第十三讲 约束最小二乘和$\boldsymbol{F}$检验

2022.11.25

给定$\mathbf{x}_1$检验$\mathbf{x}_2$的效应是否显著

$$F=\frac{n-p}{q}\times\frac{R_{2\bullet1}^2}{1-R_{2\bullet1}^2},$$

$R_{2\bullet1}^2=$控制$\mathbf{x}_1$条件下y与$\mathbf{x}_2$的部分决定系数

约束最小二乘

在有些问题，主要是假设检验问题中，回归系数满足某些线性约束,比如 $\beta_1=0,\beta_1=\beta_2,\beta_2=(\beta_1+\beta_3)/2$等，皆可表示为$A\boldsymbol{\beta}=\mathbf{c}_0$,$\mathbf{c}_0$已知,的形式。

约束最小二乘问题

假设线性模型$\mathbf{y}=X_{n\times p}\boldsymbol{\beta}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$，假设$X$列满秩，设$A_{q\times p}$为行满秩常数矩阵，$\mathbf{c}_0$ 为已知的$q\times 1$常数向量，(线性)约束最小二乘问题

$$\min_{\boldsymbol{\beta}\in R^p}\|\mathbf{y}-X\boldsymbol{\beta}\|^2 \quad s.t.\quad A\boldsymbol{\beta}=\mathbf{c}_0 \qquad (*)$$

命题1. 上述约束最小二乘问题的最优解(约束最小二乘估计)为

$$\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)$$

其中$\hat{\boldsymbol{\beta}}=(X^\top X)^{-1}X^\top\mathbf{y}$为无约束条件下的LS估计。

证明：对于约束$A\boldsymbol{\beta}=\mathbf{c}_0$，应用拉格朗日乘子法，令

$$f(\boldsymbol{\beta},\boldsymbol{\lambda})=\|\mathbf{y}-X\boldsymbol{\beta}\|^2+2\boldsymbol{\lambda}^{\mathsf{T}}(A\boldsymbol{\beta}-\mathbf{c}_0)$$

对$\boldsymbol{\beta}$求偏导并令之为0得

$$X^{\mathsf{T}}X\boldsymbol{\beta}-X^{\mathsf{T}}\mathbf{y}+A^{\mathsf{T}}\boldsymbol{\lambda}=0$$

$$\Rightarrow\tilde{\boldsymbol{\beta}}=(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\mathbf{y}-(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\boldsymbol{\lambda}=\hat{\boldsymbol{\beta}}-(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\boldsymbol{\lambda}$$

代入$A\tilde{\boldsymbol{\beta}}=\mathbf{c}_0$得$A\hat{\boldsymbol{\beta}}-A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\boldsymbol{\lambda}=\mathbf{bc}_0$，即$A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\boldsymbol{\lambda}=A\hat{\boldsymbol{\beta}}-\mathbf{c}_0$，

所以 $\boldsymbol{\lambda}=\left(A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\right)^{-1}\left(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0\right)$,

$$\Rightarrow\quad\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\left(A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\right)^{-1}\left(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0\right).\qquad\square$$

引理1. 约束$A_{q\times p}\boldsymbol{\beta}=\mathbf{0}$成立时, X张成的空间(下面称自变量空间)
定义为$V_0=\left\{X\boldsymbol{\beta}: A\boldsymbol{\beta}=0,\ \boldsymbol{\beta}\in R^p\right\}\subset V=C(X)$,则

$$V_0=C(X(X^\top X)^{-1}A^\top)^\perp\bigcap C(X)$$

进而V_0空间的投影阵为$P_{V_0}=P_X-P_{X(X^\top X)^{-1}A^\top}$.

证明：记$\boldsymbol{\mu}=X\boldsymbol{\beta}\Rightarrow\ \boldsymbol{\beta}=(X^\top X)^{-1}X^\top\boldsymbol{\mu}$，

若$\boldsymbol{\mu}\in V_0$, $\mathbf{0}=A\boldsymbol{\beta}=A(X^\top X)^{-1}X^\top\boldsymbol{\mu}\ \Rightarrow\ \boldsymbol{\mu}\in C(X(X^\top X)^{-1}A^\top)^\perp$，但$\boldsymbol{\mu}\in C(X)$,

所以$\boldsymbol{\mu}\in C(X)\bigcap C(X(X^\top X)^{-1}A^\top)^\perp$，所以$V_0\subset C(X)\bigcap C(X(X^\top X)^{-1}A^\top)^\perp$。

反之也对，所以$V_0=C(X)\bigcap C(X(X^\top X)^{-1}A^\top)^\perp$，

即$C(X(X^\top X)^{-1}A^\top)\oplus V_0=C(X)$,

所以$P_X=P_{X(X^\top X)^{-1}A^\top}+P_{V_0}$.

不妨假设$\mathbf{c}_0=0$, 这是因为 $A\boldsymbol{\beta}=\mathbf{c}_0 \Leftrightarrow$

$A\left[\boldsymbol{\beta}-A^\top(AA^\top)^{-1}\mathbf{c}_0\right]=\mathbf{0}$, 令$\boldsymbol{\beta}^*=\boldsymbol{\beta}-A^\top(AA^\top)^{-1}\mathbf{c}_0$，此时模型

$$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=X\left(\boldsymbol{\beta}^*+A^\top(AA^\top)^{-1}\mathbf{c}_0\right)+\boldsymbol{\varepsilon}=X\boldsymbol{\beta}^*+XA^\top(AA^\top)^{-1}\mathbf{c}_0+\boldsymbol{\varepsilon}$$

即 $\mathbf{y}^*\triangleq\mathbf{y}-XA^\top(AA^\top)^{-1}\mathbf{c}_0=X\boldsymbol{\beta}^*+\boldsymbol{\varepsilon}$.

