高级计量经济学

第一次作业

邓皓天　2023310114

1、 在满足古典假设下，设定多元回归模型

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + \mu_i \quad (i = 1, 2, \cdots, n)$$

采用OLS对对以上模型进行参数估计，所得参数估计量记为$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$。现在在原来$n$个样本的情况下再增加一个样本，使得样本容量由原来的n变为现在的$n+1$，并再采用OLS对原始模型进行参数估计，所得参数估计量记为$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$。试问：

（1）$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$和$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$相等吗？为什么？

（2）如果$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$和$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$不相等，那么二者在满足什么样的条件下相等？

（3）请讨论第n个样本点对$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$的影响。

答：

（1）$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$和$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$不一定相等。当样本容量为$n$时，待估参数估计值的正规方程组为

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_{i,1} + \hat{\beta}_2 \sum_{i=1}^n X_{i,2} + \cdots + \hat{\beta}_k \sum_{i=1}^n X_{i,k} = \sum_{i=1}^n Y_i \\ \hat{\beta}_0 \sum_{i=1}^n X_{i,j} + \hat{\beta}_1 \sum_{i=1}^n X_{i,j} X_{i,1} + \hat{\beta}_2 \sum_{i=1}^n X_{i,j} X_{i,2} \\ + \cdots + \hat{\beta}_k \sum_{i=1}^n X_{i,j} X_{i,k} = \sum_{i=1}^n X_{i,j} Y_i \quad (j = 1, 2, \cdots, k) \end{cases}$$

这$k+1$个方程组成的线性方程组可表示为以下矩阵形式：

$$\begin{pmatrix} n & \sum X_{i,1} & \cdots & \sum X_{i,k} \\ \sum X_{i,1} & \sum X_{i,1}^2 & \cdots & \sum X_{i,1} X_{i,k} \\ \vdots & \vdots & & \vdots \\ \sum X_{ik} & \sum X_{ik} X_{i1} & \cdots & \sum X_{i,k}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{1,1} & X_{2,1} & \cdots & X_{n,1} \\ \vdots & \vdots & & \vdots \\ X_{1,k} & X_{2,k} & \cdots & X_{nk} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

新增一个样本点后，正规方程组变为

$$\begin{cases} (n+\underline{1})\hat{\beta}_0+\hat{\beta}_1\left[(\sum_{i=1}^{n}X_{i,1})+\underline{X_{(n+1),1}}\right]+\hat{\beta}_2\left[(\sum_{i=1}^{n}X_{i,2})+\underline{X_{(n+1),2}}\right] \\ +\cdots+\hat{\beta}_k\left[(\sum_{i=1}^{n}X_{i,k})+\underline{X_{(n+1),k}}\right]=\sum_{i=1}^{n}Y_i+\underline{Y_{n+1}} \\ \hat{\beta}_0\left[(\sum_{i=1}^{n}X_{i,j})+\underline{X_{(n+1),j}}\right]+\hat{\beta}_1\left[(\sum_{i=1}^{n}X_{i,j}X_{i,1})+\underline{X_{(n+1),j}X_{(n+1),1}}\right] \\ +\hat{\beta}_2\left[(\sum_{i=1}^{n}X_{i,j}X_{i,2})+\underline{X_{(n+1),j}X_{(n+1),2}}\right]+\cdots+\hat{\beta}_k\left[(\sum_{i=1}^{n}X_{i,j}X_{i,k})+\underline{X_{(n+1),j}X_{(n+1),k}}\right] \\ =\sum_{i=1}^{n}X_{ij}Y_i+\underline{X_{(n+1),j}Y_{n+1}} \end{cases}$$

同理，这$k+1$个方程组成的线性方程组可表示为矩阵形式：

$$(\boldsymbol{X'X})\,\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}=\boldsymbol{X'Y}$$

