高中课程新学案

SHUXUE

数学

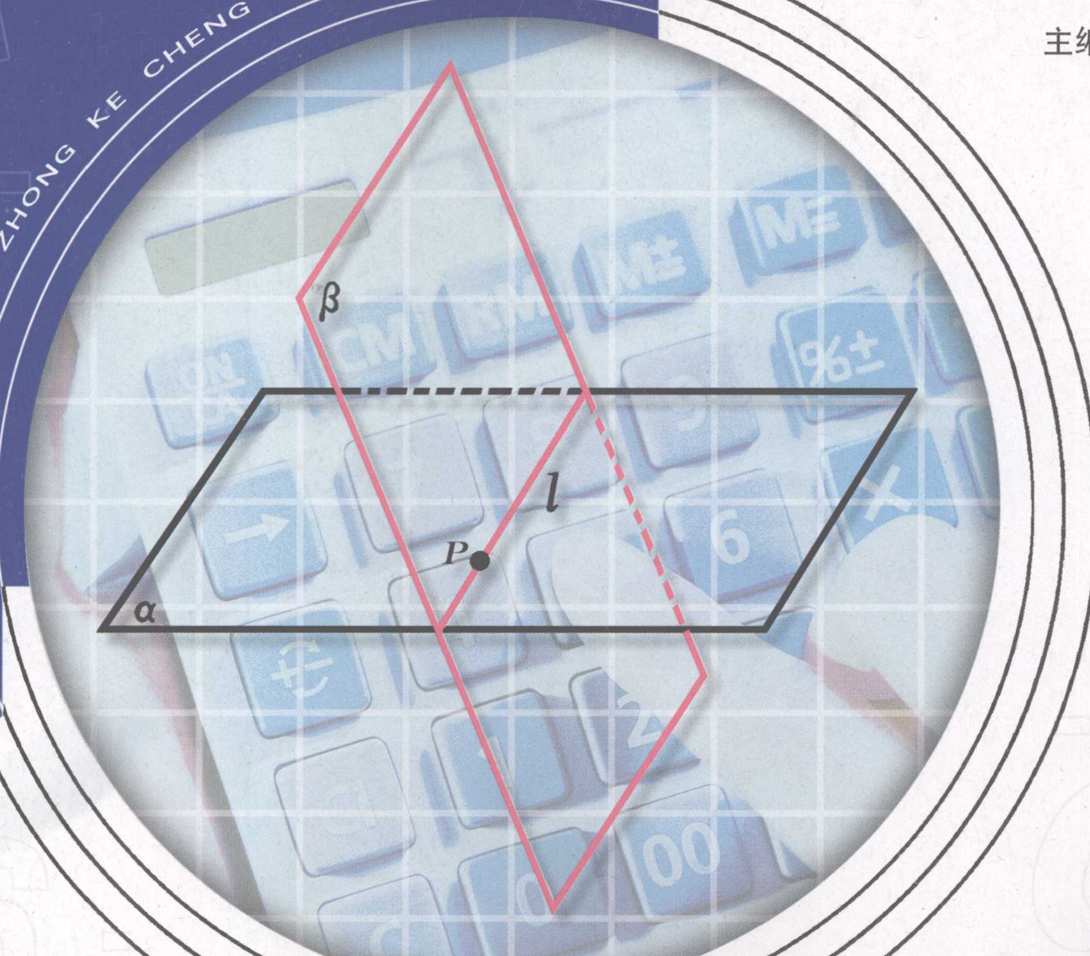
文

选修1-2

必修1、2

（复习）

GAO ZHONG KE CHENG XIN XUE AN



主编　金立村
郭允远

明天出版社
TOMORROW PUBLISHING HOUSE

高中课程新学案

GAO ZHONG KE CHENG XIN XUE AN

数学(文)

选修 1–2，必修 1、2(复习)

主　编：金立村　郭允远

副主编：王子成　王佃军

编　者：(排名不分先后)

庞　霞　李玉龙　李广之　胡文涛

王佃军　丰艳珍　李建国　沙士锦

冉祥宁　田宝运　孙玉成　杨建波

王子成　刘　杰　王中贵　许桂盛

马从吉　秦立文　潘凤生　王兆法

尚明刚　魏言玉　赵海啸

明天出版社

TOMORROW PUBLISHING HOUSE

高中课程新学案
GAO ZHONG KE CHENG XIN XUE AN

编委会名单

高中课程新学案
数学(文)
选修1-2,必修1、2(复习)
*
明天出版社出版
(济南经九路胜利大街39号)
http://www.sdpress.com.cn
http://www.tomorrowpub.com
山东省新华书店发行　　山东新华印刷厂临沂厂印刷
*
889×1194毫米　16开本　13印张　526千字
2008年1月第1版　2008年1月第1次印刷
ISBN 978-7-5332-5616-6
定价:10.40元
如有印装质量问题,请
(电话:0539—29

高中课程新学案
GAO ZHONG KE CHENG XIN XUE AN

前言

为适应基础教育课程改革的要求，推进高中教学改革的深入发展，进一步提高教学效率和质量，使教师的教与学生的学、使教学内容与教学过程、使知识传授与发展能力等，在课堂教学这一时空内的结合更加科学、和谐、完美，我们在充分搞好调查研究、总结高中学校教改经验的基础上，组织优秀骨干教师和教研人员，编写了《高中课程新学案》，供学生使用。

编写和使用《新学案》的直接目的，是为了推进课堂教学真正实现教的方式和学的方式的转变，进一步还学生以学习主人地位，更多地给学生以动手、动脑、动口的时间和空间，帮助学生打牢基础，发展能力，减轻负担，提高效率。

《新学案》高一年级本按教材顺序和新授课特点编写，高中三年级本按教材和高考考试大纲要求编写，原则上 1—2 课时一个学案，每个学案分“学海导航”、“学习探究”、“自我测评”和“拓展提高”等四个部分(答案另附)，旨在帮助学生明确学习目标，优化学习过程，以学案提供的栏目和问题为线索，理解、掌握和巩固教材的基础知识，并在自我测评和拓展提高的实战练习中发展能力。与其他资料相比，《新学案》的突出特点是：汇集群智，体例创新；以生为本，以学立意；着眼基础，适当超越。这既符合素质教育的要求，也符合高中生参加高考选拔的需要。

