回归分析 (01714601)
课程主页：http://staff.ustc.edu.cn/~ynyang/lm2020

第十七讲 一般线性假设的F检验

2020.4.15

$$F=\frac{n-p}{k}\times\frac{R^2-R_0^2}{1-R^2}=\frac{n-p}{k}\times\frac{\| P_V\mathbf{y}-P_{V_0}\mathbf{y}\|^2}{\|\mathbf{y}-P_V\mathbf{y}\|^2}$$

回归系数的显著性检验

模型 $\mathbf{y} = X_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + X_J\boldsymbol{\beta}_J + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (0, \sigma^2), \ |J| = k,$

$H_0 : \boldsymbol{\beta}_J = \mathbf{0}_{k\times 1}.$

F 或卡方检验在控制 $|I| = p-k-1$ 个自变量的前提下("偏"),检验 $|J| = k$ 个自变量与响应变量的相关性("复").

正态条件下：$F = \dfrac{n-p}{k} \times \dfrac{r^2}{1-r^2} \overset{H_0}{\sim} F_{k,n-p}, \quad r^2 = R^2_{J\bullet I}\ , \ F \geq F_{k,n-p}(1-\alpha)$

非正态假设下：$W = (n-p) \times r^2 \xrightarrow{H_0} \chi^2_k, \quad r^2 = R^2_{J\bullet I}, \ W \geq \chi^2_k(1-\alpha)$

$k = 1$的情形已经介绍过(lect3,4,8)：$t = \sqrt{n-p}\,\dfrac{r}{\sqrt{1-r^2}}, z = \sqrt{n-p} \times r$

一个参数的检验($k=1$)已经在3,4,8讲中出现过：

第3讲p7-8： $p=2 \quad (k=1)$，两个变量x, y的相关性检验($Pearson$相关系数）

第8讲p12,14：$p=2 \quad (k=1)$, 以简单回归模型为工具研究两个变量x, y的相关性.

第4讲p16： $p>2, \quad k=1$，两个变量x, y的偏相关性检验(控制其它$p-2$个变量$\mathbf{z}$);

多个参数($k \geq 1$)的检验：

第16-19讲：$p \geq 2, \ k \geq 1$, 误差正态假设下k个参数($k \geq 1$)的F检验.

下面求F检验统计量的几个等价表示，及其在原假设下的分布：

模型 $\mathbf{y} = X_{n\times p}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + X_J\boldsymbol{\beta}_J + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(0,\sigma^2),\ |J| = k.$

$H_0: \boldsymbol{\beta}_J = \mathbf{0}$

令$V = L(X),\ V_0 = L(\mathbf{1}, X_I) = L(X_{-J})$分别为全模型和$H_0$成立时的自变量空间.

$$\hat{\mathbf{y}} = P_V\mathbf{y}\,,\ \hat{\mathbf{y}}_0 = P_{V_0}\mathbf{y}.$$

$R^2 = R^2_{I\cup J}$，R_I^2分别为全模型和H_0成立时的决定系数（复相关系数平方）.

