应用数理统计

Ch5 回归分析

一元线性回归中统计量的分布

2014年6月25日

5.2.3 一元线性回归中统计量的分布

------对回归估计进行统计推断

主要结论:

$$\boxed{\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{L_{xx}}\right)}$$

$$\boxed{\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}\right)\sigma^2\right)}$$

$$\boxed{\frac{SS_e}{\sigma^2} \sim \chi^2(n-2)}$$

$$\boxed{\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}}} \sim t(n-2)}$$

$$\boxed{\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}} \sim t(n-2)}$$

定理1 $E(\hat{\beta}_1)=\beta_1$，$D(\hat{\beta}_1)=\dfrac{\sigma^2}{L_{xx}}$

证： 因为$y_i \sim N(\beta_0+\beta_1 x_i,\sigma^2)$ ，$i=1,2,\cdots,n$ ，$y_1,y_2,\cdots,y_n$ 相互独立，
所以$E(y_i)=\beta_0+\beta_1 x_i$ ，$D(y_i)=\sigma^2$，$i=1,2,\cdots,n$

$$E(\bar{y})=E(\frac{1}{n}\sum_{i=1}^{n}y_i)=\frac{1}{n}\sum_{i=1}^{n}E(y_i)=\frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1 x_i)=\beta_0+\beta_1\bar{x}$$

$$D(\bar{y})=D(\frac{1}{n}\sum_{i=1}^{n}y_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(y_i)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{\sigma^2}{n}$$

$$E(\hat{\beta}_1)=E(\frac{L_{xy}}{L_{xx}})=E[\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{L_{xx}}]=\frac{\sum_{i=1}^{n}(x_i-\bar{x})[E(y_i)-E(\bar{y})]}{L_{xx}}$$

$$=\frac{\sum_{i=1}^{n}(x_i-\overline{x})[(\beta_0+\beta_1 x_i)-(\beta_0+\beta_1\overline{x})]}{L_{xx}}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})\beta_1(x_i-\overline{x})}{L_{xx}}=\frac{\beta_1 L_{xx}}{L_{xx}}$$

即 $\hat{\beta}_1$是 β_1的 无 偏 估 计

$$D(\hat{\beta}_1)=D(\frac{L_{xy}}{L_{xx}})=D[\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{L_{xx}}]=D[\frac{\sum_{i=1}^{n}(x_i-\overline{x})y_i-\overline{y}\sum_{i=1}^{n}(x_i-\overline{x})}{L_{xx}}$$

$$=D[\frac{\sum_{i=1}^{n}(x_i-\overline{x})y_i}{L_{xx}}]=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2 D(y_i)}{L_{xx}^2}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\sigma^2}{L_{xx}^2}=\frac{\sigma^2 L_{xx}}{L_{xx}^2}=\frac{\sigma^2}{L_{xx}}$$

定理2 $\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = 0$

证： 由于$y_1, y_2, \cdots, y_n$相互独立，所以$\mathrm{Cov}(y_i, y_j) = \begin{cases} D(y_i) & i = j \\ 0 & i \neq j \end{cases}$，因此

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = \mathrm{Cov}\left(\frac{\sum_{i=1}^{n} y_i}{n}, \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{L_{xx}}\right) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})D(y_i)}{nL_{xx}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})\sigma^2}{nL_{xx}} = \frac{\left(\sum_{i=1}^{n} x_i - n\bar{x}\right)\sigma^2}{nL_{xx}} = 0$$

定理3 $E(\hat{\beta}_0)=\beta_0$，$D(\hat{\beta}_0)=\left(\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}\right)\sigma^2$

证： $E(\hat{\beta}_0)=E(\overline{y}-\hat{\beta}_1\overline{x})=E(\overline{y})-E(\hat{\beta}_1)\overline{x}=(\beta_0+\beta_1\overline{x})-\beta_1\overline{x}=\beta_0$

