第四章：假设检验与区间估计

4.1 F检验

考虑如下线性模型：

$$Y = X_{n\times p}\beta + e，\quad e \sim N(0,\sigma^2 I_n)，$$

设$rank(X)=r$，矩阵$H_{m\times p}$(已知)，线性假设

$$H_0 : H\beta = 0 \leftrightarrow H_1 : H\beta \neq 0，$$

不失一般性设$rank(H)=m$，现要检验假设H_0(作出拒绝或接受该假设的判断)。

考虑该假设的似然比(likelihood ratio)检验。设似然函数

$$L(Y;\beta,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\|Y-X\beta\|^2}{2\sigma^2}\right),$$

则似然比定义为

$$\lambda = \frac{\sup\limits_{\beta,\sigma^2} L(Y;\beta,\sigma^2)}{\sup\limits_{\substack{\beta,\sigma^2\\ H\beta=0}} L(Y;\beta,\sigma^2)}。$$

注意到当$\hat\beta=(X'X)^- X'Y$，$\tilde\sigma^2=\dfrac{\|Y-X\hat\beta\|^2}{n}$时似然比的分子达到最大，且$\sup\limits_{\beta,\sigma^2} L(Y;\beta,\sigma^2)=\left(\sqrt{\dfrac{2\pi e}{n}}\|Y-X\hat\beta\|\right)^{-n}$，当$\hat\beta_H$为约束$H\beta=0$下最小二乘估计，$\tilde\sigma_H^2=\dfrac{\|Y-X\hat\beta_H\|^2}{n}$时似然比的分母达到最大，且$\sup\limits_{\substack{\beta,\sigma^2\\ H\beta=0}} L(Y;\beta,\sigma^2)=\left(\sqrt{\dfrac{2\pi e}{n}}\|Y-X\hat\beta_H\|\right)^{-n}$。

因此似然比

$$\lambda = \left(\frac{\|Y-X\hat\beta_H\|^2}{\|Y-X\hat\beta\|^2}\right)^{\frac{n}{2}} = \left(1+\frac{k}{n-r}F\right)^{\frac{n}{2}},$$

这里

$$F=\frac{(ESS_H-ESS)/k}{ESS/(n-r)},$$

$$ESS=\|Y-X\hat\beta\|^2，\quad ESS_H=\|Y-X\hat\beta_H\|^2,$$

$$k = rank(X)+rank(H)-rank\begin{pmatrix}X\\H\end{pmatrix}。$$

ESS——error sum of squares(误差平方和)

由定义似然比$\lambda\geq 1$，直观上若H_0成立，则λ应充分接近1，故若λ偏离1很远则应该拒绝H_0，由于λ是F的单调函数，故当F很大时应该拒绝H_0。

定理 4.1.1：在本节线性模型假设下，

1. $\dfrac{ESS}{\sigma^2}\sim\chi^2_{n-r}$；
2. $\dfrac{ESS_H-ESS}{\sigma^2}\sim\chi^2_{k,\delta}$，这里 $\delta=\dfrac{\|X(\beta-E\hat\beta_H)\|^2}{\sigma^2}$；
3. ESS_H-ESS与ESS独立；
4. 在假设H_0下，$F\sim F_{k,n-r}$。

当$\mu(H')\subset\mu(X')$，即$H\beta$的每个分量都是可估的，此时$rank\begin{pmatrix}X\\H\end{pmatrix}=rank(X)$，因此$k=rank(H)=m$，$F=\dfrac{(ESS_H-ESS)/m}{ESS/(n-r)}$在假设$H_0$下服从$F_{m,n-r}$分布。由定理 3.3.2 的推论，此时

$$\hat\beta_H=\hat\beta-(X'X)^-H'\left[H(X'X)^-H'\right]^{-1}H\hat\beta。$$

$$ESS_H-ESS=\left\|X(\hat\beta_H-\hat\beta)\right\|^2=(H\hat\beta)'\left[H(X'X)^-H'\right]^{-1}(H\hat\beta)。$$