《新学案》是近几年高中教学改革的一项新成果，是广大教师集体智慧的结晶，它的使用，必将对中学教学模式的转变和教学质量的提高产生积极的影响。但由于它是新事物，限于我们的认知水平，必定还会有不足和缺陷，恳请广大师生提出宝贵意见和建议。

编　者

2008年1月

目录

选修 1－2

第一章　统计案例

§1.1　回归分析的基本思想及其初步应用

（第一课时　随机误差）

学海导航

【知识要点】 1. 回归分析；2. 回归方程；3. 散点图；4. 随机误差

【学习要求】 1. 了解线性回归模型与函数模型的差异；2. 复习最小二乘估计的思想及计算公式；3. 了解回归分析的基本思想；明确回归方程、散点图定义；理解随机误差含义．4. 鼓励学生经历数据处理的全过程．要尽量使用统计图直观展示两个变量的关系，培养学生对数据的直觉与敏感．有条件的学校可以利用统计软件画散点图，求回归直线方程并画出回归直线．

学习探究

【要点分析】

1. 回归分析　回归分析是对具有相关关系的两个变量进行统计分析的一种常用方法. 它主要解决以下问题：

（1）确定几个特定的变量之间是否存在相关关系，如果存在的话找出它们之间适合的数学表达式.

（2）根据一个或几个变量的值预测或控制另一个变量的取值，并且要知道所能达到的精确度. 体会和理解这些问题，有利于掌握这一内容的数学思想和方法.

2. 散点图　表示具有相关关系的两个变量组成一组数据，将各组数据在平面直角坐标系中用描点的方法得到，这种图形叫散点图. 可以直观分析两变量的相关关系.

3. 回归方程

（1）回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$，其中 $\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}$，

$\hat{a}=\bar{y}-\hat{b}\bar{x}\left(\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i,\bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i\right)$. $(\bar{x},\bar{y})$称为样本点的中心.

（2）回归直线方程的求法.

借助计算器进行运算求出系数 $\hat{b}$ 与 $\hat{a}$，也可以用相应的计算机软件求出方程.

具体计算步骤：①分别计算 $\sum\limits_{i=1}^{n}x_i,\sum\limits_{i=1}^{n}y_i,\sum\limits_{i=1}^{n}x_i^2,\sum\limits_{i=1}^{n}y_i^2,\sum\limits_{i=1}^{n}x_iy_i$.

②分别计算 $x_i-\bar{x},y_i-\bar{y},\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$.

③根据公式求出 $\hat{b}$ 和 $\hat{a}$，$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}$，$\hat{a}=\bar{y}-\hat{b}\bar{x}$.

④写出回归直线方程 $\hat{y}=\hat{a}+\hat{b}x$.

（3）最小二乘法.

我们研究在一般情况下（已知 n 对数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 即 n 个点）如何推导求 $\hat{a}$ 与 $\hat{b}$ 的公式.

随机误差

$\varepsilon_i=y_i-a-bx_i,i=1,2,\cdots,n.$

假如把这些随机误差直接相加作为总的误差，是很不合理的，因为它们有正有负，相加起来可能抵消一部分. 为了不使误差之和正负抵消，我们设全部误差的平方和为

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

用 Q 的大小来度量总的误差大小. Q 是 a,b 的二元函数，可记作$Q(a,b)$，用下面的配方法可以求出 $Q(a,b)$ 达到最小值时 a,b 所取的值.

记 $\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i,\bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i$，以下为了书写方便，一律省去求和号$\Sigma$的上下标，我们有

$$\begin{aligned}Q(a,b)&=\sum\{(y_i-\bar{y})+[\bar{y}-(a+b\bar{x})]-b(x_i-\bar{x})\}^2\\&=\sum(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+b^2\sum(x_i-\bar{x})^2+2[\bar{y}-(a+b\bar{x})]\cdot\sum(y_i-\bar{y})-2b[\bar{y}-(a+b\bar{x})]\cdot\sum(x_i-\bar{x})-2b\sum(x_i-\bar{x})(y_i-\bar{y})\end{aligned}$$

$$= \sum (y_i - \bar{y})^2 + n[\bar{y} - (a + b\bar{x})]^2 + b^2 \sum (x_i - \bar{x})^2 - 2b\sum (x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum (y_i - \bar{y})^2 + n[\bar{y} - (a + b\bar{x})]^2 + \sum (x_i - \bar{x})^2 \cdot [b^2 - 2b\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}]$$

$$= \sum (y_i - \bar{y})^2 + n[\bar{y} - (a + b\bar{x})]^2 + \sum (x_i - \bar{x})^2 [b - \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}]^2 - \frac{[\sum (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2}.$$

经过略嫌冗长的推导，这些推导的目的是把 a，b“配”到含有平方项的底的中间去. 对于 n 对数据来说，$x_1, x_2, \cdots x_n$ 一般不会相等（否则这 n 对数据已经在一条平行于 y 轴的直线上了，再求回归直线已失去意义），所以

$$\sum (x_i - \bar{x})^2 \neq 0.$$

观察上面最后的表达式，其中 $y_i, \bar{y}, n, \bar{x}, x_i$ 都是已知数，含 a, b 的两项是非负数，当且仅当它们等于0时，$Q(a,b)$ 取最小值. 这就是说，当

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

时 $Q(a,b)$ 达到最小值. 上面的 $\hat{b}$ 也可以进一步推导成

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}.$$

至此我们推导出了求 a 与回归直线系数 b 的数学公式.

4. 随机误差　由散点图容易发现样本点并不是“严格地”分布在一条直线上，对于某个 x_i，由回归直线方程能确定一个 $\hat{y}_i = a + bx_i$，一般地，由于测量可能存在误差，或者受某些随机因素的影响，或者回归直线方程本身就不够精确，$\hat{y}_i$ 与测得的数据 y_i 很可能不相等，这些因素在关系式中用 e 表示，称为随机误差，即

$$y_i = \hat{y}_i + e_i (i = 1, 2, \cdots, n)$$

e 是一个随机变量，一般假定它的均值为0，即 $Ey = bx + a$，也就是 y 的期望值是 x 的一次函数. 在实际问题中，线性回归模型适用的范围要比一次函数大得多. 当随机变量恒等于0时，线性回归模型就变成一次函数模型. 因此一次函数模型是线性回归模型的特殊形式，线性回归模型是一次函数模型的一般形式.

5. 随机误差 e 的主要来源.

