第6.1节　一元线性回归分析

一、一元线性回归模型

二、未知参数的估计

三、参数估计量的分布

四、参数β的显著性检验

五、预测

0、回归分析的基本思想

变量之间的关系 → 确定性关系
变量之间的关系 → 相　关关系

$S=\pi r^2$　　确定性关系

身高和体重　　相关关系

相关关系的特征是:变量之间的关系很难用一种精确的方法表示出来.

确定性关系和相关关系的联系

由于存在测量误差等原因,确定性关系在实际问题中往往通过相关关系表示出来;另一方面,当对事物内部规律了解得更加深刻时,相关关系也有可能转化为确定性关系.

回归分析——处理变量之间的相关关系的一种数学方法,它是最常用的数理统计方法.

- 回归分析
  - 线　性回归分析
    - 一元线性回归分析
    - 多元线性回归分析
  - 非线性回归分析

一、一元线性回归的数学模型

回归分析的任务——根据试验数据估计回归函数;讨论回归函数中参数的点估计、区间估计;对回归函数中的参数或者回归函数本身进行假设检验;利用回归函数进行预测与控制等等.

问题的一般提法

对x的一组不相同的值$x_1, x_2, \cdots, x_n$，设$Y_1, Y_2, \cdots, Y_n$分别是在$x_1, x_2, \cdots, x_n$处对Y的独立观察结果.

称$(x_1, Y_1), (x_2, Y_2), \cdots, (x_n, Y_n)$是一个样本.

对应的样本值记为

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n).$$

利用样本来估计Y关于x的回归函数$\mu(x)$.

求解步骤

1.推测回归函数的形式

方法一　根据专业知识或者经验公式确定;

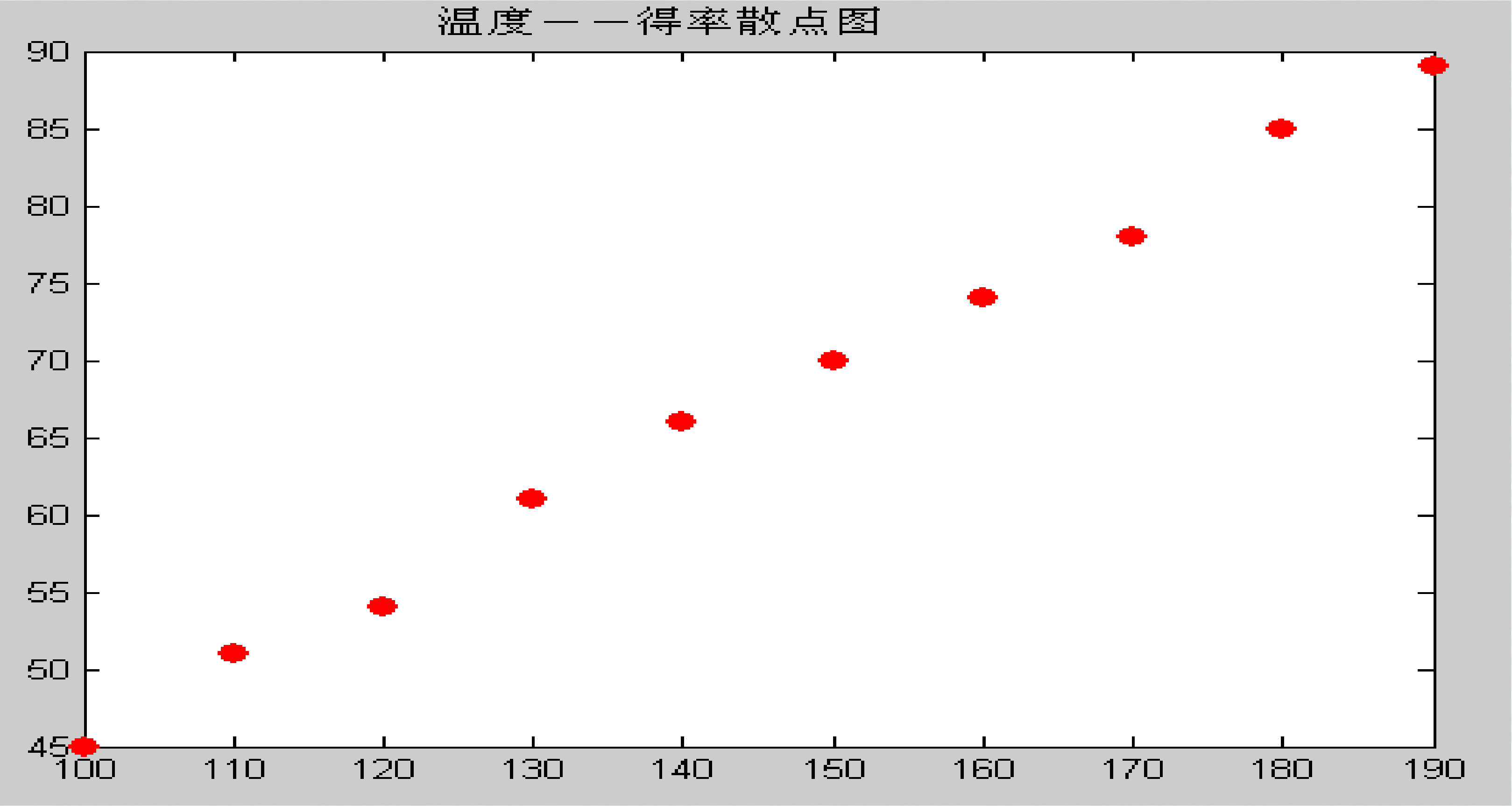
方法二　作散点图观察.