给定水平$\alpha\in(0,1)$，若$F>F_{m,n-r}(\alpha)$，则拒绝假设H_0：$H\beta=0$。此检验称为 F-检验。

关于F统计量的另一种形式。由于$ESS=Y'Y-\hat\beta'X'Y$，$ESS_H=Y'Y-\hat\beta_H'X'Y$，记$RSS=\hat\beta'X'Y$，$RSS_H=\hat\beta_H'X'Y$，则$F=\dfrac{(RSS-RSS_H)/m}{ESS/(n-r)}$。

RSS ——regression sum of squares(回归平方和)

TSS ——total sum of squares(总和)

令$TSS=Y'Y$，则有$TSS=RSS+ESS$。

下面从投影矩阵的角度来看F统计量。线性模型可表示为$Y=\theta+e,e\sim N(0,\sigma^2I_n)$，其中$\theta\in\mu(X)$。在假设$H_0$下(设$\mu(H')\subset\mu(X')$)，模型可表为$Y=\theta+e,e\sim N(0,\sigma^2I_n)$，其中$\theta\in S=\{X\beta|H\beta=0,\beta\in R^p\}$，注$S\subset\mu(X)$。

$\dim S=rank\begin{pmatrix}X\\H\end{pmatrix}-rank(H)=r-m$。记$P$，$P_S$分别为子空间$\mu(X)$，$S$上的投影矩阵，线性模型$\theta$的最小二乘估计为$\hat\theta=PY$，在假设下的最小二乘估计为$\hat\theta_H=P_SY$，由于$\theta\in\mu(X)$，$(I_n-P)\theta=0$，因此

$$SSE=\left\|Y-\hat\theta\right\|^2=Y'(I_n-P)Y=e'(I_n-P)e。$$

在假设H_0下的线性模型，$\theta\in S$，$(I_n-P_S)\theta=0$，因此$ESS_H=\left\|Y-\hat\theta_H\right\|^2=Y'(I_n-P_S)Y=e'(I_n-P_S)e$。由定理 3.2.1，$\dfrac{ESS}{\sigma^2}\sim\chi^2_{n-r}$，在假设$H_0$下$\dfrac{ESS_H}{\sigma^2}\sim\chi^2_{n+m-r}$，$ESS_H-ESS=e'(P-P_S)e$，由于$(P-P_S)(I-P)=0$，故$ESS_H-ESS$与$ESS$独立且$\dfrac{ESS_H-ESS}{\sigma^2}\sim\chi^2_m$。因此在$H_0$下，$F\sim F_{m,n-r}$。

4.2 有初始约束时的假设检验

设有初始约束的线性模型

$$\begin{cases}Y=X\beta+e,e\sim N(0,\sigma^2I_n)\\ L\beta=0,rank(L_{q\times p})=q\end{cases}$$

考虑假设H_0: $H\beta=0$的检验问题，这里$H\beta$为m个条件可估函数，即$\mu(H')\subset\mu(X'\vdots L')$。记$\hat\beta_L$，$\hat\beta_{LH}$分别为$\beta$在约束$L\beta=0$和$L\beta=0,H\beta=0$的最小二乘估计，由定理 3.3.2，$\hat\beta_L=G_{11}X'Y$(参见$G_{11}$的定义)。

$\hat{\beta}_{LH}$ 可以简单由$\begin{pmatrix} L \\ H \end{pmatrix}$代替$\hat{\beta}_L$中的$L$得到。

从而由残差平方和：

$$ESS_L = \left\| Y - X\hat{\beta}_L \right\|^2 = Y'Y - \hat{\beta}_L' X'Y,$$

$$ESS_{LH} = \left\| Y - X\hat{\beta}_{LH} \right\|^2 = Y'Y - \hat{\beta}_{LH}' X'Y。$$

此时投影子空间分别为：

$$S = \left\{ X\beta \middle| L\beta = 0, \beta \in R^p \right\},$$

$$S_1 = \left\{ X\beta \middle| L\beta = 0, H\beta = 0, \beta \in R^p \right\}。$$

$$\dim S = rank\begin{pmatrix} X \\ L \end{pmatrix} - rank(L) \equiv m_1,$$

$$\dim S_1 = rank\begin{pmatrix} X \\ L \\ H \end{pmatrix} - rank\begin{pmatrix} L \\ H \end{pmatrix} \equiv m_2。$$