①用线性回归模型近似真实模型（真实模型是客观存在的，通常我们并不知道真实模型到底是什么）所引起的误差. 可能存在非线性的函数能够更好地描述 y 与 x 之间的关系，但是现在却用线性函数来表述这种关系，结果会产生误差. 这种由模型近似所引起的误差包含在 e 中.

②忽略了某些因素的影响. 影响变量 y 的因素不只变量 x，可能还包括其他许多因素（例如在描述身高和体重关系的模型中，体重不仅受身高的影响，还会受遗传基因、饮食习惯、生长环境等其他因素的影响），它们的影响都体现在 e 中.

③观测误差. 由于测量工具等原因，导致 y 的观测值产生误差（比如一个人的体重是确定的数，不同的秤可能会得到不同的观测值，与真值之间存在误差），这样的误差也包含在 e 中.

【例题分析】

例　某产品的广告费支出 x 与销售额 y（单位：百万元）之间有如下对应数据：

x	2	4	5	6	8
y	30	40	60	50	70

（1）画出散点图；

（2）求回归直线方程.

自我测评

A组

1. 在下列各量与量的关系中：

①正方体的体积与棱长间的关系；②一块农田的水稻产量与施肥量之间的关系；③人的身高与年龄之间的关系；④家庭的支出与收入之间的关系；⑤某户家庭用电量与电价之间的关系．是相关关系的为(　　)．

(A)②③　　(B)③④

(C)④⑤　　(D)②③④

2. 为了考察两个变量 x 和 y 之间的线性相关性，甲乙两个同学各自独立的作10次和15次试验，并利用线性回归方程求得回归直线分别为 l_1 和 l_2. 已知在两个人的试验中，发现对变量 x 的观测数据的平均值恰好相等，都为 s；对变量 y 的观测数据的平均值也恰好相等，都为 t. 那么下列说法正确的是(　　)．

(A)直线 l_1 和 l_2 有交点 (s,t)

(B)直线 l_1 和 l_2 相交，但交点未必是点 (s,t)

(C)直线 l_1 和 l_2 由于斜率相等，所以必定平行

(D)直线 l_1 和 l_2 必定重合

3. 对有线性相关关系的两个变量建立的回归直线方程 $\hat{y}=a+bx$ 中，回归系数 b(　　)．

(A)可以小于0　　(B)大于0

(C)能等于0　　(D)只能小于0

4. 对于一组具有线性相关关系的数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 其回归方程中的截距为(　　)．

(A) $a=y-bx$　　(B) $a=\hat{y}-\hat{b}\bar{x}$

(C) $\hat{a}=y-bx$　　(D) $\hat{a}=\bar{y}-\hat{b}\bar{x}$

5. 已知两个变量 x 和 y 之间有线性相关性，5次试验的观测数据如下：

x	100	120	140	160	180
y	45	54	62	75	92

那么变量 y 关于 x 的回归方程是________.

B组

6. 为了表示 n 个点与相应直线在整体上的接近程度，我们常用(　　)表示．

(A) $\sum_{i=1}^{n}(y_i-\hat{y}_i)$　　(B) $\sum_{i=1}^{n}(\hat{y}_i-y_i)$

(C) $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$　　(D) $\sum_{i=1}^{n}(y_i-\bar{y})^2$

7. 自己在生活中找一下，周围有什么样的例子，可以用回归分析的知识来分析．

拓展提高

1. 观测数据不要少于5组.

2. 我们在求经验公式后，往往用来作为现实生活中两变量之间相关关系的近似关系，从而可以用来指导生活实践.

(编者：郯城二中　庞霞)

§1.1　回归分析的基本思想及其初步应用

（第二课时　相关系数）

学海导航

【知识要点】 相关系数.

【学习要求】 会用公式求相关系数，并能够根据相关系数的值判断相关性的强弱.

学习探究

【要点分析】

1. 若相应于变量 x 的取值 x_i，变量 y 的观测值 $y_i(1\leqslant i\leqslant n)$，把

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

即

$$r=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)(\sum_{i=1}^{n}y_i^2-n\bar{y}^2)}}$$

叫做变量 y 与 x 之间的样本相关系数，简称相关系数.

统计中用相关系数 r 来衡量两个变量之间线性关系的强弱. $|r|\leqslant 1$，而且 $|r|$ 越接近 1，相关程度越大；它们的散点图越接近一条直线，这时用线性回归模型拟合这组数据就越好.

$|r|$ 越接近 0，相关程度越小. $r>0$，正相关；$r<0$，负相关；$r=1$，表示两个变量为完全正相关；$r=-1$，表示两个变量为完全负相关. 通常 $|r|>0.75$ 时认为两个变量有很强的线性相关关系，从而也表明建立的回归模型是有意义的.

注意，相关系数 r 只能描述两个变量之间的变化方向及密切程度，不能揭示二者之间的本质联系.

2. 利用回归方程进行预测.

获得回归方程不是最终目的，如果建立的回归模型是有效的，我们希望用它进行预测或决策.

【例题分析】

例　已知某菜地每单位面积年平均使用氮肥量为 $x$kg 与每单位面积蔬菜年平均产量 yt 之间的关系如下数据：

年号	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
x(kg)	70	74	80	78	85	92	90	95	92	108	115	123	130	138	145
y(t)	5.1	6.0	6.8	7.8	9.0	10.2	10.0	12.0	11.5	11.0	11.8	12.2	12.5	12.8	13.0

（1）求 x 与 y 之间的相关系数，并检验是否线性相关；

（2）若线性相关，求蔬菜产量 y 与使用氮肥量之间的回归直线方程，并估计每单位面积施肥 150kg 时，每单位面积蔬菜的年平均产量.

解：（1）列出下表，并用科学计算器进行有关计算：

i	1	2	3	4	5	6	7	8
x_i	70	74	80	78	85	92	90	95
y_i	5.1	6.0	6.8	7.8	9.0	10.2	10.0	12.0
x_iy_i	357	444	544	608.4	765	938.4	900	1140
i	9	10	11	12	13	14	15	
x_i	92	108	115	123	130	138	145	
y_i	11.5	11.0	11.8	12.2	12.5	12.8	13.0	
x_iy_i	1058	1188	1357	1500.6	1625	1766.4	1885	

$\bar{x}=101$，$\bar{y}=10.1133$

$\sum_{i=1}^{15}x_i^2=161125$，$\sum_{i=1}^{15}y_i^2=1628.55$，

故 $\sum_{i=1}^{15}x_iy_i=16076.8$

蔬菜产量与施用氮肥量的相关系数

$$r=\frac{\sum_{i=1}^{15}x_iy_i-15\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{15}x_i^2-15\bar{x}^2)(\sum_{i=1}^{15}y_i^2-15\bar{y}^2)}}$$

$$=\frac{16076.8-15\times101\times10.1133}{\sqrt{(161125-15\times101^2)(1628.55-15\times10.1133^2)}}$$

≈ 0.8632

这说明蔬菜产量与氮肥量之间存在着线性相关关系.