在原模型中约束$A\boldsymbol{\beta}=\mathbf{c}_0 \Leftrightarrow$ 在模型$\mathbf{y}^*=X\boldsymbol{\beta}^*+\boldsymbol{\varepsilon}$约束$A\boldsymbol{\beta}^*=\mathbf{0}$。

命题1的几何证明：

$\mathbf{y}$在V_0上的投影$\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}$使得约束最小二乘(*)达到最小，由引理1,

$$\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}=P_V\mathbf{y}-P_{X(X^\top X)^{-1}A^\top}\mathbf{y}=X\hat{\boldsymbol{\beta}}-X(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}$$

$$=X\left[\hat{\boldsymbol{\beta}}-(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}\right]$$

$$\Rightarrow\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}。$$

注：在具体的约束LS问题中，我们通常将线性约束代入模型得到简化的无约束模型，再应用LS。例如，假设模型

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_{p-1}\beta_{p-1} + \boldsymbol{\varepsilon}$$

有约束$\beta_1 = \cdots = \beta_{p-1}$, 代入模型得$\mathbf{y} = \mathbf{1}\beta_0 + (\mathbf{x}_1 + \cdots \mathbf{x}_{p-1})\beta_1 + \boldsymbol{\varepsilon}$, 这是简单线性模型，易得$\beta_1$的LS估计。

例1(中文课本例3.3.1,p43),测量三角形的三个内角$\theta_1,\theta_2,\theta_3$得测量值$y_1,y_2,y_3$，试求$\theta_1,\theta_2,\theta_3$的最小二乘估计。

直观解法：在$\theta_1+\theta_2+\theta_3=\pi$约束下,$y_2,y_3$对于估计$\theta_1$也有帮助，我们可利用三次测量之和与$\pi$的差估算每一次测量的系统误差:$\hat{\delta}=(y_1+y_2+y_3-\pi)/3$ $=\bar{y}-\pi/3$，则θ_i的估计应为$\hat{\theta}_i = y_i - \hat{\delta} = y_i - (y_1+y_2+y_3-\pi)/3.$

我们在上述直观描述中实际上假设了线性模型

$$y_i = \delta + \theta_i + \varepsilon_i，\ \varepsilon_i \sim (0,\sigma^2), i=1,2,3，\ \theta_1+\theta_2+\theta_3=\pi$$

其中δ代表系统误差，将约束条件$\theta_3=\pi-\theta_1-\theta_2$代入模型：

$$\begin{cases} y_1 = \delta + \theta_1 + \varepsilon_1 \\ y_2 = \delta + \theta_2 + \varepsilon_2 \\ y_3 = \delta + (\pi - \theta_1 - \theta_2) + \varepsilon_3 \end{cases} \Rightarrow \hat{\delta} = \bar{y} - \pi/3，\ \hat{\theta}_i = y_i - \hat{\delta}$$

一般线性假设

科学是一种系统性的知识体系，它积累并组织可检验的有关于万物的解释或理论。科学强调理论的可证伪性。科学不是寻求绝对的真理，而是在现有基础上，摸索式地不断接近真理。假设检验是对科学理论假设进行判断的工具。

一般线性假设

假设正态线性回归模型：$\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}_{n\times 1}$， $\boldsymbol{\varepsilon}\sim N_n(0,\sigma^2 I_n)$

$H_0: A\boldsymbol{\beta}=\mathbf{c}_0$ 称为一般线性假设，其中$A,\mathbf{c}_0$已知，A行满秩。

有关$\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_{p-1})^{\mathsf{T}}$的检验问题一般都可写成上述线性假设的形式，如$p=4$时

$$H_0:\beta_1=0 \Leftrightarrow H_0:\mathbf{a}^{\mathsf{T}}\boldsymbol{\beta}=0, \mathbf{a}=(0,1,0,0)^{\mathsf{T}}$$

$$H_0:\beta_1=\beta_2=0 \Leftrightarrow H_0:A\boldsymbol{\beta}=0, A=\begin{pmatrix}0&1&0&0\\0&0&1&0\end{pmatrix}$$

但A的选取方式不唯一。上述第二个例子中也可取$A=\begin{pmatrix}0&1&0&0\\0&-1&1&0\end{pmatrix}$

一般的Wald检验

Wald检验是已知构造直观的方法。非正态总体假设下，
其原假设下的渐近分布为卡方分布（我们不证明该结论，
但当总体为正态分布时，我们可以求出精确分布）。

单参数
情形

$H_0:\theta=\theta_0(\theta_0$已知$)$, $\hat{\theta}$为θ的极大似然估计或其它 估计，Wald检验统计量

$$Z=(\hat{\theta}-\theta_0)/\sqrt{\widehat{\mathrm{var}}(\hat{\theta})}\text{或}W=Z^2=(\hat{\theta}-\theta_0)^2/\widehat{\mathrm{var}}(\hat{\theta})$$

H_0成立时，近似地$(n$足够大时$)$，$Z\sim N(0,1)$或$W=Z^2\sim\chi_1^2$。

多参数
情形

$H_0:\varphi(\boldsymbol{\theta})=\mathbf{c}_0$ ($\mathbf{c}_0$: 已知的$q\times 1$向量), Wald统计量

$$W=\left(\varphi(\hat{\boldsymbol{\theta}})-\mathbf{c}_0\right)^{\mathrm{T}}\hat{\Sigma}^{-1}\left(\varphi(\hat{\boldsymbol{\theta}})-\mathbf{c}_0\right),$$

