第九章 回归分析

现实世界中普遍地存在着的变量之间的关系。一般来说，变量之间的关系可分为两类：

1. 确定性的函数关系：已知一个（或几个）变量的值，就可以精确地求出另一个变量的值。如 $V = 4/3\pi R^3$，$S = Vt$

2. 非确定性的相关关系：几个变量之间存在着密切的关系，但不能由一个（或几个）变量的值精确地求出另一个变量的值。在相关关系中至少有一个变量是随机变量。如人的血压与年龄，环境因子与农作物的产量，树木的直径与高度，人均收入与商品的销量，商品的价格与消费者的需求量。

回归分析是研究变量之间的相关关系的一种统计方法。回归（regression）这一术语是1886年高尔顿（Galton）研究遗传现象时引进的。

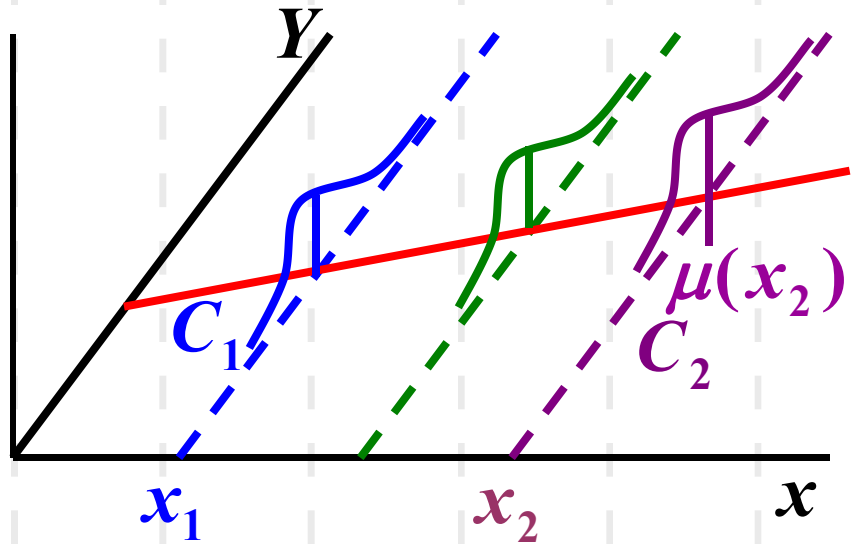
这里仅介绍一元线性回归分析。

设随机变量Y与x之间存在某种相关关系。

这里，x是可以控制或精确观测的变量（不是随机变量），如年龄、试验时的温度、施加的压力、电压与时间等。

由于Y是随机变量，对于x的每个确定值，Y有相应的分布，记其分布函数为$F(y|x)$。因此如果掌握了$F(y|x)$随着x的取值而变化的规律，也就完全掌握了Y与x之间的关系了。

然而这样做，实际中往往很难实现。作为一种近似，考察Y的数学期望$E(Y)$（假设存在），其值随x的取值而定，它是x的函数，将其记为$\mu(x)$，称为Y关于x的回归函数。于是将讨论Y与x相关关系问题转换为讨论$E(Y)=\mu(x)$与x的关系问题了。


Y
x
x_1
x_2
C_1
C_2
$\mu(x_2)$

在实际问题中，回归函数μ(x)一般是未知的，需要根据试验数据去估计。

对于x取定一组不完全相同的值$x_1, x_2, ..., x_n$，设分别在x_i处对Y作独立观察得到样本(x_i, Y_i)，$i = 1, 2, ..., n$，对应的样本观察值$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$.