偏决定系数（偏复相关系数平方）：$R^2_{J\bullet I} = \dfrac{R^2_{I\cup J} - R_I^2}{1 - R_I^2}.$

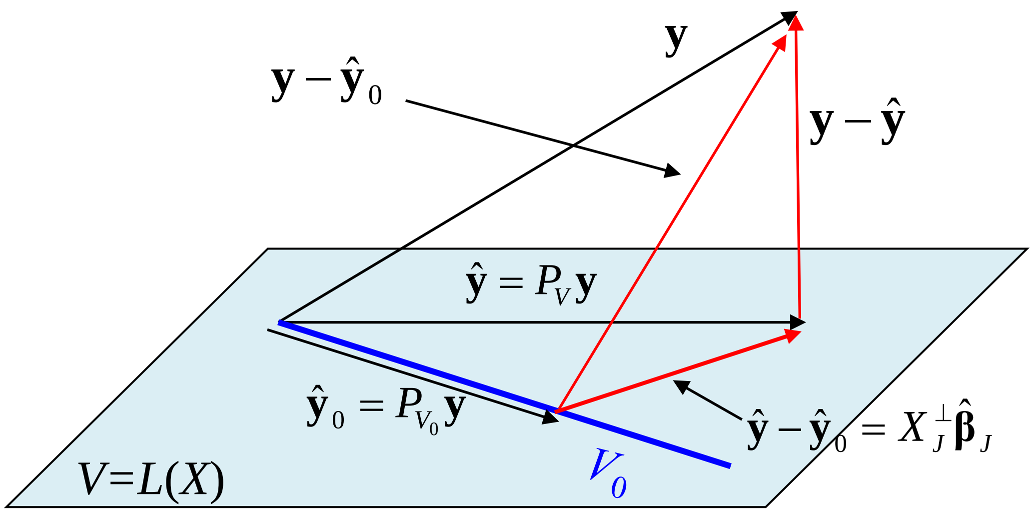
H_0的检验统计量取为：$F = \dfrac{n-p}{k} \times \dfrac{R^2_{J\bullet I}}{1 - R^2_{J\bullet I}}$

命题1. $F=\frac{n-p}{k}\times\frac{R^2_{J\bullet I}}{1-R^2_{J\bullet I}}$ 可等价地表示为

$$F=\frac{n-p}{k}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}=\frac{n-p}{k}\times\frac{\|\mathbf{e}_0\|^2-\|\mathbf{e}\|^2}{\|\mathbf{e}\|^2}=\frac{\|X_J^{\perp}\hat{\boldsymbol{\beta}}_J\|^2}{k\hat{\sigma}^2},$$

其中$\mathbf{e}_0=\mathbf{y}-\hat{\mathbf{y}}_0,\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}},X_J^{\perp}=X_J-P_{V_0}X_J=X_J-P_{X_{-J}}X_J$。

另外，原假设$H_0:\boldsymbol{\beta}_J=\mathbf{0}$成立时，$F\overset{H_0}{\sim}F_{k,n-p}$。



$$R^2_{J\bullet I}=\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2}$$

$$\frac{R^2_{J\bullet I}}{1-R^2_{J\bullet I}}=\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}\propto F$$

注1：如果$H_0: \boldsymbol{\beta}_J = \mathbf{c}_0$ ($\mathbf{c}_0$为已知的常数向量),通过令$\boldsymbol{\beta}_J^* = \boldsymbol{\beta}_J - \mathbf{c}_0, \mathbf{y}^* = \mathbf{y} - X_J\mathbf{c}_0$,
将H_0检验问题转化为对模型$\mathbf{y}^* = \mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + X_J\boldsymbol{\beta}_J^* + \boldsymbol{\varepsilon}$ 检验H_0^*： $\boldsymbol{\beta}_J^* = \mathbf{0}$

注2：我们从定义

$$F = \frac{n-p}{k} \times \frac{R_{J\bullet I}^2}{1-R_{J\bullet I}^2} \tag{1}$$

出发，得到其几何解释:

$$F = \frac{n-p}{k} \times \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2} \tag{2}$$

或以残差表示：

$$F = \frac{n-p}{k} \times \frac{\|\mathbf{e} - \mathbf{e}_0\|^2}{\|\mathbf{e}\|^2} = \frac{n-p}{k} \times \frac{\|\mathbf{e}_0\|^2 - \|\mathbf{e}\|^2}{\|\mathbf{e}\|^2} \tag{3}$$

进而与LS估计联系起来:

$$F = \frac{\|X_J^{\perp}\hat{\boldsymbol{\beta}}_J\|^2}{k\hat{\sigma}^2} \tag{4}$$

(1)-(4)都可以作为F检验的定义。不同的教材F的定义出发点可能不同，
大多教材以(4)作为F的定义

- (1)最简洁，具有一般性;
- (2),(3)几何上容易理解,具有一般性,R软件方差分析利用(3)计算F;
- (4)最容易理解,它直接度量$\hat{\boldsymbol{\beta}}_J$的大小.

