta_p^\wedge\bar x_p$, $\mathbf b^\wedge=\beta_{1\sim p}^\wedge$

$Var(\beta^\wedge)=(X'X)^{-1}\sigma^2=[[\bar X'(Z'Z)^{-1}\bar X+1/n\quad -\bar X'(Z'Z)^{-1}]\ [-(Z'Z)^{-1}\bar X\quad (Z'Z)^{-1}]]\,\sigma^2$.

$Var(\mathbf b^\wedge)=(Z'Z)^{-1}\sigma^2$, $Cov(\beta_0^\wedge,\mathbf b^\wedge)=-\bar X'Var(\mathbf b^\wedge)$

$SST=\mathbf y'(I-\mathbf{11}'/n)\mathbf y$, $SST/\sigma^2\sim\chi^2(n-1,(\beta'X'X\beta-(\mathbf 1'X\beta)^2/n)/(2\sigma^2))$

$SSR=\mathbf y'(H-\mathbf{11}'/n)\mathbf y$, $SSR/\sigma^2\sim\chi^2(p,\mathbf b'(Z'Z)\mathbf b/(2\sigma^2))$

$R^2=SSR/SST$.

	SS	Df	MS
Reg	$\mathbf b^{\wedge\prime}Z'\mathbf y$	r(X)-1	MSR=SSR/(r(X)-1)
Err	$\mathbf y'\mathbf y-\beta^{\wedge\prime}X'\mathbf y$	n-r(X)	MSE=SSE/(n-r(X))
Tot	$\mathbf y'(I-\mathbf{11}'/n)\mathbf y$	n-1	

$F=MSR/MSE\sim F(r(X)-1,n-r(X),\mathbf b'(Z'Z)\mathbf b/(2\sigma^2))$

$\sim|_{H0:\,\mathbf b=\mathbf 0}\sim F(r(X)-1,n-r(X),0)$

模型误用 Model Misspecification

误差结构：$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ T: $Cov(\mathbf{y})=\sigma^2V$, F: $Cov(\mathbf{y})=\sigma^2 I$

$E(\boldsymbol{\beta}^{\hat{}*})=\boldsymbol{\beta}$ Unbiased. $Var(\boldsymbol{\beta}^{\hat{}*})=(X'X)^{-1}X'VX(X'X)^{-1}\geq\sigma^2X'V^{-1}X$

均值：

欠拟合 Under-fit: T: $\mathbf{y}=[X_1\ X_2][\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']'+\boldsymbol{\varepsilon}$, F: $\mathbf{y}=X_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}$

$\boldsymbol{\beta}_1^{\hat{}*}=(X_1'X_1)^{-1}X_1'\mathbf{y}$. $(\boldsymbol{\beta}_1\in R^{(p+1)}, \boldsymbol{\beta}_2\in R^{(k-p)})$

$E(\boldsymbol{\beta}_1^{\hat{}*})=\boldsymbol{\beta}_1+(X_1'X_1)^{-1}X_1'X_2\boldsymbol{\beta}_2$. $Cov(\boldsymbol{\beta}_1^{\hat{}*})=\sigma^2(X_1'X_1)^{-1}$.

$S_1^2=(\mathbf{y}-X_1\boldsymbol{\beta}_1^{\hat{}*})'(\mathbf{y}-X_1\boldsymbol{\beta}_1^{\hat{}*})/(n-p-1)$

$E(S_1^2)=\sigma^2+\boldsymbol{\beta}_2'X_2'(I-X_1(X_1'X_1)^{-1}X_1')X_2\boldsymbol{\beta}_2$.

$E(\mathbf{y}^{*\hat{}})=E(\mathbf{y}^*)+(\mathbf{x}_2^*-\mathbf{x}_1^*(X_1'X_1)^{-1}X_1'X_2)\boldsymbol{\beta}_2$. All biased.

过拟合 Over-fit: T: $\mathbf{y}=X_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}$, F: $\mathbf{y}=[X_1\ X_2][\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2']'+\boldsymbol{\varepsilon}$

$E(\boldsymbol{\beta}_1^{\hat{}})=\boldsymbol{\beta}_1$, $E(\boldsymbol{\beta}_2^{\hat{}})=\mathbf{0}$, Unbiased. $Var(\boldsymbol{\beta}^{\hat{}})=\sigma^2(X'X)^{-1}$

$Var(\boldsymbol{\beta}_1^{\hat{}})=\sigma^2(X_1'X_1)^{-1}(I+CBC'(X_1'X_1)^{-1})\geq Var(\boldsymbol{\beta}_1^{\sim})$.

