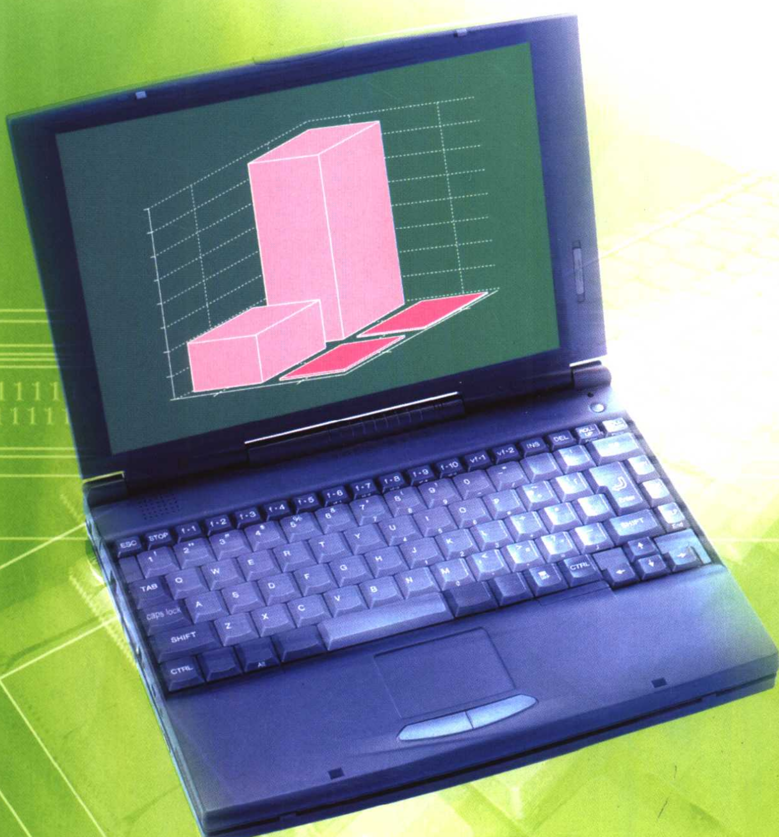
经全国中小学教材审定委员会
2005 年初审通过

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
中学数学课程教材研究开发中心　编著



人民教育出版社

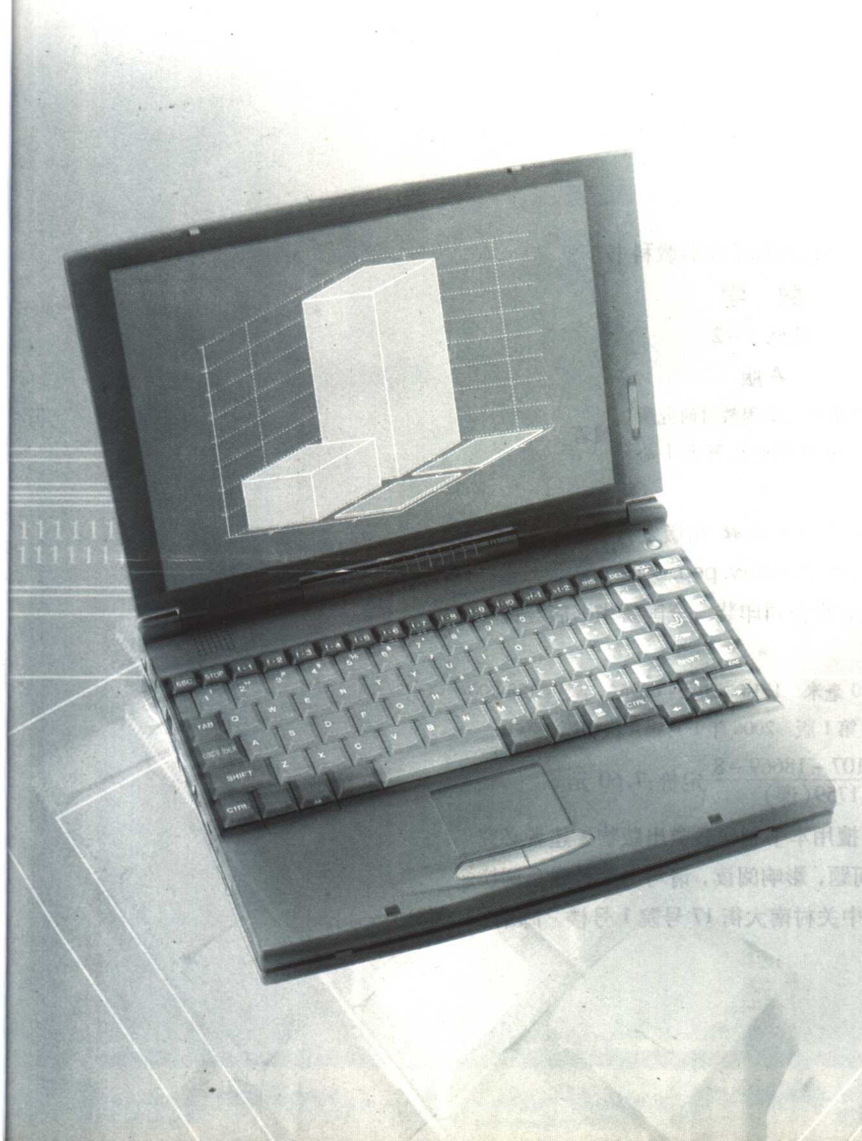
A版

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
中学数学课程教材研究开发中心　编著



人民教育出版社

A版

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 出版发行

网址：http：//www.pep.com.cn

北京市白帆印务有限公司印装 全国新华书店经销

*

开本：890 毫米×1 240 毫米 1/16 印张：6.75 字数：128 000

2005 年 6 月第 1 版 2006 年 1 月第 4 次印刷

ISBN 7－107－18669－8
G·11759（课） 定价：7.60 元

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，探索和提供解决一些问题的思路和方向，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——直接证明和间接证明，通过数学实例说明它们的思考过程和特点等．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．框图包括流程图和结构图，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和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，启发和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

$\sum\limits_{i=1}^{n}(y_i-\bar{y})^2$	总偏差平方和
$\hat{e}$	残差
$\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2$	残差平方和
i	虚数单位
$\mathbf{C}$	复数集
z，$a+b\mathrm{i}$	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 …… 14