将每对观察值(x_i, y_i)在直角坐标系中描出它相应的点（称为散点图），可以粗略看出$\mu(x)$的形式。

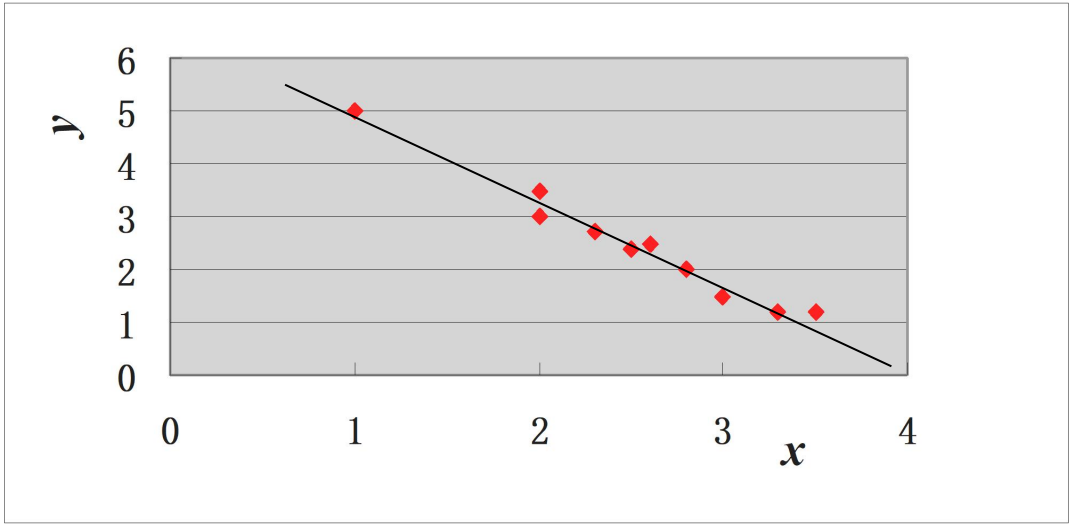
一元线性回归

例1 以家庭为单位，某商品年需求量与其价格之间的调查数据如下：

价格x（元）	1	2	2	2.3	2.5	2.6	2.8	3	3.3	3.5
需求量y（500g）	5	3.5	3	2.7	2.4	2.5	2	1.5	1.2	1.2

1. x与y之间是相关关系，不能用解析表达式$y=f(x)$表示。

2. 作散点图。发现这些点分布在一条直线附近。



故假设$\mu(x)$为线性函数：$\mu(x)=a+bx$,此时估计$\mu(x)$的问题称为求一元线性回归问题。

基本假设：$\begin{cases} Y = a + bx + \varepsilon \\ \varepsilon\text{是随机误差，不可控制，} \\ E(\varepsilon) = 0, D(\varepsilon) = \sigma^2, \\ a, b\text{(回归系数)}, \sigma^2\text{未知.} \end{cases}$

正态假设：$\varepsilon \sim N(0,\ \sigma^2)$.

对x的一组不全相同的值，得到样本$(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$

一元线性回归模型：$\begin{cases} Y_i = a + bx_i + \varepsilon_i, i = 1, 2, \ldots, n, \\ \varepsilon_i\text{相互独立，} \\ E(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2, \\ a, b\text{(回归系数)}, \sigma^2\text{未知.} \end{cases}$

正态假设：$\varepsilon_i \sim N(0,\ \sigma^2)$，相互独立，$i = 1, 2, \ldots, n$.

一元线性回归要解决的问题：

(1) a, b的估计；

(2) σ^2的估计；

(3) 线性假设的显著性检验；

(4) 回归系数b的置信区间；

(5) 回归函数$\mu(x) = a + bx$的点估计和置信区间；

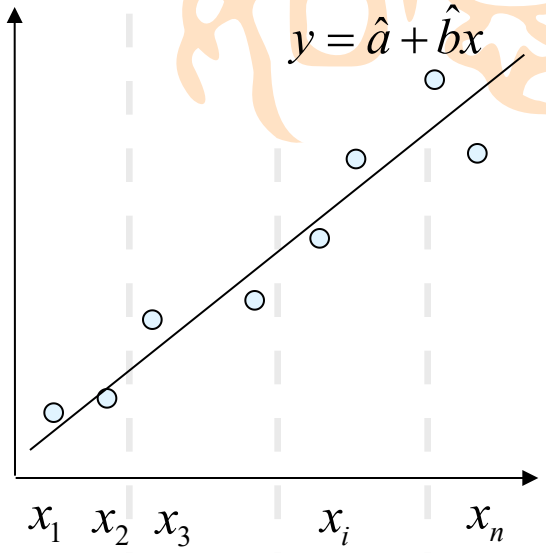
(6) Y的观察值的点预测和区间预测。

（二）a, b的估计——最小二乘估计

$$Q(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

求估计$\hat{a},\hat{b}$,

使$Q(\hat{a},\hat{b})=\min_{a,b}Q(a,b)$。



$$\frac{\partial Q}{\partial a}=-2\sum_{i=1}^{n}(y_i-a-bx_i)=0,$$

$$\frac{\partial Q}{\partial b}=-2\sum_{i=1}^{n}(y_i-a-bx_i)x_i=0.$$

整理得 $na+(\sum_{i=1}^{n}x_i)b=\sum_{i=1}^{n}y_i$,

$(\sum_{i=1}^{n}x_i)a+(\sum_{i=1}^{n}x_i^2)b=\sum_{i=1}^{n}x_iy_i$.——正规方程组

$$na+(\sum_{i=1}^{n}x_i)b=\sum_{i=1}^{n}y_i,$$

$$(\sum_{i=1}^{n}x_i)a+(\sum_{i=1}^{n}x_i^{\,2})b=\sum_{i=1}^{n}x_iy_i.$$

记号：$\bar{y}=\frac{1}{n}\sum_i y_i,\ \bar{x}=\frac{1}{n}\sum_i x_i,\ S_{xx}=\sum_i(x_i-\bar{x})^2,$