假设检验 Hypothesis Testing

H_0: $K'\boldsymbol{\beta}=\boldsymbol{\mu}$. $K'\in R^{q*(p+1)}$，行满秩

$Q=(K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})'(K'(X'V^{-1}X)^{-1}K)^{-1}(K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})=SSH$

$F(H)=(Q/q)/(SSE/(n-r(X))=Q/(q\sigma^{2\sim})\sim F(q, n-r(X)$,

$(K\boldsymbol{\beta}-\boldsymbol{\mu})'(K'(X'X)^{-1}K)^{-1}(K'\boldsymbol{\beta}-\boldsymbol{\mu})/(2\sigma^2))\sim|_{H0}F(q,n-r(X))$

p-value$=P(F(q,n-r(X))>F(H))$

$K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu}\sim N(K'\boldsymbol{\beta}-\boldsymbol{\mu}, K'(X'X)^{-1}K\sigma^2)$.

$Q/\sigma^2\sim\chi^2(q, (K'\boldsymbol{\beta}-\boldsymbol{\mu})'(K'(X'V^{-1}X)^{-1}K)^{-1}(K'\boldsymbol{\beta}-\boldsymbol{\mu})/(2\sigma^2))$

考虑在 K'a=μ限制下的模型 $\mathbf{y}\sim N(X\mathbf{a}, \sigma^2 I)$

$\min(\mathbf{y}-X\mathbf{a})'(\mathbf{y}-X\mathbf{a})+2\boldsymbol{\lambda}'(K'\mathbf{a}-\boldsymbol{\mu})$

$\mathbf{a}^{\sim}=\boldsymbol{\beta}^{\hat{}}-(X'X)^{-1}K(K'(X'X)^{-1}K)^{-1}(K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})$.是模型的 b.l.u.e.

$SSE=(\mathbf{y}-X\boldsymbol{\beta}^{\hat{}})'V^{-1}(\mathbf{y}-X\boldsymbol{\beta}^{\hat{}})\sim N(0,$

$SSE_{H0}=(\mathbf{y}-X\mathbf{a}^{\sim})'(\mathbf{y}-X\mathbf{a}^{\sim})=SSE+(\boldsymbol{\beta}^{\hat{}}-\mathbf{a}^{\sim})'X'X(\boldsymbol{\beta}^{\hat{}}-\mathbf{a}^{\sim})$

$=SSE+Q\geq SSE$.

特殊情况 Special Cases

(1) H_0: $\boldsymbol{\beta}=\boldsymbol{\beta}_0$.

$Q=(\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\beta}_0)'X'X(\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\beta}_0)$, $F(H)=Q/((p+1)\sigma^{2\hat{}})$

$F(H)\sim F(p+1, n-p-1)$, $\mathbf{a}^{\sim}=\boldsymbol{\beta}_0$.

(2) H_0: $\boldsymbol{\lambda}'\boldsymbol{\beta}=m$.

$Q=(\boldsymbol{\lambda}'\boldsymbol{\beta}^{\hat{}}-m)^2/(\boldsymbol{\lambda}'(X'X)^{-1}\boldsymbol{\lambda})$, $F(H)=Q/(\sigma^{2\hat{}})\sim_{H0}F(1, n-r(X))$

$\sqrt{F(H)}\sim t(n-r(X))$

$\mathbf{a}^{\sim}=\boldsymbol{\beta}^{\hat{}}-(\boldsymbol{\lambda}'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})(X'X)^{-1}\boldsymbol{\lambda}/(\boldsymbol{\lambda}'(X'X)^{-1}\boldsymbol{\lambda})$

(3) H_0: $\boldsymbol{\beta}=[[\boldsymbol{\beta}_{1\ (p+1-h)*1}]\ [\boldsymbol{\beta}_{2\ h*1}=\mathbf{0}]]$

$K'_{h*(p+1)}=[0_{h*(p+1-h)}\mid I_h]\to K'\boldsymbol{\beta}=\mathbf{0}$.