此时假设H_0的似然比检验F-统计量为：

$$F = \frac{(ESS_{LH} - ESS_L)/(m_1 - m_2)}{ESS_L/(n - m_1)},$$

当 H_0成立时$F \sim F_{m_1 - m_2, n - m_1}$分布。

4.3 一般线性假设检验

至此都是讨论假设检验为$H\beta = 0$，对一般线性假设检验$H_{m\times p}\beta = d$，$rank(H) = m$且方程$H\beta = d$是相容的，先考虑线性假设$H\beta$每个分量都是可估的，即$\mu(H') \subset \mu(X')$。

设$rank(X) = r$，β_0为方程$H\beta = d$某一特解即 $H\beta_0 = d$，对线性模型$Y = X\beta + e$，$e \sim N(0, \sigma^2 I_n)$，令$Z = Y - X\beta_0$，$\theta = \beta - \beta_0$，得到线性模型：

$$Z = X\theta + e，\ e \sim N(0, \sigma^2 I_n),$$

此时，对该模型要检验的假设变为$H\theta = 0$。因此对于假设$H\beta = d$，$rank(H) = m$，$\mu(H') \subset \mu(X')$，检验的F-统计量为

$$F = \frac{\left(H\hat{\beta} - d\right)' \left[H(X'X)^- H'\right]^{-1} \left(H\hat{\beta} - d\right)/m}{\left\| Y - X\hat{\beta} \right\|^2 /(n - r)},$$

在假设$H\beta = d$的条件下$F \sim F_{m, n-r}$，给定水平$\alpha \in (0,1)$，若$F > F_{m, n-r}(\alpha)$，则拒绝该假设。

更一般的，若线性假设$H\beta$包含有不可估函数，即$\mu(H') \not\subset \mu(X')$，不失一般性，把$H$剖分成$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$，其中$H_1\beta$可估，$H_2\beta$不可估。考虑线性模型

$$Y = \theta + e，\ \theta \in \mu(X)，\ e \sim N(0, \sigma^2 I_n)$$

的如下两个假设检验问题：

$$H_0 \ ：\ H\beta = 0,$$

$$H_{01} \ ：\ H_1\beta = 0。$$

从正交投影来看，两个假设相当于

$$H_0：\ \theta \in S = \left\{ X\beta \middle| H\beta = 0, \beta \in R^p \right\},$$

$$H_{01} \ ：\ \theta \in S_1 = \left\{ X\beta \middle| H_1\beta = 0, \beta \in R^p \right\}。$$

显然有$S \subset S_1$，另一方面$H_1\beta$可估，$H_2\beta$不可估，因此

$$\dim S_1 = rank\begin{pmatrix} X \\ H_1 \end{pmatrix} - rank(H_1) = rank(X) - rank(H_1),$$

$$\dim S = rank\begin{pmatrix} X \\ H \end{pmatrix} - rank(H)$$
$$= rank\begin{pmatrix} X \\ H_2 \end{pmatrix} - rank\begin{pmatrix} H_1 \\ H_2 \end{pmatrix},$$
$$= rank(X) - rank(H_1)$$

故$S = S_1$。这表明假设$H_0: \theta \in S$与假设H_{01}: $\theta \in S_1$是一样的。因此对于不可估的假设是无法检验的，称为不可检验的假设(non-testable hypothesis)。

例 4.3.1：检验两样本是否同一线性模型

$y_i = \beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \cdots \beta_{p-1}^{(1)} x_{i,p-1} + e_i, i = 1,\cdots,n_1$,

$y_i = \beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \cdots \beta_{p-1}^{(2)} x_{i,p-1} + e_i, i = n_1+1,\cdots,n_1+n_2$,

