

经全国中小学教材审定委员会
2005 年初审通过

普通高中课程标准实验教科书

数 学

选 修 1-2

人民教育出版社 课程教材研究所 编著
中学数学课程教材研究开发中心



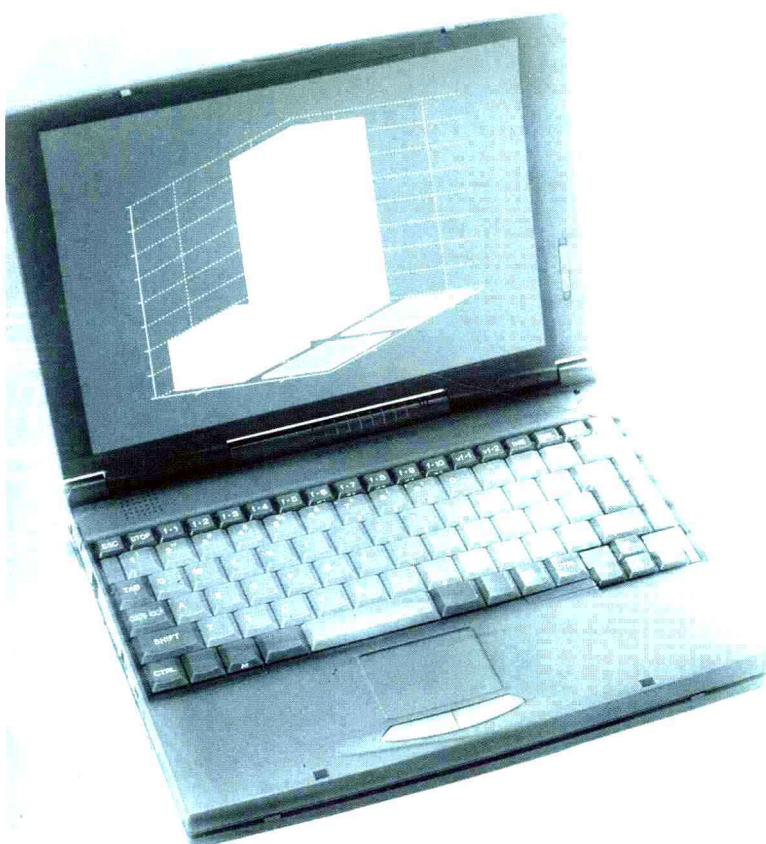
人民教育出版社
版

普通高中课程标准实验教科书

数学

选修 1-2

人民教育出版社 课程教材研究所 编著
中学数学课程教材研究开发中心



普通高中课程标准实验教科书

数 学

选修 1-2

A 版

人民教育出版社 课程教材研究所
中学数学课程教材研究开发中心 编著

*

人民教育出版社 出版

(北京市海淀区中关村南大街 17 号院 1 号楼 邮编: 100081)

网址: <http://www.pep.com.cn>

浙江省出版总社代印

浙江省新华书店发行

浙江新华数码印务有限公司印装

*

开本: 890 毫米×1 240 毫米 1/16 印张: 6 字数: 133 000

2007 年 1 月第 2 版 2011 年 5 月浙江第 5 次印刷

印数: 518 501-606 500 册

ISBN 978-7-107-20256-8 定价: 5.67 元
G·13306 (课)

著作权所有·请勿擅用本书制作各类出版物·违者必究
如发现印、装质量问题, 请与本厂联系。电话: 0571-85155604

主 编：刘绍学

副 主 编：钱珮玲 章建跃

本册主编：李建华

主要编者：李 勇 张淑梅 宋莉莉 杨照宇 蒋佩锦 李建华

责任编辑：宋莉莉

美术编辑：王俊宏 王 艾

本 册 导 引

我们根据《普通高中数学课程标准（实验）》编写了这套实验教科书。本书是高中数学选修课程5个模块中的一个，包括“统计案例”“推理与证明”“数系的扩充与复数的引入”和“框图”四章内容。

