内外学生化残差

在多元回归分析中，寻找"异常值"一般思路是：如果预测值$\hat{y}_i$与真实值y_i的残差e_i过大/过小，我们就认为$\mathrm{X_i}$是异常值。那么怎么衡量残差e_i是否过大/过小呢？

一个自然的想法时通过构建统计量，获取e_i的分布，从而进行假设检验，当p值小于给定的范围时，我们认为e_i可能是异常的，即$\mathrm{X_i}$是异常值。

我们知道，当给定X时，

$$\begin{gathered}\mathrm{E}(\mathrm{e})=\mathrm{ME}(\epsilon)=0\\ var(\mathrm{e})=var(\mathrm{M}\epsilon)=\mathrm{E}(\mathrm{M}\epsilon\epsilon^{\mathrm{T}}\mathrm{M}^{\mathrm{T}})=\sigma^2\mathrm{M}\equiv\sigma^2(\mathrm{I}_N-\mathrm{H})\end{gathered}$$

可知

$$\begin{gathered}var(e_i)=\sigma^2(1-\mathrm{H}_{ii})\\ \frac{e_i}{\sqrt{\sigma^2(1-\mathrm{H}_{ii})}}\sim N(0,1)\end{gathered}$$

由于我们不知道总体方差σ^2，我们需要想一个方式将σ^2消去。考虑到

$$\frac{\mathrm{e}^{\mathrm{T}}\mathrm{e}}{\sigma^2}=\left(\frac{\epsilon}{\sigma}\right)^{\mathrm{T}}\mathrm{M}\left(\frac{\epsilon}{\sigma}\right)\sim\chi^2(N-K)$$

我们就可以构造t统计量：

$$r_i\equiv\frac{e_i}{\sqrt{\sigma^2(1-\mathrm{H}_{ii})}}/\sqrt{\frac{\mathrm{e}^{\mathrm{T}}\mathrm{e}}{\sigma^2}/(N-K)}=\frac{e_i}{\sqrt{\frac{\mathrm{e}^{\mathrm{T}}\mathrm{e}}{N-K}(1-\mathrm{H}_{ii})}}\overset{a}{\sim}t(N-K)$$

该渐进t统计量r_i被称为学生化内残差。

注意r_i并不严格地服从t分布，因为$\frac{e_i}{\sqrt{\sigma^2(1-\mathrm{H}_{ii})}}$和$\frac{\mathrm{e}^{\mathrm{T}}\mathrm{e}}{\sigma^2}$不是独立的（$e_i$包含在e内）。$r_i$的精确分布是（见陈希孺《近代回归分析》P100）：

$$\frac{r_i^2}{N-K}\sim B\left(\frac{1}{2},\frac{N-K-1}{2}\right)$$

为了构造更严格的统计量，我们考虑从$\mathrm{e}^{\mathrm{T}}\mathrm{e}\equiv\sum_{i=1}^{N}e_i^2$中剔除$e_i$的影响。我们从数据集$\mathrm{X},\mathrm{y}$中剔除$\mathrm{X}_i,y_i$，得到含$N-1$个观测的新数据集$\mathrm{X}_{(i)},\mathrm{y}_{(i)}$，将$\mathrm{y}_{(i)}$对$\mathrm{X}_{(i)}$做回归，可以得到$N-1$个新残差的和：

$$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}=\sum_{j=1,j\neq i}^{N}e_j^2$$

同理有

$$\frac{\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}}{\sigma^2}\sim\chi^2((N-1)-K)$$

构造出新的t统计量：

$$t_i \equiv \frac{e_i}{\sqrt{\sigma^2(1-\mathrm{H}_{ii})}} \Big/ \sqrt{\frac{\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}}{\sigma^2}/(N-1-K)} = \frac{e_i}{\sqrt{\frac{\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}}{N-K-1}(1-\mathrm{H}_{ii})}} \overset{a}{\sim} t(N-K-1)$$

该渐进t统计量t_i被称为学生化（外）残差。遗憾的是，虽然我们简单地从分母中剔除了e_i，但是t_i仍然不是严格服从$t(N-K-1)$，当然，这已经比r_i的分布要更接近了，在实际应用中用t_i作为t统计量是可以的。

接下来我们从代数方面对$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}$做一些简化。我们的目标是：尽量用第一次回归的信息去表示$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(i)}$，不做二次回归，从而节省计算复杂度。

由定义知

$$\mathrm{X} = \begin{bmatrix} \mathrm{X}_{(i)} \\ \mathrm{X}_i^{\mathrm{T}} \end{bmatrix} \quad \mathrm{y} = \begin{bmatrix} \mathrm{y}_{(i)} \\ y_i \end{bmatrix}$$

$$\mathrm{e}^{\mathrm{T}}\mathrm{e} = \mathrm{y}^{\mathrm{T}}\mathrm{y} - \hat{\beta}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{X}\hat{\beta}$$

$$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(\mathrm{i})} = \mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(\mathrm{i})} - \hat{\beta}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}\hat{\beta}_{(i)}$$

$$\hat{\beta} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

$$\hat{\beta}_{(i)} = (\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)}$$