其中$\hat{\boldsymbol{\theta}}$为$\boldsymbol{\theta}$的极大似然估计或其它估计，$\Sigma=\mathrm{var}(\varphi(\hat{\boldsymbol{\theta}}))$,

$\hat{\Sigma}=\widehat{\mathrm{var}}(\varphi(\hat{\boldsymbol{\theta}}))$为$\Sigma$的估计。$H_0$成立时，近似地$(n$足够大时$)$,

$$W\sim\chi_q^2, n\to\infty$$

$$\hat{\Sigma}^{-1/2}(\varphi(\hat{\boldsymbol{\theta}})-\mathbf{c}_0)\overset{\text{近似}}{\sim}N(\mathbf{0},I_q)$$

线性模型的Wald检验和F检验

Wald检验

模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$，$\boldsymbol{\varepsilon}\sim(0,\ \sigma^2 I_n)$，$H_0:\ A_{q\times p}\boldsymbol{\beta}=\mathbf{c}_0$（已知)，$A$行满秩常数矩阵。

按照一般的Wald检验构建方式，H_0的Wald检验定义为：

$$W=(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)^{\mathsf{T}}\left(A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)/\hat{\sigma}^2 \overset{H_0,\text{近似}}{\sim} \chi_q^2$$

注意其中$A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}=\mathrm{var}(A\hat{\boldsymbol{\beta}}\mid X)$.

如果假设误差服从正态分布：$\boldsymbol{\varepsilon}\sim N(0,\ \sigma^2 I_n)$)，若定义$F=W/q$，

则后面的命题2表明$F\overset{H_0}{\sim}F_{q,n-p}$ (所以$W\overset{H_0}{\sim}qF_{q,n-p}$）。

F检验

模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$，假设误差服从正态分布 $\boldsymbol{\varepsilon}\sim N(0,\ \sigma^2 I_n)$，

原假设$H_0:\ A_{q\times p}\boldsymbol{\beta}=\mathbf{c}_0$（$\mathbf{c}_0$已知，$A$已知的行满秩矩阵）的 F检验定义为：

$$F=W/q=(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)^{\mathsf{T}}\left(A(X^{\mathsf{T}}X)^{-1}A^{\mathsf{T}}\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)/(q\hat{\sigma}^2).$$

为了证明原假设下$F\sim F_{q,n-p}$，需要引理2：

引理2. 假设$\mathbf{x} \sim N(0, I_p)$，即$\mathbf{x}$的各个分量$iid \sim N(0,1)$,
假设B是投影矩阵，则
(1) $\mathbf{x}^\top B\mathbf{x} = \| B\mathbf{x} \|^2 \sim \chi_r^2, r = rank(B)$
(2) 若$AB = 0$, 则$A\mathbf{x}$与$B\mathbf{x}$独立，特别地，$A\mathbf{x}$与$\mathbf{x}^\top B\mathbf{x}$独立。

证明：假设B的谱分解为$B = O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top$, 令$\mathbf{y} = O^\top \mathbf{x} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N(0, I_p)$,

(1) $\mathbf{x}^\top B\mathbf{x} = \mathbf{x}^\top O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top \mathbf{x} = (\mathbf{y}_1{}^\top, \mathbf{y}_2{}^\top)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \mathbf{y}_1{}^\top \mathbf{y}_1 \sim \chi_r^2.$

(2) $B\mathbf{x} = O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top \mathbf{x} = O\begin{pmatrix} \mathbf{y}_1 \\ 0 \end{pmatrix} = O_1 \mathbf{y}_1$，$O_1$为$O = (O_1, O_2)$的前r列。

$A\mathbf{x} = AO\mathbf{y} = (C, D)\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$,其中划分$AO = (C, D)$

由$AB = 0 \Rightarrow AO\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top = 0 \Rightarrow (C, D)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = 0 \Rightarrow C = 0,$

所以$A\mathbf{x} = D\mathbf{y}_2$,所以$A\mathbf{x}$与$B\mathbf{x}$独立。

推论. 假设 P是一个$n\times n$投影矩阵，$r=\operatorname{rank}(P)=\operatorname{trace}(P)$,

若随机向量 $\mathbf{x}\sim N_n(0,I_n)$, 则

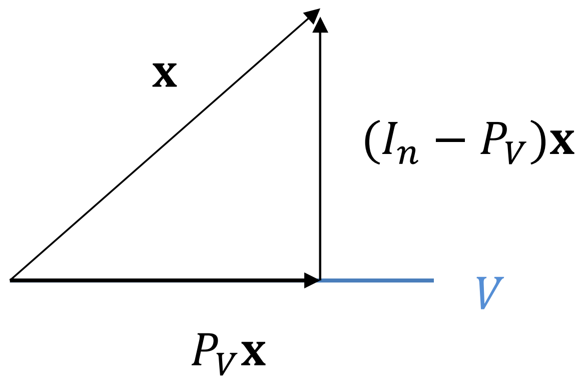
(0) $\|\mathbf{x}\|^2\sim\chi_n^2$

(1) $\|P\mathbf{x}\|^2\sim\chi_r^2$

(2) $\|(I_n-P)\mathbf{x}\|^2\sim\chi_{n-r}^2$, 且与 $\|P\mathbf{x}\|^2$独立

(3) $\dfrac{\|P\mathbf{x}\|^2/r}{\|(I_n-P)\mathbf{x}\|^2/(n-r)}\sim F_{r,n-r}$

$\dfrac{\|P\mathbf{x}\|^2}{\|\mathbf{x}\|^2}\sim\operatorname{Beta}\left(\dfrac{r}{2},\dfrac{n-r}{2}\right)$



$$
\begin{array}{ccccc}
\|\mathbf{x}\|^2 & = & \|P_V\mathbf{x}\|^2 & + & \|(I_n-P_V)\mathbf{x}\|^2 \\
\chi_n^2 & = & \chi_{\dim(V)}^2 & + & \chi_{n-\dim(V)}^2
\end{array}
$$