（2）设所求的回归直线方程为 $\hat{y}=\hat{a}x+\hat{b}$，则

$$\hat{b}=\frac{\sum_{i=1}^{15}x_iy_i-15\bar{x}\bar{y}}{\sum_{i=1}^{15}x_i^2-15\bar{x}^2}\approx 0.0931,$$

$\hat{a}=\bar{y}-\hat{b}\bar{x}\approx 0.7095$，则 $\hat{y}=0.0931x+0.7095$.

当 $x=150$，y 的估计值 $\hat{y}=0.0931\times150+0.7095=14.6745$(t).

由于回归系数是 0.0931，[illegible]的意义知，使用氮肥量每增加 1kg，[illegible]平均产量就增加 0.0931 吨.

分析：由回归系数[illegible]的公式可见，本题的计算工[illegible]同学应尽量借助计算器来完[illegible]高中生使

用的科学函数计算器普遍都设置了进入回归计算的专用按键，使用起来十分方便．我们希望同学们在掌握了回归分析的基本思想方法后，只要有条件，应该使用计算器等现代技术手段来处理数据．计算机中有许多应用软件，特别是表格类软件是提供统计计算函数的，用起来都非常方便，下面以 $fx-82TL$ 型计算器举例说明计算器计算 $r,\hat{a},\hat{b}$ 的值的过程：$(\hat{y}=\hat{b}x+\hat{a})$

MODE　3　1　（进入回归计算模式）
SHIFT　SCL =（清除计算器记忆器内容）
70,5.1　DT　74,6.0　DT
80,6.8　DT　78,7.8　DT
85,9.0　DT　92,10.2　DT
90,10.0　DT　95,12.0　DT
92,11.5　DT　108,11.0　DT
115,11.8　DT　123,12.2　DT
130,12.5　DT　138,12.8　DT
145,13.0　DT
SHIFT　A =　（计算系数 a = 0.7095）
SHIFT　B =　（计算系数 b = 0.0931）
SHIFT　r =　（计算相关系数 r = 0.8632）

自我测评

A 组

1．许多因素都会影响贫穷，教育也许是其中之一．在研究这两个因素的关系时，收集了美国 50 个州的成年人受过 9 年或更少教育的百分比 (x) 和收入低于官方规定的贫困线的人数占本州人数的百分比 (y) 的数据，建立的回归直线方程如下：

$$y=0.8x+4.6.$$

斜率的估计等于 0.8 说明________________；成年人受过 9 年或更少教育的百分比 (x) 和收入低于官方的贫困线的人数占本州人数的百分比 (y) 之间的相关系数________（填充“大于 0”或“小于 0”）．

2．在一次实验中，测得 (x,y) 的四组值分别为 $A(1,2)$，$B(2,3)$，$C(3,4)$，$D(4,5)$，则 y 与 x 之间的回归直线方程为（　　）．

(A) $\hat{y}=x+1$　　(B) $\hat{y}=x+2$
(C) $\hat{y}=2x+1$　　(D) $\hat{y}=x-1$

3．已知 x、y 之间的数据如下表所示，则 y 与 x 之间的线性回归方程过点（　　）．

x	1.08	1.12	1.19	1.28
y	2.25	2.37	2.40	2.55

(A) $(0,0)$　　(B) $(\bar{x},0)$
(C) $(0,\bar{y})$　　(D) $(\bar{x},\bar{y})$

4．已知回归直线方程为 $\hat{y}=0.50x-0.81$，则 $x=25$ 时，y 的估计值为________．

B 组

5．对于相关系数 r，以下说法正确的是（　　）．
(A) 它是描述两个变量之间相关程度的量
(B) 它是衡量两个变量之间线性相关程度的量
(C) 若 $r=0$，说明两个变量不线性相关
(D) 两个变量相关程度越高，则 $|r|$ 越大

6．由一组数据 (x_1,y_1)，(x_2,y_2)，…，(x_n,y_n) 得到的回归直线方程 $\hat{y}=bx+a$，则下列说法不正确的是（　　）．

(A) 直线 $\hat{y}=bx+a$ 必过点 $(\bar{x},\bar{y})$

(B) 直线 $\hat{y}=bx+a$ 至少经过点 (x_1,y_1)，(x_2,y_2)，…，(x_n,y_n) 中的一个点

(C) 直线 $\hat{y}=bx+a$ 的斜率为 $\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$

(D) 直线 $\hat{y}=bx+a$ 和各点 (x_1,y_1)，(x_2,y_2)，…，(x_n,y_n) 的偏差是该坐标平面上所有直线与这些点的偏差中最小的直线

7．线性回归方程 $\hat{y}=bx+a$ 中，b 的意义是________．

8．利用回归方程 $\hat{y}=0.849x-85.712$，预测身高为 172cm 女大学生的体重，只要把 $x=172$ 代入方程中即可得到体重的预测值 60.316kg，身高为 172cm 的女大学生的体重一定是 60.316kg 吗？若不是，那么这个预测值的含义是什么？

拓展提高

1．作相关性检验，有时也可以用散点图，观察所给的数据列成的点是否在一条直线的附近，这样做既直观又方便，因而对解决相关性检验问题比较常用，但在作图中，由于存在误差，有时又很难说这

些点是不是分布在一条直线的附近，这时就很难判断两个变量之间是否具有相关关系. 这时就必须利用样本相关系数对其进行相关性检验. 计算中应该特别注意细心，不要出现计算错误，也可借助于计算器等进行计算.

2. 思维探究

为什么相关系数能检验两个随机变量之间具有相关关系呢？

这是因为我们用最小二乘法所求出的回归直线方程就是要使残差平方和 $Q(\alpha,\beta)$ 最小，而当

$$\beta=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\text{且 }\alpha=\bar{y}-\beta\bar{x}\text{ 时，}$$

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

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

$$=(1-r^2)\sum_{i=1}^{n}(y_i-\bar{y})^2.$$

这表明，残差平方和 $Q(\alpha,\beta)$ 的大小与 r 密切相关. 因此 $Q\geqslant 0$.

