普通高中课程标准实验教材

高中数学

选修 1-2

GAOZHONG SHUXUE

新课标新精编

XINKEBIAO
XINJINGBIAN

主编 胡建军

配人教 A 版

浙江教育出版社
ZHEJIANG JIAOYU CHUBANSHE

普通高中课程标准实验教材

高中数学

选修1-2

新课标新精编

XINKEBIAO XINJINGBIAN

顾　问　岑　申　王而冶　金才华　许芬英
主　编　胡建军
编　者　朱恒元　陈　荣　戴三红　黄琪锋

浙江教育出版社
ZHEJIANG JIAOYU CHUBANSHE

图书在版编目(CIP)数据

新课标新精编:人教版.数学.1-2:选修 / 胡建军编.
—杭州: 浙江教育出版社,2007.10
ISBN 978-7-5338-7196-3

Ⅰ.新... Ⅱ.胡... Ⅲ.数学课 - 高中 - 教学参考资料 Ⅳ.G634

中国版本图书馆 CIP 数据核字(2007)第 145593 号

责任编辑 金馥菊　　封面设计 韩 波
责任校对 余晓克　　责任印务 温劲风

普通高中课程标准实验教材

新课标 新精编　高中数学 选修 1-2

- 主　　编 胡建军

- 出版发行 浙江教育出版社
  (杭州市天目山路 40 号 邮编 310013)
- 图文制作 杭州富春电子印务有限公司
- 印　　刷 杭州富春印务有限公司
- 开　　本 880 × 1230　1/16
- 印　　张 5.25
- 字　　数 165 000
- 印　　数 0 001—5 000
- 版　　次 2007 年 10 月第 1 版
- 印　　次 2007 年 10 月第 1 次
- 标准书号 ISBN 978-7-5338-7196-3
- 定　　价 7.30 元

联系电话:0571-85170300-80928
e-mail:zjjy@zjcb.com　　网址:www.zjeph.com

前　言

高中课程改革正在全国各地逐步展开。其中，高中数学新课程旨在提高学生的科学素养，改变学生的学习方式，从知识与技能、过程与方法、情感态度与价值观三个方面培养学生。为了深入贯彻新课程标准的精神，配合人民教育出版社《普通高中课程标准实验教科书·数学》的顺利使用，帮助学生实现高中数学课程的教育目标，我们组织了教学第一线的数学特级教师和优秀中青年教师，在深入研究了《高中数学课程标准》及其各种版本实验教科书的基础上，编写了这套《新课标新精编高中数学》丛书。

本丛书的编写以“讲求循序渐进，重视科学思想与科学方法，强调实践意识与探究精神，渗透情感态度与价值观的教育”为原则，与人民教育出版社《普通高中课程标准实验教科书·数学》配套。它具有以下几个鲜明的特点：

1. 同步性。本丛书的例题和练习均以课时为基本单位，根据新课程教学的要求和学生学习的特点进行编写，与教学同步，便于教师的教学和学生的使用。

2. 科学性。本丛书根据新课标学习的需要，设置了“学法指导”、“基础例说·基本训练”、“应用·拓展·综合训练”、“自我评估”、“高考链接”等栏目。“学法指导”帮助学生深刻理解教材的重点、难点和目标要求。“基础例说·基本训练”分“例说”和“训练”两部分，“例说”以典型例题为载体，教给学生思考问题、分析问题和解决问题的策略和方法；“训练”目的在于让学生通过训练，巩固所学知识，发展思维能力。“应用·拓展·综合训练”纵览全章，起到复习、巩固、拓展、加强应用和综合训练的作用。“自我评估”为全章知识的综合评估，分A，B两份试卷，其中A卷为基本要求，B卷为较高要求。“高考链接”选取近几年有代表性的高考真题，让学生试做，以同步了解高考命题的基本特点。

3. 层次性。为了适应不同学习水平的学生的不同要求以及学生在不同学习阶段的不同要求，本丛书选编的训练题都分为“A组”和“B组”两组，分别反映了课程的基础性目标和发展性目标，使不同层次的学生都能够充分获益，也符合循序渐进的学习原则。

4. 新颖性。本丛书力求体现新课程的理念，突出数学探究、联系实际，注重激发学生学习的兴趣，力求反映近年来高中数学教学和命题研究的最新成果，所选习题无论是在内容上，还是在形式上，都具有一定的新颖性。

由于时间匆促，加上作者对新课程的认识有待进一步提高，本丛书在编写时难免出现一些不足之处，敬请广大师生指正。

浙江教育出版社

2007年9月

目 录

MULU

第一章　统计案例

学法指导

本章主要内容有：求回归方程和回归分析，独立性检验及其初步应用.

学习目标

1. 通过对实际问题的分析，了解回归分析的必要性与回归分析的一般步骤；会求回归直线方程，了解随机误差的概念及其对预报变量的影响，能进行简单回归分析.

2. 通过典型案例的探究，进一步了解回归分析的基本思想、方法及初步应用. 会根据观测数据的特点来选择回归模型，通过探究有些非线性模型通过变换可以转化为线性回归模型，初步体会不同模型拟合数据的效果.

3. 通过典型案例“吸烟是否与患肺癌有关系”的探究，理解独立性检验的基本思想，并学会运用样本数据的列联表、柱形图和条形图，亲身体验独立性检验的实施步骤与必要性.

重点、难点

重点是理解与运用回归思想，如何求回归方程并会判断方程有无意义，了解线性回归模型与函数模型的差异，了解判断刻画模型拟合效果的方法——相关指数和残差分析；通过探究体会有些非线性模型通过变换可以转化为线性回归模型，了解在解决实际问题的过程中寻找更好的模型的方法；理解独立性检验的基本思想及实施步骤，能运用变量 K^2 来检验.

难点是残差变量的含义；选择不同的函数模型进行回归分析，并通过比较相关指数或残差平方和对不同的模型进行比较；了解独立性检验的基本思想，了解随机变量 K^2 的含义.

主要概念、定理、公式及规律

1. 主要概念

(1) 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.

(2) 线性回归模型 $y=bx+a+e$，其中 a,b 为模型的未知参数，e 称为随机误差.

(3) 样本值与回归值的差叫残差，即 $\hat{e}_i=y_i-\hat{y}_i$.

(4) 通过残差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.

(5) 以残差为纵坐标，以样本编号，或身高数据，或体重估计值等为横坐标，作出的图形称为残差图. 观察残差图，如果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，模型拟合精度越高，回归方程的预报精度越高.

