天善智能
TIANSHAN SOFT

专注于商业智能BI和大数据的垂直社区平台

一元线性回归模型的假设检验

Allen

课程目录

- 拟合优度检验
- 回归系数的显著性检验——t检验
- 回归方程的显著性检验——F检验
- 小结

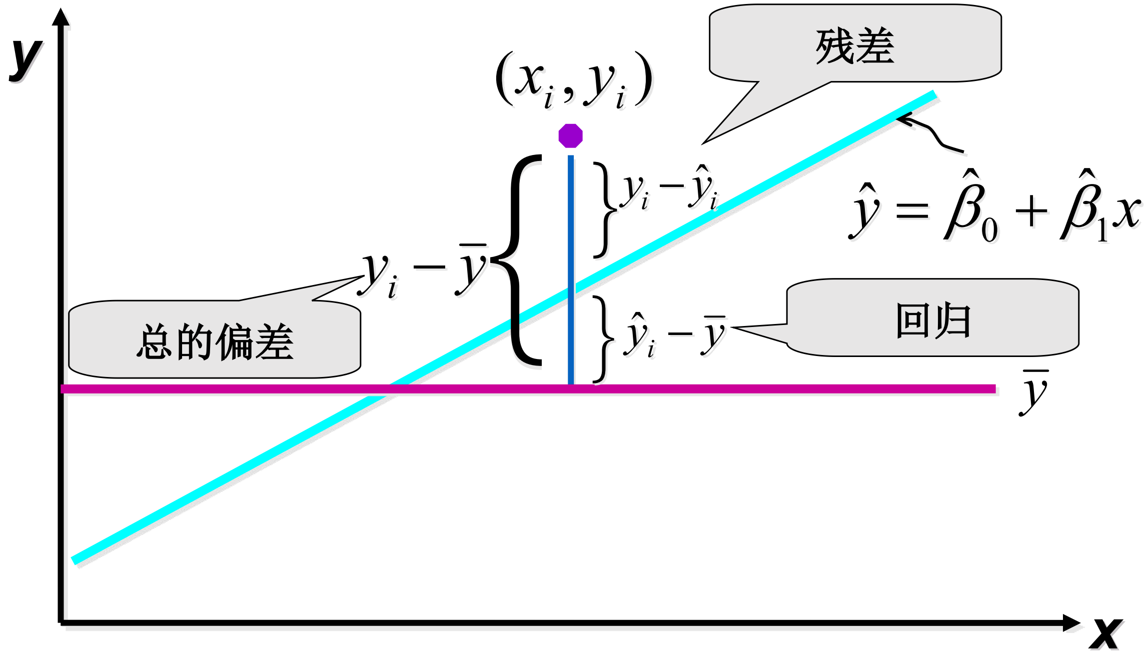
天善智能
TIANSHAN SOFT

拟合优度检验

- 拟合优度检验：对样本回归直线和样本观测值直接的拟合程度进行检验
- 度量拟合优度检验的指标：判定系数 R^2 ，也可以用总的偏差平方和来衡量数据波动的大小：$S_T = \sum (y_i - \bar{y})^2$
- 问题：采用了最小二乘估计法，已经保证了模型最好拟合观测值，为什么还要检验拟合程度？

天善智能 TIANSHAN SOFT

拟合优度检验——总的偏差平方和的分解



- 从上图中很直观地看出来自残差和来自回归的和等于总的偏差

拟合优度检验——总的偏差平方和的分解

- 将总的偏差平方和化简：

$$S_T=\sum\left(y_i-\overline{y}\right)^2 \rightarrow =\sum\left(y_i-\hat{y}_i+\hat{y}_i-\overline{y}\right)^2 \rightarrow =\sum\left(y_i-\hat{y}_i\right)^2+\sum\left(\hat{y}_i-\overline{y}\right)^2+2\sum\left(y_i-\hat{y}_i\right)\left(\hat{y}_i-\overline{y}\right)$$

- 由于：$\sum\left(y_i-\hat{y}_i\right)\left(\hat{y}_i-\overline{y}\right)=0$

- 所以：$S_T=\sum\left(y_i-\overline{y}\right)^2=\sum\left(y_i-\hat{y}_i\right)^2+\sum\left(\hat{y}_i-\overline{y}\right)^2$　记为：$S_T=S_E+S_R$