即$\hat{\beta}_0$是β_0的无偏估计

$$D(\hat{\beta}_0)=D(\overline{y}-\hat{\beta}_1\overline{x})=\mathrm{Cov}(\overline{y}-\hat{\beta}_1\overline{x}\,,\,\overline{y}-\hat{\beta}_1\overline{x})$$

$$=\mathrm{Cov}(\overline{y}\,,\,\overline{y})-2\mathrm{Cov}(\overline{y}\,,\,\hat{\beta}_1)\overline{x}+\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_1)\overline{x}^2=D(\overline{y})-0+D(\hat{\beta}_1)\overline{x}^2$$

$$=\frac{\sigma^2}{n}+\frac{\sigma^2}{L_{xx}}\overline{x}^2=\left(\frac{1}{n}+\frac{\overline{x}^2}{L_{xx}}\right)\sigma^2$$

定理4 $\hat{\beta}_1 \sim N(\beta_1, \dfrac{\sigma^2}{L_{xx}})$，$\hat{\beta}_0 \sim N(\beta_0, \left(\dfrac{1}{n} + \dfrac{\overline{x}^2}{L_{xx}}\right)\sigma^2)$

证：因为$\hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}} = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{L_{xx}}$和$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$都是$y_1, y_2, \cdots, y_n$的线性函数，而$y_1, y_2, \cdots, y_n$相互独立，都服从正态分布，所以$\hat{\beta}_1$和$\hat{\beta}_0$也都服从正态分布。

由定理1可知，$E(\hat{\beta}_1) = \beta_1$，$D(\hat{\beta}_1) = \dfrac{\sigma^2}{L_{xx}}$，因此$\hat{\beta}_1 \sim N(\beta_1, \dfrac{\sigma^2}{L_{xx}})$

由定理3可知，$E(\hat{\beta}_0) = \beta_0$，$D(\hat{\beta}_0) = \left(\dfrac{1}{n} + \dfrac{\overline{x}^2}{L_{xx}}\right)\sigma^2$，因此

$$\hat{\beta}_0 \sim N(\beta_0, \left(\frac{1}{n} + \frac{\overline{x}^2}{L_{xx}}\right)\sigma^2)$$

定理5 $\frac{SS_e}{\sigma^2} \sim \chi^2(n-2)$，而且$SS_e, \hat{\beta}_1, \bar{y}$相互独立

证： 因为$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$，$i = 1, 2, \cdots, n$ ，所以

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i, \quad i = 1, 2, \cdots, n$$

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = \bar{y} - \beta_0 - \beta_1 \bar{x}$$

$$\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2 = \sum_{i=1}^{n}[(y_i - \beta_0 - \beta_1 x_i) - (\bar{y} - \beta_0 - \beta_1 \bar{x})]^2 = \sum_{i=1}^{n}[(y_i - \bar{y}) - \beta_1(x_i - \bar{x})]^2$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\beta_1 \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) + \beta_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= L_{yy} - 2\beta_1 L_{xy} + \beta_1^2 L_{xx} = L_{yy} - \hat{\beta}_1^2 L_{xx} + \hat{\beta}_1^2 L_{xx} - 2\beta_1 \hat{\beta}_1 L_{xx} + \beta_1^2 L_{xx}$$

$$= (L_{yy} - \hat{\beta}_1^2 L_{xx}) + (\hat{\beta}_1^2 - 2\beta_1 \hat{\beta}_1 + \beta_1^2) L_{xx}$$

$$= SS_e + (\hat{\beta}_1 - \beta_1)^2 L_{xx}$$

因为$\varepsilon_i \sim N(0,\sigma^2)$，$i=1,2,\cdots,n$，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$相互独立，所以

$\dfrac{\varepsilon_i}{\sigma} \sim N(0,1)$，$i=1,2,\cdots,n$ ，$\dfrac{\varepsilon_1}{\sigma},\dfrac{\varepsilon_2}{\sigma},\cdots,\dfrac{\varepsilon_n}{\sigma}$相互独立

$$\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2=\frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(\varepsilon_i-\bar{\varepsilon})^2+n\bar{\varepsilon}^2}{\sigma^2}=\frac{SS_e+(\hat{\beta}_1-\beta_1)^2L_{xx}+n\bar{\varepsilon}^2}{\sigma^2}$$

$$=\frac{SS_e}{\sigma^2}+\frac{(\hat{\beta}_1-\beta_1)^2L_{xx}}{\sigma^2}+\frac{n\bar{\varepsilon}^2}{\sigma^2}=Q_1+Q_2+Q_3$$