其中误差$i.i.d \sim N(0,\sigma^2)$，要检验两组数据是否来于同一模型，即检验$\beta_i^{(1)} = \beta_i^{(2)}, 0 \le i \le p-1$。

首先写成矩阵形式，有两个线性模型$i = 1,2$

$Y_i = X_i\beta_i + e_i$，$e_i \sim N_{n_i}(0,\sigma^2 I_{n_i})$，$rank(X_i) = p$

要检验 H_0：$\beta_1 = \beta_2$。

将原问题写成一个线性模型：

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

要检验假设H_0：$\begin{pmatrix} I_p & -I_p \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0$。

在假设H_0下，$\beta_1 = \beta_2 = \beta$，模型为$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\beta + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$，最小二乘估计

$$\hat{\beta}_H = (X_1'X_1 + X_2'X_2)^{-1}(X_1'Y_1 + X_2'Y_2)。$$

记$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i, i = 1,2$，则检验的$F$统计量为$F = \dfrac{(ESS_H - ESS)/p}{ESS/(n_1+n_2-2p)}$，其中

$$ESS_H = Y_1'Y_1 + Y_2'Y_2 - \hat{\beta}_H'(X_1'Y_1 + X_2'Y_2),$$
$$ESS = Y_1'Y_1 + Y_2'Y_2 - (\hat{\beta}_1'X_1'Y_1 + \hat{\beta}_2'X_2'Y_2)。$$

在假设H_0下，$F \sim F_{p,n_1+n_2-2p}$分布，给定水平$\alpha \in (0,1)$，若$F > F_{p,n_1+n_2-2p}(\alpha)$，则拒绝该假设。

4.4 置信椭球(confidence ellipsoid)

设线性模型$Y = X_{n\times p}\beta + e$，$e \sim N(0,\sigma^2 I_n)$，$rank(X) = r$，$\Phi = H_{m\times p}\beta = \begin{pmatrix} h_1'\beta \\ \vdots \\ h_m'\beta \end{pmatrix}$为$m$个独立的可估函数，即$rank(H) = m$，$\mu(H') = \mu(X')$。

令$\hat{\beta} = (X'X)^- X'Y$，则$\hat{\Phi} = H\hat{\beta}$为$\Phi$的 BLUE，且$\hat{\Phi} \sim N_m(\Phi,\sigma^2 V)$，这里$V = H(X'X)^- H' > 0$。由推论 3.2.2

$$\frac{(\hat{\Phi} - \Phi)'V^{-1}(\hat{\Phi} - \Phi)}{\sigma^2} \sim \chi_m^2。$$

由定理 3.2.1，σ^2的估计$\hat{\sigma}^2 = \dfrac{\|Y - X\hat{\beta}\|^2}{n-r}$且与$\hat{\Phi}$独立，$\dfrac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2$，从而

$$\frac{(\hat{\Phi}-\Phi)'V^{-1}(\hat{\Phi}-\Phi)}{m\hat{\sigma}^2} \sim F_{m,n-r}。$$

因此对$\forall\alpha\in(0,1)$，

$$P\left(\frac{(\hat{\Phi}-\Phi)'V^{-1}(\hat{\Phi}-\Phi)}{m\hat{\sigma}^2}\le F_{m,n-r}(\alpha)\right)=1-\alpha。$$

令

$$D=\left\{\Phi\left|(\Phi-\hat{\Phi})'V^{-1}(\Phi-\hat{\Phi})\le m\hat{\sigma}^2F_{m,n-r}(\alpha)\right.\right\},$$

是以$\hat{\Phi}$为中心的一个椭球，$P(\Phi\in D)=1-\alpha$，D称为Φ的置信系数为$1-\alpha$的置信椭球。

用$H\beta$，$H\hat{\beta}$，$H(X'X)^{-}H'$代替$\Phi,\hat{\Phi},V$，则置信椭球可写为

$$\left(H\beta-H\hat{\beta}\right)'\left[H(X'X)^{-}H'\right]^{-1}\left(H\beta-H\hat{\beta}\right)\le m\hat{\sigma}^2F_{m,n-r}(\alpha)。$$