证明：第14讲命题3的证明过程已经推导过$R^2_{J\bullet I}$，此处再推导一遍：

$$\hat{\mathbf{y}} \triangleq P_V\mathbf{y} = P_{V_0}\mathbf{y} + P_{V\backslash V_0}\mathbf{y} = P_{(\mathbf{1},X_I)}\mathbf{y} + P_{X_J^{\perp}}\mathbf{y} = \hat{\mathbf{y}}_0 + X_J^{\perp}\hat{\boldsymbol{\beta}}_J, \Rightarrow \hat{\mathbf{y}} - \hat{\mathbf{y}}_0 = X_J^{\perp}\hat{\boldsymbol{\beta}}_J$$

因为$R^2_{I\cup J} = \dfrac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$， $R^2_I = \dfrac{\|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$，所以

$$R^2_{J\bullet I} = \frac{R^2_{I\cup J} - R^2_I}{1 - R^2_I} = \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2 - \|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2 - \|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2} = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}_0\|^2}$$

$$\Rightarrow \frac{R^2_{J\bullet I}}{1 - R^2_{J\bullet I}} = \frac{R^2_{I\cup J} - R^2_I}{1 - R^2_{I\cup J}} = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2} = \frac{\|X_J^{\perp}\hat{\boldsymbol{\beta}}_J\|^2}{(n-p)\hat{\sigma}^2}$$

原假设下，$\mathbf{y} = \mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + \boldsymbol{\varepsilon}$, 因为$P_V X_I = X_I, P_{V_0} X_I = X_I$, 我们有

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_0 = \left(P_V - P_{V_0}\right)\mathbf{y} = \left(P_V - P_{V_0}\right)(\mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + \boldsymbol{\varepsilon}) = \left(P_V - P_{V_0}\right)\boldsymbol{\varepsilon},$$

$$\mathbf{y} - \hat{\mathbf{y}} = \left(I_n - P_V\right)\mathbf{y} = \left(I_n - P_V\right)(\mathbf{1}\beta_0 + X_I\boldsymbol{\beta}_I + \boldsymbol{\varepsilon}) = \left(I_n - P_V\right)\boldsymbol{\varepsilon}.$$

所以原假设下，$F=\dfrac{\mathbf{y}^{\top}\left(P_V-P_{V_0}\right)\mathbf{y}/k}{\mathbf{y}^{\top}\left(I_n-P_V\right)\mathbf{y}/(n-p)}=\dfrac{\boldsymbol{\varepsilon}^{\top}\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon}/k}{\boldsymbol{\varepsilon}^{\top}\left(I_n-P_V\right)\boldsymbol{\varepsilon}/(n-p)}$

因为

$\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n)$,

$P_V-P_{V_0}$是秩k投影阵，

I_n-P_V是秩$n-p$投影阵, 且$(P_V-P_{V_0})(I_n-P_V)=0$,

由第10讲命题4, 有：

$$\boldsymbol{\varepsilon}^{\top}\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon}\sim\sigma^2\chi_k^2,\ \ \boldsymbol{\varepsilon}^{\top}\left(I_n-P_V\right)\boldsymbol{\varepsilon}\sim\sigma^2\chi_{n-p}^2,\text{且两者独立.}$$

所以原假设下，$F\sim F_{k,n-p}$

特例1. 单个回归系数的t-检验

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

$H_0 : \beta_j = 0$

$$\mathbf{x}_{(j)}^{\perp} = \mathbf{x}_{(j)} - P_{X_{(-j)}} \mathbf{x}_{(j)}.$$

$$\hat{\beta}_j = (\mathbf{x}_{(j)}^{\perp\mathsf{T}} \mathbf{x}_{(j)}^{\perp})^{-1} \mathbf{x}_{(j)}^{\perp\mathsf{T}} \mathbf{y},$$

$$\mathrm{var}(\hat{\beta}_j \mid X) = \sigma^2 (\mathbf{x}_{(j)}^{\perp\mathsf{T}} \mathbf{x}_{(j)}^{\perp})^{-1}.$$