$(K'(X'X)^{-1}K)^{-1}=X_2'X_2-X_2'X_1(X_1'X_1)^{-1}X_1'X_2=X_2'(I-H_1)X_2=B$

$\boldsymbol{\beta}_2^{\hat{}}=B^{-1}X_2'(I-H_1)\mathbf{y}$

$Q=(K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})'(K'(X'X)^{-1}K)^{-1}(K'\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\mu})=\boldsymbol{\beta}_2^{\hat{}}{}'X_2'(I-H_1)X_2\boldsymbol{\beta}_2^{\hat{}}$

$Q/\sigma^2\sim\chi^2(h, \boldsymbol{\beta}_2'X_2'(I-H_1)X_2\boldsymbol{\beta}_2/(2\sigma^2))$

$F=Q/(h\sigma^{2\hat{}})\sim|_{H0}F(h, n-p-1)$

Likelihood Ratio Test LRT 似然比检验

设 $\mathbf{y}\sim N_n(\mathbf{xa}, \sigma^2 I)$, $r(\mathbf{x})=p+1$,则假设 H_0: $\mathbf{a}=\mathbf{0}$ 的似然比检验

$F=\mathbf{a}^{\hat{}}{}'\mathbf{x}'\mathbf{y}/(p+1)/((\mathbf{y}'\mathbf{y}-\mathbf{a}^{\hat{}}{}'\mathbf{x}'\mathbf{y})/(n-p-1))$.

$F>F(\alpha, p+1, n-p-1)$时拒绝 H_0.

置信区间/区域

$\boldsymbol{\beta}$: $(\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\beta})'X'X(\boldsymbol{\beta}^{\hat{}}-\boldsymbol{\beta})\leq(p+1)\sigma^{2\hat{}}F(\alpha, p+1, n-p-1)$.

β_j: $\beta_j^{\hat{}}\pm t(\alpha/2, n-p-1)S\{\beta_j^{\hat{}}\}$ $(S\{\beta_j^{\hat{}}\}=\sqrt{(\sigma^{2\hat{}}(X'X)^{-1}{}_{jj})})$

$\boldsymbol{\lambda}'\boldsymbol{\beta}$: $\boldsymbol{\lambda}'\boldsymbol{\beta}^{\hat{}}\pm t(\alpha/2, n-p-1)\sigma^{\hat{}}\sqrt{(\boldsymbol{\lambda}'(X'X)^{-1}\boldsymbol{\lambda})}$

$E(\mathbf{y}^*)|\mathbf{x}=\mathbf{x}^*$: $\mathbf{x}^{*\prime}\boldsymbol{\beta}^{\hat{}}\pm t(\alpha/2, n-p-1)\sigma^{\hat{}}\sqrt{(\mathbf{x}^{*\prime}(X'X)^{-1}\mathbf{x})}$

$Var(\mathbf{y}^{*\hat{}})=(\mathbf{x}^{*\prime}(X'X)^{-1}\mathbf{x})\sigma^{2\hat{}}$

Prediction: $\mathbf{x}^{*\prime}\boldsymbol{\beta}^{\hat{}}\pm t(\alpha/2, n-p-1)\sigma^{\hat{}}\sqrt{(1+\mathbf{x}^{*\prime}(X'X)^{-1}\mathbf{x})}$

σ^2: $[(n-p-1)\sigma^{2\hat{}}/\chi^2(\alpha/2, n-p-1), (n-p-1)\sigma^{2\hat{}}/\chi^2(1-\alpha/2, n-p-1)]$

模型诊断

残差 $\boldsymbol{\varepsilon}^{\hat{}}=(I-H)\mathbf{y}=(I-H)\boldsymbol{\varepsilon}=(I-X(X'X)^{-1}X')\boldsymbol{\varepsilon}$.

记 $H=X(X'X)^{-1}X'=[h_{ij}]$，则 $\varepsilon_i^{\hat{}}=\varepsilon_i-\sum_{j=1}^n h_{ij}\varepsilon_j\sim N(0, \sigma^2(1-h_{ii}))$

残差的性质

$E(\boldsymbol{\varepsilon}^{\hat{}})=\mathbf{0}$; $Cov(\boldsymbol{\varepsilon}^{\hat{}})=\sigma^2(I-H)=Cov(\boldsymbol{\varepsilon}^{\hat{}}, \mathbf{y})$, $Cov(\boldsymbol{\varepsilon}^{\hat{}}, \mathbf{Y}^{\hat{}})=O$.