例1　为研究某一化学反应过程中,温度$x(^oC)$对产品得率$Y(\%)$的影响,测得数据如下.

温度$x(^oC)$	100	110	120	130	140	150	160	170	180	190
得率$Y(\%)$	45	51	54	61	66	70	74	78	85	89

用*MATLAB*画出散点图

```
x=100:10:190;y=[45,51,54,61,66,70,74,78,85,89];
plot(x,y,'.r')
```



观察散点图,$\mu(x)$具有线性函数$\alpha+\beta x$的形式.

2.建立回归模型

$\mu(x) = \alpha + \beta x$ 一元线性回归问题

假设对于x的每一个值有$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, α, β, σ^2都是不依赖于x的未知参数.

记$\varepsilon_i = Y_i - (\alpha + \beta x_i)$, 那么

$$\boxed{\begin{array}{l} Y_i = \underbrace{\alpha + \beta x_i}_{x\text{的线性函数}} + \underbrace{\varepsilon_i}_{\text{随机误差}}, \quad \varepsilon_i \sim N(0, \sigma^2). \\ \alpha, \beta, \sigma^2\text{是不依赖于}x\text{的未知参数}. \end{array}}$$

一元线性回归模型

二、未知参数的估计

$$Y = \alpha + \beta x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

对于样本$(Y_1, x_1), (Y_2, x_2), \cdots, (Y_n, x_n)$

$Y_i = \alpha + \beta x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$, 各$\varepsilon_i$相互独立.

于是$Y_i \sim N(\alpha + \beta x_i, \sigma^2), i = 1, 2, \cdots, n.$

1. (α,β)的最小二乘估计

使得下式成立的(α,β)称为其最小二乘估计.

$$\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = \min_{\alpha, \beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2$$

利用微分法求解下式：

设 $Q(\alpha,\beta)=\sum_{i=1}^{n}(Y_i-\alpha-\beta x_i)^2$,求偏导可得

$$\frac{\partial Q(\alpha,\beta)}{\partial \alpha}=-2\sum_{i=1}^{n}(Y_i-\alpha-\beta x_i)=0$$

$$\frac{\partial Q(\alpha,\beta)}{\partial \beta}=-2\sum_{i=1}^{n}x_i(Y_i-\alpha-\beta x_i)=0$$

$$\left.\begin{aligned}\frac{\partial Q}{\partial \alpha} &= -2\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)=0\\ \frac{\partial Q}{\partial \beta} &= -2\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)x_i=0\end{aligned}\right\}$$

$$\left.\begin{aligned}n\alpha+\left(\sum_{i=1}^{n}x_i\right)\beta &= \sum_{i=1}^{n}y_i\\ \left(\sum_{i=1}^{n}x_i\right)\alpha+\left(\sum_{i=1}^{n}x_i^2\right)\beta &= \sum_{i=1}^{n}x_iy_i\end{aligned}\right\}\text{正规方程组}$$