$S_{xy}=\sum_i(x_i-\bar{x})(y_i-\bar{y}),\ S_{yy}=\sum_i(y_i-\bar{y})^2.$

将正规方程整理得：$\hat{a}+\bar{x}\hat{b}=\bar{y},\quad S_{xx}\hat{b}=S_{xy}.$

a,b的最小二乘估计：$\hat{a}=\bar{y}-\bar{x}\hat{b},\ \hat{b}=S_{xy}/S_{xx}.$

在误差为正态分布假定下，最小二乘估计等价于极大似然估计。

事实上，似然函数

$$L(a,b)=\frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2\right\}$$

对$L(a,b)$最大化等价于对$\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2$最小化，即最小二乘估计。

a,b的最小二乘估计：$\hat{a}=\bar{y}-\bar{x}\hat{b},\ \hat{b}=S_{xy}/S_{xx}$.

给定x，$\mu(x)=a+bx$的估计为：

$\hat{\mu}(x)=\hat{a}+\hat{b}x$——经验回归函数。

方程：$\hat{y}=\hat{a}+\hat{b}x$

——Y关于x的（经验）回归方程，

其图形称为回归直线。

性质：$\hat{a},\hat{b}$分别是a,b的无偏估计，从而$E(\hat{Y})=a+bx$。

证明：因为$\hat{b}=S_{xy}/S_{xx}=S_{xx}^{-1}\sum_i(x_i-\bar{x})Y_i$,

$$E(\hat{b})=S_{xx}^{-1}\sum_i(x_i-\bar{x})E(Y_i)=S_{xx}^{-1}\sum_i(x_i-\bar{x})(a+bx_i)$$

$$=bS_{xx}^{-1}\sum_i(x_i-\bar{x})x_i=bS_{xx}^{-1}\sum_i(x_i-\bar{x})^2=b$$

因为$\hat{a}=\bar{Y}-\bar{x}\hat{b}$，所以

$$E(\hat{a})=E(\bar{Y})-\bar{x}E(\hat{b})$$

$$=(a+b\bar{x})-\bar{x}b=a$$

例1 中需求量与价格的关系：

价格x_i	1	2	2	2.3	2.5	2.6	2.8	3	3.3	3.5	25
需求y_i	5	3.5	3	2.7	2.4	2.5	2	1.5	1.2	1.2	25
$x_i y_i$	5	7	6	6.21	6	6.5	5.6	4.5	3.96	4.2	54.97
x_i^2	1	4	4	5.29	6.25	6.76	7.84	9	10.89	12.25	67.28

$$\bar{x} = 2.5, \qquad \bar{y} = 2.5$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = 67.28 - 10 \times 2.5^2 = 4.78$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 54.97 - 10 \times 2.5 \times 2.5 = -7.53$$

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = -1.575$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 2.5 - (-1.575) \times 2.5 = 6.44$$

故回归方程应为 $Y = 6.44 - 1.575x$

（三）误差方差的估计

误差方差σ^2估计的意义：

(a) 误差方差σ^2的大小对模型的好坏有很大的影响。

(b) 自变量对因变量影响的大小是同误差对因变量的影响相比较的。

(c) 如果自变量对因变量的影响不能显著的超过误差对因变量的影响，就很难从这样的模型中提炼出有效的、有足够精度的信息。

误差ε_i的估计

定义：残差$e_i = y_i - \hat{y}_i,\ \ i = 1,2,\cdots,n,$

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

则 (1)$Q_e = S_{yy} - \hat{b} S_{xy}$,

(2)$\hat{\sigma}^2 = \dfrac{Q_e}{n-2}$ 是 σ^2 的无偏估计(证略).

证明：（1）$e_i = y_i - \hat{y}_i = y_i - \hat{a} - \hat{b}x_i = y_i - \bar{y} - \hat{b}(x_i - \bar{x})$

$$Q_e = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - \bar{y} - \hat{b}(x_i - \bar{x})]^2$$

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

$$= S_{yy} - 2\hat{b}S_{xy} + \hat{b}^2 S_{xx} = S_{yy} - \hat{b}S_{xy}.$$

$\hat{b} = S_{xy} / S_{xx}$

求例1中误差方差的无偏估计。

x_i	1	2	2	2.3	2.5	2.6	2.8	3	3.3	3.5	25
y_i	5	3.5	3	2.7	2.4	2.5	2	1.5	1.2	1.2	25
$x_i y_i$	5	7	6	6.21	6	6.5	5.6	4.5	3.96	4.2	54.97
x_i^2	1	4	4	5.29	6.25	6.76	7.84	9	10.89	12.25	67.28
y_i^2	25	12.25	9	7.29	5.76	6.25	4	2.25	1.44	1.44	74.68