$\bar{\varepsilon}^{\hat{}}=\boldsymbol{\varepsilon}^{\hat{}}{}'\mathbf{1}/n=0$; $\boldsymbol{\varepsilon}^{\hat{}}{}'\mathbf{y}=SSE$; $\boldsymbol{\varepsilon}^{\hat{}}{}'\mathbf{Y}=0$; $\boldsymbol{\varepsilon}^{\hat{}}{}'X=0'$

$H=X(X'X)^{-1}X'$的性质

$1/n\leq h_{ii}\leq1$; $-1/2\leq h_{ij}\leq1/2$

$h_{ii}=1/n+(X_{1i}-\bar{X}_1)'(X_c'X_c)^{-1}(X_{1i}-\bar{X}_1)$.

$tr(H)=\sum h_{ii}=rank(H)=p+1$

中心化模型 $y=[\mathbf{1}\ X_C][[\alpha][\boldsymbol{\beta}_1]]+\varepsilon$，$\boldsymbol{\beta}_1=[\beta_1\ldots\beta_p]'$, $X_C=(I-J/n)X_1$

$\alpha^{\hat{}}=\bar{y}$，$\boldsymbol{\beta}_1=(X_C'X_C)^{-1}X_C'\mathbf{y}$, $\mathbf{y}^{\hat{}}=(J/n+H_C)\mathbf{y}=H\mathbf{y}$

残差分析 (记$\sigma^{\hat{}}=S=SSE/(n-p-1)$)

对于残差的方差并非定值的情况，有

Studentized Residual $r_i=\varepsilon_i^{\hat{}}/(\sigma^{\hat{}}\sqrt{(1-h_{ii})})$ (rstandard, $|r_i|>2$)

Deleted Residual $\varepsilon_{(i)}^{\hat{}}=y_i-y_i^{\hat{}}=\varepsilon_i^{\hat{}}/(1-h_{ii})$

$\boldsymbol{\beta}_{(i)}^{\hat{}}=\boldsymbol{\beta}^{\hat{}}-\varepsilon_i^{\hat{}}(X'X)^{-1}\mathbf{x}_i/(1-h_{ii})$, $SSE_{(i)}=SSE-\varepsilon_{(i)}^{\hat{}}$

$y_{j(i)}^{\hat{}}=\mathbf{x}_j'\boldsymbol{\beta}_{(i)}^{\hat{}}$(不使用第 i 个样本得到的$\boldsymbol{\beta}$的估计)

$S_{(i)}^2=SSE_{(i)}/(n-p-2)=\sum_{j\neq i}(y_j-y_{j(i)}^{\hat{}})^2/(n-1-(p+1))$

S- Deleted R- $t_i=\varepsilon_i^{\hat{}}/(S_{(i)}\sqrt{(1-h_{ii})})$ (rstudent)

$PreSS=\sum_{i=1}^n\varepsilon_{(i)}^{2\hat{}}=\sum_{i=1}^n(\varepsilon_i^{\hat{}}/(1-h_{ii}))^2$

强影响点 Influential Observations

Leverage h_{ii}: $h_{ii}\geq2(p+1)/n$ (hatvalues)

Cook's Distance: $D_i=(\boldsymbol{\beta}_{(i)}^{\hat{}}-\boldsymbol{\beta}^{\hat{}})'X'X(\boldsymbol{\beta}_{(i)}^{\hat{}}-\boldsymbol{\beta}^{\hat{}})/((p+1)\sigma^{2\hat{}})$

$D_i\geq F(0.5, p+1, n-p-1)$

DFFITS: $(y_i^{\hat{}}-y_{(i)}^{\hat{}})/(\sigma_{(i)}^{\hat{}}\sqrt{h_{ii}})$, $y_{(i)}^{\hat{}}=\mathbf{x}_i'\boldsymbol{\beta}_{(i)}^{\hat{}}\geq2\sqrt{((p+1)/n)}$

多重共线性 Multicollineartiy

检测：方差膨胀因子 Variance Inflation Factor VIF

$(VIF)_j=1/(1-R_j^2)$, $j=1, \ldots, p$. R_j是将第 j 个自变量当作因变量 $E(x_i)=\alpha_0+\alpha_1x_1+\ldots+\alpha_{j-1}x_{j-1}+\alpha_{j+1}x_{j+1}+\ldots+\alpha_px_p$ 的模型的 COD.

$(VIF)_j\geq10$,即 $R_j^2\geq0.9$ 时认为此变量重要。

令 X'X 的特征值为$\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_r$. $(r=p+1)$

Condition Number: $\kappa=\sqrt{(\lambda_1/\lambda_r)}$

Condition Indexes: $\sqrt{(\lambda_1/\lambda_j)}$, $j=2, 3, \ldots, r$

Multicollinear Problem: Condition Number≥30.