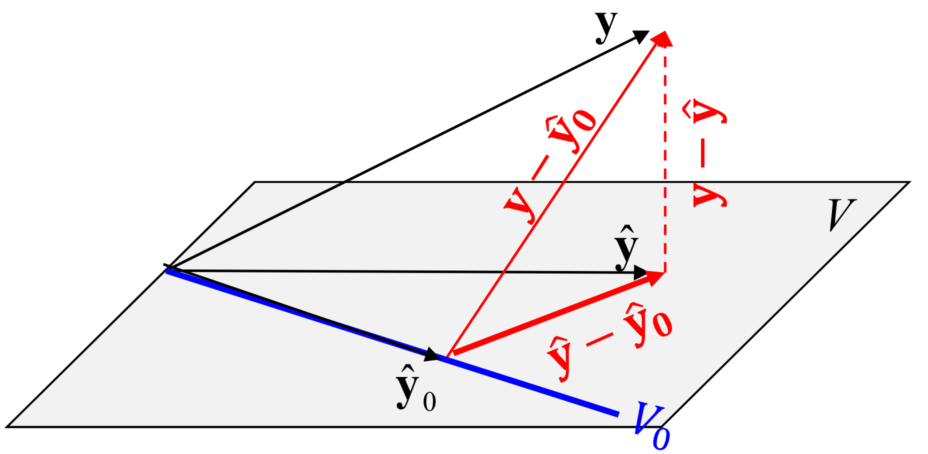
如P5所说，我们不妨仅考虑$\mathbf{c}_0=\mathbf{0}$的情形

命题2. 线性模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n)$，$H_0: A\boldsymbol{\beta}=\mathbf{0}$，其中$A_{q\times p}$行满秩已知。记$V=C(X),V_0=\{X\boldsymbol{\beta}: A\boldsymbol{\beta}=0\},\hat{\mathbf{y}}=P_V\mathbf{y},\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}$，定义

$$F=\frac{(A\hat{\boldsymbol{\beta}})^{\mathsf T}\left(A(X^{\mathsf T}X)^{-1}A^{\mathsf T}\right)^{-1}(A\hat{\boldsymbol{\beta}})}{q\hat{\sigma}^2},$$

则

$$F=\frac{n-p}{q}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}\overset{\mathrm{H}_0}{\sim}F_{q,n\text{-}p}。$$



$$\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2$$

$$\|\mathbf{y}-\hat{\mathbf{y}}\|^2=\|(I-P_V)\boldsymbol{\varepsilon}\|^2\sim\sigma^2\chi^2_{n-\mathrm{di}\ (V)}$$

$$\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2=\left\|(I-P_{V_0})\boldsymbol{\varepsilon}\right\|^2\sim\sigma^2\chi^2_{n-\dim(V_0)}$$

$$\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\left\|(P_V-P_{V_0})\boldsymbol{\varepsilon}\right\|^2\sim\sigma^2\chi^2_{\dim(V)-\dim(V_0)}$$

命题2的证明：(1) 我们首先证明F统计量的投影表达：

零假设成立时，$\boldsymbol{\beta}$满足约束$A\boldsymbol{\beta}=\mathbf{0}$,由命题1，此时约束最小二乘估计

$$\tilde{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}-(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}},$$

拟合值向量： $\hat{\mathbf{y}}_0=X\tilde{\boldsymbol{\beta}}=\hat{\mathbf{y}}-X(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}$

$\Rightarrow\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\hat{\boldsymbol{\beta}}^\top A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}$

$\Rightarrow F=\dfrac{\hat{\boldsymbol{\beta}}^\top A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}}{q\hat{\sigma}^2}=\dfrac{n-p}{q}\times\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}$。

H_0成立时,$P_{X(X^\top X)^{-1}A^\top}X\boldsymbol{\beta}=(\cdots)A\boldsymbol{\beta}=\mathbf{0}$,所以

$\hat{\mathbf{y}}-\hat{\mathbf{y}}_0=P_{X(X^\top X)^{-1}A^\top}\mathbf{y}=P_{X(X^\top X)^{-1}A^\top}(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=P_{X(X^\top X)^{-1}A^\top}\boldsymbol{\varepsilon}$

由引理4, $a=\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2\sim\sigma^2\chi_q^2$.

另一方面$\mathbf{y}-\hat{\mathbf{y}}=(I-P_V)(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=(I-P_V)\boldsymbol{\varepsilon}\quad\Rightarrow b=\|\mathbf{y}-\hat{\mathbf{y}}\|^2\sim\sigma^2\chi_{n-p}^2$

因为$(I-P_V)P_{X(X^\top X)^{-1}A^\top}=0$，$a,b$独立，所以$H_0$下$F=\dfrac{a/q}{b/(n-p)}\sim F_{q,n-p}$.