实习作业 ………………………………………… 22

小结………………………………………………… 23

复习参考题 …………………………………… 24

第二章　推理与证明………………………………… 26

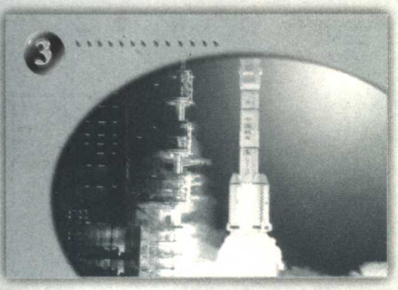
2.1 合情推理与演绎推理 ……………………… 28

阅读与思考　科学发现中的推理 …………………… 42

2.2 直接证明与间接证明 ……………………… 45

小结………………………………………………… 55

复习参考题 …………………………………… 56



第三章　数系的扩充与复数的引入 ………… 58

3.1 数系的扩充和复数的概念 ………… 60

3.2 复数代数形式的四则运算 ………… 66

小结 ………… 72

复习参考题 ………… 73

第四章　框图 ………… 74

4.1 流程图 ………… 76

4.2 结构图 ………… 85

信息技术应用　用 Word2002 绘制流程图 ………… 90

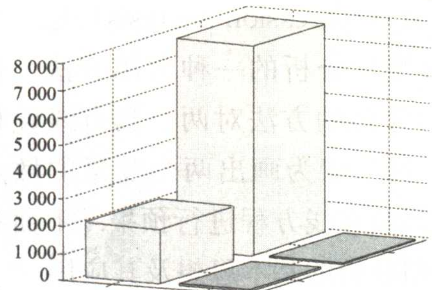
小结 ………… 93

复习参考题 ………… 94

第一章 统计案例

1.1 回归分析的基本思想及其初步应用

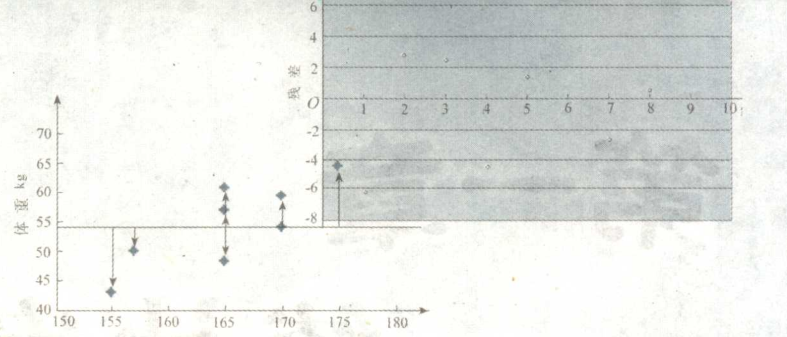
1.2 独立性检验的基本思想及其初步应用



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．



1.1 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

求根据一名女大学生的身高预报她的体重的回归方程，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）：

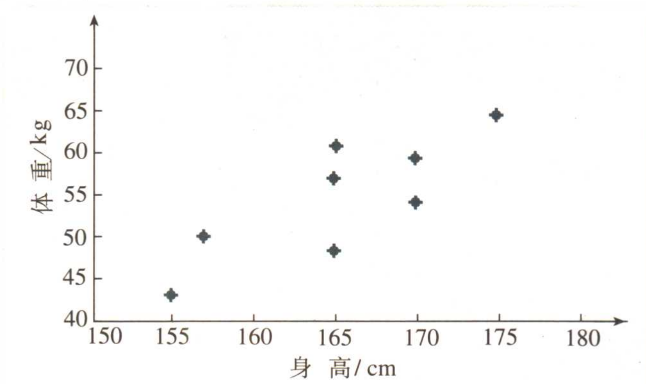


图 1.1-1

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，因此可以用线性回归方程刻画它们之间的关系.

另外，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来表示：

$$y=bx+a+e, \qquad (*)$$

其中 a 和 b 为模型的未知参数，e 称为随机误差.

思考?

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，还受许多其他因素的影响，例如饮食习惯、是否喜欢运动、度量误差等，而且我们选用的线性模型往往只是一种近似的模型，所有这些因素都会导致随机误差 e 的产生.

线性回归模型（$*$）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.