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 74.68 - 10 \times 2.5^2 = 12.18$$

$$S_{xy} = -7.53 \qquad \hat{b} = -1.575$$

$$Q_e = S_{yy} - \hat{b} S_{xy} = 0.32$$

$$\hat{\sigma}^2 = \frac{Q_e}{n-2} = \frac{0.32}{8} = 0.04$$

（四）线性假设的显著性检验

采用最小二乘法估计参数a和b，并不需要事先知道Y与x之间一定具有相关关系，即使是平面图上一堆完全杂乱无章的散点，也可以用公式求出回归方程。因此μ(x)是否为x的线性函数，一要根据专业知识和实践来判断，二要根据实际观察得到的数据用假设检验方法来判断。

即要检验假设 $H_0: b = 0, H_1: b \neq 0,$

若原假设被拒绝，说明回归效果是显著的，否则，若接受原假设，说明Y与x不是线性关系，回归方程无意义。回归效果不显著的原因可能有以下几种：

（1）影响Y取值的，除了x，还有其他不可忽略的因素；
（2）E(Y)与x的关系不是线性关系，而是其他关系；
（3）Y与x不存在关系。

检验假设 $H_0: b=0, H_1: b\neq 0$, 拒绝域形式: $|\hat{b}|\geq c$。

假定: ε_i独立同服从$N(0,\sigma^2)$分布$(i=1,2,\cdots,n)$。

则可以证明 (1) $\hat{b}\sim N\left(b, S_{xx}^{-1}\sigma^2\right)$;

$$(2)\ \frac{(n-2)\sigma^2}{\sigma^2}=\frac{Q_e}{\sigma^2}\sim\chi^2(n-2);$$

(3) $\bar{y}, \hat{b}, Q_e$相互独立。

故 $$\frac{\hat{b}-b}{\sqrt{\sigma^2/S_{xx}}}\Bigg/\sqrt{\frac{(n-2)\sigma^2}{\sigma^2}\Bigg/(n-2)}\sim t(n-2),$$

当H_0为真即$b=0$时, $t=\dfrac{\hat{b}}{\hat{\sigma}}\sqrt{S_{xx}}\sim t(n-2)$,

水平为α的检验拒绝域: $\boxed{|t|=\dfrac{1}{\hat{\sigma}}|\hat{b}|\sqrt{S_{xx}}\geq t_{\alpha/2}(n-2).}$

例3 检验例1中回归效果是否显著，取 α =0. 05。

由 例 1， 例 2 知：

$\hat{b} = -1.575,\quad S_{xx} = 4.78,\quad \sigma^2 = 0.04.$

查表得：$t_{\alpha/2}(n-2) = t_{0.025}(8) = 2.306.$

因此假设$H_0 : b = 0$的检验拒绝域为：

$$|t| = \frac{|\hat{b}|}{\hat{\sigma}}\sqrt{S_{xx}} \geq 2.306.$$

计算得，$|t| = \dfrac{1.575}{\sqrt{0.04}}\sqrt{4.78} = 17 > 2.306.$

故拒绝$H_0 : b = 0$，认为回归效果是显著的。

（五）回归系数b的置信区间

当回归效果显著时，常需要对回归系数b作区间估计。

由于 $\dfrac{\hat{b}-b}{\hat{\sigma}}\sqrt{S_{xx}}\sim t(n-2)$,

所以 $P\left(\dfrac{|\hat{b}-b|}{\hat{\sigma}}\sqrt{S_{xx}}\le t_{\alpha/2}(n-2)\right)=1-\alpha$

即 b 的置信水平 $1-\alpha$ 的置信区间：

$$\left(\hat{b}\pm t_{\alpha/2}(n-2)\times\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right).$$

例如例1中b的置信水平为0.95的置信区间为：

$$\left(-1.575\pm 2.306\times\sqrt{\frac{0.04}{4.78}}\right)=(-1.786,\ -1.364).$$

(六) 回归函数 $\mu(x)=a+bx$ 函数值的点估计和置信区间

对给定的 x_0，$\mu(x_0)=a+bx_0$ 的点估计为 $\hat{y}_0=\hat{\mu}(x_0)=\hat{a}+\hat{b}x_0$.

则有 (1) 相应的估计量 $\hat{Y}_0=\hat{a}+\hat{b}x_0$ 是 $\mu(x_0)=a+bx_0$ 无偏估计，

(2) $\mu(x_0)=a+bx_0$ 的置信水平为 $1-\alpha$ 的置信区间为：

$$\left(\hat{Y}_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

证明：(1) 因为 $E(\hat{b})=b, E(\hat{a})=a,$

所以 $E(\hat{Y}_0)=E(\hat{a}+\hat{b}x_0)=a+bx_0$. 即为无偏估计

（2）可以证明：$\hat{Y}_0 \sim N(a+bx_0, \left(\frac{1}{n}+(x_0-\bar{x})^2 S_{xx}^{-1}\right)\sigma^2)$.