命题3. (1) 记RSS_0，RSS分别为原假设下和全模型下的残差平方和，则

$$F=\frac{n-p}{q}\times\frac{(RSS_0-RSS)}{RSS}$$

(2)若$A_{q\times p}$第一列为$\mathbf{0}_{q\times 1}$，R_0^2，R^2别为原假设下和全模型下的决定系数，则

$$F=\frac{n-p}{q}\times\frac{R^2-R_0^2}{1-R^2}.$$

证明:(1)因为$\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2-\|\mathbf{y}-\hat{\mathbf{y}}\|^2=RSS_0-RSS\Rightarrow F=\dfrac{(RSS_0-RSS)/q}{RSS/(n-p)}$

（2）记$A=(\mathbf{0},B)$,则$A\boldsymbol{\beta}=(\mathbf{0},B)\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}=B\mathbf{b}$（从而约束$A\boldsymbol{\beta}=0$对$\beta_0$没有约束),

而$X\boldsymbol{\beta}=(\mathbf{1},\mathrm{Z})\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}=\mathbf{1}\beta_0+Z\mathbf{b}$，所以

$$V_0=\left\{X\boldsymbol{\beta}:A\boldsymbol{\beta}=0,\ \boldsymbol{\beta}\in R^p\right\}=\left\{\mathbf{1}\beta_0+Z\mathbf{b}:B\mathbf{b}=0,\ \mathbf{b}\in R^{p-1}\right\},$$

所以$\mathbf{1}\in V_0$（从而$\mathbf{1}\perp\hat{\mathbf{y}}-\hat{\mathbf{y}}_0$，$\mathbf{1}^{\mathrm{T}}\hat{\mathbf{y}}_0=\mathbf{1}^{\mathrm{T}}\hat{\mathbf{y}}=\mathbf{1}^{\mathrm{T}}\mathbf{y}=n\bar{y}$）

$$\text{因为} R^2 = \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}, R_0^2 = \frac{\|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$$

$$\Rightarrow R^2 - R_0^2 = \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2 - \|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2} = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$$

$$\text{而} 1 - R^2 = 1 - \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}, \Rightarrow \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2} = \frac{R^2 - R_0^2}{1 - R^2}$$

$$\Rightarrow F = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1 - R^2}\text{。证毕。}$$

注：事实上，“A的第一列为$\mathbf{0}$”$\Leftrightarrow$约束$A\boldsymbol{\beta} = \mathbf{0}$中不含截距项$\beta_0 \Leftrightarrow \mathbf{1} \in V_0$。

这是因为由引理1我们知$V_0 = C(X(X^\top X)^{-1} A^\top)^\perp \cap C(X)$,

$$\mathbf{1} \in V_0 \Leftrightarrow \mathbf{1} \perp C(X(X^\top X)^{-1} A^\top) \Leftrightarrow A(X^\top X)^{-1} X^\top \mathbf{1} = A\begin{pmatrix} 1/n \\ 0 \end{pmatrix} = 0 \Leftrightarrow A\text{的第一列为}\mathbf{0}.$$

参见第12讲P18的分块求逆公式

$$(X^\top X)^{-1} = \begin{pmatrix} a & -\bar{\mathbf{x}}^\top A^{-1} \\ -A^{-1}\bar{\mathbf{x}} & A^{-1} \end{pmatrix}, \text{其中} A = (n-1)S_{\mathbf{xx}}, \quad a = 1/n + \bar{\mathbf{x}}^\top A^{-1}\bar{\mathbf{x}}$$

注解：在实际应用中，一般线性假设中的A矩阵，自变量空间V_0和F检验并不一定需要套用命题1中的公式，而是通过具体问题容易得到，例如：

(1) $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$，其中$X = (X_1, X_2)$，$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$, $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$

H_0成立时 $X\boldsymbol{\beta} = X_1\boldsymbol{\beta}_1$,所以$V_0 = L(X_1)$.

(2) 模型$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + \mathbf{x}_1\beta_1 + .. + \mathbf{x}_{p-1}\beta_{p-1} + \boldsymbol{\varepsilon}$，$H_0 : \beta_1 = ... = \beta_{p-1}$

H_0成立时$X\boldsymbol{\beta} = \mathbf{1}\beta_0 + \mathbf{x}_1\beta_1 + .. + \mathbf{x}_{p-1}\beta_{p-1} = \mathbf{1}\beta_0 + (\mathbf{x}_1 + .. + \mathbf{x}_{p-1})\beta_1$,

所以$V_0 = L(\mathbf{1},\ \mathbf{x}_1 + ... + \mathbf{x}_{p-1})$.

Recap：F检验的构造及零分布

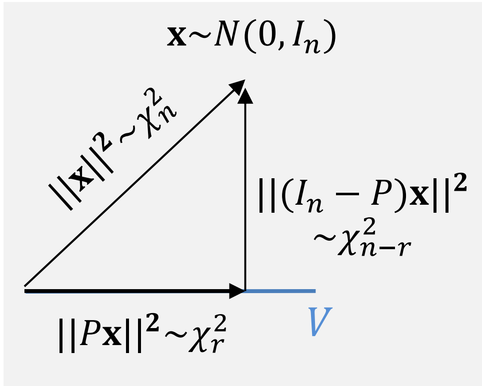
$H_0: \theta = \theta_0$的Wald 检验$W = (\hat{\theta} - \theta_0)^T [var(\hat{\theta})]^{-1} (\hat{\theta} - \theta_0)$

$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$，线性假设$H_0: A\boldsymbol{\beta} = \mathbf{0}_{q\times 1}$，$F$检验构造如下：

- Wald 检验：$W = (A\widehat{\boldsymbol{\beta}})^T [var(A\widehat{\boldsymbol{\beta}})]^{-1} (A\widehat{\boldsymbol{\beta}})$
- F检验：$F = W/q$
- 由约束最小二乘结果，以投影表示F

$$F = \frac{n-p}{q} \times \frac{||\hat{\mathbf{y}} - \hat{\mathbf{y}}_0||^2}{||\mathbf{y} - \hat{\mathbf{y}}||^2}, \quad \hat{\mathbf{y}} = P_V \mathbf{y},\ \hat{\mathbf{y}}_0 = P_{V_0} \mathbf{y}$$