其中，$(\boldsymbol{X'X})$为

$$\begin{pmatrix} n+\underline{1} & (\sum_{i=1}^{n}X_{i,1})+\underline{X_{(n+1),1}} & \cdots & (\sum_{i=1}^{n}X_{i,k})+\underline{X_{(n+1),k}} \\ (\sum_{i=1}^{n}X_{i,1})+\underline{X_{(n+1),1}} & \sum X_{i,1}^2+\underline{X_{(n+1),1}^2} & \cdots & \sum X_{i,1}X_{i,k}+\underline{X_{(n+1),1}X_{(n+1),k}} \\ \vdots & \vdots & & \vdots \\ (\sum_{i=1}^{n}X_{i,k})+\underline{X_{(n+1),k}} & \sum X_{i,k}X_{i,1}+\underline{X_{(n+1),k}X_{(n+1),1}} & \cdots & \sum X_{i,k}^2+\underline{X_{(n+1),k}^2} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots\\ \hat{\beta}_k\end{pmatrix},\boldsymbol{X'Y}=\begin{pmatrix} 1 & 1 & \cdots & 1 & \underline{1} \\ X_{1,1} & X_{2,1} & \cdots & X_{n,1} & \underline{X_{(n+1),1}} \\ \vdots & \vdots & & \vdots & \vdots \\ X_{1,k} & X_{2,k} & \cdots & X_{n,k} & \underline{X_{(n+1),k}} \end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\\ \underline{Y_{(n+1)}}\end{pmatrix}$$

显然，增加一个样本点后方程组的系数矩阵$(\boldsymbol{X'X})$ 和增广矩阵$\boldsymbol{X'Y}$与之前均不同，因此方程的解$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$和$\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$也不一定相等。

（2）若$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}=\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}$，则两组正规方程组中的系数$\beta_k$均相等。将两组正规方程组相相减得

$$\begin{cases} \hat{\beta}_0+\hat{\beta}_1X_{(n+1),1}+\hat{\beta}_2X_{(n+1),2}+\cdots+\hat{\beta}_kX_{(n+1),k}=Y_{n+1} \\ \hat{\beta}_0X_{(n+1),j}+\hat{\beta}_1X_{(n+1),j}X_{(n+1),1}+\hat{\beta}_2X_{(n+1),j}X_{(n+1),2} \\ +\cdots+\hat{\beta}_kX_{(n+1),j}X_{(n+1),k}=X_{(n+1),j}Y_{n+1}\quad(j=1,2,\cdots,k) \end{cases}$$

若$X_{(n+1),j}=0$，虽然增加前后的$\boldsymbol{X'Y}$结果未发生变化，但系数矩阵$(\boldsymbol{X'X})$第一行第一列的元素从n变为$n+1$，因此这种情况下$\boldsymbol{\hat{\beta}_n} \neq \boldsymbol{\hat{\beta}_{n+1}}$，因此$X_{(n+1),j}\neq 0$。将上述方程组的第2至$k+1$个等式左右两边同除$X_{(n+1),j}$得

$$\hat{\beta}_0+\hat{\beta}_1X_{(n+1),1}+\hat{\beta}_2X_{(n+1),2}+\cdots+\hat{\beta}_kX_{(n+1),k}=Y_{n+1}$$

显然，这也与上述方程组的第1个等式相同。因此，若$\boldsymbol{\hat{\beta}_n}=\boldsymbol{\hat{\beta}_{n+1}}$，则必然意味着新样本点$(X_{(n+1),j},Y_{(n+1)})$落在参数为$\boldsymbol{\hat{\beta}_n}$的样本回归线上。

（3）若第n个样本点在样本回归线上，则其对$\boldsymbol{\hat{\beta}_n}$无影响；若第$n$个样本点在在样本回归线上侧，则样本回归线向上移动，靠近该样本点；反之同理。

2、 在经典线性回归条件下，设$Y_i=\beta_0+\beta_1X_{i,1}+\beta_2X_{i,2}+\cdots+\beta_kX_{i,k}+\mu_i \quad (i=1,2,\cdots,n)$，采用OLS 估计参数。模型的拟合值为$\hat{Y}_i$，设定拟合值与样本观测值的相关系数可以表示为：

$$\rho=\frac{\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})}{\sqrt{\sum(Y_i-\bar{Y})^2\sum(\hat{Y}_i-\bar{Y})^2}}$$