$$F = W = \frac{\| \mathbf{x}_{(j)}^{\perp} \hat{\beta}_j \|^2}{\hat{\sigma}^2} = \frac{\hat{\beta}_j^{\,2}}{\hat{\sigma}^2 / \| \mathbf{x}_{(j)}^{\perp} \|^2} \sim_{H_0} F_{1,n-p},$$

$$t = \frac{\hat{\beta}_j}{\hat{\sigma} \ / \| \mathbf{x}_j^{\perp} \|} = \frac{\hat{\beta}_j}{\sqrt{\widehat{\mathrm{var}}(\hat{\beta}_j)}} \sim_{H_0} t_{n-p},$$

$$F = t^2$$

特例2. 回归方程的显著性检验

$\mathbf{y}=\mathbf{1}\beta_0+Z\boldsymbol{\gamma}+\boldsymbol{\varepsilon}, \beta_0$是截距, $Z_{(c)}=Z^{\perp}=Z-P_{\mathbf{1}}Z$中心化矩阵

$$H_0:\boldsymbol{\gamma}=\mathbf{0}_{(p-1)\times 1}$$

$V_0=L(\mathbf{1}), \hat{\mathbf{y}}_0=\mathbf{1}\bar{y}$, 全模型下的决定系数$R^2$，原假设下的$R_0^2=0$

回归方程显著性检验：

$$F=\frac{n-p}{p-1}\times\frac{R^2}{1-R^2}=\frac{n-p}{p-1}\times\frac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{\mathrm{y}}\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}=\frac{\|Z^{\perp}\hat{\boldsymbol{\gamma}}\|^2}{(p-1)\hat{\sigma}^2}\sim_{H_0}F_{p-1,n-p}$$

R软件中的部分回归系数的F检验

R软件回归分析标准输出结果中包含了所有单个回归系数的t-检验和回归方程的显著性F-检验：

> summary (lm.out)

anova： ANalysis Of Variance，方差分析

一般的线性假设（比如同时检验多个担不是回归系数）的检验可由函数 **anova** 得到

> anova (null.model, full.model)

其中**null.model, full.model** 分别是零模型和全模型下的**lm**函数输出结果。

例1. 待遇上的性别歧视案例 (数据集salary：alr3).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
```

回归方程的显著性检验

单个回归系数的t检验

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

$$H_0 : \beta_2 = \beta_3 = 0$$