特别若$m=1$，由于$F_{1,n-r}$与${t^2}_{n-r}$分布一致，令$t_{n-r}(\alpha/2)$为上$\alpha/2$分位点，则此时可估函数$h'\beta$的$1-\alpha$置信区间为：

$$h'\hat{\beta}\pm t_{n-r}(\alpha/2)\hat{\sigma}\sqrt{h'(X'X)^{-}h},$$

或$h'\hat{\beta}\pm t_{n-r}(\alpha/2)\hat{\sigma}_{h'\hat{\beta}}$，其中$\hat{\sigma}^2_{h'\hat{\beta}}=\hat{\sigma}^2h'(X'X)^{-}h$为$Var(h'\beta)$的估计。

4.5 同时置信区间与 Bonferroni t-区间

往往要对若干个可估函数同时给出区间估计。设有m个可估函数$\phi_1=h_1'\beta,\cdots,\phi_m=h_m'\beta$，用上一节的方法可以对每个$\phi_i$作一个置信水平为$1-\alpha$的$t$-区间估计$\hat{\phi}_i\pm t_{n-r}(\alpha/2)\hat{\sigma}_{\hat{\phi}_i}$，$1\le i\le m$。由不等式

$$P\left(\bigcap_{i=1}^{m}A_i\right)=1-P\left(\bigcup_{i=1}^{m}\bar{A}_i\right)\ge 1-\sum_{i=1}^{m}P(\bar{A}_i),\quad (*)$$

若每个事件发生的概率$P(A_i)=1-\alpha$，则

所有事件同时发生的概率不是$1-\alpha$，而是只能保证$P\left(\bigcap_{i=1}^{m}A_i\right)\ge 1-m\alpha$。例如$\alpha=0.05$，$m=10$，则只能保证$\ge 0.5$。一般来说虽然每个置信区间置信系数是$1-\alpha$，同时置信区间的置信系数比$1-\alpha$要低。若要确保 m 个联合置信区间同时成立的概率达到名义上的$1-\alpha$，一个可供选择的办法是把每个置信区间的置信系数提高到$1-\frac{\alpha}{m}$。

上述作法当m不太大($m\le 5$)，效果还可以，但总的来说过于保守，特别当 m很大时，此时每个置信区间都太宽以致没有多大的实际意义。切合实际的折衷办法是增大α。

由于不等式(*)称为 Bonferroni 不等式，基于用$\frac{\alpha}{m}$替换 α的方法得到的同时置信区间 $h_i'\hat{\beta}\pm t_{n-r}(\alpha/2m)\hat{\sigma}_{h_i'\hat{\beta}},i=1,\cdots,m$ 称为 Bonferroni t-区间。

4.6 最大模 t-区间

对 m个独立可估函数$h_1'\beta,\cdots,h_m'\beta$作同时区间估计，令 $H_{m\times p}=\begin{pmatrix}h_1'\\ \vdots\\ h_m'\end{pmatrix}$，$rank(H)=m$，

$V=(v_{ij})_{m\times m}=H(X'X)^{-}H'$，则$H\hat{\beta}\sim N_m(H\beta,\sigma^2V)$，$H\hat{\beta}$与$\hat{\sigma}^2$独立且$\frac{(n-r)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-r}$分布。令

$$x_i=\frac{h_i'\hat{\beta}-h_i'\beta}{\sqrt{v_{ii}}},$$

$X=(x_1,\cdots,x_m)'$，则 $X\sim N_m(0,\sigma^2 R)$，其中 $R=(r_{ij})_{m\times m}$，$r_{ij}=\dfrac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}$。因此令$t=(t_1,\cdots,t_m)'$，$t_i=\dfrac{h_i'\hat\beta-h_i'\beta}{\hat\sigma\sqrt{v_{ii}}}$，则

$$t\sim t_m(0,R,n-r)(\text{多元 } t\text{-分布})。$$

令$t_m^{\alpha/2}(0,R,n-r)$使得

$$P\left(-t_m^{\alpha/2}(0,R,n-r)\le t_i\le t_m^{\alpha/2}(0,R,n-r),1\le i\le m\right)=1-\alpha$$