其中 $V = C(X)$和V_0 分别为全模型下和原假设的自变量空间。



- 由引理2及F的投影表示，H_0成立时，$F \sim \chi^2_{q,n-p}$。
- F的其它表示：$F = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS} = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1 - R^2} = \frac{n-p}{q} \times \frac{R_p^2}{1 - R_p^2}$，
  即F检验比较全模型和原假设下的残差平方和或比较决定系数，
  $R_p^2 = \frac{R^2 - R_0^2}{1 - R_0^2}$为部分决定系数（参见P19）。
- $\mathbf{y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$，$H_0: \boldsymbol{\beta}_2 = \mathbf{0}_{q\times 1}$，$F = ||X_2^{\perp}\widehat{\boldsymbol{\beta}}_2||^2 / q\hat{\sigma}^2$

部分回归系数的F检验

模型：$\mathbf{y}_{n\times1}=X_{n\times p}\boldsymbol{\beta}_{p\times1}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+X_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n),\ V=C(\mathbf{1},X_1,X_2)$

原假设$H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times1}$，记$H_0$成立时的子模型为$\mathbf{y}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon},\ V_0=C(\mathbf{1},X_1)$。

由命题2，H_0的F检验统计量

$$F=\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2/q}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2/(n-p)},\quad \hat{\mathbf{y}}=P_V\mathbf{y},\quad \hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}$$

命题4. 假设模型$\mathbf{y}_{n\times1}=X_{n\times p}\boldsymbol{\beta}_{p\times1}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+X_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$，$H_0$：$\boldsymbol{\beta}_2=\mathbf{0}_{q\times1}$，记$X_2^{\perp}=X_2-P_{V_0}X_2=X_2-P_{\mathbf{1},X_1}X_2$，则$H_0$的$F$检验统计量为

$$F=\frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2}{q\hat{\sigma}^2}\overset{H_0}{\sim}F_{q,n-p},$$

其中$\hat{\boldsymbol{\beta}}_2=(X_2^{\perp\mathrm{T}}X_2^{\perp})^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}$为$\boldsymbol{\beta}_2$的LS估计。

证明：注意到 $\hat{\mathbf{y}} = P_{(1,X_1,X_2)}\mathbf{y} = P_{(1,X_1,X_2{}^{\perp})}\mathbf{y} = \hat{\mathbf{y}}_0 + P_{X_2{}^{\perp}}\mathbf{y} = \hat{\mathbf{y}}_0 + X_2{}^{\perp}\hat{\boldsymbol{\beta}}_2^{\perp}$

$\Rightarrow \hat{\mathbf{y}} - \hat{\mathbf{y}}_0 = X_2{}^{\perp}\hat{\boldsymbol{\beta}}_2^{\perp}$

另外，$\hat{\sigma}^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2/(n-p)$，所以

$$F = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2/q}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2/(n-p)} = \frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2/q}{\hat{\sigma}^2}$$

定义：假设R^2是全模型 $\mathbf{y}_{n\times1} = X_{n\times p}\boldsymbol{\beta}_{p\times1} + \boldsymbol{\varepsilon}$下的决定系数，$R_0^2$是子模型（$H_0$：$\boldsymbol{\beta}_2 = 0$成立时的模型）下的决定系数，定义部分(partial)决定系数

$$R_p^2 = R^2_{y\mathbf{x}_2\bullet\mathbf{x}_1} = \frac{R^2 - R_0^2}{1 - R_0^2}$$