```
> m1   = lm(Salary~Sex+Rank+Year, data=salary)    #full model
> m0   = lm(Salary~Sex, data=salary)               #null model
> anova(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 ***
```

$$F = \frac{n-p}{k} \times \frac{RSS_0 - RSS}{RSS} = \frac{52-4}{2} \times \frac{1671623638 - 276016717}{276016717} = 121.3$$

$$F = \frac{n-p}{k} \times \frac{R^2 - R_0^2}{1-R^2} = \frac{52-4}{2} \times \frac{0.8454 - 0.0639}{1-0.8454} = 121.3$$

一般线性假设

线性假设指的是原假设为回归系数的一些线性约束，比如$\beta_1 = 0$, $\beta_1 = \beta_2, \beta_2 = (\beta_1 + \beta_3)/2$等等，皆可表示为 $H_0 : A\boldsymbol{\beta} = \mathbf{c}_0$ （$\mathbf{c}_0$已知）。

$H_0 : A\boldsymbol{\beta} = \mathbf{c}_0$ （$A, \mathbf{c}_0$已知）称为一般线性假设。 不妨设$\mathbf{c}_0 = 0$。

这是因为 $H_0 : A\boldsymbol{\beta} = \mathbf{c}_0 \Leftrightarrow H_0 : A\left[\boldsymbol{\beta} - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0\right] = \mathbf{0} \Leftrightarrow H_0 : A\boldsymbol{\beta}^* = \mathbf{0}$,
其中我们令$\boldsymbol{\beta}^* = \boldsymbol{\beta} - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0$. 此时模型

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

$\Leftrightarrow \mathbf{y}^* = X\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$，其中我们令$\mathbf{y}^* = \mathbf{y} - XA^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0$.

在模型$\mathbf{y}^* = X\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$中检验$H_0$：$A\boldsymbol{\beta}^* = \mathbf{0}$, 等价于在原模型中检验$H_0 : A\boldsymbol{\beta} = \mathbf{c}_0$。

- 最重要的特殊情况H_0：$\boldsymbol{\beta}_J=\mathbf{0},\quad J\subset\{1,\ldots,p-1\}$

比如$A=(\mathbf{0},I_k),\boldsymbol{\beta}=\begin{pmatrix}\boldsymbol{\beta}_{(1)}\\ \boldsymbol{\beta}_{(2)}\end{pmatrix},A\boldsymbol{\beta}=\boldsymbol{\beta}_{(2)},$

- 一般线性假设H_0：$A\,\boldsymbol{\beta}=\mathbf{c}_0$的表示形式不唯一 ($A$不唯一).

比如$A\,\boldsymbol{\beta}=\mathbf{0}\Leftrightarrow BA\boldsymbol{\beta}=\mathbf{0}$,对任何$k\times k$可逆矩阵$B$.

$\boldsymbol{\beta}=(\beta_{0,}\beta_1,\beta_2,\beta_3)^T,H_0:\beta_1=\beta_2=\beta_3$可表示为$H_0:A\boldsymbol{\beta}=0,$

$$A=\begin{pmatrix}0&1&-1&0\\0&0&1&-1\end{pmatrix},\begin{pmatrix}0&1&0&-1\\0&0&1&-1\end{pmatrix},\begin{pmatrix}0&-1&2&-1\\0&-1&-1&2\end{pmatrix}$$

零空间：零假设下自变量张成的空间

原假设$H_0: A_{k\times p}\boldsymbol{\beta} = \mathbf{0}$成立时$\boldsymbol{\beta}$处于$k$个约束之下，
$V_0 = \left\{X\boldsymbol{\beta} : A\boldsymbol{\beta} = 0\right\} \subset V = L(X)$称为零空间.

例2. (1) $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = X_1\boldsymbol{\beta}_{(1)} + X_2\boldsymbol{\beta}_{(2)} + \boldsymbol{\varepsilon}$，其中$X = (X_1, X_2)$，$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \boldsymbol{\beta}_{(2)} \end{pmatrix}$,

$H_0: \boldsymbol{\beta}_{(2)} = \mathbf{0}_{k\times 1}$，$H_0$成立时 $X\boldsymbol{\beta} = X_1\boldsymbol{\beta}_{(1)} + X_2\boldsymbol{\beta}_{(2)} = X_1\boldsymbol{\beta}_{(1)}$,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$V_0 = L(\mathbf{1},\ \mathbf{x}_1 + ... + \mathbf{x}_{p-1})$