变量选择 Variable Selection

目标：寻找由独立变量构成的最优的集合关联的模型。

要求：移除无关变量后更容易解释，共线性概率降低。

方法：Stepwise Regression/Best-subset/LASSO...

Elastic Net: $\min(\mathbf{y}-X\boldsymbol{\beta})'(\mathbf{y}-X\boldsymbol{\beta})+\lambda_1\sum_{j=1}^p|\beta_j|+\lambda_2\sum_{j=1}^p\beta_j^2$

Qualitative Independeng Variables 离散自变量

引入 Dummy Variable

Interaction Regression Models 交互回归模型 $X_3=X_1X_2$

Polynomial- 多项式回归模型 $y_i=\beta_0+\beta_1x_i+\beta_2x_i^2+\varepsilon_i$, $x_i=X_i-\bar{X}$

Coef. of Partial Determination

$R^2_{Y_j.(Other\ var.)}=(SSE_{reduced}-SSE_{full})/SSE_{reduced}$

不满秩的/降秩模型 (X'X 不可逆，记 $G=(X'X)^-$)

OLSE: $\mathbf{b}^0=GX'\mathbf{y}$, 不唯一

$E(\mathbf{b}^0)=GX'X\mathbf{b}\triangle=A\mathbf{b}$($A\mathbf{b}$ 的无偏估计，不是 $\mathbf{b}$ 的)

$Var(\mathbf{b}^0)=GX'XG'\sigma^2$; $\mathbf{y}^{\hat{}}=XGX'\mathbf{y}$. (XGX'不变); $E(\mathbf{y}^{\hat{}})=X\mathbf{b}$.

$SSE=\mathbf{y}'(I-XGX')\mathbf{y}$. $SST=\mathbf{y}'(I-\mathbf{11}'/n)\mathbf{y}$

$E(SSE)=\sigma^2(n-r(X))$ $(\to\sigma^{2\hat{}}=SSE/(n-r(X))$无偏)

$SSR=SST-SSE=\mathbf{y}'(XGX'-J/n)\mathbf{y}$.

$\mathbf{y}\sim N(X\mathbf{b}, \sigma^2 I)$; $\mathbf{b}^0=GX'\mathbf{y}\sim N(GX'X\mathbf{b}, GX'XG'\sigma^2)$. $\mathbf{b}^0$ 独立于$\sigma^{2\hat{}}$.

$SSE/\sigma^2\sim\chi^2(n-r(X))$, $SSR/\sigma^2\sim\chi^2(r(X)-1, \mathbf{b}'X'(I-J/n)X\mathbf{b}/(2\sigma^2))$

SSE 与 SSR 独立

$F(R)=SSR/(r(X)-1)/(SSE/(n-r(X))\sim$

$F(r(X)-1, n-r(X), \mathbf{b}'X'(I-\mathbf{11}'/n)X\mathbf{b}/(2\sigma^2))$

Remedy over-parametrization $y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$

(1) $y_{ij}=\alpha_i+\varepsilon_{ij}$; (2) $y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$

(3) $\alpha_1=0$, α_i=Difference between i^{th} group and 1^{st} group.

可估函数 Estimable Functions

如果存在 $\mathbf{t}$ 使得 $E(\mathbf{t}'\mathbf{y})=\mathbf{q}'\mathbf{b}$，$\forall\mathbf{b}$，则 $\mathbf{q}'\mathbf{b}$ 被称作可估的. 其中 $\mathbf{y}=X\mathbf{b}+\boldsymbol{\varepsilon}$是降秩的。如果 $\mathbf{q}'\mathbf{b}$ 可估，则存在 $\mathbf{t}$ 使得 $\mathbf{t}'X=\mathbf{q}'$.

可估函数 $\mathbf{q}'\mathbf{b}$ 的 b.l.u.e.是 $\mathbf{q}'\mathbf{b}^0$.