由于 $\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} \neq 0,$

则 $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x},\qquad \hat{\beta} = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$

其中$\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i, \overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i.$

2. (α,β)的最大似然估计

根据$Y_1, Y_2, \cdots, Y_n$的独立性可得到联合密 度函数为

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right]$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right].$$

L取最大值等价于

$$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

取最小值．这就回到了最小二乘估计的情形。

$\hat{Y} = \hat{\alpha} + \hat{\beta}x$ 称其为Y关于x的线性回归方程

由于$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$,

$$\hat{y} = \bar{y} + \hat{\beta}(x - \bar{x})$$

回归直线通过散点图的 几何中心$(\bar{x}, \bar{y})$.

注 在高斯假设下参数的最小二乘估计和最大似然估计等价

3. 未知参数σ^2的估计

$$Y=\alpha+\beta x+\varepsilon,\quad \varepsilon\sim N(0,\sigma^2).$$

$$E\{[Y-(\alpha+\beta x)]^2\}=E(\varepsilon^2)=D(\varepsilon)+[E(\varepsilon)]^2=\sigma^2$$

则$\hat{\sigma}^2$的估计为 $$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}[Y_i-(\hat{\alpha}+\hat{\beta}x_i)]^2$$

显然σ^2越小,用回归函数$\mu(x)=\alpha+\beta x$作为Y的近似导致的均方误差就越小.

$$\hat{y}_i=\hat{y}\big|_{x=x_i}=\hat{\alpha}+\hat{\beta}x_i$$

$y_i-\hat{y}_i$　　x_i处的残差

$$Q_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\hat{\alpha}-\hat{\beta}x_i)^2$$ 残差平方和

$$Q_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}[y_i-\overline{y}-\hat{\beta}(x_i-\overline{x})]^2$$

$$=\sum_{i=1}^{n}(y_i-\overline{y})^2+\hat{\beta}^2\sum_{i=1}^{n}(x_i-\overline{x})^2-2\hat{\beta}\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})$$

由于β的估计值为$\hat{\beta}=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$，则

$$Q_e=\sum_{i=1}^{n}(y_i-\overline{y})^2-\hat{\beta}^2\sum_{i=1}^{n}(x_i-\overline{x})^2$$

因而 $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y}_i)^2-\hat{\beta}^2\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$

例2(p198例6.2) 设父亲和他们长子的身高分别为 $(x_i, y_i)(i=1,2,\cdots,12)$,其观测数据为

父亲身高x	65	63	67	64	68	62	70	66	68	67	69	71
长子身高y	68	66	68	65	69	66	68	65	71	67	68	70

求Y关于x的线性回归方程

解 回归方程为$\hat{Y}=\hat{\alpha}+\hat{\beta}x$ 将观测值代入正规方程

$$\begin{cases} n\hat{\alpha}+(\sum_{i=1}^{n}x_i)\hat{\beta}=\sum_{i=1}^{n}y_i \\ (\sum_{i=1}^{n}x_i)\hat{\alpha}+(\sum_{i=1}^{n}x_i^2)\hat{\beta}=\sum_{i=1}^{n}x_iy_i \end{cases}$$

$$\begin{cases} 12\hat{\alpha} + 800\hat{\beta} = 811 \\ 800\hat{\alpha} + 53418\hat{\beta} = 54107 \end{cases}$$

求解得

$$\hat{\alpha} = 35.82 \qquad \hat{\beta} = 0.476$$

则Y关于x的线性回归方程为

$$\hat{Y} = 35.82 + 0.476x$$

这个例子表明：高个子的先代会有高个子的后代，但后代的增高并不与先代的增高等量。例如父亲身高超过祖父身高**6in**,则儿子的身高超过父亲的身高大约为**3in**

称这种现象为向平常高度的回归，回归一词即来源于此.