又有，$\frac{(n-2)\hat{\sigma}^2}{\sigma^2}=\frac{Q_e}{\sigma^2}\sim \chi^2(n-2)$; 且$\hat{Y}_0$与$Q_e$独立。

于是 $$\frac{\hat{Y}_0-(a+bx_0)}{\sigma\sqrt{\frac{1}{n}+(x_0-\bar{x})^2 S_{xx}^{-1}}} \Big/ \sqrt{\frac{(n-2)\sigma^2}{\sigma^2}\Big/(n-2)} \sim t(n-2),$$

即 $$\frac{\hat{Y}_0-(a+bx_0)}{\hat{\sigma}\sqrt{\frac{1}{n}+(x_0-\bar{x})^2 S_{xx}^{-1}}} \sim t(n-2),$$

所以，$\mu(x_0)=a+bx_0$的置信水平为$1-\alpha$的置信区间为：

$$\left(\hat{Y}_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

（七）Y的观察值的点预测和预测区间

考虑对指定点$x = x_0$处因变量Y的观察值Y_0的预测问题。由于在$x = x_0$处并未进行观察，或暂时无法观察。经验回归函数的重要应用是，可利用它对因变量Y的新观察值Y_0进行点预测和区间预测。

设Y_0是在$x = x_0$处对Y的观察结果。则

$$Y_0 = a + bx_0 + \varepsilon_0, \varepsilon_0 \sim N(0, \sigma^2).$$

(1) Y_0的点预测为：$\hat{Y}_0 = \hat{a} + \hat{b}x_0$.

(2) Y_0的置信水平为$1-\alpha$的预测区间为：

$$\left(\hat{Y}_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right).$$

证明：因Y_0是将要做的独立试验结果，因此，它与已得到的试验结果$Y_1, Y_2, ..., Y_n$相互独立。

又$\hat{Y}_0 = \bar{Y} + \hat{b}(x_0 - \bar{x})$是$Y_1, Y_2, ..., Y_n$的线性组合，

故Y_0与$\hat{Y}_0$相互独立。

$$Y_0 \sim N(a + bx_0, \sigma^2), \hat{Y}_0 \sim N\left(a + bx_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right).$$

所以，$\hat{Y}_0 - Y_0 \sim N\left(0, \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right)$,

又$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \frac{Q_e}{\sigma^2} \sim \chi^2(n-2)$; 且$Y_0$, $\hat{Y}_0$, Q_e相互独立。

于是$$\frac{\hat{Y}_0 - Y_0}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \Bigg/ \sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \Big/ (n-2)} \sim t(n-2),$$