$\mathbf{q}'\mathbf{b}^0=\mathbf{q}'GX'\mathbf{y}$; $\mathbf{q}'\mathbf{b}^0\sim N(\mathbf{q}'\mathbf{b}; \mathbf{q}'G\mathbf{q}\sigma^2)$(最小方差)

定理 9.1 当且仅当 $\mathbf{q}'A=\mathbf{q}'$时，$\mathbf{q}'\mathbf{b}$ 可估，其中 $A=GX'X$

可测假设 Testable Hypothesis H_0: $K'\mathbf{a}=\boldsymbol{\mu}$. $K\in R^{r*(p+1)}$,$K'\mathbf{a}$ 可估

$\mathbf{a}^0=(X'X)^-X'\mathbf{Y}$，则 $E(K'\mathbf{a}^0)=K'\mathbf{a}$, $Var(K'\mathbf{a}^0)=K'GK\sigma^2$

$\mathbf{y}\sim N(X\mathbf{a}, \sigma^2 I)$, $\mathbf{a}^0\sim N(GX'X\mathbf{a}, GX'XG'\sigma^2)$

$K'\mathbf{a}^0-\boldsymbol{\mu}\sim N(K'\mathbf{a}-\boldsymbol{\mu}, K'GK\sigma^2)$.

$Q=(K'\mathbf{a}^0-\boldsymbol{\mu})'(K'GK)^{-1}(K'\mathbf{a}^0-\boldsymbol{\mu})$,

$Q/\sigma^2\sim\chi^2(r, (K'\mathbf{a}-\boldsymbol{\mu})'(K'(X'X)^{-1}K)^{-1}(K'\mathbf{a}-\boldsymbol{\mu})/(2\sigma^2))$

$F(H)=Q/r/(SSE/(n-r(X))\sim F(r, n-r(X))$,

$(K'\mathbf{a}-\boldsymbol{\mu})'(K'GK)^{-1}(K'\mathbf{a}-\boldsymbol{\mu})/(2\sigma^2))\sim|_{H0}F(r, n-r(X))$

一因素布局（仅考虑一个变量）One-way Layout

$\mathbf{y}=X\mathbf{a}+\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}\sim N(\mathbf{0}, \sigma^2 I)$, $(y_{ij}=\mu+a_i+\varepsilon_{ij})$.

$X'\mathbf{y}=[\sum_{i,j}y_{ij}\ \sum_j y_{1j}\ \ldots\ \sum_j y_{mj}]$, $\mathbf{a}$ 不可估

$\mathbf{a}^0=GX'\mathbf{y}=[0\ \bar{y}_1\ \bar{y}_2\ \ldots\ \bar{y}_m]$

$\mathbf{a}^{0\prime}X'\mathbf{y}=\sum_{i=1}^m y_i^2/n_i$.

$SSR=\sum_{y=1}^m y_i^2/n_i-n\bar{y}^2$. $SST=\sum_{i=1}^m\sum_{j=1}^{n\_i}y_{ij}^2-n\bar{y}^2$.

$SSE=\sum_{i=1}^m\sum_{j=1}^{n\_i}y_{ij}^2-\sum_{y=1}^m y_i^2/n_i$

	SS	D.f.	MS	F
Reg.	Q	r	Q/r	Q/r/(SSE/(n-r(X))
Res.	SSE	n-r(X)	SSE/(n-r(X))	

$F\sim|_{H0}F(r, n-r(X))$ $(H_0: a_1=\ldots=a_m=0)$

-假设检验

H_0:$a_1=\ldots=a_m$.则 $K=[[0\ 1\ -1\ \ldots\ 0]\ [0\ 1\ 0\ -1\ \ldots0]\ \ldots\ [0\ 1\ \ldots\ 0\ -1]]$

$Q=(K'\mathbf{a}-\boldsymbol{\mu})'(K'(X'X)^-K)^{-1}(K'\mathbf{a}-\boldsymbol{\mu})$,

$F=Q/(m-1)/(SSE/(n-m))\sim F(m-1, n-m)$ $(r=m-1, r(X)=m)$

If A, B and A+PBQ are nonsingular, then

$(A+PBQ)^{-1}=A^{-1}-A^{-1}PB(B+BQA^{-1}PB)^{-1}BQA^{-1}$

例 $\mathbf{X}\sim N_3(\boldsymbol{\mu}, \Sigma)$，$\boldsymbol{\mu}'=[3\ -2\ 0]$，$\mathbf{X}'=[X_1\ X_2\ X_3]$，$\Sigma=[[5\ 0\ -3]\ [0\ 9\ 0]\ [-3\ 0\ 2]]$

a. X_2 和 $2X_1-X_3$ 是否独立？ b. 找到$(2X_1-5X_3, X_1+X_2)$的分布.

c. 给定 $X_1=1$，$X_2=-2$，求 X_3 的条件分布.