即 $P\left(\max\limits_{1\le i\le m}|t_i|\le t_m^{\alpha/2}(0,R,n-r)\right)=1-\alpha$。

这样 m个区间

$$h_i'\hat\beta\pm\hat\sigma\sqrt{v_{ii}}t_m^{\alpha/2}(0,R,n-r),i=1,\cdots,m$$

为置信系数$1-\alpha$的同时置信区间。由于$t_m^{\alpha/2}(0,R,n-r)$是由 m个 t分布变量取最大模分布确定的，故上同时置信区间称为最大模 t 区间(maximum modulus t-intervals)。

计算上区间的关键是求出$t_m^{\alpha/2}(0,R,n-r)$，一般是比较困难。Sidak(1968)证明了

$$t_m^{\alpha/2}(0,R,n-r)\le t_m^{\alpha/2}(0,I_m,n-r),$$

因此

$$P\left(h_i'\beta\in h_i'\hat\beta\pm\hat\sigma\sqrt{v_{ii}}t_m^{\alpha/2}(0,I_m,n-r),1\le i\le m\right)\ge 1-\alpha,$$

而$t_m^{\alpha/2}(0,I_m,n-r)$是易求出的。

4.7 Scheffe 区间和置信带

引理 4.7.1：设$A_{n\times n}>0$，则$\sup\limits_{b\ne 0}\dfrac{(a'b)^2}{b'Ab}=a'A^{-1}a$。

定理 4.7.1：设线性模型 $Y=X\beta+e$，$e\sim N(0,\sigma^2 I_n)$，$rank(H_{m\times p})=m$，$\mu(H')\subset\mu(X')$，则对任意可估$l'\beta$，$l\in\mu(H')$，其置信系数为$1-\alpha$的同时置信区间为

$$l'\hat\beta\pm\left[mF_{m,n-r}(\alpha)\right]^{1/2}\hat\sigma\left[l'(X'X)^{-}l\right]^{1/2}。$$

上述同时置信区间是 Scheffe(1953)提出来的，称为 Scheffe 区间。注意 Scheffe 区间不是有限多个可估函数的同时置信区间，它是所有$l'\beta$，$l\in\mu(H')$的同时置信区间。对有限可估函数同时置信区间，Scheffe 方法不一定最好，其长度可能会偏长，但 Scheffe 方法可以用于所有线性模型而无需对设计矩阵做任何限制。

当$m=r$即$\mu(H')=\mu(X')$时，可以得到所有可估函数$l'\beta$的同时$1-\alpha$置信区间。

若$rank(X_{n\times p})=p$，此时任何线性函数$l'\beta$都是可估的，此时

$$P\left(l'\beta\in l'\hat\beta\pm\left[pF_{p,n-p}(\alpha)\right]^{1/2}\hat\sigma\left[l'(X'X)^{-}l\right]^{1/2},\forall l\in R^p\right)=1-\alpha。$$

当 l变化时，区间

$$l'\hat\beta\pm\left[pF_{p,n-p}(\alpha)\right]^{1/2}\hat\sigma\left[l'(X'X)^{-}l\right]^{1/2}$$

也变化，形成一个区域，称为置信带(confidence band)，其宽度为$2\left[pF_{p,n-p}(\alpha)\right]^{1/2}\hat\sigma\left[l'(X'X)^{-}l\right]^{1/2}$。

例 4.7.1：设简单线性模型

$y_i=\beta_0+\beta_1x_i+e_i$，$i=1,\cdots,n$，$e_i\sim N(0,\sigma^2)$独立同分布。$\hat\beta=\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\end{pmatrix}$，其中

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

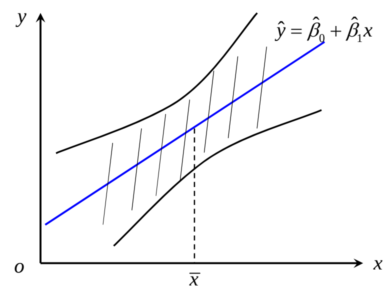
$\hat\beta_0=\bar y-\hat\beta_1\bar x$，$\bar x=\sum_{i=1}^n x_i/n$，$\bar y=\sum_{i=1}^n y_i/n$，