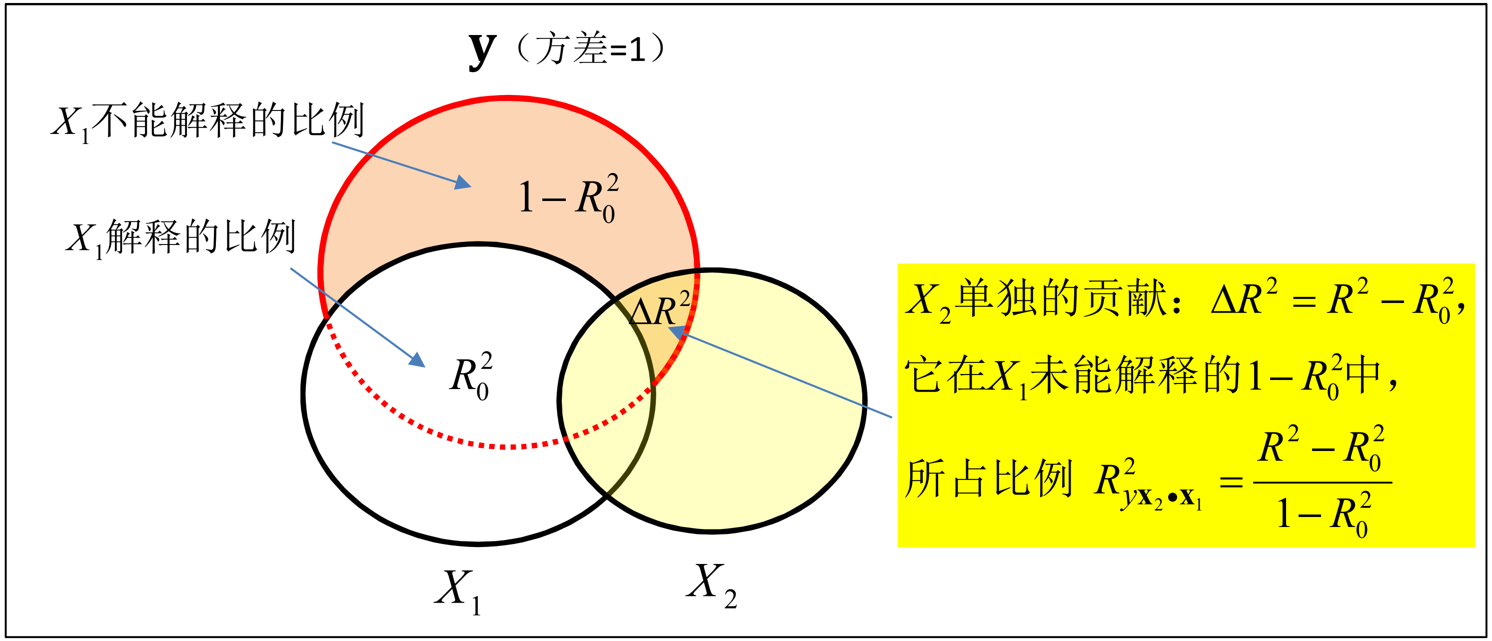
代表$\mathbf{y}$中X_1不能解释的部分中，X_2所能解释的比例。

参见第4讲P12总体的部分决定系数：

$q = 1$时(y是随机变量)，$\Phi^{\perp}$是介于0,1之间的实数，记作 $R^2_{y\mathbf{x}\bullet\mathbf{z}}$

$$R^2_{y\mathbf{x}\bullet\mathbf{z}} = \frac{\Sigma_{y^{\perp}\mathbf{x}^{\perp}}\Sigma^{-1}_{\mathbf{x}^{\perp}\mathbf{x}^{\perp}}\Sigma_{\mathbf{x}^{\perp}y^{\perp}}}{\Sigma_{y^{\perp}y^{\perp}}} = \frac{\Sigma_{y\mathbf{x}\bullet\mathbf{z}}\Sigma^{-1}_{\mathbf{x}\mathbf{x}\bullet\mathbf{z}}\Sigma_{\mathbf{x}y\bullet\mathbf{z}}}{\Sigma_{yy\bullet\mathbf{z}}}$$

我们称之为控制随机向量 $\mathbf{z}$条件下，响应变量 y和随机向量 $\mathbf{x}$之间的决定系数。



上图的具体解释如下：

原假设成立$(H_0: \boldsymbol{\beta}_2 = 0)$时，子模型$\mathbf{y} = \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$的自变量$X_1$能解释响应方差的百分比为 R_0^2，不能解释的比例为$1-R_0^2$。

增加变量X_2后，模型$\mathbf{y} = \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$对响应的解释能力增加 了$\Delta R^2 = R^2 - R_0^2$，其在子模型不能解释 的部分$1-R_0^2$中的占比，

$$R^2_{y\mathbf{x}_2 \bullet \mathbf{x}_1} = \frac{\Delta R^2}{1-R_0^2} = \frac{R^2 - R_0^2}{1-R_0^2},$$

它度量了X_2的重要性，可用来检验 $\boldsymbol{\beta}_2$是否等于0。

$$\boxed{\text{命题5. } R^2_{y\mathbf{x}_2\bullet\mathbf{x}_1}=\frac{S_{y2\bullet1}S^{-1}_{22\bullet1}S_{2y\bullet1}}{S_{yy\bullet1}}\text{。}}$$

证明：$\hat{\mathbf{y}}-\hat{\mathbf{y}}_0=X_2^{\perp}\hat{\boldsymbol{\beta}}_2 \Rightarrow \|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\mathbf{y}^{\top}X_2^{\perp}(X_2^{\perp\top}X_2^{\perp})^{-1}X_2^{\perp\top}\mathbf{y}=(n-1)S_{y2\bullet1}S^{-1}_{22\bullet1}S_{2y\bullet1}$,

另外，$\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2=\|\mathbf{y}-\mathbf{1}\bar{y}\|^2-\|\hat{\mathbf{y}}_0-\mathbf{1}\bar{y}\|^2=(n-1)S_{yy}-(n-1)S_{y1}S^{-1}_{11}S_{1y}=(n-1)S_{yy\bullet1}$

所以$R^2_{y\mathbf{x}_2\bullet\mathbf{x}_1}=\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2}=\dfrac{S_{y2\bullet1}S^{-1}_{22\bullet1}S_{2y\bullet1}}{S_{yy\bullet1}}$

结合上述命题3，4有

命题5. 正态模型$\mathbf{y}_{n\times1}=X_{n\times p}\boldsymbol{\beta}_{p\times1}+\boldsymbol{\varepsilon}$中，$X=(\mathbf{1},X_1,X_2)$，$\boldsymbol{\beta}=(\beta_0,\boldsymbol{\beta}_1^{\top},\boldsymbol{\beta}_2^{\top})$，

$H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times1}$的F检验统计量$F=\dfrac{n-p}{q}\times\dfrac{R_p^2}{1-R_p^2}\overset{H_0}{\sim}F_{q,n-p}$，

其中$R_p^2\overset{\Delta}{=}R^2_{y\mathbf{x}_2\bullet\mathbf{x}_1}=\dfrac{S_{y2\bullet1}S^{-1}_{22\bullet1}S_{2y\bullet1}}{S_{yy\bullet1}}$为控制$X_1$时$\mathbf{y}$与$X_2$的部分决定系数。

我们已经看到，F检验有若干表示形式，对比如下

$H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times 1}$

$F=\frac{n-p}{q}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}$ 几何解释

$F=\frac{n-p}{q}\times\frac{RSS_0-RSS}{RSS}$ 几何解释,软件使用

$F=\frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2}{q\hat{\sigma}^2}$ 直观