在必修课程中，同学们已经学习了最基本的获取样本数据的方法，从样本数据中提取信息的一些统计方法，其中包括用样本估计总体分布及数字特征、线性回归等内容。在本书的第一章中，同学们将通过对典型案例的讨论，了解一些最常用的统计思想方法和统计模型，如回归分析和独立性检验等，进而体会统计思想在解决实际问题中的作用。理解和利用这些统计思想方法和统计模型，对同学们处理未来生活和工作中的某些问题是非常有用的。

第二章将通过生活实例和数学实例，介绍合情推理和演绎推理的涵义，并学习如何利用合情推理去猜测和发现一些新结论，探索 and 提供解决一些问题的思路 and 方向，如何利用演绎推理去进行一些简单的推理，证明一些数学结论，等等。这一章还将介绍证明的两类基本方法——直接证明 and 间接证明，通过数学实例说明它们的思考过程 and 特点等。通过这一章的学习，同学们不仅可以学到如何猜测，也可以学到如何证明。

在第三章中，同学们将在问题情境中了解数系扩充的过程以及引入复数的必要性，学习复数的一些基本知识，体会人类理性思维在数系扩充中的作用。

在必修课程中，同学们已经学习了程序框图，了解了程序框图是表达算法的一种重要方法。实际上，一般意义下的框图在实际生活和科学技术上有着更广泛的应用。框图包括流程图 and 结构图，流程图通常用来描述动态过程，结构图一般用来表达系统结构。在第四章中，同学们将通过丰富的实例，进一步学习和了解框图，体会它在直观清晰地表达 and 交流思想过程中的重要作用。

学习始于疑问。在本书中，我们将通过适当的问题情境，引出需要学习的数学内容，然后在“观察”“思考”“探究”等活动中，引导同学们自己发现问题、提出问题，通过亲身实践、主动思维，经历不断的从具体到抽象、从特殊到一般的抽象概括活动来理解和掌握数学基础知识，打下坚实的数学基础。

学而不思则罔。只有通过自己的独立思考，并掌握科学的思维方法才能真正学会数学。在本书中，我们将利用数学内容之间的内在联系，特别是蕴涵在数学知识中的数学思想方法，启发 and 引导同学们学习类比、推广、特殊化、化归等数学思考的常用逻辑方法，使大家学会数学思考与推理，不断提高数学思维能力。

学习的目的在于应用. 在本书中, 我们将努力为同学们提供应用数学知识解决各种数学内外问题的机会, 以使同学们加深对数学概念本质的理解, 认识数学知识与实际的联系, 并学会用数学知识和方法解决一些实际问题. 另外, 我们还开辟了“观察与猜想”“阅读与思考”“探究与发现”“信息技术应用”等拓展性栏目, 为大家提供选学素材, 有兴趣的同学可以自主选择其中的一些内容进行探究.

祝愿同学们通过本册书的学习, 不但学到更多的数学知识, 而且在数学能力、用数学解决问题的能力等方面都有较大的提高, 并培养起更高的数学学习兴趣, 形成对数学的更加全面的认识.

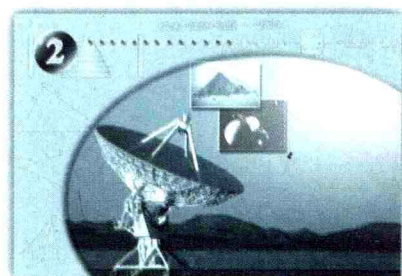
本书部分数学符号

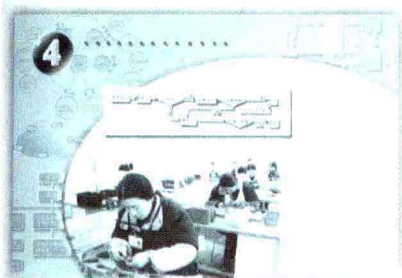
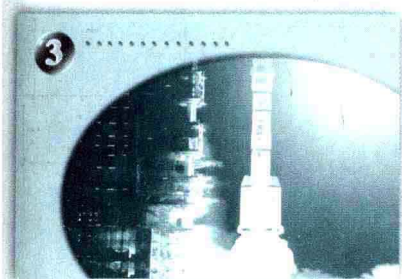
$\hat{e}$	残差
i	虚数单位
$\mathbf{C}$	复数集
$z, a+bi$	复数 z , 实部为 a , 虚部为 b 的复数