这种提法最早是由高登提出的，一直沿用至今.当时高登、皮尔逊、**Lee**研究了**1078**个家庭，得到的回归方程为:

$$\hat{Y} = 0.516x + 33.73$$

三、参数估计的分布

为了对参数估计量进行检验，需要讨论它们的分布

1. $\hat{\beta}$的分布

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}Y_i$$

$$=\sum_{i=1}^{n}a_iY_i$$

其中$a_i=\dfrac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$,由于$Y_i$相互独立，而且

$$Y_i \sim N(\alpha+\beta x_i, \sigma^2)$$

因而$\hat{\beta}$服从正态分布，其期望值为

$$E\hat{\beta}=\sum_{i=1}^{n} a_i EY_i=\sum_{i=1}^{n} a_i(\alpha+\beta x_i)=\beta\frac{\sum_{i=1}^{n}(x_i-\overline{x})x_i}{\sum_{i=1}^{n}(x_i-\overline{x})^2}=\beta$$

$$D\hat{\beta}=\sum_{i=1}^{n} a_i^2 DY_i=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\sigma^2}{\{\sum_{i=1}^{n}(x_i-\overline{x})^2\}^2}=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

则 $\hat{\beta} \sim N(\beta, \sigma^2 / \sum_{i=1}^{n}(x_i-\overline{x})^2)$

2. $\hat{\alpha}$的分布

$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{x}=\sum_{i=1}^{n}\left[\frac{1}{n}-\frac{(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]Y_i$$

因而$\hat{\alpha}$服从正态分布，其期望值为

$$E\hat{\alpha}=E\bar{Y}-E\hat{\beta}\bar{x}=\frac{1}{n}\sum_{i=1}^{n}(\alpha+\beta x_i)-\beta\bar{x}=\alpha$$

$$D\hat{\alpha}=\sum_{i=1}^{n}\left[\frac{1}{n}-\frac{(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2 DY_i$$

$$=\left[\frac{1}{n}+\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\,\bar{x}^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}\right]\sigma^2$$

$$=\left[\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\sigma^2$$

则 $\hat{\alpha}\sim N\left(\alpha,\left[\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\sigma^2\right)$

3. 对$x=x_0$,回归方程$\hat{Y}_0=\hat{\alpha}+\hat{\beta}x_0$的分布

$$\hat{Y}_0=\hat{\alpha}+\hat{\beta}x_0=\sum_{i=1}^{n}\left[\frac{1}{n}-\frac{(x_i-\overline{x})\overline{x}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]Y_i+\frac{\sum_{i=1}^{n}(x_i-\overline{x})x_0}{\sum_{i=1}^{n}(x_i-\overline{x})^2}Y_i$$

$$=\sum_{i=1}^{n}\left[\frac{1}{n}+\frac{(x_i-\overline{x})(x_0-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]Y_i$$

因而$\hat{Y}_0$服从正态分布，其期望值为

$$E\overline{Y}_0=E(\hat{\alpha}+\hat{\beta}x_0)=\alpha+\beta x_0$$

$$D(\hat{Y}_0)=\sum_{i=1}^{n}\left[\frac{1}{n}+\frac{(x_i-\overline{x})(x_0-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^2 DY_i$$

$$=\left[\frac{1}{n}+\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2(x_0-\overline{x})^2}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}\right]\sigma^2$$

$$=\left[\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]\sigma^2$$

则 $\hat{Y}_0 \sim N\left(\alpha+\beta x_0, \left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\sigma^2\right)$