解：a.令 $B=[[0\ 1\ 0][2\ 0\ -1]]$，则 $B\mathbf{X}=[[X_2][2X_1-X_3]]$.

$B\mathbf{X}\sim N(B\boldsymbol{\mu}, B\Sigma B')=N([-2\ 6]', [[9\ 0][0\ 34]])$.

注意到 $Cov(X_2, 2X_1-X_3)=0$，所以它们独立.

b. 令 $C=[[2\ 0\ -5][1\ 1\ 0]]$，则 $B\mathbf{X}=[[2X_1-5X_3][X_1+X_2]]$.

$B\mathbf{X}\sim N([6\ 1]', [[130\ 25][25\ 14]])$.

c. $X_3|X_1=1, X_2=2\sim N(0+[-3\ -0][[5\ 0][0\ 9]]^{-1}([1\ 2]'-[3\ -2])'$, $[[2]]-[-3\ -0][[5\ 0][0\ 9]]^{-1}[-3\ -0]')$

$=N((6/5, 1/5)$.

例 Let y be the aggregate personal saving divided by disposable income, and x1, x2, x3, x4 be percentage of population under 15, disposable income per capita (in 100 US dollars), percentage of population over 75 and the percentage rate of change in per capita disposable income respectively. We first use a multiple regression model (errors assumed to be distributed as independent $N(0, \sigma^2)$) with covariates x3 and x4. Based on a dataset collected from 50 countries, we have the following R outcome.

```
Call: lm(formula = y ~ x3 + x4)
Coef.:        Est.    Std.    t      Pr(>|t|)
(Intercept) 5.4695 1.4101 3.879   0.000325***
x3          1.0726 ?        2.351   0.022992*
x4          0.4636 0.2052 2.259   0.028562*

Residual standard error: 4.122 on 47 degrees of freedom
Multiple R-Squared: 0.1883, Adjusted R-squared: 0.1538
F-statistic: ? on ? and ? DF, p-value ?

(X'X)^-1:     (Intercept)   x3            x4
(Intercept)   0.1170        -0.02758      -0.008997
x3            -0.02758      0.01226       -0.0001396
x4            -0.008997     -0.0001396    0.002479
```

(a) 求 x3 的系数估计，计算方差并构建 95% CI(use α=0.05, t(0.025, 47)=2.011741, t(0.05, 47)=1.677927, t(0.025, 49)=2.00975, t(0.05, 49)=1.676551).

Answer: $E(\beta_3^{\hat{}})=1.0726$; $\sigma^{\hat{}}=4.122$; $Var(\beta_3^{\hat{}})=4.122^2*0.01226$;

CI: $[\beta_3^{\hat{}}\pm t(\alpha/2, 50-2-1)\cdot\sqrt{Var(\beta_3^{\hat{}})}]=(0.1545, 1.9907)$

(b) Using the fact that SST = 983.63, SSE = 798.40, conduct a test of overall model utility (α=0.05, F(0.025, 2, 47)=3.994171, F(0.05,2,47)=3.195056, F(0.025, 2, 45)=4.008502, F(0.05, 2, 45)=3.204317).

Answer: SSR=SST-SSE=185.23,

MSR=SSR/(3-1)=92.615, MSE=SSE/(50-3)=16.987,

F=92.615/16.987, F=5.452>F(0.05, 2, 47)=3.195, Reject H0.

(c) Let y^* be the predicted value when x3=1.06, x4=3.08, find y^* and its variance.

Answer: $y^{\hat{}*}=8.0343$, $Var(y^{\hat{}*})=\mathbf{x}^*(X'X)^{-1}\mathbf{x}^*\sigma^{2\hat{}}=0.6711$.

(d) Determine the 95% confidence interval for the mean value of y when x3 =1.06, x4=3.08.

Answer: (6.3863, 9.6823).

(e) Write down the details of the test on whether we need to include covariates x1, x2 in the model. You may use the following results.

```
>g1=lm(y~x3+x4+x1+x2)
>anova(g2, g1)
   Res. Df. RSS       Df.    Sum of Sq. F Pr(>F)
1   47       798.40
2   45       650.71 ?(2)    ?(Q)        ? 0.01003 *
```

Answer: $Q=SSE_{H0}-SSE=RSS_{H0}-RSS=147.69$;

$F=Q/2/(SSE/(n-4-1))=5.1068$; p-value $=Pr(F(2, 45)\geq5.1068)=0.01003$

例 Consider the model $y_{ij}=\mu+\tau_i+\varepsilon_{ij}$, $i=1,2,3$, $j=1,2$.