(6) 变量的不同“值”表示个体所属的不同类别的变量称为分类变量. 分类变量的取值一定是离散的，而且不同的取值仅表示个体所属的类别. 如性别变量，只取男、女两个值，商品的等级变量只取一级、二级、三级等. 分类变量的取值有时可用数字来表示，但这时的数字除了分类以外没有其他的含义. 如用“0”表示“男”，用“1”表示“女”.

(7) 列联表是分类变量的汇总统计表(频数表). 一般只研究每个分类变量只取两个值，这样的列联表称为 2×2 列联表. 一般地，对于两个研究对象Ⅰ和Ⅱ，Ⅰ有两类取值，即类 A 和 B(如吸烟与不吸烟)；Ⅱ也有两类取值，即类 1 和 2(如患病与不患病). 于是得到下列联表所示的抽样数据：

	类 1	类 2	总计
类 A	a	b	$a+b$
类 B	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

要推断“Ⅰ和Ⅱ有关系”，可按下面的步骤进行：

第一步，提出假设检验问题. 提出假设 H_0：Ⅰ和Ⅱ没有关系. 如 H_0：吸烟与患肺癌没有关系↔H_1：吸烟与患肺癌有关系.

第二步，选择检验的指标. 根据 2×2 列联表与公式计算 K^2 的值，$K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$.

第三步，查对临界值，作出判断. K^2 越小，原假设“H_0：吸烟与患肺癌没有关系”成立的可能性越大；K^2 越大，备择假设“H_1：吸烟与患肺癌有关系”成立的可能性越大.

$P(K^2>k)$	0.50	0.40	0.25	0.15	0.10
k	0.455	0.708	1.323	2.072	2.706
$P(K^2>k)$	0.05	0.025	0.010	0.005	0.001
k	3.84	5.024	6.635	7.879	10.83

(8) 利用随机变量 K^2 来确定在多大程度上可以认为“两个分类变量有关系”的方法称为两个分类变量的独立性检验.

2. 主要公式及规律

(1) 线性回归方程 $\hat{y}=\hat{b}x+\hat{a}$，其中

$$\begin{cases}\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})\sum\limits_{i=1}^{n}(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\\ =\dfrac{n\sum\limits_{i=1}^{n}x_iy_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{n\sum\limits_{i=1}^{n}x_i^2-(\sum\limits_{i=1}^{n}x_i)^2},\\ \hat{a}=\bar{y}-\hat{b}\bar{x}.\end{cases}$$

求回归直线方程的一般步骤：

①作出散点图(由样本点是否呈条状分布来判断两个量是否具有线性相关关系)，判断它们之间是否存在线性相关关系；

②求回归系数$\hat{a},\hat{b}$；

③写出回归直线方程$\hat{y}=\hat{a}+\hat{b}x$，并利用回归直线方程进行预测说明.

(2) 相关系数 $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})\sum\limits_{i=1}^{n}(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}$

$=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sqrt{(\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2)(\sum\limits_{i=1}^{n}y_i^2-n\bar{y}^2)}}$，$|r|$越接近于1时，线性相关程度越强；$|r|$越接近于0时，线性相关程度越弱. 相关系数的绝对值越接近于1，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精确，此时建立的线性回归模型是有意义的.

(3) 总偏差平方和：所有单个样本值与样本均值差的平方和，即 $\sum\limits_{i=1}^{n}(y_i-\bar{y})^2$.

(4) 残差平方和：回归值与样本值差的平方和，即 $\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2$.

(5) 回归平方和：相应回归值与样本均值差的平方和，即 $\sum\limits_{i=1}^{n}(\hat{y}_i-\bar{y})^2$.

(6) 回归平方和=总偏差平方和-残差平方和.

(7) 回归效果的相关指数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}$.

(8) $K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$.

学习方法指导

1. 注意 $y_i,\hat{y}_i,\bar{y}$三者的区别.

2. 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，即

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

3. 当总偏差平方和相对固定时，残差平方和越小，则回归平方和越大，此时模型拟合的效果越好.

4. 对于多个不同的模型，引入相关指数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}$ 来刻画回归的效果，它表示解释变量对预报变量变化的贡献率. R^2 的值越大，说明残差平方和越小，也就是说模型拟合的效果越好.

5. 在确定具有线性关系后，需建立回归模型. 建立回归模型的基本步骤是：

①确定研究对象，明确哪个变量是解释变量，哪个变量是预报变量；

②画出确定好的解释变量和预报变量的散点图，观察它们之间的关系(线性关系)；

③由经验确定回归方程的类型；

④按一定规则估计回归方程中的参数$\hat{a},\hat{b}$(最小二乘法)；

⑤得出结论后分析残差图是否异常，若存在异常，则检查数据是否有误，或模型是否合适等.

6. 独立性检验的基本思想就是利用小概率事件不会发生的事实来解释的，而它却偏偏发生了，从而否定前面的假设.

基础例说·基本训练

1.1 回归分析的基本思想及其初步应用

第1课时 求线性回归直线方程及回归分析

例说

例1 给出下列关系：

①人的年龄与他(她)的体重之间的关系；

②曲线上的点与该点的坐标之间的关系；

③苹果的产量与气候之间的关系；

④森林中的同一种树木，其断面直径与高度之间的关系；

⑤学生与他(她)的学号之间的关系.

其中有相关关系的是_____.

解 有相关关系的是①③④.

例2 某调查者从调查中获知某公司近年来科研费用支出与公司所获得利润的统计资料如下表：

科研费用支出与利润统计表 (单位：万元)

年 份	科研费用支出	利 润
2001	5	31
2002	11	50

年　份	科研费用支出	利　润
2003	4	30
2004	5	34
2005	3	25
2006	2	20
合　计	30	190

(1) 作利润与科研费用支出的散点图，根据该图猜想它们之间的关系；

(2) 建立科研费用支出为解释变量，利润为预报变量的回归模型；

(3) 根据得到的回归模型，预报科研费用支出为 13 万元时的利润；

(4) 作出残差图，并计算出残差平方和；

(5) 根据得到的回归模型，你认为这个模型能较好地刻画利润和科研费用支出的关系吗？请说明理由.

解　(1) 利润与科研费用支出的散点图如图 1-1 所示.