$\hat\sigma^2=\sum_{i=1}^n(y_i-\hat\beta_0-\hat\beta_1x_i)^2/(n-2)$。

对任线性函数$\beta_0+\beta_1x$，其$1-\alpha$置信带为

$(\hat\beta_0+\hat\beta_1x)\pm\left[2F_{2,n-2}(\alpha)\right]^{1/2}\hat\sigma\sqrt{\frac{1}{n}+\frac{(x-\bar x)^2}{\sum_{i=1}^n(x_i-\bar x)^2}}$。

置信带关于经验直线$\hat y=\hat\beta_0+\hat\beta_1x$对称，当$x=\bar x$时带宽最小。



4.8 预测问题

假定响应变量 y与协变量$x_1,\cdots,x_p$存在线性关系$y=\beta_1x_1+\cdots+\beta_px_p+e$，设有 n次观测，则未知参数$\beta_1,\cdots,\beta_p$可以由前面的结果来作出估计，设估计为$\hat\beta_1,\cdots,\hat\beta_p$，这样得到经验模型$\hat y=\hat\beta_1x_1+\cdots+\hat\beta_px_p$，该经验模型近似的描述了响应变量 y与协变量$x_1,\cdots,x_p$之间的关系，给定协变量的值，可以对响应变量作出预测。

点预测

考虑线性模型$y_i=x_i'\beta+e_i$，$i=1,\cdots,n$，写成矩阵形式：

$Y=X\beta+e$，$Ee=0$，$Cov(e)=\sigma^2\Sigma$，$rank(X_{n\times p})=r$，$\Sigma>0$(已知)。现有m个点$x_i=(x_{i1},\cdots,x_{ip})'$，$i=n+1,\cdots,n+m$，感兴趣的问题是由此预测响应变量$y$的$m$个值$y_{n+1},\cdots,y_{n+m}$。

令$X_0=\begin{pmatrix}x'_{n+1}\\ \vdots\\ x'_{n+m}\end{pmatrix}$，$Y_0=\begin{pmatrix}y_{n+1}\\ \vdots\\ y_{n+m}\end{pmatrix}$，$e_0=\begin{pmatrix}e_{n+1}\\ \vdots\\ e_{n+m}\end{pmatrix}$，则

$Y_0=X_0\beta+e_0$，$Ee_0=0$，$Cov(e_0)=\sigma^2\Sigma_0$。

假设$\mu(X_0')\subset\mu(X')$。

首先考虑被预测量 Y_0与历史数据 Y不相关，此时$Cov(e_0,e)=0$，为预测 Y_0，一个直观的方法就是用$EY_0=X_0\beta$的估计作为预测。即用

$$Y_0^* = X_0\beta^* = X_0(X'\Sigma^{-1}X)^- X'\Sigma^{-1}Y$$

来预测Y_0，这里β^*为广义最小二乘估计。由于$\mu(X_0')\subset\mu(X')$，所以Y_0^*与广义逆的选取无关。预测的偏差$Z = Y_0^* - Y_0$，则$EZ = 0$，即预测Y_0^*为Y_0的一个无偏估计。偏差的协方差阵$Cov(Z) = \sigma^2\left[\Sigma_0 + X_0(X'\Sigma^{-1}X)^- X_0'\right]$。

以上传统预测方法假定被预测量 Y_0与历史数据 Y不相关，但在一些情形 Y_0与 Y 相关。

设Y_0与 Y相关程度$Cov(e_0,e) = \sigma^2 V_{m\times n}$(V已知)，设$Y_0^* = C_{m\times n}Y$为 Y_0的一个线性预测，评价预测 Y_0^*好坏目前常用的度量是广义预测均方误差(generalized prediction mean squared error，简写 PMSE)。给定$A > 0$，$PMSE(Y_0^*) = E(Y_0^* - Y_0)'A(Y_0^* - Y_0)$。若线性预测是无偏的且广义预测均方误差最小，则称该线性预测是最优线性无偏预测(best linear unbiased predictor，简写 BLUP)。