4. $\hat{\sigma}^2$的分布

$$E\hat{\sigma}^2 = \frac{1}{n}\left[E\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - E(\hat{\beta}^2)\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}EY_i^2 - nE(\bar{Y})^2 - (D\hat{\beta}+E^2(\hat{\beta}))\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}(DY_i+(EY_i)^2)-n(D\overline{Y}+(E\overline{Y})^2)\right.$$

$$\left.-(D\hat{\beta}+E^2(\hat{\beta}))\sum_{i=1}^{n}(x_i-\overline{x})^2\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2+(\alpha+\beta x_i)^2)-n\left(\frac{\sigma^2}{n}+(\alpha+\beta\overline{x})^2\right)-\right.$$

$$\left.\left(\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}+\beta^2\right)\sum_{i=1}^{n}(x_i-\overline{x})^2\right]$$

$$=\frac{1}{n}\left[(n-1)\sigma^2+\beta^2\sum_{i=1}^{n}(x_i-\overline{x})^2-\sigma^2-\beta^2\sum_{i=1}^{n}(x_i-\overline{x})^2\right]$$

$$=\frac{1}{n}(n-2)\sigma^2=\frac{(n-2)}{n}\sigma^2$$

设 $$\hat{\sigma}^{*2}=\frac{n}{n-2}\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{\alpha}-\hat{\beta}x_i)^2$$

$$=\frac{1}{n-2}\sum_{i=1}^{n}(Y_i-\overline{Y}_i)^2-\hat{\beta}^2\frac{1}{n-2}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

则 $E\hat{\sigma}^{*2}=\sigma^2$

定理6.1 假设(Y_i, x_i)满足线性回归模型的条件，则

$$\frac{(n-2)}{\sigma^2}\hat{\sigma}^{*2} \sim \chi^2(n-2)$$

而且$\hat{\sigma}^{*2}$分别与$\hat{\alpha}, \hat{\beta}$独立，其中$\hat{\sigma}^{*2}$是$\sigma^2$的修正估计.

证明参见下一节多元回归理论

四、参数β的显著性检验

根据前三小节的理论，给定一组观测值，就可以得其相应的回归方程。但是二者是否具有此种相关关系，需要进行必要的检验.

通常检验一元线性回归模型是否成立，需要检验：

(1) 给定x时，Y服从正态分布且方差相等;

(2) 对于给定的范围，EY是x的线性函数;

(3) $Y_1, Y_2, \cdots, Y_n$相互独立.

本节主要讨论第二类问题，也就是等价于β是否为0.

设 $Y=\alpha+\beta x+\varepsilon,\quad \varepsilon\sim N(0,\sigma^2).$

检验假设: $H_0:\beta=0,\quad H_1:\beta\neq 0.$

$$\hat{\beta}\sim N\left(\beta,\sigma^2\Big/\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\qquad \frac{(n-2)\hat{\sigma}^{*2}}{\sigma^2}=\frac{Q_e}{\sigma^2}\sim\chi^2(n-2).$$

并且$\hat{\beta}$,Q_e相互独立,因此

$$T=\frac{\hat{\beta}-\beta}{\hat{\sigma}^*}\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sim t(n-2).$$

当H_0为真时$\beta=0$,此时$T=\dfrac{\hat{\beta}}{\hat{\sigma}^*}\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sim t(n-2),$

则，H_0的拒绝域为

$$|t| = \frac{|\hat{\beta}|}{\hat{\sigma}^*}\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \geq t_{\alpha/2}(n-2)$$

拒绝$H_0: \beta \neq 0$,认为回归效果显著.

接受$H_0: \beta = 0$,认为回归效果不显著.

回归效果不显著的原因分析

(1)影响Y取值的,除x及随机误差外还有其他不可忽略的因素;

(2)$E(Y)$与x的关系不是线性的;

(3)Y与x不存在关系.

对于给定的显著性水平α,可构造检验步骤如下：

(1)$H_0:\beta=0$;

(2)构造检验统计量$T=\hat{\beta}\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}/\hat{\sigma}^*$;

(3)对于给定的α,查分位数$t_{\alpha/2}(n-2)$;

(4)对给定的一组回归观测值，带入检验统计量计算得t,

如果$|t|\geq t_{\alpha/2}(n-2)$，则拒绝$H_0$，否则接受$H_0$.

例3(p197例6.3)　检验例2中的回归效果是否显著，取显著性水平为0.05.