目 录

第一章 统计案例	1
1.1 回归分析的基本思想及其初步应用	2
1.2 独立性检验的基本思想及其初步应用	10
实习作业	17
小结	18
复习参考题	19

第二章 推理与证明	21
2.1 合情推理与演绎推理	22
阅读与思考 科学发现中的推理	33
2.2 直接证明与间接证明	36
小结	45
复习参考题	46





第三章 数系的扩充与复数的引入 49

3.1 数系的扩充和复数的概念 50

3.2 复数代数形式的四则运算 56

小结 62

复习参考题 63

第四章 框图 65

4.1 流程图 66

4.2 结构图 74

信息技术应用 用 Word2002 绘制流程图 79

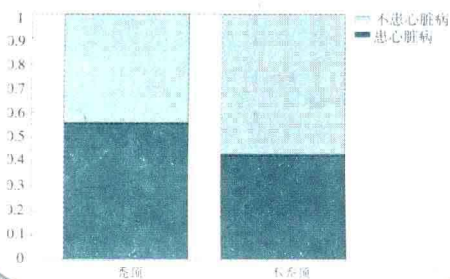
小结 82

复习参考题 83

第一章 统计案例

1.1 回归分析的基本思想及其初步应用

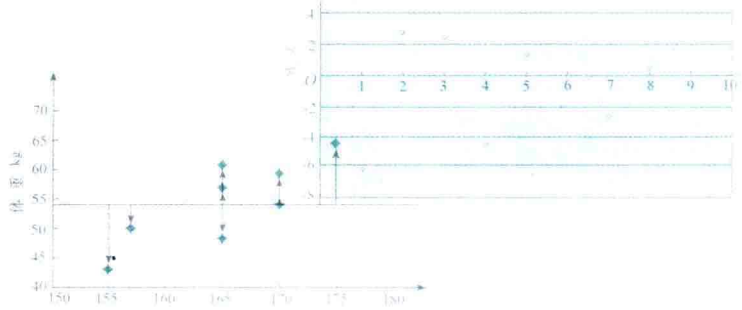
1.2 独立性检验的基本思想及其初步应用



在现实中，我们经常会遇到类似下面的问题：肥胖是影响人类健康的一个重要因素，标准的身高和体重之间是否存在线性相关关系？肺癌是严重威胁人类生命的一种疾病，吸烟与患肺癌有关系吗？等等。

为了回答这些问题，必须明确问题涉及的对象（总体）是什么，用怎样的量来描述要解决的问题，并确定获取变量值（数据）的方法，然后用恰当的统计方法分析数据，以得到最可靠的结论。

在必修模块中，我们学习过抽样、用样本估计总体、线性回归等基本知识。本章中，我们将在此基础上，通过对典型案例的讨论，进一步学习线性回归分析方法及其应用，并初步了解独立性检验的基本思想，认识统计方法在决策中的作用。



回归分析的基本思想及其初步应用

我们知道，函数关系是一种确定性关系，而相关关系是一种非确定性关系．回归分析（regression analysis）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．在《数学 3》中，我们利用回归分析的方法对两个具有线性相关关系的变量进行了研究，其步骤为画出两个变量的散点图，求回归直线方程，并用回归直线方程进行预报．下面我们通过案例，进一步学习回归分析的基本思想及其应用．

例 1 从某大学中随机选取 8 名女大学生，其身高和体重数据如表 1-1 所示．

表 1-1

编号	1	2	3	4	5	6	7	8
身高/cm	165	165	157	170	175	165	155	170
体重/kg	48	57	50	54	64	61	43	59

