普通高中课程标准实验教材
PUTONG GAOZHONG KECHENG BIAOZHUN SHIYAN JIAOCAI

随堂纠错

SUITANGJIUCUO

超级练

CHAOJI LIAN

数学

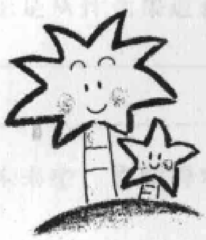
选修1-2

浙 江 教 育 出 版 社

图书在版编目(CIP)数据

随堂纠错超级练. 数学. 1-2：选修 / 郑日锋编.
—杭州：浙江教育出版社，2007
配人教版
ISBN 978-7-5338-7198-7
Ⅰ. 随… Ⅱ. 郑… 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 145592 号



随堂纠错超级练

数学 1-2 选修

主　　编　郑日锋
出　　版　浙江教育出版社
（杭州市天目山路 40 号　邮编 310013）
发　　行　浙江省新华书店集团有限公司
总 策 划　邱连根
责任编辑　金馥菊
装帧设计　韩　波
责任校对　胡　星
责任印务　吴梦菁
图文制作　杭州富春电子印务有限公司
印刷装订　富阳美术印刷有限公司

开　　本　890×1240　1/16
印　　张　7
字　　数　200 000
版　　次　2007 年 9 月第 1 版
印　　次　2007 年 9 月第 1 次
印　　数　0 001—7 000
标准书号　ISBN 978-7-5338-7198-7
定　　价　9.50 元

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

《随堂纠错超级练》丛书编委会

（以姓氏笔画为序）

方青稚（台州中学）
冯任儿（湖州中学）
史定海（鄞州中学）
刘　岩（杭州第十四中学）
庄志琳（桐乡高级中学）
许军国（宁波市教育局教研室）
朱建国（杭州外国语学校）
朱恒元（义乌中学）
任学宝（杭州学军中学）
任美琴（台州回浦中学）
任富强（慈溪中学）
伊建军（杭州高级中学）
沈玉荣（杭州学军中学）
沈金林（平湖中学）
沈骏松（嘉兴市教育研究院）
杨志敏（杭州市教育局教研室）
李兆田（嘉兴高级中学）
郑日锋（杭州学军中学）
郑青岳（玉环县教育局教研室）
林金法（温岭中学）
苗金德（绍兴鲁迅中学）
周业宇（丽水市教育局教研室）
周　红（杭州学军中学）
施丽华（宁波效实中学）
姜水根（宁波效实中学）
赵一兵（杭州高级中学）
赵力红（富阳中学）
胡　辛（杭州第二中学）
胡伯富（杭州市教育局教研室）
枯　荣（绍兴市教育局教研室）
高　宁（杭州第四中学）
徐　勤（杭州学军中学）
鄢伟友（金华市教育局教研室）
潘健男（湖州第二中学）