从而 $0\leqslant r^2\leqslant 1$，即 $|r|\leqslant 1$. 并且，当样本相关系数 r 的绝对值愈接近 1 时，$Q(\alpha,\beta)$ 愈接近于零，说明线性回归的效果就越好. 反之，当 $|r|$ 愈接近于 0 时，$Q(\alpha,\beta)$ 愈大，回归的效果就越差.

3. 也可以利用柯西不等式 $(a_1b_1+a_2b_2+\cdots+a_nb_n)^2\leqslant(a_1{}^2+a_2{}^2+\cdots+a_n{}^2)(b_1{}^2+b_2{}^2+\cdots+b_n{}^2)$ 来证明相关系数 $|r|\leqslant 1$.

证明：根据柯西不等式：

$[(x_1-\bar{x})(y_1-\bar{y})+(x_2-\bar{x})(y_2-\bar{y})+(x_3-\bar{x})(y_3-\bar{y})+\cdots+(x_n-\bar{x})(y_n-\bar{y})]^2\leqslant[(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2][(y_1-\bar{y})^2+(y_2-\bar{y})^2+(y_3-\bar{y})^2+\cdots+(y_n-\bar{y})^2]$

即 $\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2\leqslant\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2$

$$\therefore\quad r^2=\frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}\leqslant 1$$

即　$|r|\leqslant 1$

（编者：郯城二中　庞霞）

§1.1　回归分析的基本思想及其初步应用

（第三课时　残差分析）

学海导航

【知识要点】 1. 总偏差平方和；2. 残差；3. 残差平方和；4. 回归平方和；5. 残差分析；6. 残差图；7. 相关指数 R^2.

【学习要求】 1. 了解总偏差平方和、残差、残差平方和、回归平方和公式及其意义；2. 了解判断刻画模型拟合效果的方法，理解相关指数和残差分析的概念；3. 了解偏差平方和分解的思想；4. 残差分析公式多且复杂，一般不要求记，计算量很大，现在一般用计算机解决，只求能明白原理即可.

学习探究

【要点分析】

1. 预报变量的值受解释变量和随机误差的影响，观测值与平均值的差称为解释变量和随机误差的组合效应，数学上，把每个效应（观测值减去总的平均值）的平方加起来，即用 $\sum_{i=1}^{n}(y_i-\bar{y})^2$ 表示总的效应，称为总偏差平方和，它代表了解释变量和随机误差的总效应.

2. 数据点和它在回归直线上相应位置的差异 $(y_i-\hat{y}_i)$ 是随机误差的效应，称 $\hat{e}_i=y_i-\hat{y}_i$ 为残差，$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ 称为残差平方和，它代表了随机误差的效应.

3. 总偏差平方和与残差平方和的差称为回归平方和即 $\sum_{i=1}^{n}(y_i-\bar{y})^2-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$，它代表了解释变量的效应.

4. 偏差平方和分解公式如下：

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2,$$

即总的偏差平方和 = 回归平方和 + 残差平方和. 偏差平方和分解公式的推导过程如下，假设观测数据为 (x_i,y_i)，$i=1,2,\cdots,n$，则

$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum$ [illegible]

[illegible] $(y_i-\hat{y}_i)^2+$

$$2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)\quad (*)$$

而$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)$

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

$$=\sum_{i=1}^{n}\hat{b}(x_i-\bar{x})[(y_i-\hat{a}-\hat{b}\bar{x}-\hat{b}(x_i-\bar{x})]$$

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

$$=\hat{b}(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-\hat{b}\sum_{i=1}^{n}(x_i-\bar{x})^2=0,$$

代入(＊)式可得偏差平方和分解公式.

因此,可以把平方和分解公式解释为:预报变量的变化程度可以分解为由解释变量引起的变化程度与残差变量的变化程度之和.

5. 由平方和分解公式,可得

$$1=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2},$$

移项得

$$\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=R^2,$$

即 $R^2=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

上式说明了在线性回归模型中相关指数与平方和分解公式之间的关系.

用相关指数 R^2 来刻画回归的效果,$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$;$R^2$ 的值越大,说明残差平方和越小,也就是说模型的拟合效果越好. 在线性回归模型中,R^2 表示解释变量对预报变量变化的贡献率. R^2 越接近于1,表示回归的效果越好. 某组数据可能采取几种不同的回归方程进行回归分析,则可以通过比较 R^2 的值来做出选择,即选择 R^2 大的模型作为这组数据的模型.

6. 在研究两个变量间的关系时,首先要根据散点图来粗略判断它们是否线性相关,是否可以用线性回归模型来拟合数据. 然后,通过残差 $\hat{e}_1,\hat{e}_2,\cdots,\hat{e}_n$ 来判断模型拟合的效果,判断原始数据中是否存在可疑数据,这方面的分析工作称为残差分析.

7. 作残差图时,纵坐标为残差,横坐标可以选为编号、样本数据或估计值等,但分析出的结果是一致的.

8. 进行回归分析需注意:

(1)回归方程只适用于所研究的样本的总体;

(2)所建立的回归方程一般都有时间性;

(3)样本的取值范围会影响回归方程的适用范围;

(4)不能期望回归方程得到的预报值就是预报变量的精确值. 事实上,它是预报变量的可能取值的平均值.

【例题分析】

例1　一个车间为了规定工时定额,需要确定加工零件所花费的时间,为此进行了10次试验,测得数据如下:

零件数 x(个)	10	20	30	40	50	60	70	80	90	100
加工时间 y(min)	62	68	75	81	89	95	102	108	115	122

(1)画出散点图,并初步判断是否线性相关;

(2)若线性相关,求出回归直线方程;

(3)计算总偏差平方和、残差及残差平方和;

(4)求出相关指数;

(5)作出残差图;

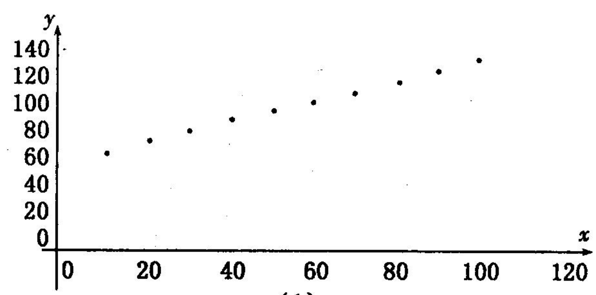
(6)进行残差分析.