求根据女大学生的身高预报体重的回归方程，并预报一名身高为 172 cm 的女大学生的体重．

解：由于问题中要求根据身高预报体重，因此选取身高为自变量 x ，体重为因变量 y ．作散点图（图 1.1-1）：

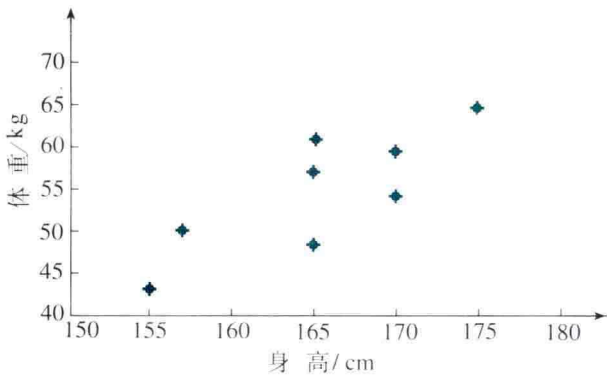


图 1.1-1

从图 1.1-1 中可以看出，样本点呈条状分布，身高和体重有比较好的线性相关关系，因此可以用回归直线 $y = bx + a$ 来近似刻画它们之间的关系．

由《数学 3》的知识可知，未知参数 b 和 a 的最小二乘估计分别为 $\hat{b}$ 和 $\hat{a}$ ，其计算公式如下：

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

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

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. $(\bar{x}, \bar{y})$ 称为样本点的中心.

回归直线过
样本点的中心.

在本例中, 根据上面的公式, 可以得到

$$\hat{b} = 0.849, \hat{a} = -85.712.$$

于是得到线性回归方程

$$\hat{y} = 0.849x - 85.712.$$

$\hat{b} = 0.849$ 是回归直线的斜率的估计值, 说明身高 x 每增加一个单位, 体重 y 就增加 0.849 个单位, 这表明体重与身高具有正的线性相关关系.

所以, 对于身高为 172 cm 的女大学生, 由回归方程可以预报其体重为

$$\hat{y} = 0.849 \times 172 - 85.712 = 60.316 (\text{kg}).$$



身高为 172 cm 的女大学生的体重一定是 60.316 kg 吗? 如果不是, 你能解释一下原因吗?

显然身高为 172 cm 的女大学生的体重不一定是 60.316 kg, 但一般可以认为她的体重在 60.316 kg 左右. 图 1.1-2 中的样本点和回归直线的相互位置说明了这一点.

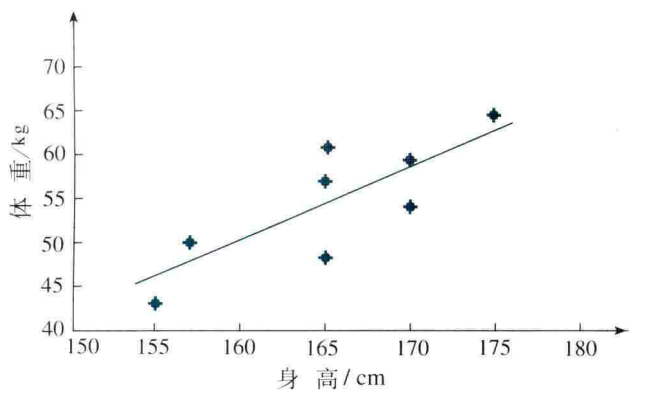


图 1.1-2

从散点图中还看到, 样本点散布在某一条直线的附近, 而不是在一条直线上, 所以不能用一次函数

$$y = bx + a$$

来描述它们之间的关系. 这时我们把身高和体重的关系用下面的线性回归模型

$$y = bx + a + e \quad (1)$$

来表示, 其中 a 和 b 为模型的未知参数, e 称为随机误差 (random error).



产生随机误差项 e 的原因是什么?

实际上, 一个人的体重除了受身高的影响外, 还受其他许多因素的影响, 例如饮食习惯、是否喜欢运动、度量误差等. 另一方面, 没有人知道身高和体重之间的真正关系是什么, 现在只是利用线性回归方程来近似这种关系. 而这种近似和上面提到的影响因素都会导致随机误差 e 的产生.

线性回归模型 (1) 与我们熟悉的一次函数模型的不同之处是增加了随机误差项 e , 因变量 y 的值由自变量 x 和随机误差 e 共同确定, 即自变量 x 只能解释部分 y 的变化. 在统计中, 我们也把自变量 x 称为解释变量, 因变量 y 称为预报变量.