其中$Q_1=\dfrac{SS_e}{\sigma^2}=\dfrac{\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}{\sigma^2}$是$n$项的平方和，

但这n项又满足2个线性关系式

$$\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)=\sum_{i=1}^{n}y_i-n\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{n}x_i=0$$

$$\sum_{i=1}^{n}x_i(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)=\sum_{i=1}^{n}x_iy_i-\hat{\beta}_0\sum_{i=1}^{n}x_i-\hat{\beta}_1\sum_{i=1}^{n}x_i^2=0$$

（因为$\hat{\beta}_0,\hat{\beta}_1$是正规方程$\begin{cases} n\beta_0+\beta_1\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i \\ \beta_0\sum_{i=1}^{n}x_i+\beta_1\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_iy_i \end{cases}$的解。）

所以Q_1的自由度$f_1=n-2$

$$Q_2=\frac{(\hat{\beta}_1-\beta_1)^2L_{xx}}{\sigma^2}=\left(\frac{\hat{\beta}_1-\beta_1}{\sigma}\sqrt{L_{xx}}\right)^2$$是1项的平方和，

所以Q_2的自由度$f_2=1$

$$Q_3=\frac{n\bar{\varepsilon}^2}{\sigma^2}=\left(\frac{\bar{\varepsilon}\sqrt{n}}{\sigma}\right)^2$$是1项的平方和，

所以Q_3的自由度$f_3=1$

因为$f_1+f_2+f_3=(n-2)+1+1=n$所以由$Cochran$定理可知：

$$Q_1=\frac{SS_e}{\sigma^2}\sim\chi^2(n-2),\quad Q_2=\left(\frac{\hat{\beta}_1-\beta_1}{\sigma}\sqrt{L_{xx}}\right)^2\sim\chi^2(1),\quad Q_3=\left(\frac{\bar{\varepsilon}\sqrt{n}}{\sigma}\right)^2\sim\chi^2(1)$$

而Q_1，Q_2，Q_3相互独立，即$SS_e, \hat{\beta}_1, \overline{y}$相互独立

$$\Rightarrow \begin{cases} E(SS_e) = E(\dfrac{SS_e}{\sigma^2}\sigma^2) = E(\dfrac{SS_e}{\sigma^2})\sigma^2 = (n-2)\sigma^2 \\ E(\sigma^2) = E(\dfrac{SS_e}{n-2}) = \dfrac{E(SS_e)}{n-2} = \dfrac{(n-2)\sigma^2}{n-2} = \sigma^2 \end{cases} ---定理6$$

即$\sigma^2 = \dfrac{SS_e}{n-2}$是$\sigma^2$的无偏估计

定理7 $\dfrac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}\sim t(n-2),\ \dfrac{\hat{\beta}_0-\beta_0}{\hat{\sigma}\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{L_{xx}}}}\sim t(n-2)$

证： 由定理4可知$\hat{\beta}_1\sim N(\beta_1,\dfrac{\sigma^2}{L_{xx}})$，即有$\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{\sigma^2/L_{xx}}}\sim N(0,1)$

又由定理5可知$\dfrac{SS_e}{\sigma^2}\sim\chi^2(n-2)$，而且$\hat{\beta}_1$与$SS_e$相互独立，

即$\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{\sigma^2/L_{xx}}}$与$\dfrac{SS_e}{\sigma^2}$相互独立

所以由t分布的定义便可推出

$$\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}=\frac{\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{\sigma^2/L_{xx}}}}{\sqrt{\dfrac{SS_e}{\sigma^2}\Big/(n-2)}}\sim t(n-2)$$

由定理5可知$\dfrac{SS_e}{\sigma^2}\sim\chi^2(n-2)$，而且$\hat{\beta}_1$，$\overline{y}$，$SS_e$相互独立，即$\hat{\beta}_0$与$SS_e$相互独立，$\dfrac{\hat{\beta}_0-\beta_0}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{L_{xx}}}}$与$\dfrac{SS_e}{\sigma^2}$相互独

