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

第十六讲 F 检验

2020.4.10

$$F=\frac{n-p}{k}\frac{r^2}{1-r^2}\overset{H_0:\beta_J=\mathbf{0}_k}{\sim}F_{k,n-p},\ r^2=R^2_{J\bullet I}$$

回归系数的显著性检验

回归系数的假设有多种,其中最要和最常见的是检验某些自变量没有效应，即某一部分回归系数为0.

假设下标集合$I \subset \{1,2,\ldots,p-1\}$, $J=\{1,2,\ldots,p-1\}\setminus I$, $|J|=k$. 划分X和$\boldsymbol{\beta}$:

$$X=(\mathbf{1},X_I,X_J),\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}_I\\ \boldsymbol{\beta}_J\end{pmatrix}$$

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

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

复偏相关系数平方 $R^2_{J\bullet I}=\dfrac{R^2_{I\cup J}-R^2_I}{1-R^2_I}$ 度量了排除X_I影响之后，X_J与响应变量的关联性大小，因此$R^2_{J\bullet I}$度量了$\boldsymbol{\beta}_J$的大小，可用来检验$H_0:\boldsymbol{\beta}_J=\mathbf{0}$。

回归系数的F检验和卡方检验

$H_0:\boldsymbol{\beta}_J=\mathbf{0}_{k\times 1}$的$F$检验和卡方检验分别为

正态条件下：

$$F=\frac{n-p}{k}\times\frac{r^2}{1-r^2}\overset{H_0}{\sim}F_{k,n-p},\quad r^2=R^2_{J\bullet I}$$

$$k=1\text{时为}t\text{检验}:\ t=\sqrt{n-p}\,\frac{r}{\sqrt{1-r^2}}\sim t_{n-p}$$

非正态假设下：

$$W=(n-p)\times r^2\overset{H_0}{\to}\chi^2_k,\quad r^2=R^2_{J\bullet I},$$

$$k=1\text{时为}z\text{检验}:\ z=\sqrt{n-p}\times r\sim N(0,1)$$

或更简洁地，$\tilde{W}=nr^2\overset{H_0}{\to}\chi^2_k$

上述问题中共有p个变量(1个响应变量, $p-1$个自变量)， F或卡方检验研究的是在控制其它$|I|=p-1-k$个自变量前提下，响应变量与$|J|=k$个自变量的复相关性。 "复"体现在$k\ge 1$, "偏"体现在$|I|=p-1-k\ge 1$, $p\ge k+2$.

我们将在后面的命题1证明上述正态假设下结果。

统计中所有检验统计量的分布只有三种：
(1)正态假设下的(精确)F检验;
(2)无正态假设下的(近似)卡方检验
(3)非参数型的精确检验(如置换检验）。

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.