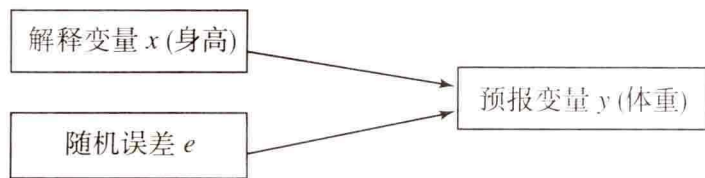


图 1.1-3



在线性回归模型中, e 是用 $bx+a$ 预报真实值 y 的随机误差, 它是一个不可观测的量, 那么应该怎样研究随机误差呢?

在实际应用中, 我们用回归方程 $\hat{y} = \hat{b}x + \hat{a}$ 中的 $\hat{y}$ 估计 (1) 中的 $bx+a$. 由于随机误差 $e = y - (bx+a)$, 所以 $\hat{e} = y - \hat{y}$ 是 e 的估计量. 对于样本点

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

而言, 它们的随机误差为

$$e_i = y_i - bx_i - a, i = 1, 2, \dots, n,$$

其估计值为

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{b}x_i - \hat{a}, i = 1, 2, \dots, n,$$

$\hat{e}_i$ 称为相应于点 (x_i, y_i) 的残差 (residual).



如何发现数据中的错误? 如何衡量模型的拟合效果?

可以通过残差发现原始数据中的可疑数据, 判断所建立模型的拟合效果. 表 1-2 列出了女大学生身高和体重的原始数据以及相应的残差数据.

表 1-2

编号	1	2	3	4	5	6	7	8
身高/cm	165	165	157	170	175	165	155	170
体重/kg	48	57	50	54	64	61	43	59
残差 $\hat{e}$	-6.373	2.627	2.419	-4.618	1.137	6.627	-2.883	0.382

我们可以利用图形来分析残差特性. 作图时纵坐标为残差, 横坐标可以选为样本编号、或身高数据, 或体重估计值等, 这样作出的图形称为残差图. 图 1.1-4 是以样本编号为横坐标的残差图.

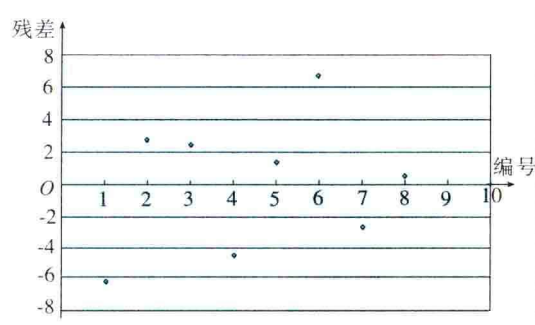


图 1.1-4

从图 1.1-4 中可以看出, 第 1 个样本点和第 6 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. 如果数据采集有错误, 就予以纠正, 然后再重新利用线性回归模型拟合数据; 如果数据采集没有错误, 则需要寻找其他的原因. 另外, 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, 回归方程的预报精度越高.

另外, 我们可以用 R^2 来刻画回归的效果, 其计算公式是

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

对于已经获取的样本数据, R^2 表达式中的 $\sum_{i=1}^n (y_i - \bar{y})^2$ 为确定的数. 因此 R^2 越大, 意味着残差平方和 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 越小, 即模型的拟合效果越好; R^2 越小, 残差平方和越大, 即模型的拟合效果越差. 在线性回归模型中, R^2 表示解释变量对于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. 在例 1 中, $R^2 \approx 0.64$, 表明“女大学生的身高解释了 64% 的体重变化”, 或者说“女大学生的体重差异有 64% 是由身高引起的”. R^2 是常用的选择模型的指标之一, 在实际应用中应该尽量选择 R^2 大的回归模型.

用身高预报体重时, 需要注意下列问题:

1. 回归方程只适用于我们所研究的样本的总体. 例如, 不能用女大学生的身高和体重之间的回归方程, 描述女运动员的身高和体重之间的关系. 同样, 不能用生长在南方多雨地区的树木的高与直径之间的回归方程, 描述北方干旱地区的树木的高与直径之间的关系.

2. 我们所建立的回归方程一般都有时间性. 例如, 不能用 20 世纪 80 年代的身高、体

重数据所建立的回归方程,描述现在的身高和体重之间的关系.