命题2. 假设线性模型$\mathbf{Y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$，设$A$为$k\times p$行满秩常数矩阵，则一般线性假设$H_0: A\boldsymbol{\beta}=\mathbf{0}_{k\times 1}$成立时自变量空间

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

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$\boldsymbol{\mu}=X\boldsymbol{\beta}\in L(X)\Rightarrow \boldsymbol{\beta}=(X^\top X)^{-1}X^\top\boldsymbol{\mu}$，所以原假设成立时，

$$A\boldsymbol{\beta}=A(X^\top X)^{-1}X^\top\boldsymbol{\mu}=0$$

$$\Leftrightarrow \boldsymbol{\mu}\perp L(X(X^\top X)^{-1}A^\top)\Rightarrow \boldsymbol{\mu}\in L(X(X^\top X)^{-1}A^\top)^\perp,$$

但$\boldsymbol{\mu}\in L(X)$,所以$\boldsymbol{\mu}\in L(X)\bigcap L(X(X^\top X)^{-1}A^\top)^\perp$,

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

线性假设的子空间表示

模型 $\mathbf{y}=X_{n\times p}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$，记$V=L(X)=\{X\boldsymbol{\beta}\mid\boldsymbol{\beta}\in R^{p}\}$,

无论关于$\boldsymbol{\beta}$的线性假设H_0具有何种具体的约束形式，在H_0成立时，

$\left\{X\boldsymbol{\beta}\mid\boldsymbol{\beta}\in R^{p},\boldsymbol{\beta}\text{满足原假设}H_0\right\}\in V_0\subset V$,$V_0$为线性假设下的自变量子空间.

通常,$\mathbf{1}\in V_0$

因此如果记$\boldsymbol{\mu}=E(\mathbf{y}\mid X)$,模型 $\mathbf{y}=X_{n\times p}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$,可以表示为

$$\mathbf{y}=\boldsymbol{\mu}+\boldsymbol{\varepsilon},\quad \boldsymbol{\mu}\in V=L(X),$$

而线性原假设可表示为

$$H_0:\boldsymbol{\mu}\in V_0\subset V$$

偏决定系数与F统计量

全模型V： $\hat{\mathbf{y}} = P_V \mathbf{y}$, 决定系数(复相关系数平方) $R^2 = \dfrac{\| P_V \mathbf{y} - \mathbf{1}\bar{y} \|^2}{\| \mathbf{y} - \mathbf{1}\bar{y} \|^2} = \dfrac{\| \hat{\mathbf{y}} - \mathbf{1}\bar{y} \|^2}{\| \mathbf{y} - \mathbf{1}\bar{y} \|^2}$;

原假设V_0： $\hat{\mathbf{y}}_0 = P_{V_0} \mathbf{y}$, 决定系数 $R_0^2 = \dfrac{\| P_{V_0} \mathbf{y} - \mathbf{1}\bar{y} \|^2}{\| \mathbf{y} - \mathbf{1}\bar{y} \|^2} = \dfrac{\| \hat{\mathbf{y}}_0 - \mathbf{1}\bar{y} \|^2}{\| \mathbf{y} - \mathbf{1}\bar{y} \|^2}$.

定义: 偏决定系数 $R_{\bullet 0}^2 = \dfrac{R^2 - R_0^2}{1 - R_0^2}$.

定义： 一般线性假设$H_0 : \boldsymbol{\mu} \in V_0 \subset V$的$F$ 统计量

$$F = \frac{n-p}{k} \times \frac{R_{\bullet 0}^2}{1 - R_{\bullet 0}^2} = \frac{n-p}{k} \times \frac{R^2 - R_0^2}{1 - R^2}$$

其中$k = \dim(V \setminus V_0)$.

类似于命题1的证明,容易验证F具有如下各种等价表达：

$$\boxed{R_{\bullet 0}^2 = \frac{R^2 - R_0^2}{1 - R_0^2} = \frac{\| P_V \mathbf{y} - P_{V_0} \mathbf{y} \|^2}{\| \mathbf{y} - P_{V_0} \mathbf{y} \|^2}.}$$

F 统计量具有如下各种等价表示：

$$F = \frac{n-p}{k} \times \frac{R_{\bullet 0}^2}{1 - R_{\bullet 0}^2} = \frac{n-p}{k} \times \frac{R^2 - R_0^2}{1 - R^2} = \frac{n-p}{k} \times \frac{\| P_V \mathbf{y} - P_{V_0} \mathbf{y} \|^2}{\| \mathbf{y} - P_V \mathbf{y} \|^2} = \frac{n-p}{k} \times \frac{\mathrm{RSS}_0 - \mathrm{RSS}}{\mathrm{RSS}},$$