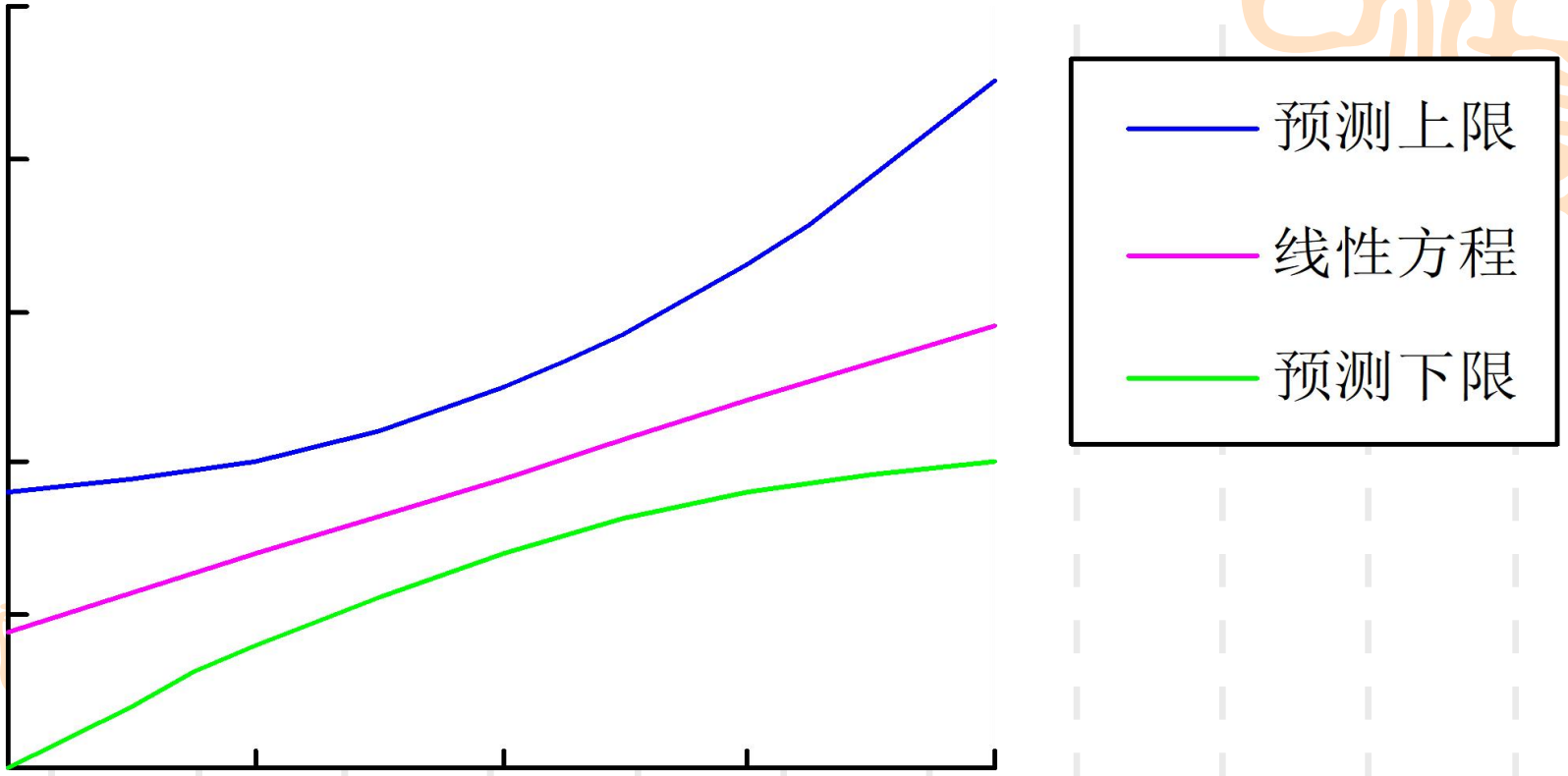
即 $$\frac{\hat{Y}_0 - Y_0}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}} \sim t(n-2),$$

所以，Y_0的置信水平为$1-\alpha$的预测区间为：

$$\left(\hat{Y}_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right).$$

注1，这一预测区间的长度随$|x_0-\bar{x}|$的增加而增加，当$x_0=\bar{x}$时最短。

注2，在相同的置信水平下，$\mu(x_0)$的置信区间要比Y_0的预测区间短。这是因为$Y_0=a+bx_0+\varepsilon_0$比$\mu(x_0)=a+bx_0$多了一项ε_0的缘故。