分析:本题涉及公式多且复杂,计算量也很大,只求能了解公式明白原理.

解:由 x,y 数据得散点图(1). 由散点图可以认为样本点大致分布在某条直线的附近,因此可以用线性回归模型来拟合. 设回归模型为 $y=a+bx$

由数据计算得 $b=0.668,a=54.93$,则回归直线方程为:

$$\hat{y}=0.668x+54.93$$



(1)

将数据代入相应公式可得如下数据表

零件个数 x(个)	10	20	30	40	50	60	70	80	90	100
加工时间 y(min)	62	68	75	81	89	95	102	108	115	122
$(y_i-\bar{y})^2$	882.1	551.7	278.9	114.5	7.29	10.9	106.1	265.7	542.9	918.1
$\hat{y}=0.668x+54.93$	61.6	68.3	74.9	81.6	88.3	95.0	101.7	108.4	115.1	121.73
残差	0.39	−0.29	0.03	−0.65	0.67	−0.01	0.31	−0.37	−0.05	0.27

$\bar{x}=55\quad \bar{y}=91.7$

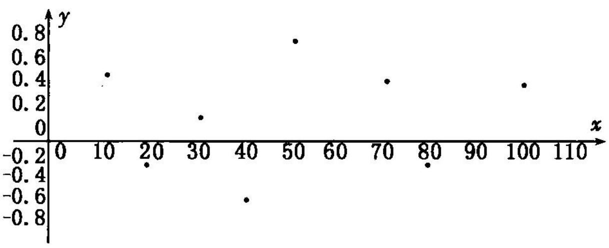
$r=0.9998$

总偏差平方和为 3678.1

残差平方和为 1.417

$R^2=0.9996$

作出残差图(2),横坐标为零件个数,纵坐标为残差



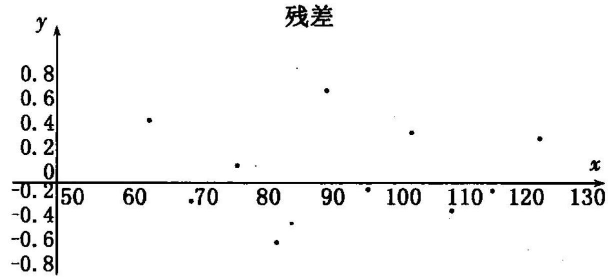
(2)

由 r 的值和散点图都说明 x 与 y 有很强的相关性,由 R^2 的值可以看出回归效果很好,也说明用线性回归模型拟合数据效果很好.

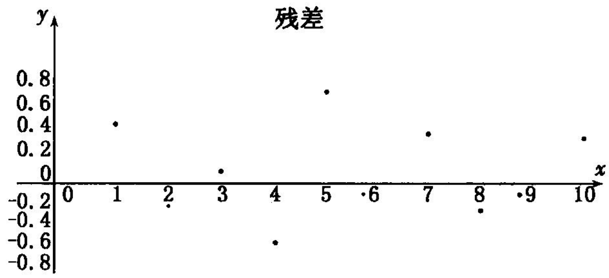
由残差图也可以观察到,第 4 个样本点和第 5 个样本点的残差比较大,需要确认在采集这两个样本点的过程中是否有人为的错误.

点评:(1)残差图的横坐标可以选为样本编号,或加工时间等,如图:

加工时间	62	68	75	81	89
残　差	0.39	−0.29	0.03	−0.65	0.67
加工时间	95	102	108	115	122
残　差	−0.01	0.31	−0.37	−0.05	0.27



序号	1	2	3	4	5	6	7	8	9	10
残差	0.39	−0.29	0.03	−0.65	0.67	−0.01	0.31	−0.37	−0.05	0.27



从上面三个残差图可以观察到,虽然横坐标不同,但分析结果一致.

(2)残差点比较均匀地落在水平的带状区域中,说明选用的模型比较合适.这样的带状区域的宽度越窄,说明模型拟合精度越高,回归方程的预报精确度越高.

例 2　关于 x 与 y 有如下数据:

x	2	4	5	6	8
y	30	40	60	50	70

有如下的两个线性模型:

(1)$\hat{y}=6.5x+17.5$;(2)$\hat{y}=7x+17$.

试比较哪一个拟合效果更好.

解:由(1)得,$y_i-\hat{y}_i$ 与 $y_i-\bar{y}$ 的关系如下表:

$y_i-\hat{y}_i$	−0.5	−3.5	10	−6.5	0.5
$y_i-\bar{y}$	−20	−10	10	0	20

$$\therefore \sum_{i=1}^{5}(y_i-\hat{y}_i)^2=(-0.5)^2+(-3.5)^2+10^2+(-6.5)^2+0.5^2=155,$$

$$\sum_{i=1}^{5}(y_i-\bar{y})^2=(-20)^2+(-10)^2+10^2+0^2+20^2=1\,000.$$

$$\therefore R_1^2=1-\frac{\sum_{i=1}^{5}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{5}(y_i-\bar{y})^2}=1-\frac{155}{1\,000}=0.845.$$

由(2)可得 $y_i-\hat{y}_i$ 与 $y_i-\bar{y}$ 的关系如下表:

$y_i-\hat{y}_i$	-1	-5	8	-9	-3
$y_i-\bar{y}$	-20	-10	10	0	20

$$\therefore \sum_{i=1}^{5}(y_i-\hat{y}_i)^2=(-1)^2+(-5)^2+8^2+(-9)^2+(-3)^2=180,$$

$$\sum_{i=1}^{5}(y_i-\bar{y})^2=(-20)^2+(-10)^2+10^2+0^2+20^2=1\,000.$$

$$\therefore R_2^2=1-\frac{\sum_{i=1}^{5}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{5}(y_i-\bar{y})^2}=1-\frac{180}{1\,000}=0.82,$$

由于 $R_1^2=0.845$，$R_2^2=0.82$，$0.845>0.82$，

$\therefore R_1^2>R_2^2$.

$\therefore$ (1)的拟合效果好于(2)的拟合效果.