其中$k = \dim(V \setminus V_0), \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}, \mathbf{e}_0 = \mathbf{y} - \hat{\mathbf{y}}_0$.

一般线性假设的F检验

定理1. 假设 模型 $\mathbf{y}=\boldsymbol{\mu}+\boldsymbol{\varepsilon}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\ \boldsymbol{\mu}\in V=L(X_{n\times p}),\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n),$

$H_0:\boldsymbol{\mu}\in V_0\subset V$,则$H_0$成立的条件下，$F\sim F_{k,n-p}$.

证明：$$F=\frac{n-p}{k}\times\frac{\|P_V\mathbf{y}-P_{V_0}\mathbf{y}\|^2}{\|\mathbf{y}-P_V\mathbf{y}\|^2}=\frac{\mathbf{y}^{\top}\left(P_V-P_{V_0}\right)\mathbf{y}/k}{\mathbf{y}^{\top}\left(I_n-P_V\right)\mathbf{y}/(n-p)}.$$

原假设成立时，$\boldsymbol{\mu}=X\boldsymbol{\beta}\in V_0,\left(P_V-P_{V_0}\right)\boldsymbol{\mu}=0$,所以

$\left(P_V-P_{V_0}\right)\mathbf{y}=\left(P_V-P_{V_0}\right)(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon},$

$\left(I_n-P_V\right)\mathbf{y}=\left(I_n-P_V\right)X\boldsymbol{\beta}+\left(I_n-P_V\right)\boldsymbol{\varepsilon}=\left(I_n-P_V\right)\boldsymbol{\varepsilon}$

$$\Rightarrow F=\frac{\boldsymbol{\varepsilon}^{\top}\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon}/k}{\boldsymbol{\varepsilon}^{\top}\left(I_n-P_V\right)\boldsymbol{\varepsilon}/(n-p)}$$

因为$\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n),P_V-P_{V_0},I_n-P_V$分别时秩为$k$和$n-p$的投影矩阵，且

$(I_n-P_V)(P_V-P_{V_0})=0$,所以$\boldsymbol{\varepsilon}^{\top}\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon}\sim\sigma^2\chi_k^2,\boldsymbol{\varepsilon}^{\top}\left(I_n-P_V\right)\boldsymbol{\varepsilon}\sim\sigma^2\chi_{n-p}^2,$

且两者独立，所以$F=\dfrac{\boldsymbol{\varepsilon}^{\top}\left(P_V-P_{V_0}\right)\boldsymbol{\varepsilon}/k}{\boldsymbol{\varepsilon}^{\top}\left(I_n-P_V\right)\boldsymbol{\varepsilon}/(n-p)}\sim F_{k,n-p}$

推论1. 假设线性模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$，设$A$为$k\times p$行满秩常数矩阵，则一般线性假设 $H_0:A\boldsymbol{\beta}=\mathbf{0}_{k\times 1}$ 的F检验统计量

$$F=(A\hat{\boldsymbol{\beta}})^\top\left[A(X^\top X)^{-1}A^\top\right]^{-1}(A\hat{\boldsymbol{\beta}})/k\hat{\sigma}^2\sim_{H_0} F_{k,n-p}.$$

注$:A=(0,I_k)$即得命题1.

证明：$H_0:A\boldsymbol{\beta}=\mathbf{0}_{k\times 1}$的自变量空间$V_0=L(X(X^\top X)^{-1}A^\top)^{\perp}\cap L(X), P_{V_0}=P_V-P_{X(X^\top X)^{-1}A^\top}$,

$$\begin{aligned}
&\|P_V\mathbf{y}-P_{V_0}\mathbf{y}\|^2=\mathbf{y}^\top(P_V-P_{V_0})\mathbf{y}\\
&=\mathbf{y}^\top X(X^\top X)^{-1}A^\top\left[A(X^\top X)^{-1}A^\top\right]^{-1}A(X^\top X)^{-1}X^\top\mathbf{y}\\
&=(A\hat{\boldsymbol{\beta}})^\top\left[A(X^\top X)^{-1}A^\top\right]^{-1}(A\hat{\boldsymbol{\beta}})
\end{aligned}$$

所以$F=\dfrac{n-p}{k}\times\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}=\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{k\hat{\sigma}^2}=(A\hat{\boldsymbol{\beta}})^\top\left[A(X^\top X)^{-1}A^\top\right]^{-1}(A\hat{\boldsymbol{\beta}})/k\hat{\sigma}^2.$

LS估计的分布