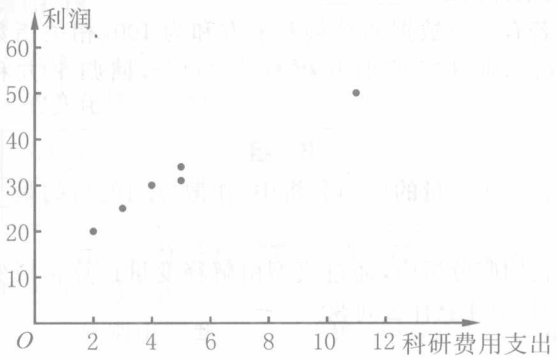


图 1-1

根据图 1-1 可以看出，各点大致在一条直线附近，利润与科研费用支出有大致的线性相关关系.

(2) 科研费用支出为解释变量 x，利润为预报变量 y，设线性回归模型直线方程为 $\hat{y}=\hat{a}+\hat{b}x$.

$\because\ \bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{6}x_i=\dfrac{30}{6}=5$,

$\bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{6}y_i=\dfrac{190}{6}\approx 31.67$,

$$\hat{b}=\frac{n\sum\limits_{i=1}^{n}x_iy_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2}$$

$$=\frac{6\times 1\,110-30\times 190}{6\times 200-30^2}=3.2,$$

$\hat{a}=\bar{y}-\hat{b}\bar{x}=31.67-3.2\times 5=15.67$,

$\therefore$ 利润 y 对科研费用支出 x 的线性回归模型直线方程为 $\hat{y}=15.67+3.2x$.

(3) 把 $x=13$ 代入 $\hat{y}=15.67+3.2x$，得

$\hat{y}=15.67+3.2\times 13=57.27$,

即当科研费用支出为 13 万元时，利润为 57.27 万元.

(4) 残差为

年份	2001	2002	2003	2004	2005	2006
$\hat{y}_i$	31.67	50.87	28.47	31.67	25.27	22.07
y_i	31	50	30	34	25	20
$\hat{e}_i$	0.67	0.87	−1.53	−2.33	0.27	2.07

残差平方和 $=\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum\limits_{i=1}^{n}\hat{e}_i^{\,2}=13.33$.

残差图如图 1-2 所示.

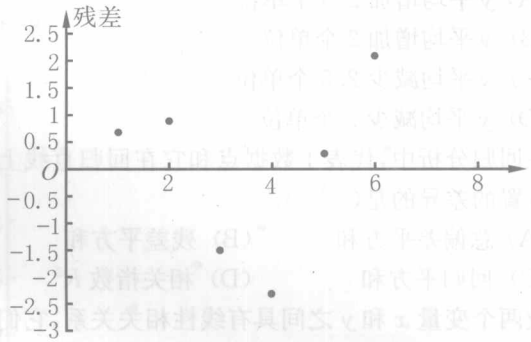


图 1-2

(5) 相关系数为 $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})\sum\limits_{i=1}^{n}(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}\approx 0.987\,2$，这表明利润与科研费用支出有很强的线性相关关系. 同时也可以从残差图看出，由于带状区域比较窄，说明模型拟合精度比较高.

注意　(1) 求回归直线方程的一般步骤：①作出散点图(由样本点是否呈条状分布来判断两个量是否具有线性相关关系)；②若存在线性相关关系，求回归系数 $\hat{a}$，$\hat{b}$；③写出回归直线方程 $\hat{y}=\hat{a}+\hat{b}x$，并利用回归直线方程进行预测说明.

(2) 当相关系数的绝对值越接近于 1，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. 残差平方和越小或残差图的带状区域比较窄，说明模型拟合的效果越好.

训　练

A　组

1. 给出下列结论：

①函数关系是一种确定性关系；

②相关关系是一种非确定性关系；

③回归分析是对具有函数关系的两个变量进行统计分析的一种常用方法；

④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.

其中正确的是(　　).

(A) ①②　　(B) ①②③

(C) ①②④　　(D) ①②③④

2. 在画两个变量的散点图时,下列叙述正确的是(　　).

(A) 预报变量在 x 轴上,解释变量在 y 轴上

(B) 解释变量在 x 轴上,预报变量在 y 轴上

(C) 可以选择两个变量中任意一个在 x 轴上

(D) 可以选择两个变量中任意一个在 y 轴上

3. 设一个回归方程为 $\hat{y}=2-2.5x$,解释变量 x 增加 1 个单位时,则(　　).

(A) y 平均增加 2.5 个单位

(B) y 平均增加 2 个单位

(C) y 平均减少 2.5 个单位

(D) y 平均减少 2 个单位

4. 在回归分析中,代表了数据点和它在回归直线上相应位置的差异的是(　　).

(A) 总偏差平方和　　(B) 残差平方和

(C) 回归平方和　　(D) 相关指数 R^2

5. 设两个变量 x 和 y 之间具有线性相关关系,它们的相关系数是 r,y 关于 x 的回归直线的斜率是 b,纵截距是 a,那么必有(　　).

(A) a 与 r 的符号相反　　(B) a 与 r 的符号相同

(C) b 与 r 的符号相反　　(D) b 与 r 的符号相同

6. 已知回归直线的斜率的估计值是 1.23,样本点的中心为(4,5),则回归直线的方程是(　　).

(A) $\hat{y}=1.23x+4$　　(B) $\hat{y}=1.23x+5$

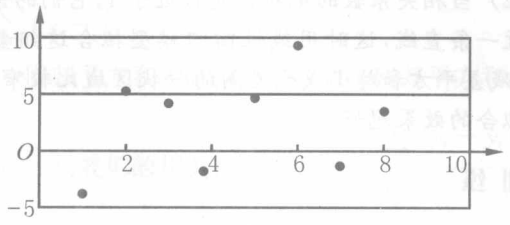
(C) $\hat{y}=1.23x+0.08$　　(D) $\hat{y}=0.08x+1.23$

7. 在回归分析中,相关指数 R^2 的值越大,说明残差平方和(　　).