为了用e表示$\mathrm{e}_{(i)}$，我们需要用β表示$\hat{\beta}_{(i)}$。由矩阵之和的求逆公式知

$$\begin{aligned}
(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} &= \left(\begin{bmatrix} \mathrm{X}_{(i)}^{\mathrm{T}} & \mathrm{X}_i \end{bmatrix} \begin{bmatrix} \mathrm{X}_{(i)} \\ \mathrm{X}_i^{\mathrm{T}} \end{bmatrix} \right)^{-1} = (\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)} + \mathrm{X}_i\mathrm{X}_i^{\mathrm{T}})^{-1} \\
&= (\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1} - \frac{(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}}{1+\mathrm{X}_i(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i^{\mathrm{T}}}
\end{aligned}$$

为简便起见，记矩阵$(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1} \equiv \mathrm{Q}^{-1}$，标量$\mathrm{X}_i(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i^{\mathrm{T}} = \mathrm{X}_i\mathrm{Q}^{-1}\mathrm{X}_i^{\mathrm{T}} \equiv m$，则有

$$\begin{aligned}
\hat{\beta} &= \left(\mathrm{Q}^{-1} - \frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}}{1+m} \right) \begin{bmatrix} \mathrm{X}_{(i)}^{\mathrm{T}} & \mathrm{X}_i \end{bmatrix} \begin{bmatrix} \mathrm{y}_{(i)} \\ y_i \end{bmatrix} \\
&= \left(\mathrm{Q}^{-1} - \frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}}{1+m} \right) (\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)} + \mathrm{X}_i y_i) \\
&= \mathrm{Q}^{-1}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)} + \mathrm{Q}^{-1}\mathrm{X}_i y_i - \frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)}}{1+m} - \frac{\mathrm{Q}^{-1}\mathrm{X}_i(\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i)y_i}{1+m}
\end{aligned}$$

注意到

$$\mathrm{Q}^{-1}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)} = (\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)} \equiv \hat{\beta}_{(i)}$$

于是

$$\begin{aligned}
\hat{\beta} &= \hat{\beta}_{(i)} + \mathrm{Q}^{-1}\mathrm{X}_i y_i - \frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}}{1+m} - \frac{m\mathrm{Q}^{-1}\mathrm{X}_i y_i}{1+m} \\
&= \hat{\beta}_{(i)} + \frac{\mathrm{Q}^{-1}\mathrm{X}_i y_i}{1+m} - \frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}}{1+m} \\
&= \hat{\beta}_{(i)} + \frac{\mathrm{Q}^{-1}\mathrm{X}_i(y_i - \mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)})}{1+m}
\end{aligned}$$

我们需要进一步将$y_i - \mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}$用第一次回归的元素表示：

$$
\begin{aligned}
y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)} &= y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}+\mathrm{X}_i^{\mathrm{T}}(\hat{\beta}-\hat{\beta}_{(i)}) \\
&= e_i+\frac{\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i(y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)})}{1+m} \\
&= e_i+\frac{m(y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)})}{1+m}
\end{aligned}
$$

直接解出$y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}$整体：

$$
y_i-\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}=e_i(1+m)
$$

代入上式得：

$$
\hat{\beta}=\hat{\beta}_{(i)}+e_i\mathrm{Q}^{-1}\mathrm{X}_i
$$

将$\hat{\beta}$和$\hat{\beta}_{(i)}$的关系代入$\mathrm{e}^{\mathrm{T}}\mathrm{e}$：

$$
\begin{aligned}
\mathrm{e}^{\mathrm{T}}\mathrm{e} &= \mathrm{y}^{\mathrm{T}}\mathrm{y}-\hat{\beta}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{X}\hat{\beta} \\
&= \begin{bmatrix}\mathrm{y}_{(i)}^{\mathrm{T}} & y_i\end{bmatrix}\begin{bmatrix}\mathrm{y}_{(i)} \\ y_i\end{bmatrix}-\hat{\beta}^{\mathrm{T}}\begin{bmatrix}\mathrm{X}_{(i)}^{\mathrm{T}} & \mathrm{X}_i\end{bmatrix}\begin{bmatrix}\mathrm{X}_{(i)} \\ \mathrm{X}_i^{\mathrm{T}}\end{bmatrix}\hat{\beta} \\
&= \mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)}+y_i^2-\hat{\beta}^{\mathrm{T}}(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}+\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}})\hat{\beta} \\
&= \mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)}+(\mathrm{X}_i^{\mathrm{T}}\hat{\beta}+e_i)^2-\hat{\beta}^{\mathrm{T}}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}\hat{\beta}-\hat{\beta}^{\mathrm{T}}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta} \\
&= \mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(i)}+e_i^2+2e_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}-\hat{\beta}^{\mathrm{T}}\mathrm{Q}\hat{\beta}
\end{aligned}
$$