- 其中S_T称为总的偏差平方和，S_E称为残差平方和，S_R称为回归平方和

天善智能 TIANSHAN SOFT

拟合优度检验——总的偏差平方和的分解

- 总的偏差平方和S_T
  - 反映因变量的 n 个观察值与其均值的总离差

- 残差平方和S_E
  - 反映除 x 以外的其他因素对 y 取值的影响，也称为不可解释的平方和或剩余平方和

- 回归平方和S_R
  - 反映自变量 x 的变化对因变量 y 取值变化的影响，或者说，是由于 x 与 y 之间的线性关系引起的 y 的取值变化，也称为可解释的平方和

拟合优度检验——总的偏差平方和的分解

- 总的来说y的变动是由x得变动以及除x外的其他因素所引起的，如果x的变动在y的总变动中占很大比例，那么x就能很好地解释y，否则x不能很好地解释y

- 所以拟合优度的判定系数为：回归平方和/总的偏差平方和

- 即：$R^2=\frac{S_R}{S_T}=1-\frac{S_E}{S_T}$

判定系数越接近1，说明拟合优度越高，实际观测点离样本线越近

回归系数的显著性检验——t检验

- 变量的显著性检验指的是对模型中的被解释变量与某个解释变量之间的线性关系是否显著成立做出推断。为决定某个解释变量是否留在模型中提供参考依据
- 之前求得的最小二乘估计量的期望和方差，可知 $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum\left(x_i-\bar{x}\right)^2}\right)$
- 由于σ^2未知，所以使用无偏估计量 $\hat{\sigma}^2=\sum e_i^2/(n-2)$替代，构造统计量如下：

$$t=\frac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\sigma}^2 / \sum\left(x_i-\bar{x}\right)^2}}=\frac{\hat{\beta}_1-\beta_1}{S_{\hat{\beta}_1}} \sim t(n-2)$$

回归系数的显著性检验——t检验步骤

- 1.对参数提出假设： $H_0:\beta_1=0 \quad H_1:\beta_1\neq 0$
- 2.由原假设构造t检验统计量：$t=\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\sigma}^2 \big/ \sum\left(x_i-\bar{x}\right)^2}}=\dfrac{\hat{\beta}_1-\beta_1}{S_{\hat{\beta}_1}}\sim t(n-2)$
- 3.给定显著性水平 α 查t分布表，求出临界域 $t_{\alpha/2}(n-2)$
- 4.根据样本观测值计算并比较： 若 $|t|>t_{\alpha/2}(n-2)$ 则拒绝原假设

  若 $|t|\leq t_{\alpha/2}(n-2)$ 则接受原假设

天善智能 TIANSHAN SOFT

回归方程的显著性检验——F检验

- 回归方程的显著性检验指的是对模型中的被解释变量与解释变量之间的线性关系在总体上是否显著成立做出检验，即检验该模型的有关参数的总体是否显著为0

- 将回归均方同残差均方加以比较，如果比值较大，说明自变量X对因变量Y的解释程度高。应用F检验来分析二者之间的差别是否显著
  - 回归均方：回归平方和除以相应的自由度(自变量的个数p)
  - 残差均方：残差平方和除以相应的自由度(n-p-1)

天善智能 TIANSHAN SOFT

回归方程的显著性检验——F检验步骤

- 1.对参数提出假设： $H_0:\beta_1=0 \quad H_1:\beta_1\neq 0$
- 2.由原假设构造F检验统计量 :$F=\dfrac{S_R/1}{S_E/(n-2)}\sim F(1,n-2)$
- 3.给定显著性水平 α 查F分布表，求出临界域 $F_\alpha(1,n-2)$
- 4.根据样本观测值计算并比较： 若 $F>F_\alpha(1,n-2)$ 则拒绝原假设

  若 $F<F_\alpha(1,n-2)$ 则接受原假设

天善智能 TIANSHAN SOFT

小结

- 拟合优度检验
- 回归系数的显著性检验——t检验
- 回归方程的显著性检验——F检验
- 小结