3. 样本取值的范围会影响回归方程的适用范围. 例如, 我们的回归方程是由女大学生身高和体重的数据建立的, 那么用它来描述一个人幼儿时期的身高和体重之间的关系就不恰当. (在回归方程中, 解释变量 x 的样本的取值范围为 $155 \sim 175$ cm, 而用这个方程计算 $x=70$ cm 时的 y 值, 显然不合适.)

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.

一般地, 建立回归模型的基本步骤为:

- (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) 画出解释变量和预报变量的散点图, 观察它们之间的关系 (如是否存在线性关系等).
- (3) 由经验确定回归方程的类型 (如我们观察到数据呈线性关系, 则选用线性回归方程).
- (4) 按一定规则 (如最小二乘法) 估计回归方程中的参数.
- (5) 得出结果后分析残差图是否有异常 (如个别数据对应残差过大, 残差呈现不随机的规律性等). 若存在异常, 则检查数据是否有误, 或模型是否合适等.

例 2 一只红铃虫的产卵数 y 和温度 x 有关, 现收集了 7 组观测数据列于表 1-3 中, 试建立 y 关于 x 的回归方程.

表 1-3

温度 $x/^{\circ}\text{C}$	21	23	25	27	29	32	35
产卵数 y / 个	7	11	21	24	66	115	325

解: 根据收集的数据, 作散点图:

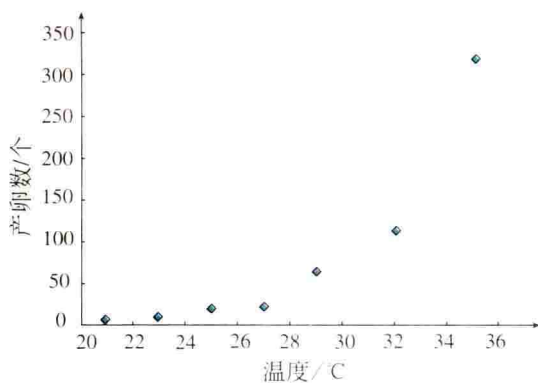


图 1.1-5

在图 1.1-5 中, 样本点并没有分布在某个带状区域内, 因此两个变量不呈线性相关关系, 不能直接利用线性回归方程来建立两个变量之间的关系. 根据已有的函数知识, 可以发现样本点分布在某一条指数函数曲线 $y=c_1 e^{c_2 x}$ 的周围, 其中 c_1 和 c_2 是待定参数.

现在, 问题变为如何估计待定参数 c_1 和 c_2 . 我们可以通过对数变换把指数关系变为线性关系. 令 $z = \ln y$, 则变换后样本点应该分布在直线

$$z = bx + a \quad (a = \ln c_1, b = c_2)$$

的周围. 这样, 就可以利用线性回归模型来建立 y 关于 x 的非线性回归方程^①了.

由表 1-3 的数据可以得到变换后的样本数据表 1-4, 图 1.1-6 给出了表 1-4 中数据的散点图. 从图 1.1-6 中可以看出, 变换后的样本点分布在一条直线的附近, 因此可以用线性回归方程来拟合.

① 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.

表 1-4

x	21	23	25	27	29	32	35
z	1.946	2.398	3.045	3.178	4.190	4.745	5.784

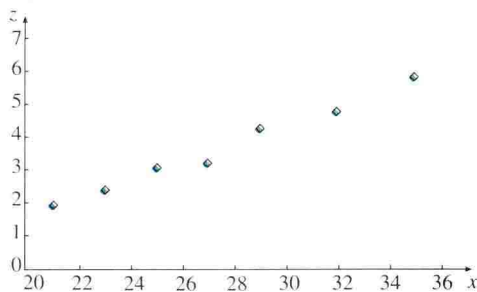


图 1.1-6

由表 1-4 中的数据得到线性回归方程

$$\hat{z} = 0.272x - 3.849.$$

因此红铃虫的产卵数对温度的非线性回归方程为

$$\hat{y}^{(1)} = e^{0.272x - 3.849}. \quad (2)$$

另一方面, 可以认为图 1.1-5 中样本点集中在某二次曲线 $y = c_3x^2 + c_4$ 的附近, 其中 c_3 和 c_4 为待定参数. 因此可以对温度变量做变换, 即令 $t = x^2$, 然后建立 y 与 t 之间的线性回归方程, 从而得到 y 与 x 之间的非线性回归方程.