请推导ρ和拟合优度$R^2=1-\frac{RSS}{TSS}$的函数关系。

答:

欲证明

$$\rho^2=\frac{(\sum y_i\hat{y}_i)^2}{\sum y_i^2\sum\hat{y}_i^2}=R^2=1-\frac{RSS}{TSS}=\frac{ESS}{TSS}=\frac{\sum\hat{y}_i^2}{\sum y_i^2}$$

即需证明

$$\frac{(\sum y_i\hat{y}_i)^2}{\sum y_i^2\sum\hat{y}_i^2}=\frac{\sum\hat{y}_i^2}{\sum y_i^2}\Rightarrow\sum y_i\hat{y}_i=\sum\hat{y}_i^2$$

其中

$$\sum y_i\hat{y}_i=\sum(\hat{y}_i+e_i)\hat{y}_i=\sum\hat{y}_i^2+\sum e_i\hat{y}_i$$

而

$$\sum e_i\hat{y}_i=\sum e_i(\hat{Y}_i)-\bar{Y}=\sum e_i(\hat{\beta}_0+\hat{\beta}_1X_{i,1}+\cdots+\hat{\beta}_kX_{i,k}-\bar{Y})$$

根据正规方程组可知$\sum e_i = 0$，$\sum e_i X_{i,j} = 0$，因此$\sum e_i \hat{y}_i = 0$。所以

$$\sum y_i \hat{y}_i = \sum \hat{y}_i^2$$

得证

$$\rho^2 = R^2$$

3、 将拟合优度定义为：$R^2 = 1 - \frac{RSS}{TSS}$。试问：

（1）R^2一定是在0到1之间吗？如果是，为什么？

（2）如果R^2不一定在0到1之间，请写出相关计量经济学模型表达式，并从理论上回答为什么？

答：

（1）R^2不一定在0到1之间。

（2）无截距项的回归模型的R^2不一定在0到1之间。

$$\begin{aligned}
R^2 &= 1 - \frac{RSS}{TSS} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} = \frac{\sum (Y_i - \bar{Y})^2 - \sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \\
&= \frac{(\sum Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 - \sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \\
&= \frac{\sum \left(Y_i - \hat{Y}_i\right)^2 + 2\sum \left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \bar{Y}\right) + \sum \left(\hat{Y}_i - \bar{Y}\right)^2 - \sum \left(Y_i - \hat{Y}_i\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \\
&= \frac{2\sum \left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \bar{Y}\right) + \sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} = \frac{2\sum e_i \left(\hat{Y}_i - \bar{Y}\right) + \sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \\
&= \frac{2\sum e_i \left(\hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,2} + \cdots + \hat{\beta}_k x_{i,k} - \bar{Y}\right) + \sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \\
&= \frac{2\sum_{j=1}^{k} \hat{\beta}_j \sum_{i=1}^{n} e_i X_{i,j} - 2\bar{Y}\sum e_i + \sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}
\end{aligned}$$

其中，仅有$\sum e_i X_{i,j} = 0$，因此

$$R^2 = \frac{-2\bar{Y}\sum e_i + \sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

当$2\bar{Y}\sum e_i > \sum(\hat{Y}_i - \bar{Y})^2$时，$R^2 < 0$; 当$\sum e_i(Y_i + \hat{Y}_i) < 0$时，$R^2 > 1$。

4、 在满足古典假设情况下，设定计量经济学模型：

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_k X_{i,k} + \mu_i \quad (i = 1, 2, \cdots, n)$$

如果参数$|\beta_2| > |\beta_1|$，那么X_2对于Y的影响大于X_1对于Y的影响，这种说法正确吗？为什么？请从理论上加以回答。

答：

不正确。$|\beta_j| = \left|\frac{\partial y}{\partial x_{i,j}}\right|$代表在控制了其他解释变量后，$X_{i,j}$对$Y_i$的直接影响，而并未考虑其他间接影响。也即某个变量可能通过其他变量间接对被解释变量产生影响，而这是无法单独通过参数β来表示的。另外，两个变量可能存在量纲不同的问题，不同的度量单位会导致β成比例变化。