注：在预测时，一般要求 x_0要落在已有的的数据范围内部，否则预测常常没有意义。

在例1中，当x_0=1.5时，回归函数的水平为0.95的置信区间为

$$(4.1 \pm 0.256)$$

$$= (3.844,\ \ 4.356)$$

在例1中，当x_0=1.5时，Y_0的概率为0.95的预测区间为

$$(4.1 \pm 0.528)$$

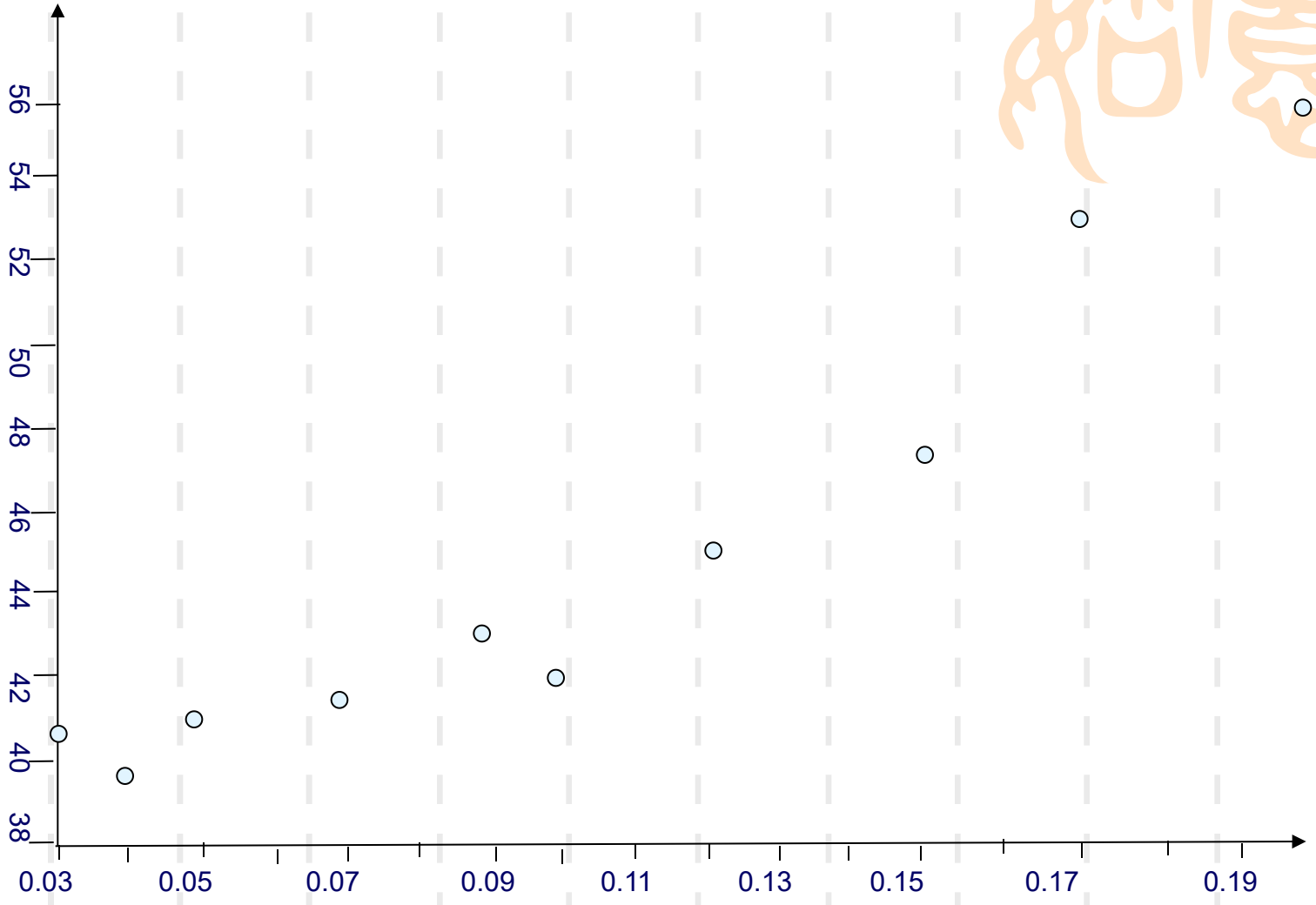
$$= (3.572,\ \ 4.628)$$

例2 合金钢的强度y与钢材中碳的含量x有密切关系。为了冶炼出符合要求强度的钢常常通过控制钢水中的碳含量来达到目的，为此需要了解y与x之间的关系。其中x：碳含量（%） y：钢的强度（kg/mm^2）数据见下：

x	0.03	0.04	0.05	0.07	0.09	0.10	0.12	0.15	0.17	0.20
y	40.5	39.5	41.0	41.5	43.0	42.0	45.0	47.5	53.0	56.0

（（1）画出散点图；（2）设 $\mu(x)=a+bx$, 求a, b的估计；（3）求误差方差的估计；（4）检验回归效果是否显著（取显著性水平 $\alpha=0.05$）；（5）求回归系数b的95%置信区间；（6）求在x=0.06点，回归函数的点估计和95%置信区间；（7）求在x=0.06点，Y的点预测和95%区间预测。

（1）合金钢的强度y与钢材中碳的含量x的散点图



x	0.03	0.04	0.05	0.07	0.09	0.10	0.12	0.15	0.17	0.20
y	40.5	39.5	41.0	41.5	43.0	42.0	45.0	47.5	53.0	56.0

(2)计算得：

$$\sum_i y_i = 449, \sum_i x_i = 1.02,$$

$$\sum_i x_i^2 = 0.1338, \sum_i x_i y_i = 48.555, \quad \hat{a} = \bar{y} - \bar{x}\hat{b},$$