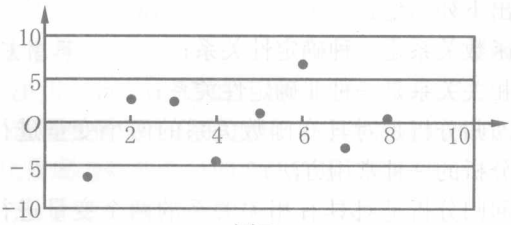
(A) 越小　　(B) 越大

(C) 可能大也可能小　　(D) 以上都不对

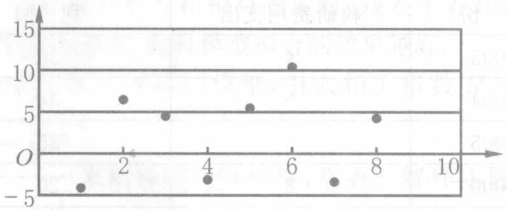
8. 下面是四个回归方程的残差图,它们中模型拟合精度最高的是(　　).



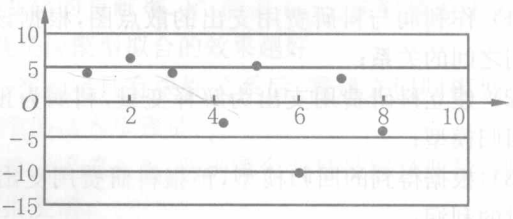
图甲



图乙



图丙



图丁

(第 8 题)

(A) 图甲　　(B) 图乙

(C) 图丙　　(D) 图丁

9. 设一个回归方程为 $\hat{y}=0.5x-0.81$,则当 $x=25$ 时,y 的估计值是_____.

10. 若有一组数据的总偏差平方和为 100,相关指数为 0.5,则其残差平方和为_____,回归平方和为_____.

B 组

11. 在两个变量的回归分析中,作散点图的目的是_____.

12. 在回归分析中,通过模型由解释变量计算预报变量时,应注意什么问题?

13. 某 5 名学生的数学成绩和化学成绩如下表:

数学成绩 x/分	88	76	73	66	63
化学成绩 y/分	78	65	71	64	61

(1) 画出散点图;

(2) 若 x,y 成线性相关,求:

①y 对 x 的线性回归方程;

②x 对 y 的线性回归方程.

14. 为研究质量 x(g)对弹簧长度 y(cm)的影响，对挂有不同质量的 6 根弹簧进行测量，数据如下表：

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) 对 y，x 两个变量进行相关性检验；

(4) 画出残差图，并说明它是否异常.

第 2 课时　求非线性回归直线方程及回归分析

例说

例 3　为了研究某种细菌随时间 x 的变化，其繁殖个数 y 的变化情况，收集数据如下：

天数 x/天	1	2	3	4	5	6
繁殖个数 y/个	6	12	25	49	95	190

(1) 用天数作解释变量，繁殖个数作预报变量，作出这些数据的散点图；

(2) 分别用模型 $y=c_1e^{c_2x}$ 和 $y=c_3x^2+c_4$ 来拟合 y 和 x 之间的关系；

(3) 运用残差分析来比较两者的拟合效果.

分析　非线性回归问题有时并不给出经验公式，可以画出已有数据的散点图，把它与必修模块《数学 1》中学过的各种函数(幂函数、指数函数、对数函数、二次函数等)图象比较，挑选一种跟这些点拟合最好的函数，然后采取适当的置换，把问题化为线性回归问题，使其得到解决.

解　(1) 散点图如图 1-3 所示.

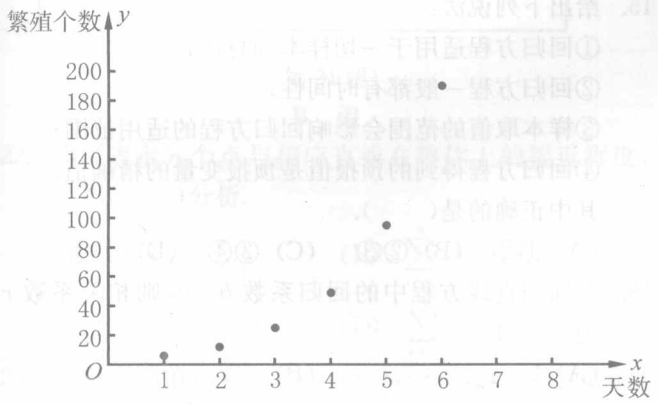


图 1-3

(2) 从散点图可以看出，样本点并没有分布在某条直线附近，因此两个变量不呈线性相关关系，所以不能直接利用线性回归方程来建立两个变量之间的关系. 样本点分布在某一条指数函数 $y=c_1e^{c_2x}$ 周围. 令 $z=\ln y$，$a=\ln c_1$，$b=c_2$，则 $z=bx+a$. 得

x	1	2	3	4	5	6
$\ln y$	1.79	2.48	3.22	3.89	4.55	5.25

由图 1-4 可以看出，变换后的样本点分布在一条直线的附近，得到 $\hat{z}=0.690\,2x+1.115\,8$，因此细菌繁殖个数与天数的非线性回归方程为 $\hat{y}=e^{0.690\,2x+1.115\,8}$，且 $R^2=0.999\,8$.

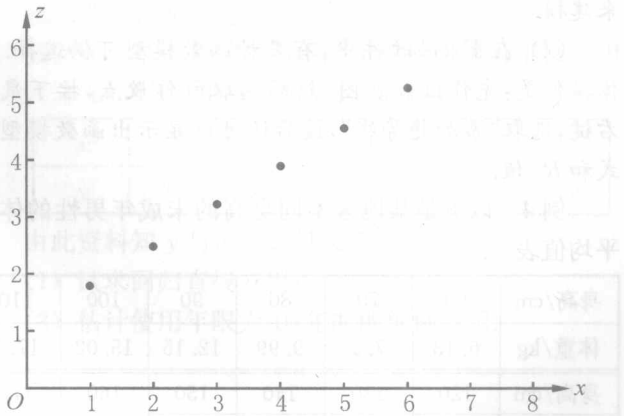


图 1-4

另一方面，令 $t=x^2$，则

t	1	4	9	16	25	36
y	6	12	25	49	95	190

散点图如图 1-5 所示.

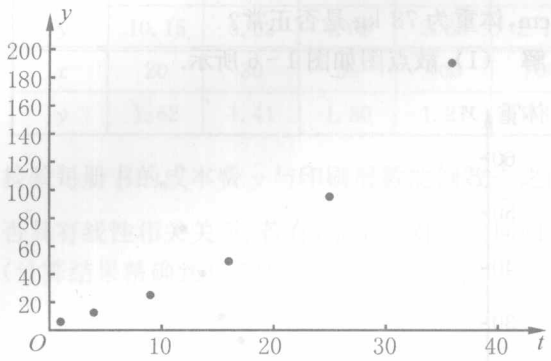


图 1-5

先可以通过上面的数据算出 y 与 x 的二次回归模型 $y=5.096\,1x^2-14.458$.