表 1-5 是红铃虫的产卵数和对应的温度的平方, 图 1.1-7 是相应的散点图.

表 1-5

t	441	529	625	729	841	1 024	1 225
y	7	11	21	24	66	115	325

从图 1.1-7 中可以看出, y 与 t 的散点图并不分布在一条直线的周围, 因此不宜用线性回归方程来拟合它, 即不宜用二次函数 $y = c_3x^2 + c_4$ 来拟合 y 和 x 之间的关系. 这个结论还可以通过下面的残差分析得到.

为比较两个不同模型的残差, 需要建立两个相应的回归方程. 前面我们已经建立了 y 关于 x 的指数回归方程 (2), 下面建立 y 关于 x 的二次回归方程. 用线性回归模型拟合表

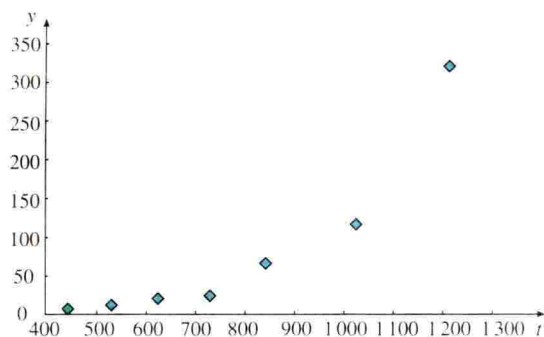


图 1.1-7

1-5中的数据, 得到 y 关于 t 的线性回归方程

$$\hat{y}^{(2)} = 0.367t - 202.543,$$

即 y 关于 x 的二次回归方程为

$$\hat{y}^{(2)} = 0.367x^2 - 202.543. \quad (3)$$

可以通过残差来比较两个回归方程 (2) 和 (3) 的拟合效果. 用 x_i 表示表 1-3 中第 1 行第 $(i+1)$ 列的数据, 则回归方程 (2) 和 (3) 的残差计算公式分别为

$$\hat{e}_i^{(1)} = y_i - \hat{y}_i^{(1)} = y_i - e^{0.272x_i - 3.849}, \quad i = 1, 2, \dots, 7;$$

$$\hat{e}_i^{(2)} = y_i - \hat{y}_i^{(2)} = y_i - 0.367x_i^2 + 202.543, \quad i = 1, 2, \dots, 7.$$

表 1-6 给出了原始数据及相应的两个回归方程的残差, 从表中的数据可以看出模型 (2) 的残差的绝对值显然比模型 (3) 的残差的绝对值小, 因此模型 (2) 的拟合效果比模型 (3) 的拟合效果好.

表 1-6

x	21	23	25	27	29	32	35
y	7	11	21	24	66	115	325
$\hat{e}^{(1)}$	0.557	-0.101	1.875	-8.950	9.230	-13.381	34.675
$\hat{e}^{(2)}$	47.696	19.400	-5.832	-41.000	-40.104	-58.265	77.968

在一般情况下, 比较两个模型的残差比较困难. 原因是在某些样本点上一个模型的残差的绝对值比另一个模型的小, 而另一些样本点的情况则相反. 这时可以用 R^2 来比较两个模型的拟合效果, R^2 越大, 模型的拟合效果越好. 由表 1-6 容易算出模型 (2) 和 (3) 的 R^2 分别约为 0.98 和 0.80, 因此模型 (2) 的拟合效果好于模型 (3).

练习

1. 在两个变量的回归分析中, 作散点图的目的是什么?
2. 在回归分析中, 分析残差能够帮助我们解决哪些问题?
3. 如果发现散点图中所有的样本点都落在一条斜率为非 0 实数的直线上, 请回答下列问题:
 - (1) 解释变量和预报变量的关系是什么?
 - (2) R^2 是多少?