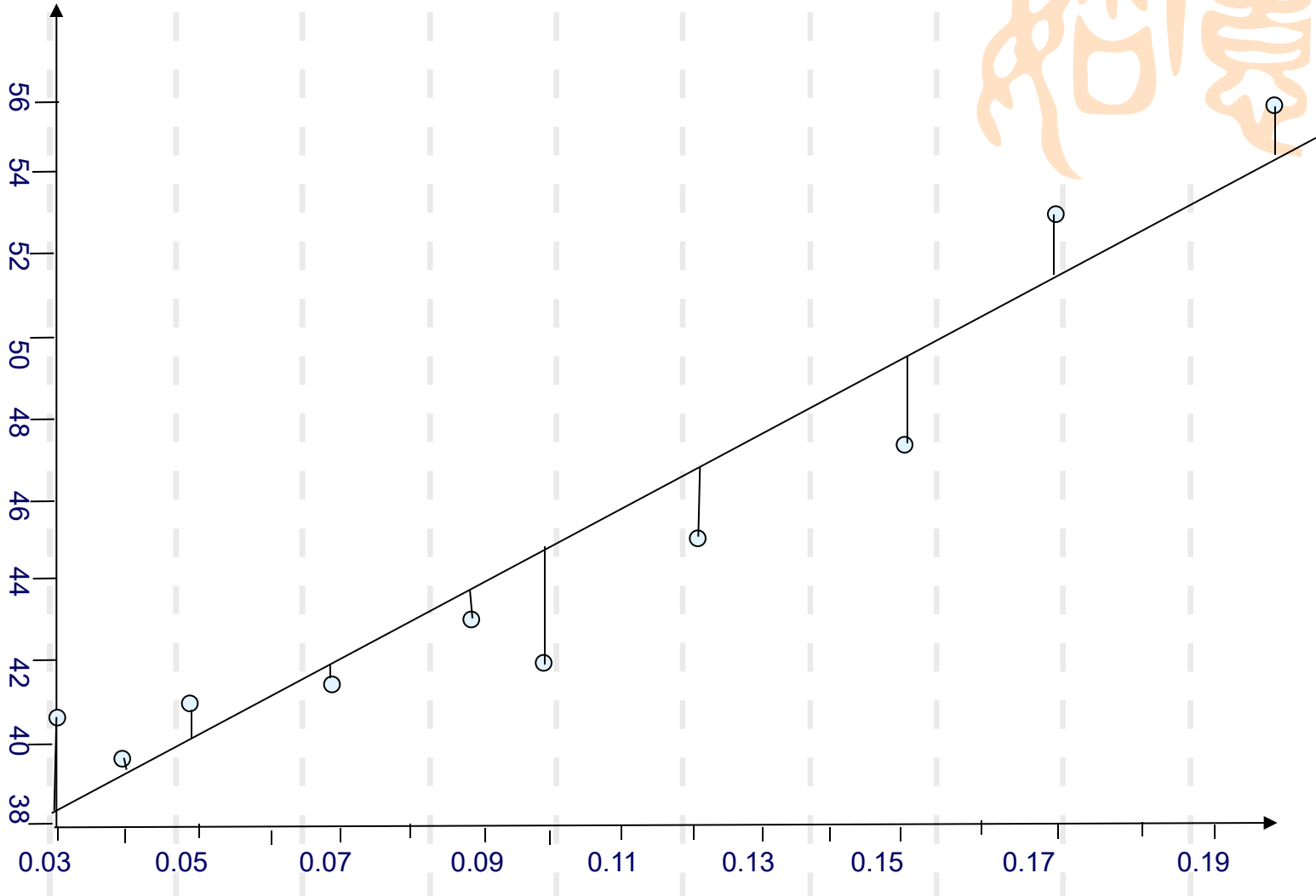
$$S_{xx} = 0.02976, S_{xy} = 2.757. \quad \hat{b} = S_{xy} / S_{xx}.$$

a,b的最小二乘估计：$\hat{a} = 35.4506,\ \hat{b} = 92.6411$

回归方程：$\hat{y} = 35.4506 + 92.6411x.$

或写成：$\hat{y} = 44.9 + 92.6411(x - 0.102).$

合金钢的强度y与钢材中碳的含量x的回归直线图



(3)计算得：

$$\sum_i y_i = 449, \sum_i y_i^2 = 20443, S_{yy} = 282.9.$$

又已知 $S_{xy} = 2.757, \hat{b} = 92.6411.$

$$Q_e = S_{yy} - \hat{b}S_{xy} = 27.4884,$$

所以，σ^2的无偏估计$\sigma^2 = Q_e/(n-2) = 3.436.$

(4)检验假设$H_0: b=0, H_1: b\neq 0$的显著性水平

为α的检验拒绝域：$$|t|=\frac{|\hat{b}|}{\hat{\sigma}}\sqrt{S_{xx}}\geq t_{\alpha/2}(n-2)。$$

经计算

$$|t|=\frac{92.6411}{\sqrt{3.436}}\sqrt{0.02976}=8.6217\geq t_{0.025}(8)=2.306,$$

拒绝原假设，认为合金钢强度与炭含量的回归效果显著。

(5)回归系数b的置信水平95%的置信区间：

$$\left(\hat{b}\pm t_{\alpha/2}(n-2)\times\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)=(67.8629,\ 117.4193).$$

（6）当$x_0 = 0.06$时，$\hat{y}_0 = \hat{a} + \hat{b}x_0 = 41.0091$

$$t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}} = 2.306\times\sqrt{3.436}\sqrt{\frac{1}{10}+\frac{(0.06-0.102)^2}{0.02976}} = 1.706$$

所以，$\mu(0.06)$的0. 95的置信区间为:$(39.303, 42.715)$.

(7) $x_0 = 0.06$时，Y_0的置信水平为0.95的预测区间为：

$(36.407,\ 45.611)$.

其中$t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}} = 4.602.$