从图 1-5 可以看出 y 与 t 的散点图并不分布在一条直线的附近，因此不宜用线性回归方程来拟合它，即不宜用二次曲线 $y=c_3x^2+c_4$ 来拟合 y 和 x 之间的关系.

(3) 下面通过残差分析来比较两个不同拟合模型的拟合效果.

x	1	2	3	4	5	6
y	6	12	25	49	95	190
$\hat{e}^{(1)}$	−0.086	−0.136	0.800	0.742	−1.232	−1.898
$\hat{e}^{(2)}$	15.362	6.074	−6.407	−18.080	−17.945	20.998

残差的平方和分别为$\hat{Q}^{(1)}=6.337$，$\hat{Q}^{(2)}=1\,542.405$.

利用残差分析可以明显地看出指数模型比二次函数模型拟合的程度更好.

注意 (1) 若散点图中的点分布在一个直线状带形区域，则可以选线性回归模型来建模；若散点图中的点分布在一个曲线状带形区域，则需选择非线性回归模型来建模.

(2) 在Excel软件中，有多种函数模型可供选择. 具体操作是：先作出散点图，然后选取所作散点，按下鼠标右键，选取“添加趋势线”，最后还可以显示出函数模型公式和R^2值.

例4 以下是某地区不同身高的未成年男性的体重平均值表.

身高/cm	60	70	80	90	100	110
体重/kg	6.13	7.9	9.99	12.15	15.02	17.5
身高/cm	120	130	140	150	160	170
体重/kg	20.92	26.86	31.11	38.85	47.25	55.05

(1) 作出散点图；

(2) 分别用线性回归模型和指数模型进行拟合，哪个回归方程拟合效果更好？

(3) 若体重超过相同身高男性平均值的1.2倍为偏胖，低于0.8为偏瘦，则该地区某中学一男生身高为175 cm，体重为78 kg是否正常？

解 (1) 散点图如图1-6所示.

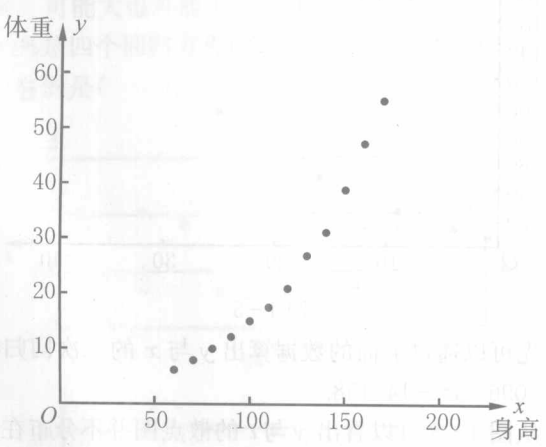


图1-6

(2) 利用Excel可以画出拟合的线性回归模型和指数模型(如图1-7、图1-8所示)，从而求出两个函数模型的表达式.

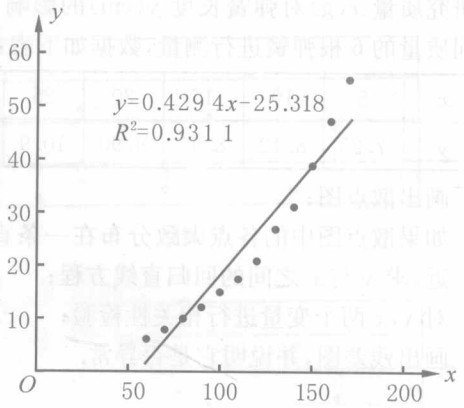


图1-7

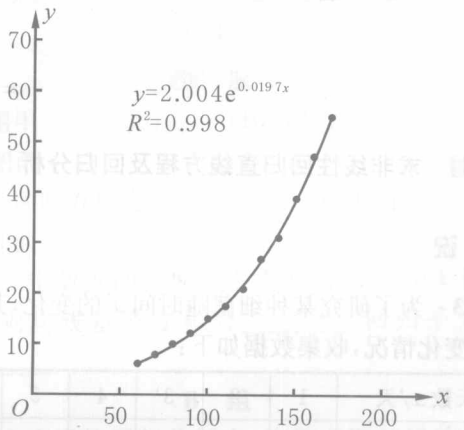


图1-8

两个回归方程分别是$y=0.429\,4x-25.318$，$y=2.004\mathrm{e}^{0.019\,7x}$，

对应的相关指数分别是$R_1^2=0.931\,1$，$R_2^2=0.998$.

从散点图和相关指数都可以看出，指数模型$y=2.004\mathrm{e}^{0.019\,7x}$的拟合效果好.

(3) 把$x=175$代入$y=2.004\mathrm{e}^{0.019\,7x}$，得$y=62.97$.

由于$78\div62.97\approx1.24>1.2$，因此这名男生体型偏胖.

训练

A 组

15. 给出下列说法：

①回归方程适用于一切样本和总体；

②回归方程一般都有时间性；

③样本取值的范围会影响回归方程的适用范围；

④回归方程得到的预报值是预报变量的精确值.

其中正确的是(　　).

(A) ①②　(B) ②③　(C) ③④　(D) ①③

16. 若回归直线方程中的回归系数$b=0$，则相关系数r为(　　).

(A) 1　　(B) −1

(C) 0　　(D) 无法确定

17. 一位母亲记录了儿子3岁～9岁的身高，数据略，由

此建立的身高与年龄的回归模型为 $y=7.19x+73.93$. 若用这个模型预测该孩子 10 岁时的身高，则下列叙述正确的是(　　).

(A) 身高一定是 145.83 cm

(B) 身高在 145.83 cm 以上

(C) 身高在 145.83 cm 左右

(D) 身高在 145.83 cm 以下

18. 在两个变量 y 与 x 的回归模型中，分别选择了 4 个不同的模型，它们的相关指数 R^2 如下，其中拟合效果最好的模型是(　　).

(A) 模型 1 的相关指数 R^2 为 0.98

(B) 模型 2 的相关指数 R^2 为 0.80

(C) 模型 3 的相关指数 R^2 为 0.50

(D) 模型 4 的相关指数 R^2 为 0.25

19. 已知 x 与 y 之间的一组数据如下：

x	0	1	2	3
y	1	3	5	7

则 y 与 x 的线性回归方程为 $y=bx+a$ 必过(　　).