自我测评

A组

1. 为研究重量 x（单位：克）对弹簧长度 y（单位：厘米）的影响，对不同重量的6根弹簧进行测量，得如下数据：

x	5	10	15	20	25	30
y	7.25	8.12	8.95	9.90	10.9	11.8

(1)画出散点图；

(2)如果散点图中的各点大致分布在一条直线的附近，求 y 与 x 之间的回归直线方程；

(3)求出残差，进行残差分析.

2. 如果发现散点图中所有样本点都在同一条直线上，请回答下列问题：

(1)解释变量和预报变量的关系是什么？残差平方和是多少？

(2)解释变量和预报变量之间的相关指数多少？

3. 在回归分析中，分析残差能解决哪些问题？

B组

搜集数据，整理你班中加班加点学习与成绩优秀的关系，如果有，在多大程度上相关.

拓展提高

1. 统计已成为人们常识，它几乎渗透到每一个学科中，哪里有实验，哪里有数据，哪里就少不了统计，不懂统计，就无法应付大量信息.

2. 现代社会是信息社会，学会搜集、测量、评价信息作出决策是一个人成功必备的素质.

3. 回归分析就是将相关关系用函数关系近似的描述的方法，就是寻找相关关系中非确定性关系的某种确定性.

4. 相关指数 R^2 与相关系数 r 的关系

在含有一个解释变量的线性回归模型中，相关指数 R^2 恰好等于相关系数 r 的平方，推导如下：

$$R^2=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{\sum_{i=1}^{n}(\hat{b}\cdot x_i+\hat{a}-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

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

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

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

$$=\frac{\sum_{i=1}^{n}((x_i-\bar{x})(y_i-\bar{x}))^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}=r^2.$$

进一步地，由上式以及线性相关系数的性质知：在线性回归模型中有 $0\leqslant R^2\leqslant 1$. 因此，在一元线性回归模型中，相关指数和两个变量的相关系数都能刻画用线性回归模型拟合数据的效果. 相关系数的绝对值越大，相关指数就越大，用线性回归模型拟合数据的效果就越好.

（编者：郯城二中　庞霞）

§1.1 回归分析的基本思想及其初步应用

（第四课时 两个变量的非线性相关）

学海导航

【知识要点】 1. 两个变量的非线性相关；2. 函数的线性化问题

【学习要求】 1. 了解两个变量的非线性相关关系；2. 会将非线性相关的两个变量的关系变换为线性相关关系.

学习探究

【要点分析】

1. 研究两个变量的关系，我们常常根据实验数据作出散点图. 类似于两个变量的线性相关关系. 观察散点图中点的分布，从整体看如果大致不在某一条直线附近，我们就称这两个变量之间不具有线性相关关系.

2. 当回归方程不是形如 $y=bx+a(a、b\in\mathbf{R})$ 时，称之为非线性回归方程.

3. 我们知道，变量的线性相关关系用线性函数 $y=bx+a(a,b\in\mathbf{R})$ 来表示，其图形是直线，研究起来比较方便些，但是在许多自然科学中所遇到的变量关系，并不都是像线性函数关系那样简单. 因此人们在有些情况下常设法用变量代换法，把较复杂的函数化为线性函数，把曲线化为直线，以便于对函数进行计算和研究，这就是所谓函数的线性化问题.

4. 常遇到的幂函数和指数函数线性化问题介绍如下：

（1）将幂函数 $y=ax^m$（a 为正的常数，x,y 取正值）化为线性函数.

如果将 $y=ax^m$ 两边取对数，则有

$\lg y=m\lg x+\lg a$

令 $u=\lg y,v=\lg x,\lg a=b$，代入上式，得

$u=mv+b$

其中 m,b 是常数. 这是 u,v 的线性函数. 如果以 u 为纵坐标，v 为横坐标，则 $u=mv+b$ 的图形就是一直线.

（2）将指数函数 $y=ca^x$（$a>0,c>0$ 常数）化为线性函数.

将 $y=ca^x$ 两边以 10 为底取对数，有

$\lg y=x\lg a+\lg c$，

令 $\lg y=u,\lg a=k,\lg c=b$，得

$u=kx+b.$

其中 k 和 b 是常数，这就是 u 为 x 的线性函数，它的图形是直线，与幂函数不同的是 x 仍保持原来的，只是用 y 的对数 $\lg y$ 代替了 y.

【例题分析】

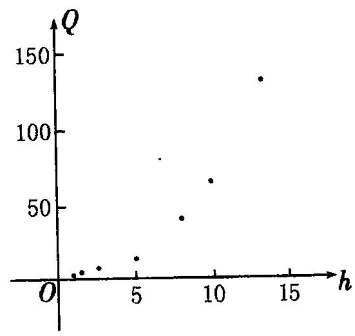
例 1 有一测量水流的实验装置——量水堰，测得试验数据如下表：

i	1	2	3	4	5	6	7
水高 h_i（厘米）	0.7	1.1	2.5	4.9	8.1	10.2	13.5
流量 Q_i（升/分）	0.082	0.25	1.8	11.2	37.5	66.5	134

根据表中数据，建立 Q 与 h 之间回归方程.

分析：首先考查水高与流量接近哪一类函数（如一次函数、二次函数、幂函数、指数函数等等）. 为此最好先作出散点图来观察. 再根据经验选择函数模型.

解：根据测得的数据作出散点图如图：



根据已有的函数知识，可以发现样本点分布在某一条幂函数曲线 $Q=\alpha h^{\beta}$（$\alpha、\beta$ 是待定正的常数）①的周围.

为此将

$Q=\alpha h^{\beta}$ 两边取对数，得到

$\lg Q=\beta\lg h+\lg\alpha$ ②

令 $\lg Q=y,\lg h=x$

于是②式就可化为

$y=\beta x+\lg\alpha$

这样 y 就是 x 的线性函数了. 可以利用线性回归模型来建立 y 和 x 之间的非线性回归方程

$y=bx+a(\beta=b,\lg\alpha=a)$ 了.