定理 4.8.1：对于本节线性模型，若$Cov(e_0,e) = \sigma^2 V_{m\times n}$ (V已知)且$X_0\beta$可估，则

$$Y_0^* = X_0\beta^* + V\Sigma^{-1}(Y - X\beta^*)$$

为Y_0的最优线性无偏预测。特别若 $V = 0$，则Y_0的最优线性无偏预测为$Y_0^* = X_0\beta^*$。

区间预测

以下假设误差为正态分布，即$e \sim N_n(0,\sigma^2\Sigma)$，$e_0 \sim N_n(0,\sigma^2\Sigma_0)$，为简单起见，只考虑 Y与 Y_0不相关情形，即 $V = 0$。在正态误差假设下，偏差$Z = Y_0^* - Y_0$服从正态分布$N\left(0,\sigma^2\left[\Sigma_0 + X_0(X'\Sigma^{-1}X)^- X_0'\right]\right)$。设$\mu(X_0')\subset\mu(X')$，$rank(X_{n\times p}) = r$，

$\sigma^{*2} = \dfrac{(Y - X\beta^*)'\Sigma^{-1}(Y - X\beta^*)}{n - r}$，$\Sigma_0 = \left(\sigma_{ij}^{(0)}\right)_{n+1\le i,j\le n+m}$，

则类似 4.5 节，对每个$i = n+1,\cdots,n+m$，y_i的$1-\alpha$预测区间为

$$x_i'\beta^* \pm t_{n-r}(\alpha/2)\sigma^*\left[\sigma_{ii}^{(0)} + x_i'(X'\Sigma^{-1}X)^- x_i\right]^{1/2},$$

利用 Bonferroni 方法，$y_{n+1},\cdots,y_{n+m}$的一个置信系数至少$1-\alpha$同时预测区间为

$$x_i'\beta^* \pm t_{n-r}(\alpha/2m)\sigma^*\left[\sigma_{ii}^{(0)} + x_i'(X'\Sigma^{-1}X)^- x_i\right]^{1/2},$$

$n+1\le i\le n+m$。

也可以由 Scheffe 方法得到$y_{n+1},\cdots,y_{n+m}$的一个置信系数至少$1-\alpha$的同时预测区间：

$x_i'\beta^* \pm \left[mF_{m,n-r}(\alpha)\right]^{1/2}\sigma^*\left[\sigma_{ii}^{(0)} + x_i'(X'\Sigma^{-1}X)^- x_i\right]^{1/2}$，

$n+1\le i\le n+m$。

例 4.8.1：简单线性回归模型$y_i = \beta_0 + \beta_1 x_i + e_i$，$e_i \sim N(0,\sigma^2)$独立同分布。现感兴趣的是同时预测$y_{n+1},\cdots,y_{n+m}$(设相互独立)。对每个$i = n+1,\cdots,n+m$，$y_i$的点预测为：$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$。

$y_{n+1},\cdots,y_{n+m}$的一个置信系数至少$1-\alpha$同时预测 Bonferroni 区间为：

$$(\hat{\beta}_0+\hat{\beta}_1 x_i) \pm t_{n-r}\left(\frac{\alpha}{2m}\right)\hat{\sigma}\left[1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_{k=1}^{n}(x_k-\bar{x})^2}\right]^{1/2}$$

，$n+1 \le i \le n+m$。

$y_{n+1},\cdots,y_{n+m}$的一个置信系数至少$1-\alpha$同时预测 Scheffe 区间为：

$$(\hat{\beta}_0+\hat{\beta}_1 x_i) \pm \left[mF_{m,n-r}(\alpha)\right]^{1/2}\hat{\sigma}\left[1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_{k=1}^{n}(x_k-\bar{x})^2}\right]^{1/2}$$

，$n+1 \le i \le n+m$。