(A) 点(2,2)　　(B) 点(1.5,0)

(C) 点(1,2)　　(D) 点(1.5,4)

20. 在回归分析中，残差平方和的值越大，说明相关指数 R^2 (　　).

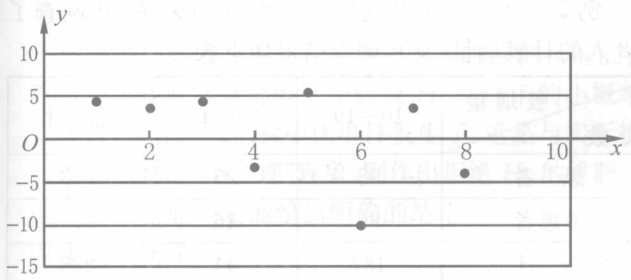
(A) 越小　　(B) 越大

(C) 可能大也可能小　　(D) 以上都不对

21. 通过残差图发现在采集样本点过程中，样本点数据不准确的是(　　).

(A) 第 4 个　　(B) 第 5 个

(C) 第 6 个　　(D) 第 8 个



(第 21 题)

B 组

22. 为了表示 n 个点与相应直线在整体上的接近程度，常用(　　)分析.

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

23. 一组观测值 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 之间满足 $y_i=bx_i+a+e_i$ $(i=1,2,\cdots,n)$. 若 e_i 恒为 0，则 R^2 为_____.

24. 关于 x 与 y，有如下数据：

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

现有两个模型：(1) $\hat{y}=6.5x+17.5$；(2) $\hat{y}=7x+17$. 通过残差分析发现第(1)个线性模型比第(2)个拟合效果好，则 R_1^2 _____ R_2^2，Q_1 _____ Q_2(用“>”、“<”填空，R^2，Q 分别是相关指数和残差平方和).

25. 若发现散点图中所有的样本点都在一条直线上，则残差平方和等于_____，解释变量和预报变量之间的相关系数等于_____.

26. 已知相关指数 $R^2=0.75$，随机误差平方和为 100，则残差平方和为_____.

27. 假设关于某设备的使用年限 x 的所支出的维修费用 y(万元)有如下的统计数据：

x	2	3	4	5	6
y	2.2	3.8	5.5	6.5	7.0

由此资料知 y 与 x 呈线性关系.

(1) 试求回归直线方程；

(2) 估计使用年限为 10 年时的维修费用.

28. 某种书每册的成本费 y(元)与印刷册数 x(千册)有关，经统计得到数据如下：

x	1	2	3	5	10
y	10.15	5.52	4.08	2.85	2.11
x	20	30	50	100	200
y	1.62	1.41	1.30	1.21	1.15

检验每册书的成本费 y 与印刷册数的倒数 $\frac{1}{x}$ 之间是否具有线性相关关系. 若有，求出 y 对 x 的回归方程(计算结果精确到 0.001).

1.2 独立性检验的基本思想及其初步应用

例说

例1 在研究色盲与性别的关系调查中，调查了男性480人，其中有38人患色盲，调查的520个女性中6人患色盲.

(1) 根据以上的数据建立一个2×2的列联表；

(2) 画出列联表的三维柱形图、二维条形图和等高条形图；

(3) 若认为“性别与患色盲有关系”，则出错的概率会是多少？

解 (1) 2×2的列联表如下：

	患色盲	不患色盲	总 计
男	38	442	480
女	6	514	520
总 计	44	956	1 000

(2) 三维柱形图如图1-9所示.

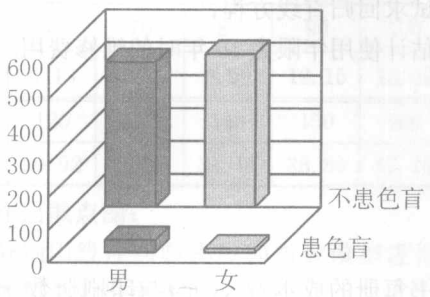


图1-9

二维条形图如图1-10所示.

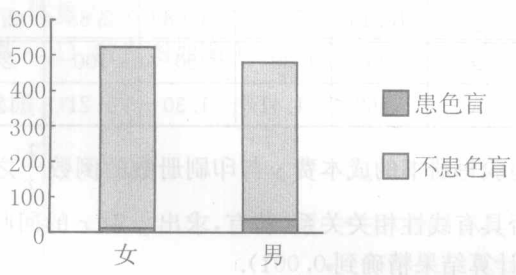


图1-10

等高条形图如图1-11所示.

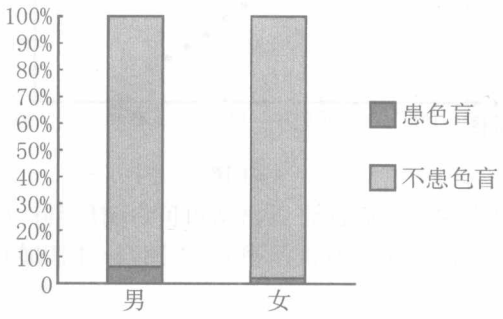


图1-11

从二维条形图和等高条形图可以看出男性患色盲的百分比高.

(3) 假设 H_0：“性别与患色盲没有关系”.

构造一个随机变量 $K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中 $n=a+b+c+d$.

先算出 K^2 的观测值：

$k=\dfrac{1\,000\times(38\times514-442\times6)^2}{480\times520\times44\times956}\approx27.14$，

则有 $P(K^2\geqslant10.808)=0.001$，

即 H_0 成立的概率不超过0.001.

若认为“性别与患色盲有关系”，则出错的概率为0.001.

注意 (1) 列联表中的数据是样本数据，它只是总体的代表，具有随机性，故需要用列联表检验的方法确认所得结论在多大程度上适用于总体.

(2) 独立性检验的原理(与反证法类似)：

反证法	假设检验
要证明结论 A	备择假设 H_1
在 A 不成立的前提下进行推理	在 H_1 不成立的条件下，即 H_0 成立的条件下进行推理
推出矛盾，意味着结论 A 成立	推出有利于 H_1 成立的小概率事件(概率不超过 α 的事件)发生，意味着 H_1 成立的可能性(可能性为 $1-\alpha$)很大
没有找到矛盾，不能对 A 下任何结论，即反证法不成功	推出有利于 H_1 成立的小概率事件不发生，接受原假设

(3) 由于抽样的随机性，由样本得到的推断有可能正确，也有可能错误. 利用 K^2 进行独立性检验，可以对推断的正确性的概率作出估计，样本量 n 越大，估计越准确.