所以，由t分布的定义可推出

$$\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{L_{xx}}}}=\frac{\dfrac{\hat{\beta}_0-\beta_0}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{L_{xx}}}}}{\sqrt{\dfrac{SS_e}{\sigma^2}\Big/(n-2)}}\sim t(n-2)$$

一元回归中的区间估计

$$\boxed{\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}} \sim t(\text{n-2})}$$

对于给定的置信水平$1-\alpha$ ，从t分布的分位数表可以查到$t_{1-\alpha/2}(n-2)$ ，使得

$$P\{\left|\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}\right| \le t_{1-\alpha/2}(n-2)\}=1-\alpha$$

$$P\{\hat{\beta}_1-t_{1-\alpha/2}(n-2)\frac{\hat{\sigma}}{\sqrt{L_{xx}}} \le \beta_1 \le \hat{\beta}_1+t_{1-\alpha/2}(n-2)\frac{\hat{\sigma}}{\sqrt{L_{xx}}}\}=1-\alpha$$

一元回归中的区间估计

$$\boxed{\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{L_{xx}}}} \sim t(\text{n}-2)}$$

给定的置信水平$1-\alpha$，从t分布的分位数表可以查到$t_{1-\alpha/2}(n-2)$，使得

$$P\{\left|\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{L_{xx}}}}\right| \le t_{1-\alpha/2}(n-2)\} = 1-\alpha$$

$$P\{\hat{\beta}_0 - t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{L_{xx}}} \le \beta_0 \le \hat{\beta}_0 + t_{1-\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{L_{xx}}}\} = 1-\alpha$$

即β_0和β_1置信水平为$1-\alpha$的置信区间上下限分别是：

$$\hat{\beta}_0 \pm \hat{\sigma} t_{1-\frac{\alpha}{2}}(n-2)\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}}$$

$$\hat{\beta}_1 \pm t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}/\sqrt{L_{xx}}$$

一元回归中的假设检验

检验x与y之间是否统计线性相关，相当于$H_0:\beta_1=0.$

如果H_0不真，即$\beta_1\neq 0$ ，则x与y 线性相关；如果假设H_0为真，$\beta_1=0$，则x与y无关。

检验方法一（*t*检验） $$\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}\sim t(n-2)$$

若H_0：$\beta_1=0$为真，则$\dfrac{\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}=0$，有$T=\dfrac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}}=\dfrac{\hat{\beta}_1-\beta_1}{\hat{\sigma}}\sqrt{L_{xx}}\sim t(n-2)$

从观测数据求出统计量$T=\dfrac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}}$的值，当$|T|>t_{1-\alpha/2}(n-2)$

拒绝H_0，否则接受H_0

检验方法二（*F*检验）

取一个统计量$F=T^2=(\frac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}})^2=\frac{\hat{\beta}_1^2 L_{xx}}{\hat{\sigma}^2}=\frac{L_{yy}-SS_e}{SS_e/(n-2)}$

若H_0：$\beta_1=0$为真，则有$T\sim t(n-2)$，这时$F=T^2\sim F(1,n-2)$

若H_0：$\beta_1=0$不真，则有T的绝对值会偏大，这时$F=T^2$ 的值也会偏大

因此可以得到如下检验方法：

从观测数据$F=\frac{L_{yy}-SS_e}{SS_e/(n-2)}$的值，对于给定显著性水平$\alpha$，

从F的分布表查出分为数$F_{1-\alpha}(1,n-2)$，使得$P\{F>F_{1-\alpha}(1,n-2)\}=\alpha$

将统计量F与分为数作比较，当$F>F_{1-\alpha}(1,n-2)$时拒绝H_0，否则接受H_0

在前面的例1中$n=5$，$L_{xx}=2.5$，$L_{yy}=10.173$，$\hat{\beta}_1=2.01$

$SS_e=0.07275$，$\hat{\sigma}=0.1557$。要检验H_0：$\beta_1=0$（显著性水平$\alpha=0.05$）

解： 上面介绍了两种不同的检验方法，下面分别用它们来检验一下

t分布检验：$T=\dfrac{\hat{\beta}_1}{\hat{\sigma}}\sqrt{L_{xx}}=\dfrac{2.01}{0.1557}\sqrt{2.5}=20.41$