解　对$\alpha = 0.05, n-2=10$,查表得

查表得$t_{0.05/2}(n-2) = t_{0.025}(10) = 2.2281$

$$|t| = 3.128$$

$$|t| > t_{0.025}(10).$$

拒绝$H_0: \beta = 0$,认为回归效果显著.

五、预测

1. 系数β的置信区间

当回归效果显著时,对系数β作区间估计.

系数β的置信水平为$1-\alpha$的置信区间为

$$\left(\hat{\beta} \pm t_{\alpha/2}(n-2)\times\frac{\hat{\sigma}^*}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\right).$$

2. Y的观察值的点预测和预测区间

设Y_0是在$x=x_0$处对Y的观察结果.

$$\hat{Y}_0=\hat{\mu}(x_0)=\hat{\alpha}+\hat{\beta}x_0$$ 此式为Y_0的点预测

又因为 $Y_0-\hat{Y}_0=Y_0-(\hat{\alpha}+\hat{\beta}x_0)$

由于Y_0与$\hat{Y}_0$相互独立，而且都服从正态分布，因而$Y_0-\hat{Y}_0$亦服从正态分布，其期望值为

$$E(Y_0-\hat{Y}_0)=EY_0-E(\hat{\alpha}+\hat{\beta}x_0)=\alpha+\beta x_0-\alpha-\beta x_0=0$$

$$D(Y_0-\hat{Y}_0)=DY_0+D(\hat{Y}_0)$$

$$=\left[1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]\sigma^2$$

$$Y_0-\hat{Y}_0\sim N\left(0,\left(1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)\sigma^2\right)$$

$\dfrac{(n-2)\hat{\sigma}^{*2}}{\sigma^2}\sim\chi^2(n-2)$, σ^{*2}, $Y_0-\hat{Y}_0$相互独立, 则

$$T=\frac{Y_0-(\hat{\alpha}+\hat{\beta}x_0)}{\hat{\sigma}^*\sqrt{1+\dfrac{1}{n}+\dfrac{(x_0-\overline{x})^2}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}}}\sim t(n-2),$$

给定置信水平为$1-\alpha$, Y_0的预测区间为

$$\left(\hat{\alpha}+\hat{\beta}x_0\pm t_{\alpha/2}(n-2)\hat{\sigma}^*\sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}}\right)$$

设 $\delta(x_0)=t_{\alpha/2}(n-2)\hat{\sigma}^* \sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}$

于是 给定置信水平为$1-\alpha$, Y_0的预测上下限为

$$y_1(x_0)=\hat{Y}_0-\delta(x_0)$$

$$y_2(x_0)=\hat{Y}_0+\delta(x_0)$$

则这两条曲线形成带状域包含回归曲线.

例4(p198例6.4) (续例2)

(1)设$x_0=65.5, 1-\alpha=0.95$

(2)设$x_0=70.3, 1-\alpha=0.95$

试求出两种情形下，Y_0的置信上下限.

解 已知$n-2=10$, $t_{0.025}(10)=2.2281$

(1) $\hat{y}_0=35.8+0.476x_0=35.8+0.476\cdot 65.5=66.998$

$$\delta(x_0)=\delta(66.5)=2.2281\times 1.40\times\sqrt{1+\frac{1}{12}+\frac{(66.5-800/12)^2}{(2.66)^2}}$$

$$\approx 2.2281 \times 1.40 \times 1.129 = 3.522$$

于是 给定置信水平为0.95, Y_0的预测区间为

$$(\hat{y}_0 - \delta(x_0), \hat{y}_0 + \delta(x_0)) = (63.467, 70.52)$$

(2) 同理可得

$x_0 = 70.5$, 给定置信水平为0.95, Y_0的预测区间为

$$(\hat{y}_0 - \delta(x_0), \hat{y}_0 + \delta(x_0)) = (63.832, 74.924)$$

四、小结

1.回归分析的任务

研究变量之间的相关关系

2.一元线性回归的步骤

(1)推测回归函数;　　(2)建立回归模型;
(3)估计未知参数;　　(4)进行假设检验;
(5)预测与控制.

Thank You!