例2 为了研究患气管炎与吸烟的关系，共调查了228人的日吸烟量(支)，调查结果如下表：

人数 \ 吸烟量	10～19	20～40	合 计
患 者	98	25	123
非患者	89	16	105
合 计	187	41	228

由“X 与 Y 有关系”的可信程度为

$P(K^2\geqslant k)$	0.50	0.40	0.25	0.15	0.10
k	0.455	0.708	1.323	2.072	2.706
$P(K^2\geqslant k)$	0.05	0.025	0.010	0.005	0.001
k	3.841	5.024	6.635	7.879	10.828

试说明患气管炎与吸烟的关系.

解 假设 H_0：“患气管炎与吸烟没有关系”.

由表中数据计算得 $K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}\approx0.994$.

因为 $K^2 \approx 0.994$ 及 $K^2 < 1.323$，查表可得 H_0 成立的概率超过 0.25，所以没有充分的理由认为“患气管炎与吸烟有关系”.

例 3　在对人们休闲方式的一次调查中，共调查了 124 人，其中女性 70 人，男性 54 人. 女性中有 43 人主要的休闲方式是看电视，另外 27 人主要的休闲方式是运动；男性中有 21 人主要的休闲方式是看电视，另外 33 人主要的休闲方式是运动.

(1) 根据以上数据建立一个 2×2 列联表；

(2) 判断性别与休闲方式是否有关系.

解　(1) 2×2 列联表为

性别 \ 休闲方式	看电视	运　动	总　计
女	43	27	70
男	21	33	54
总　计	64	60	124

(2) 假设“休闲方式与性别无关”，

计算 $k=\dfrac{124\times(43\times33-27\times21)^2}{70\times54\times64\times60}\approx6.201$.

因为 $k \geqslant 5.024$，所以有理由认为假设“休闲方式与性别无关”是不合理的，即有 97.5% 的把握认为“休闲方式与性别有关”.

训　练

A　组

1. 独立性检验中的统计假设就是假设相关事件 A, B (　　).

(A) 互斥　　(B) 不互斥

(C) 相互独立　　(D) 不独立

2. 给出下列说法：

①独立性检验的基本思想是带有概率性质的反证法；

②独立性检验就是选取一个假设 H_0 条件下的小概率事件，若在一次试验中该事件发生了，这是与实际推断相抵触的“不合理”现象，则作出拒绝 H_0 的推断；

③独立性检验一定能给出明确的结论.

其中正确的是(　　).

(A) ①②　　(B) ①③

(C) ②③　　(D) ①②③

3. 独立性检验中的“小概率事件”是指某事件发生的概率(　　).

(A) 小于 4%　　(B) 小于 5%

(C) 小于 6%　　(D) 小于 8%

4. 给出假设 H_0，下列结论不能对 H_0 成立与否作出明确判断的是(　　).

(A) $K^2=2.536$　　(B) $K^2=7.278$

(C) $K^2=11.341$　　(D) $K^2=20.135$

5. 研究某新药的疗效，给 50 个患者服用此药，跟踪调查后得如下表的数据. 设 H_0：“服用此药的效果与患者的性别无关”，则 $K^2=$_____，从而得出结论_____.

	无　效	有　效	合　计
男性患者	15	35	50
女性患者	4	46	50
合　计	19	81	100

6. 在性别与吃零食这两个分类变量的计算中，下列说法正确的是_____.

①若 K^2 的观测值为 $k=6.635$，我们有 99% 的把握认为吃零食与性别有关系，那么在 100 个吃零食的人中必有 99 人是女性；

②从独立性检验可知有 99% 的把握认为吃零食与性别有关系时，我们说某人吃零食，那么此人是女性的可能性为 99%；

③若从统计量中求出有 99% 的把握认为吃零食与性别有关系，是指有 1% 的可能性使得出的判断出现错误.

7. 下列关于 K^2 的说法，正确的是_____.

①K^2 在任何相互独立问题中都可以用于检验是否相关；

②K^2 越大，两个事件的相关性越大；

③K^2 是用来判断两个相互独立事件相关与否的一个统计量，它可以用来判断两个事件是否相关这一类问题.

8. 某医疗机构为了了解肝病与酗酒是否有关，对成年人进行了一次随机抽样调查，结果如下表. 画出列联表的三维柱形图、二维条形图和等高条形图，从直观上你能得到哪些结论？

	患肝病	未患肝病	合　计
酗　酒	30	170	200
不酗酒	20	280	300
合　计	50	450	500

B 组

9. 对196个接受心脏搭桥手术的病人和196个接受血管清障手术的病人进行3年跟踪研究，调查他们是否发作过心脏病，调查结果如下表：

	发作过心脏病	未发作过心脏病	合　计
心脏搭桥手术	39	157	196
血管清障手术	29	167	196
合　计	68	324	392

试根据上述数据比较两种手术对病人发作心脏病的影响有没有差别.

10. 在对人们饮食习惯的一次调查中，共调查了124人，其中60岁以上的70人，60岁以下的54人. 60岁以上的人中有43人的饮食以蔬菜为主，另外27人则以肉类为主；60岁以下的人中有21人的饮食以蔬菜为主，另外33人则以肉类为主.

(1) 根据以上数据建立一个2×2的列联表；

(2) 判断人的饮食习惯是否与年龄有关.

11. 在500人身上试验某种血清预防感冒作用，把他们一年中的感冒记录与另外500名未用血清的人的感冒记录作比较，结果如下表：

	未感冒	感　冒	合　计
使用血清	258	242	500
未使用血清	216	284	500
合　计	474	526	1 000

则该种血清能否起到预防感冒的作用?

应用·拓展·综合训练

例说

例1 为了调查患慢性气管炎是否与吸烟有关，调查了339名50岁以下的人，调查结果如下表：

	患慢性气管炎	未患慢性气管炎	合　计
吸　烟	43	162	205
不吸烟	13	121	134
合　计	56	283	339

试讨论患慢性气管炎是否与吸烟有关.

分析 独立性检验问题：抽取样本→提出统计假设→运用 K^2 检验.

解 根据2×2列联表，得

$$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}\approx 7.469>6.635,$$

查表可知，有99%的把握认为“患慢性气管炎与吸烟有关”.