对$\alpha=0.05$，查t分布的分位数表，可得$t_{1-\alpha/2}(n-2)=t_{0.975}(3)=3.1824$

因为$|T|=|20.41|=20.41>3.1824$，所以H_0：$\beta_1=0$，

说明自变量与因变量之间有显著的统计线性相关关系。

分布检验：$F=\dfrac{L_{yy}-SS_e}{SS_e/(n-2)}=\dfrac{10.173-0.07275}{0.07275/(5-2)}=416.5$

对$\alpha=0.05$，查F分布的分位数表，可得$F_{1-\alpha}(1,n-2)=F_{0.95}(1,3)=10.1$

因为$F=416.5>10.1$，所以结论也是拒绝H_0：$\beta_1=0$。

一元回归中的预测

点预测值即为回归方程计算所得回归值.

已知 x_0, 对应预测因变量y的取值为 y_0 则:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

实际值 与其预测值之间有预测误差 $y_0 - \hat{y}_0$

$$E(y_0 - \hat{y}_0) = 0 \qquad D(y_0 - \hat{y}_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{L_{xx}}\right)$$

证明: $y_0 - \hat{y}_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0$

$$= (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_0 + \varepsilon_0$$

$$E(y_0 - \hat{y}_0) = (\beta_0 - \beta_0) + (\beta_1 - \beta_1)x_0 + 0 = 0$$

$$y_0 - \hat{y}_0 = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_0 + \varepsilon_0$$

$$= [\beta_0 - (\overline{y} - \hat{\beta}_1 \overline{x})] + (\beta_1 - \hat{\beta}_1)x_0 + \varepsilon_0$$

$$= (\beta_0 + \beta_1 x_0) - \overline{y} - \hat{\beta}_1 (x_0 - \overline{x}) + \varepsilon_0$$

$$D(y_0 - \hat{y}_0) = 0 + \frac{\sigma^2}{n} + (x_0 - \overline{x})^2 D\hat{\beta}_1 + \sigma^2$$

$$= \sigma^2 (1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{L_{xx}})$$

$$D(y_0 - \hat{y}_0) = \sigma^2 (1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{L_{xx}})$$

易见：如果要降低 $D(y_0 - \hat{y}_0)$,可以采取如下措施

（1）增大样本容量n；

（2）增大样本中自变量的分散性　（即增大 L_{xx}）

（3）减少 x_0 与自变量样本均值 $\overline{x}$ 之间的距离。

由 $y_0 - \hat{y}_0 = (\beta_0 + \beta_1 x_0) - \overline{y} - \hat{\beta}_1 (x_0 - \overline{x}) + \varepsilon_0$

知 $y_0 - \hat{y}_0$ 也 服 从 正 态 分 布

($\overline{y}, \hat{\beta}_1, \varepsilon_0$ 独 立 ，都 服 从 正 态 分 布)

$$\frac{\boxed{N(0,1)}}{\sqrt{\boxed{\chi^2}/\boxed{\text{自由度}}}} = \frac{y_0 - \hat{y}_0\text{的标准化}}{\sqrt{\dfrac{SS_e/\sigma^2}{n-2}}} = \frac{y_0 - \hat{y}_0}{\hat{\sigma}\sqrt{1 + \dfrac{1}{n} + \dfrac{(x_0 - \overline{x})^2}{L_{xx}}}} \sim t(n-2)$$

y_0的置信水平为 1-α的置信区间的上下限为:

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{L_{xx}}}$$

预测区间说明

当样本容量充分大时，y_0 的预测区间可简化：

对于一元线性回归模型 $y=\beta_0+\beta_1 x+\varepsilon$ ，其中误差项满足正态性，独立性，及方差齐性的条件， 给定 x_0，则对应 y_0 的点估计为 $\hat{y}_0=\hat{\beta}_0+\hat{\beta}_1 x_0$ ；当 n 充分大时，y_0 置信水平为$1-\alpha$ 的置信区间可近似表示为 $[\hat{y}_0-\hat{\sigma}u_{1-\frac{\alpha}{2}}，\hat{y}_0+\hat{\sigma}u_{1-\frac{\alpha}{2}}]$