$F=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2}$ 形式统一

- $q=1$情形：
  - $p=2,q=1$ 两个变量的相关性检验$F=t^2=(n-2)\frac{r^2}{1-r^2}\overset{H_0}{\sim}F_{1,n-2}$
  - $p>2,q=1$ 两个变量的偏相关性检验$F=t^2=(n-p)\frac{r_{partial}^2}{1-r_{partial}^2}\overset{H_0}{\sim}F_{1,n-p}$
- $q>1$情形：
  - $q=p-1$: 回归方程的显著性检验 $F=\frac{n-p}{q}\times\frac{R^2}{1-R^2}\overset{H_0}{\sim}F_{p-1,n-p}$
  - $1<q<p-1$: $F=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2}$

特例：回归方程的显著性检验

R输出结果summary中包括了两种重要情形：

- 所有自变量的F检验（回归方程的显著性检验）
- 每个回归系数的t检验

回归方程的显著性检验

<u>检验所有自变量是否显著：</u>

假设模型 $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\mathbf{b} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n),$

零假设为 $H_0 : \mathbf{b} = \mathbf{0}_{(p-1)\times 1}$，回归方程的显著性$F$检验：

$$F = \frac{\| Z_c \hat{\mathbf{b}} \|^2}{(p-1)\hat{\sigma}^2} = \boxed{\frac{n-p}{p-1} \times \frac{R^2}{1-R^2}} \sim_{H_0} F_{p-1,n-p}$$

其中Z_c为Z的中心化，$\hat{\mathbf{b}}$，$\hat{\sigma}^2$为LS估计，R^2为决定系数。

$F \geq F_{p-1,n-p}(1-\alpha)$时在水平$\alpha$下否定原假设.

零假设下的决定系数$R_0^2 = 0$

回归方程显著性检验由决定系数唯一决定： $F = \frac{n-p}{q} \times \frac{R^2}{1-R^2}$

特例：单个回归系数的t检验

单个回归系数的t检验
($\sqrt{F}$检验)

假设模型$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + \mathbf{x}_1\beta_1 + \ldots + \mathbf{x}_{p-1}\beta_{p-1} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$,

记$\mathbf{x}_k^{\perp} = \mathbf{x}_k - P_{X_{(-k)}}\mathbf{x}_k$，$X_{(-k)}$为$X$除第$k$列之外的构成的矩阵，

LS估计$\hat{\beta}_k = (\mathbf{x}_k^{\perp\top}\mathbf{x}_k^{\perp})^{-1}\mathbf{x}_k^{\perp\top}\mathbf{y}$，$\mathrm{var}(\hat{\beta}_k \mid X) = \sigma^2(\mathbf{x}_k^{\perp\top}\mathbf{x}_k^{\perp})^{-1} = \sigma^2 / \|\mathbf{x}_k^{\perp}\|^2$.

$H_0: \beta_k = 0$，$k \geq 1$的F检验：

$$F = \frac{\|\mathbf{x}_k^{\perp}\hat{\beta}_k\|^2}{\hat{\sigma}^2} = \left(\frac{\hat{\beta}_k}{\hat{\sigma} \ / \|\mathbf{x}_k^{\perp}\|}\right)^2 \sim_{H_0} F_{1,n-p}$$

这等价于t检验

$$t = \pm\sqrt{F} = \frac{\hat{\beta}_k}{\hat{\sigma} \ / \|\mathbf{x}_k^{\perp}\|} \sim_{H_0} t_{n-p}。$$

t = 估计值 ÷ 标准差

例2. 待遇上的性别歧视案例 (数据集salary：alr4).有大学女教师反映女性在待遇上受到了歧视，为此收集了一批数据，包括每个大学教师的工资Salary，性别Sex, 职称Rank和工龄Year。Salary~Sex得到Sex的效应是接近显著的(p=0.07)，但控制其它变量时Sex不再显著(p=0.46):

$$\text{Salary} = \beta_0 + \beta_1\text{Sex} + \beta_2\text{Rank} + \beta_3\text{Year} + \varepsilon$$

```
lm(formula = Salary ~ Sex + Rank + Year, data = salary)

Coefficients:
              Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)   11011.76   966.95      11.388    3.03e-15 ***
Sex           603.77     811.20      0.744     0.46
Rank          4747.18    452.58      10.489    5.18e-14 ***
Year          393.86     74.53       5.285     3.04e-06 ***
---
Residual standard error: 2398 on 48 degrees of freedom
Multiple R-squared: 0.8454,    Adjusted R-squared: 0.8358
F-statistic: 87.51 on 3 and 48 DF,  p-value: < 2.2e-16
```

t = 估计值 ÷ 标准差

t=603.77/811.20
=0.744

回归方程的显著性检验

单个回归系数的t检验

在R软件中，summary没有输出的检验需要使用anova函数，比较完全模型和零模型

例如，同时检验Rank, Year的显著性：$H_0: \beta_2 = \beta_3 = 0$。

```
> m1  = lm(Salary~Sex+Rank+Year, data=salary)   #完全模型
> m0  = lm(Salary~Sex, data=salary)              #零模型
> anova(m0, m1)  #方差分析anova：比较m0,m1
Analysis of Variance Table

Model 1: Salary ~ Sex
Model 2: Salary ~ Sex + Rank + Year
  Res.Df      RSS        Df    Sum of Sq      F        Pr(>F)
1    50      1671623638
2    48      276016717    2    1395606921  121.35    < 2.2e-16 ***
```

$$F = \frac{n-p}{k} \times \frac{RSS_0 - RSS}{RSS} = \frac{52-4}{2} \times \frac{1671623638 - 276016717}{276016717} = 121.35$$