(a) Write X, X'X, X'y, and the normal equations;

Answer:

$$X=\begin{bmatrix}1&1&&\\1&1&&\\1&&1&\\1&&1&\\1&&&1\\1&&&1\end{bmatrix}, X'X=\begin{bmatrix}6&2&2&2\\2&2&&\\2&&2&\\2&&&2\end{bmatrix}, K=\begin{bmatrix}0&1&-1&0\\0&0&1&-1\end{bmatrix}, a=\begin{bmatrix}\mu\\\tau_1\\\tau_2\\\tau_3\end{bmatrix}, \vec{\mu}=\begin{bmatrix}0\\0\end{bmatrix}$$

Normal equation: X'Xa=X'y.

(b) 适当重参数化使估计唯一，并求解

Answer: 令μ=0

(c) Show that H_0: $\tau_1=\tau_2=\tau_3$ are testable.

(d) Derive the detailed test procedure to test the above H_0.

例 考虑模型: $Y_i=\beta_0+\beta_1X_i+\varepsilon_i$, $i=1, 2, \ldots, n$.其中$\varepsilon_i\sim^{i.i.d.}N(0, \sigma^2)$. 令 $\bar{Y}$为 $Y_1, \ldots, Y_n$的均值. 设$\beta_0^{\hat{}}$和$\beta_1^{\hat{}}$是β_0和β_1 $(\neq0)$的最小二乘估计. $\sigma^{\hat{}2}$是σ^2的无偏估计.

(a)令$\delta=(\beta_0+\beta_1\bar{X})/\beta_1$, $\bar{X}$是 $X_1,\ldots,X_n$ 的均值,求 $\bar{Y}-\delta\beta_1^{\hat{}}$的分布.

Answer: $\beta_1^{\hat{}}\sim N(\beta_1, \sigma^2/S_{xx})$. $\bar{Y}\sim N(\beta_0+\beta_1\bar{X}, \sigma^2/n)$.

$\to\bar{Y}-\delta\beta_1^{\hat{}}\sim N(0, \sigma^2(1/n+\delta^2/S_{xx}))$.

(b) 求$(\bar{Y}-\delta\beta_1^{\hat{}})^2/\sigma^{\hat{}2}$的分布.

Answer: F(1, n-2).

(c) 若 H_0: $\beta_1=0$ 被一个双边α水平的 t 检验拒绝. 写出得到$-\beta_0/\beta_1$的 1-α置信区间过程.

Answer: From (b), we have

$Pr((\bar{Y}-\delta\beta_1^{\hat{}})^2/(\sigma^{\hat{}2}(1/n+\delta^2/S_{xx}))\leq F_{\alpha,1,n-2})=1-\alpha$

$\to Pr(*\leq0)=1-\alpha$. * = 0 has 2 solutions of δ: say g_1 and g_2.

Given that H_0: $\beta_1=0$ was rejected at level $\alpha\to\beta_1^{\hat{}2}-F(\alpha, 1, n-2)\sigma^{\hat{}2}/S_{xx}\geq0$

Thus, $Pr(g_1\leq\delta\leq g_2)=Pr(g_1\leq-(-\beta_0/\beta_1+\bar{X})\leq g_2)=1-\alpha$

100(1-α)% CI: $(\bar{X}-g_2, \bar{X}-g_1)$.

Simultaneous Intervals

Bonferroni CI: m intervals with $\alpha_c=\alpha_f/m$

For m linear functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_m'\boldsymbol{\beta}$, 100(1-α)% CI:

$\boldsymbol{\lambda}_i'\boldsymbol{\beta}^{\hat{}}\pm t(\alpha/2, m, n-p-1)\sigma^{\hat{}}\sqrt{(\boldsymbol{\lambda}_i'(X'X)^{-1}\boldsymbol{\lambda}_i)}$

Scheffe CI: For all possible linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$, 100(1-α)% CI:

$\boldsymbol{\lambda}'\boldsymbol{\beta}^{\hat{}}\pm\sigma^{\hat{}}\sqrt{((p+1)F(\alpha, p+1, n-p-1)\boldsymbol{\lambda}'(X'X)^{-1}\boldsymbol{\lambda})}$