将$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(\mathrm{i})}$展开：

$$
\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(\mathrm{i})}=\mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(\mathrm{i})}-\hat{\beta}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)}\hat{\beta}_{(i)}=\mathrm{y}_{(i)}^{\mathrm{T}}\mathrm{y}_{(\mathrm{i})}-\hat{\beta}_{(i)}^{\mathrm{T}}\mathrm{Q}\hat{\beta}_{(i)}
$$

于是

$$
\begin{aligned}
\Delta e &\equiv \mathrm{e}^{\mathrm{T}}\mathrm{e}-\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(\mathrm{i})} \\
&= e_i^2+2e_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}-\hat{\beta}^{\mathrm{T}}\mathrm{Q}\hat{\beta}+\hat{\beta}_{(i)}^{\mathrm{T}}\mathrm{Q}\hat{\beta}_{(i)} \\
&= e_i^2+2e_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}+2e_i^2\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i-(\hat{\beta}_{(i)}+e_i\mathrm{Q}^{-1}\mathrm{X}_i)^{\mathrm{T}}\mathrm{Q}(\hat{\beta}_{(i)}+e_i\mathrm{Q}^{-1}\mathrm{X}_i)+\hat{\beta}_{(i)}^{\mathrm{T}}\mathrm{Q}\hat{\beta}_{(i)} \\
&= e_i^2+2e_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}+2me_i^2-2e_i\mathrm{X}_i^{\mathrm{T}}\hat{\beta}_{(i)}-e_i^2\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i \\
&= (1+m)e_i^2 \\
&\equiv (1+\mathrm{X}_i(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i^{\mathrm{T}})e_i^2
\end{aligned}
$$

说明从数据集中去掉一个观测之后，单准减掉该观测的残差会高估目前回归的残差（去掉一个观测之后，回归会拟合地更好，导致另外$N-1$个残差也有不同程度的下降，这也是t_i分子分母不独立的原因）。

由于我们希望仅由第一次回归的数据就计算出两次残差的差值，于是我们将$(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}$和$(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}$联系起来：

$$
\begin{aligned}
\mathrm{H} &= \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}} \\
&= \begin{bmatrix}\mathrm{X}_{(i)} \\ \mathrm{X}_i^{\mathrm{T}}\end{bmatrix}\left((\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}-\frac{(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}}{1+\mathrm{X}_i(\mathrm{X}_{(i)}^{\mathrm{T}}\mathrm{X}_{(i)})^{-1}\mathrm{X}_i^{\mathrm{T}}}\right)^{-1}\begin{bmatrix}\mathrm{X}_{(i)}^{\mathrm{T}} & \mathrm{X}_i\end{bmatrix} \\
&= \begin{bmatrix}\mathrm{X}_{(i)} \\ \mathrm{X}_i^{\mathrm{T}}\end{bmatrix}\left(\mathrm{Q}^{-1}-\frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}}{1+m}\right)\begin{bmatrix}\mathrm{X}_{(i)}^{\mathrm{T}} & \mathrm{X}_i\end{bmatrix}
\end{aligned}
$$

可知

$$\begin{aligned}
\mathrm{H}_{ii} &= \mathrm{X}_i^{\mathrm{T}}\left(\mathrm{Q}^{-1}-\frac{\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}}{1+m}\right)\mathrm{X}_i \\
&= m-\frac{\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}\mathrm{Q}^{-1}\mathrm{X}_i}{1+m} \\
&= m-\frac{m^2}{1+m} \\
&= \frac{m}{1+m}
\end{aligned}$$

于是我们可以得到

$$\mathrm{e}_{(i)}^{\mathrm{T}}\mathrm{e}_{(\mathrm{i})}=\mathrm{e}^{\mathrm{T}}\mathrm{e}-\frac{e_i^2}{1-\mathrm{H}_{ii}}$$

代入t_i的表达式：

$$t_i \equiv \frac{e_i}{\sqrt{1-\mathrm{H}_{ii}}}\Big/\sqrt{\left(\mathrm{e}^{\mathrm{T}}\mathrm{e}-\frac{e_i^2}{1-\mathrm{H}_{ii}}\right)/(N-1-K)} \overset{a}{\sim} t(N-K-1)$$

如此，我们在计算学生化（外）残差时，不需要额外的计算，只需要原有的H_{ii}即可。