例2 一个人从出生到死亡，在每个生日都测量身高，并作出这些数据的散点图，这些点将不会落在一条直线上，但在一段时间内的增长数据有时可以用线性回归来分析. 下表是一位母亲给儿子作的成长记录：

年龄/周岁	3	4	5	6	7	8	9
身高/cm	90.8	97.6	104.2	110.9	115.6	122.0	128.5
年龄/周岁	10	11	12	13	14	15	16
身高/cm	134.2	140.8	147.6	154.2	160.9	167.5	173.0

(1) 年龄和身高之间具有怎样的相关关系?

(2) 若年龄相差5岁，则身高有多大差异(3岁～16岁)?

(3) 若身高相差20 cm，其年龄相差多少?

解 (1) $y=6.317x+71.984$.

(2) 若年龄相差5岁，

身高差异 $\Delta y=6.317\times 5=31.585$(cm).

(3) 若身高相差20 cm，

年龄相差 $\Delta x=\frac{20}{6.317}\approx 3$(岁).

例3 某城市理论预测2005年到2010年人口总数与年份的关系如下表：

年份 x	2005	2006	2007	2008	2009	2010
人口数 y/万	50	69	88	110	190	350

(1) 画出散点图，试建立 y 与 x 之间的回归方程；

(2) 据此估计2011年人口总数；

(3) 计算相关指数 R^2、残差、残差平方和.

解 (1) 散点图如图1-12所示.

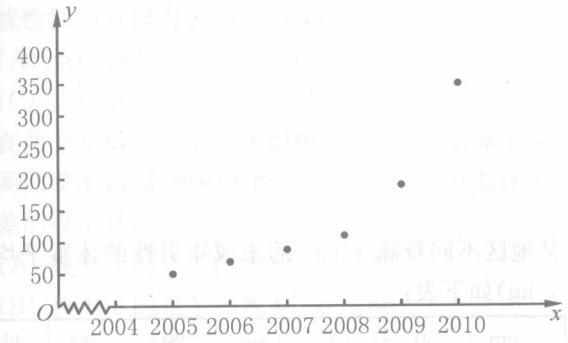


图 1-12

由图 1-12 知，样本点分布在某一条指数函数曲线 $y=c_1e^{c_2x}$ 的周围. 令 $z=\ln y$，则 $z=bx+a$ （$a=\ln c_1$，$b=c_2$），得到变换后的数据如下表：

x	2005	2006	2007	2008	2009	2010
z	3.912	4.234	4.477	4.700	5.247	5.858

作散点图如图 1-13 所示.

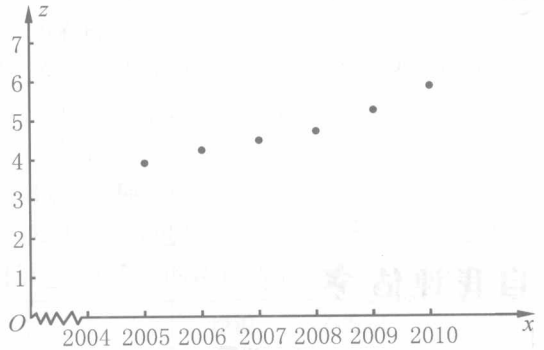


图 1-13

即知变换后的样本分布在一条直线附近，故可以用线性回归方程来拟合.

由计算器计算得线性回归方程为

$\hat{z}=0.371\ 2x-740.45$，

因此，年份 x 与人口数 y（万）之间的非线性回归方程为 $\hat{y}=e^{0.371\ 2x-740.45}$.

（2）估计 2011 年人口总数应为 417.047 万.

（3）相关指数 $R^2=0.9\ 571$，

年份 x	2005	2006	2007	2008	2009	2010
人口数 y/万	50	69	88	110	190	350
$\hat{y}_i$	44.970	65.183	94.481	137.948	198.502	287.724
$\hat{e}_i$	5.030	3.817	−6.481	−27.948	−8.502	62.276

即知残差平方和为 4 758.705.

注意 回归分析：抽取样本→提出统计假设→运用 r 检验.

相关系数公式：$r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}$

$=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sqrt{(\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2)(\sum\limits_{i=1}^{n}y_i^2-n\bar{y}^2)}}$.

r 的性质：$|r|\leqslant 1$，若 $|r|$ 越接近于 1，则线性相关程度越强；若 $|r|$ 越接近于 0，则线性相关程度越弱.

综合训练

A 组

1. 身高与体重有关系可以用（　　）分析来解决.
   （A）残差　　（B）回归
   （C）二维条形图　　（D）独立检验
2. 下列关系为线性相关关系的是（　　）.
   （A）圆周长与半径的关系
   （B）体重与身高的关系
   （C）物体加速度与力的关系
   （D）头发长短与成绩的关系
3. 在两个变量 y 与 x 的回归模型中，分别选择了 4 个不同模型，它们的残差平方和 σ^2 如下，其中拟合效果最好的模型是（　　）.
   （A）模型 1 的 σ^2 为 5.98
   （B）模型 2 的 σ^2 为 32.16
   （C）模型 3 的 σ^2 为 120.51
   （D）模型 4 的 σ^2 为 245.65
4. 工人月工资（元）依劳动生产率（千元）变化的回归直线方程为 $\hat{y}=60+90x$，下列判断正确的是（　　）.
   （A）劳动生产率为 1 000 元时，工资为 50 元
   （B）劳动生产率提高 1 000 元时，工资提高 150 元
   （C）劳动生产率提高 1 000 元时，工资提高 90 元左右
   （D）劳动生产率为 1 000 元时，工资为 90 元
5. 为研究变量 x 和 y 的线性相关性，甲、乙两人分别作了研究，利用线性回归方法得到回归直线 l_1 和 l_2，两人计算知 $\bar{x}$ 相同，$\bar{y}$ 也相同. 下列正确的是（　　）.
   （A）直线 l_1 与直线 l_2 重合
   （B）直线 l_1 与直线 l_2 一定平行
   （C）直线 l_1 与直线 l_2 相交于点 $(\bar{x},\bar{y})$
   （D）无法判断直线 l_1 和直线 l_2 是否相交
6. 考察棉花种子经过处理跟得病之间的关系得到如下表数据：

	种子处理	种子未处理	合　计
得病棉花种子粒数	32	101	133
不得病棉花种子粒数	61	213	274
合　计	93	314	407