定理2. 假设正态线性模型 $\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}_{n\times 1}$， $\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n)$,则

(1) $\hat{\boldsymbol{\beta}}\,|\,X\sim N\left(\boldsymbol{\beta},\sigma^2(X^\top X)^{-1}\right)$,

(2) $\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-p}$，且$\hat{\boldsymbol{\beta}},\hat{\sigma}^2$独立。

证明: (1) 由$\mathbf{y}\,|\,X\sim N(X\boldsymbol{\beta},\sigma^2 I_n)$知$\hat{\boldsymbol{\beta}}=(X^\top X)^{-1}X^\top\mathbf{y}\sim N(\boldsymbol{\beta},\sigma^2(X^\top X)^{-1})$。

(2) 因为$\mathbf{e}=(I_n-P_X)\mathbf{y}=(I_n-P_X)(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=(I_n-P_X)\boldsymbol{\varepsilon}$,

所以 $\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2}=\dfrac{\|\mathbf{e}\|^2}{\sigma^2}=(\boldsymbol{\varepsilon}/\sigma)^\top(I_n-P_X)(\boldsymbol{\varepsilon}/\sigma)$,

因为$\boldsymbol{\varepsilon}/\sigma\sim N(0,I_n)$, $rank(I_n-P_X)=tr(I_n-P_X)=n-p$,所以$\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-p}$.

另外，$\hat{\boldsymbol{\beta}}=(X^\top X)^{-1}X^\top\mathbf{y}=(X^\top X)^{-1}X^\top(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\boldsymbol{\beta}+(X^\top X)^{-1}X^\top\boldsymbol{\varepsilon}\triangleq\boldsymbol{\beta}+B\boldsymbol{\varepsilon}$

$\mathbf{e}=(I_n-P_X)\boldsymbol{\varepsilon}$, 由$B\ (I_n-P_X)=0$,知$\hat{\boldsymbol{\beta}}$与$\mathbf{e}$独立，进而与$\hat{\sigma}^2$独立。

F检验的代数构建

由定理2，$A\hat{\boldsymbol\beta} \mid X \sim N_k\left(A\boldsymbol\beta, \sigma^2 A(X^\top X)^{-1}A^\top\right)$,

$$\Rightarrow u = \left(A\hat{\boldsymbol\beta} - A\boldsymbol\beta\right)^\top\left[\sigma^2 A(X^\top X)^{-1}A^\top\right]^{-1}\left(A\hat{\boldsymbol\beta} - A\boldsymbol\beta\right) \sim \chi_k^2,$$

以及 $v = \dfrac{(n-p)\hat\sigma^2}{\sigma^2} \sim \chi_{n-p}^2$, 且$v$与$\hat{\boldsymbol\beta}$独立, 进而与$A\hat{\boldsymbol\beta}$独立,

故由F分布的定义

$$\frac{u/k}{v/(n-p)} = \left(A\hat{\boldsymbol\beta} - A\boldsymbol\beta\right)^\top\left[A(X^\top X)^{-1}A^\top\right]^{-1}\left(A\hat{\boldsymbol\beta} - A\boldsymbol\beta\right)/(k\hat\sigma^2) \sim F_{k,n-p}.$$

因此线性假设$H_0 : A\boldsymbol\beta = \mathbf{c}_{0_{k\times1}}$的检验可以构造如下：

$$F = \frac{\left(A\hat{\boldsymbol\beta} - \mathbf{c}_0\right)^\top \left(A(X^\top X)^{-1}A^\top\right)^{-1}\left(A\hat{\boldsymbol\beta} - \mathbf{c}_0\right)}{k\hat\sigma^2},\ H_0\text{成立时，} F \sim F_{k,n-p}$$

总之，一般线性假设 $H_0: A\boldsymbol{\beta}=\mathbf{0}$ 的F检验有如下各种表示，体现了不同角度下全模型和零模型的差异

$$F=\frac{n-p}{k}\times\frac{R_{\bullet 0}^2}{1-R_{\bullet 0}^2}$$

$$=\frac{n-p}{k}\times\frac{R^2-R_0^2}{1-R^2}$$

拟合优度之差

$$=\frac{n-p}{k}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}$$

投影之差

$$=\frac{n-p}{k}\times\frac{\mathrm{RSS}_0-\mathrm{RSS}}{\mathrm{RSS}}$$

残差平方和之差

$$=(A\hat{\boldsymbol{\beta}})^{\mathrm{T}}\left[A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\right]^{-1}(A\hat{\boldsymbol{\beta}})/k\hat{\sigma}^2$$