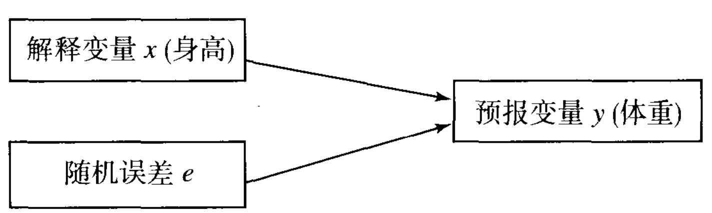


图 1.1-2

如何估计模型（$*$）中的未知参数 a 和 b? 由《数学 3》的知识可知，最小二乘估计 $\hat{a}$ 和 $\hat{b}$ 就是未知参数 a 和 b 的最好估计，其计算公式如下：

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

$$\hat{a}=y-\hat{b}x, \tag{2}$$

其中$x=\frac{1}{n}\sum_{i=1}^{n}x_i$，$y=\frac{1}{n}\sum_{i=1}^{n}y_i$．$(x，y)$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

$$\hat{a}=-85.712，\hat{b}=0.849.$$

于是得到线性回归方程

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

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

$$\hat{y}=0.849\times72-85.712=60.316(\text{kg}).$$

探究

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 1.1-3 中的样本点和回归直线的相互位置说明了这一点.

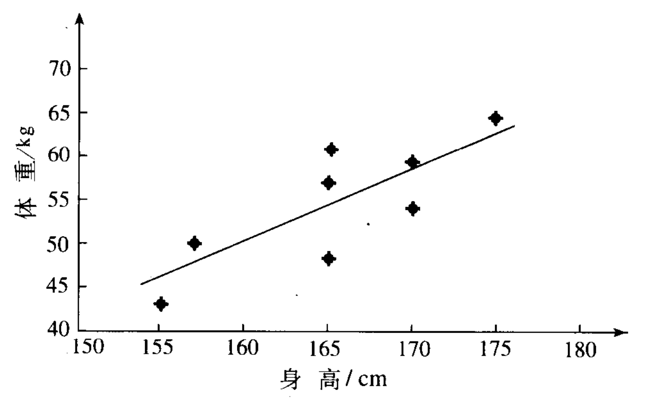


图 1.1-3

$\hat{b}=0.849$ 是斜率的估计值，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．如何描述它们之间线性相关关系的强弱？

在《数学 3》中，我们介绍了用相关系数 r 来衡量两个变量之间线性相关关系的方法．样本相关系数的具体计算公式为

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

当 $r>0$ 时，表明两个变量正相关；当 $r<0$ 时，表明两个变量负相关. r 的绝对值越接近 1，表明两个变量的线性相关性越强；r 的绝对值接近于 0 时，表明两个变量之间几乎不存在线性相关关系. 通常当 r 大于 0.75 时，认为两个变量有很强的线性相关关系.

在本例中，可以计算出 $r=0.798$，这表明体重与身高有很强的线性相关关系，从而也表明我们建立的回归模型是有意义的.

假设身高和随机误差的不同不会对体重产生任何影响，那么所有人的体重将相同. 在体重不受任何变量影响的假设下，设 8 名女大学生的体重都是她们体重的平均值，即 8 个人的体重都为 54.5 kg（表 1-2).