先作出下面的数据表：

i	h_i	Q_i	$x_i=\lg h$	$y_i=\lg Q$	x_i^2	x_iy_i
1	0.7	0.082	−0.1549	−1.0862	0.024	0.1683
2	1.1	0.25	0.0414	−0.6021	0.0017	−0.0249
3	2.5	1.8	0.3979	0.2553	0.1583	0.1016
4	4.9	11.2	0.6902	1.0492	0.4764	0.7242
5	8.1	37.5	0.9085	1.5740	0.8254	1.4300
6	10.2	66.5	1.0086	1.8228	1.0173	1.8385
7	13.5	134	1.1303	2.1271	1.2776	2.4043
Σ			$\sum_{i=1}^{7}x_i=4.022$	$\sum_{i=1}^{7}y_i=5.1401$	$\sum_{i=1}^{7}x_i^2=3.7807$	$\sum_{i=1}^{7}x_iy_i=6.642$

由上表得到$\beta=2.5097$，$\lg\alpha=-0.7077$，则$\alpha=0.1960$，

于是所得的回归方程为：

$Q=0.196h^{2.51}$.

点评：从以上例子看到，在建立经验公式时，选择合适的函数类型是十分重要的. 通常是根据实验数据，描出一系列的点，依次把这些点联成平滑的曲线（或者很靠近这些点），从中观察其变化规律，并与已知函数的图形对比，看接近于什么函数，根据实践经验来决定选取公式的类型. 所选的类型是否符合实际，还需通过实践来检验.

自我测评

1. 将函数$y=10x^2(x>0)$变换为线性函数，并作图.

2. 将函数$y=2\cdot10^{\frac{x}{2}}$变换为线性函数，并作图.

3. 将函数$y=a+\dfrac{b}{x}$变换为线性函数.

拓展提高

在实践中，两个变量间的关系可能是线性相关，但在很多情况下呈现出一种“曲线关系”，例如在生物医学的毒理学试验中，毒物的浓度与动物死亡率之间的关系、污染物与其跟污染源距离之间的关系等都是“曲线关系”.

（编者：郯城二中　庞霞）

§1.1　回归分析的基本思想及其初步应用

（第五课时　非线性回归方程及其残差分析）

学海导航

【知识要点】 1. 非线性回归方程；2. 残差分析．

【学习要求】 1. 会求非线性回归方程；2. 了解通过比较残差平方和或相关指数 R^2 进行比较哪种回归方程拟合效果好；3. 了解通过求相关系数进行检验所求回归方程是否有价值；4. 由于计算量大，只求能明白原理，条件允许，可鼓励学生自己进行数据处理.

学习探究

【要点分析】

1. 根据收集的数据，作出散点图，若样本点没有分布在某个带状区域内，则不能直接利用线性回归方程来建立两个变量之间的关系，应根据已有的函数知识，选择一种曲线模型来拟合，应先把非线性关系转化为线性关系，利用线性回归模型来建立 y 和 x 之间的非线性回归方程.

2. 若选用多种模型拟合，可以通过比较两个模型的残差平方和，或相关指数 R^2 来比较拟合效果.

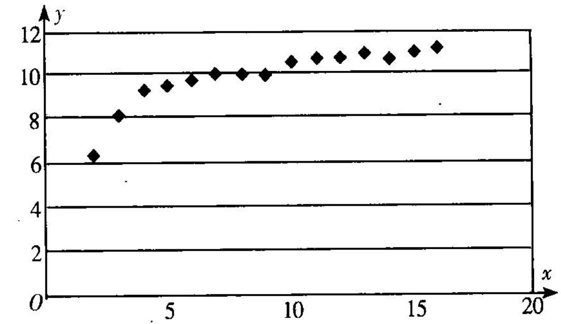
残差平方和越小的拟合效果越好；相关指数 R^2 较大的拟合效果越好.

【例题分析】

例　出钢时所用的盛钢水的钢包，由于钢水对耐火材料的侵蚀，容积不断增大．我们希望找出使用次数与增大容积之间的关系．试验数据如下：

使用次数(t)	2	3	4	5	6	7	8	9
增大容积(y)	6.42	8.2	9.3	9.5	9.7	10	9.93	9.99
使用次数(t)	10	11	12	13	14	15	16	
增大容积(y)	10.49	10.59	10.6	10.8	10.6	10.9	11.11	

解：作散点图如图，由图可知，x、y 之间不存在线性相关关系，但仔细分析一下会发现，钢包侵蚀速度开始较快，然后逐渐减慢．显然钢包容积不会无限增加，它必有一条平行于 x 的渐近线．根据这些特点，我们选用双曲线 $\frac{1}{y}=a+\frac{b}{x}$ 模型来模拟．如果令 $y'=\frac{1}{y}$，$x'=\frac{1}{x}$，那么问题就化为 $y'=a+bx'$ 的线性情形．



次数 x	容积 y	$1/x$	$1/y$	x'^2_1	y'^2_1	$x'_i y'_i$
2	6.42	0.5	0.155 763	0.25	0.024 262	0.077 882
3	8.2	0.333 333	0.121 951	0.111 111	0.014 872	0.040 65
4	9.3	0.25	0.107 527	0.062 5	0.011 562	0.026 882
5	9.5	0.2	0.105 263	0.04	0.011 08	0.021 053
6	9.7	0.166 667	0.103 093	0.027 778	0.010 628	0.017182
7	10	0.142 857	0.1	0.020 408	0.01	0.014 286
8	9.93	0.125	0.100 705	0.015 625	0.010 141	0.012 588
9	9.99	0.111 111	0.100 1	0.012 346	0.010 02	0.011 122
10	10.49	0.1	0.095 329	0.01	0.009 088	0.009 533
11	10.59	0.090 909	0.094 429	0.008 264	0.008 917	0.008 584
12	10.6	0.083 333	0.094 34	0.006 944	0.008 9	0.007 862
13	10.8	0.076 923	0.092 593	0.005 917	0.008 573	0.007 123
14	10.6	0.071 429	0.094 34	0.005 102	0.008 9	0.006 739
15	10.9	0.066 667	0.091 743	0.004 444	0.008 417	0.006 116
16	11.11	0.062 5	0.090 009	0.003 906	0.008 102	0.005 626

由上表计算可得：$\bar{x}'=0.158\ 75$，$\bar{y}'=0.103\ 146$，$\sum_{i=1}^{15}x'^2_i=0.584347$，$\sum_{i=1}^{15}y'^2_i=0.163\ 462$，$\sum_{i=1}^{15}x'_iy'_i=0.273\ 226$. 代入公式可得 $b=0.133\ 977$，$a=0.081\ 881$.

$\therefore$ 其线性回归方程为 $y'=0.133\ 977x'+0.081\ 881$，将 $y'=\frac{1}{y}$，$x'=\frac{1}{x}$ 代入得所求的回归方程