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	54.5	54.5	54.5	54.5	54.5	54.5	54.5	54.5

在根据表 1-2 的数据作出的散点图中，所有的点应该落在同一条水平直线上，但是观测到的数据并非如此，它们的散布情况如图 1.1-4 所示. 这就意味着预报变量（体重）的值受解释变量（身高）或随机误差的影响.

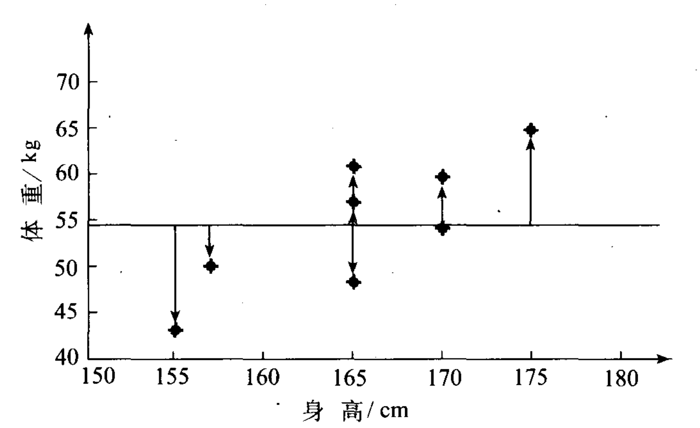


图 1.1-4

思考？

如何刻画预报变量（体重）的变化？这个变化在多大程度上与解释变量（身高）有关？在多大程度上与随机误差有关？

例如，编号为 6 的女大学生的体重并没有落在水平直线上，她的体重为 61 kg. 解释变量（身高）和随机误差共同把这名学生的体重从 54.5 kg“推”到了 61 kg，相差 6.5 kg，所以 6.5 kg 是解释变量和随机误差的组合效应. 类似地，编号为 3 的女大学生的体重也没有落在水平直线上，她的体重为 50 kg，这时解释变量和随机误差的组合效应为 -4.5. 用这种方法可以对所有预报变量计算组合效应.

如何把所有这些效应合并在一个数中呢？在数学上，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**（corrected total sum of squares）. 在例 1 中，总偏差平方和为 354.

那么，在这个总的效应（总偏差平方和）中，有多少来自于解释变量（身高）？有多少来自于随机误差？假设随机误差对体重没有影响，也就是说，体重仅受身高的影响，那么散点图中所有的点将完全落在回归直线上. 但是，在图 1.1-3 中，数据点并没有完全落在回归直线上. 这些点散布在回归直线附近，所以一定是随机误差把这些点从回归直线上“推”开了.

因此，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**（residual）. 例如，对于编号为 6 的女大学生，计算随机误差的效应为

$$61-(0.849\times165-85.712)=6.627.$$

对每名女大学生计算这个差异，然后分别将所得的值平方后加起来，用数学符号表示为

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

称为**残差平方和**（residual sum of squares），它代表了随机误差的效应. 在例 1 中，残差平方和约为 128.361.

由于解释变量和随机误差的总效应（总偏差平方和）为354，而随机误差的效应为128.361，所以解释变量的效应为

$$354-128.361=225.639,$$

这个值称为**回归平方和**（regression sum of squares）.

我们可以用相关指数 R^2 来刻画回归的效果，其计算公式是

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

显然，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（因为 R^2 越接近于1，表示解释变量和预报变量的线性相关性越强）．如果某组数据可能采取几种不同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.

从表1-3中可以看出，解释变量对总效应约贡献了64%，即 $R^2\approx 0.64$，可以叙述为“身高解释了64%的体重变化”，而随机误差贡献了剩余的36%．所以，身高对体重的效应比随机误差的效应大得多.

表1-3

来源	平方和	比例
随机误差	225.639	0.64
残差变量	128.361	0.36
总计	354	1.00

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，我们可以通过残差

$$\hat{e}_1,\ \hat{e}_2,\ \cdots,\ \hat{e}_n$$

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**．表1-4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 1-4

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 1.1-5 是以样本编号为横坐标的残差图．

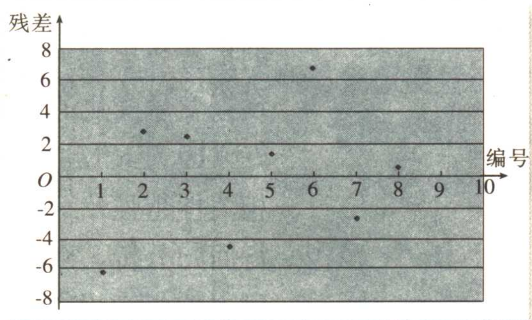


图 1.1-5

从图 1.1-5 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．

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．