丛书总策划　邱连根

栏目设置及使用说明

教材解读
归纳学习要点,梳理知识脉络,方便理解与记忆。

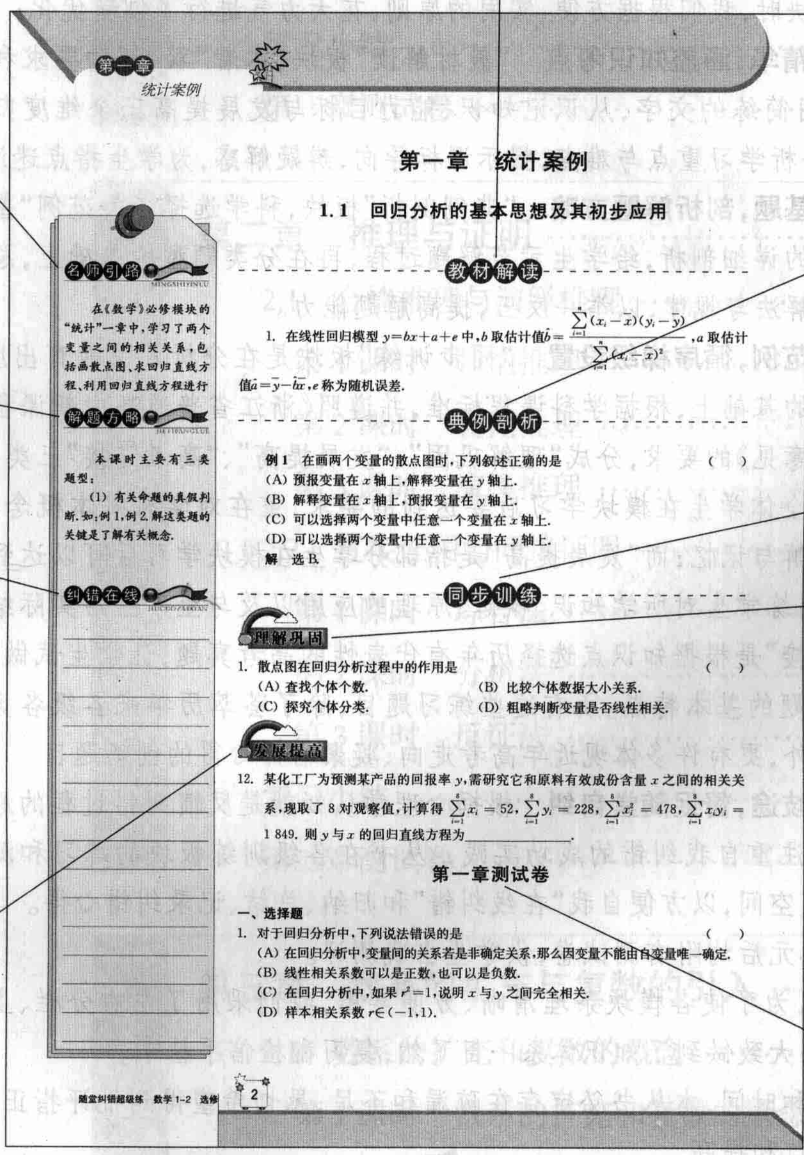
名师引路
揭示重点,剖析难点,点拨学法,提供学习心理辅导。

典例剖析
选择"基题",分析解题思路与方法,提供表达示范。

解题方略
分类题型,总结问题解决的一般规律,并揭示解题技巧。

同步训练
理解巩固
提供理解、巩固基本知识和技能的基础题。覆盖教材要点,强化重点,题量适宜,注重有效。

纠错在线
让学生记录做题过程中出现的错误,提倡学生随时总结自己不足的学习习惯。


第一章
统计案例

第一章 统计案例

1.1 回归分析的基本思想及其初步应用

名师引路 MINGSHIYINLU

在《数学》必修模块的"统计"一章中,学习了两个变量之间的相关关系,包括画散点图、求回归直线方程、利用回归直线方程进行

解题方略 JIETIFANGLUE

本课时主要有三类题型:

(1) 有关命题的真假判断.如:例1,例2.解这类题的关键是了解有关概念.

纠错在线 JIUCUOZAIXIAN

教材解读

1. 在线性回归模型 $y=bx+a+e$ 中,b 取估计值 $\hat{b}=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$,$a$ 取估计值 $\hat{a}=\bar{y}-\hat{b}\bar{x}$,$e$ 称为随机误差.

典例剖析

例1 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变量在 x 轴上.
(D) 可以选择两个变量中任意一个变量在 y 轴上.
解 选 B.

同步训练

理解巩固

1. 散点图在回归分析过程中的作用是 ( )
(A) 查找个体个数. (B) 比较个体数据大小关系.
(C) 探究个体分类. (D) 粗略判断变量是否线性相关.

发展提高

12. 某化工厂为预测某产品的回报率 y,需研究它和原料有效成份含量 x 之间的相关关系,现取了 8 对观察值,计算得 $\sum_{i=1}^{8}x_i=52$,$\sum_{i=1}^{8}y_i=228$,$\sum_{i=1}^{8}x_i^2=478$,$\sum_{i=1}^{8}x_iy_i=1\,849$. 则 y 与 x 的回归直线方程为________.

第一章测试卷

一、选择题
1. 对于回归分析中,下列说法错误的是 ( )
(A) 在回归分析中,变量间的关系若是非确定关系,那么因变量不能由自变量唯一确定.
(B) 线性相关系数可以是正数,也可以是负数.
(C) 在回归分析中,如果 $r^2=1$,说明 x 与 y 之间完全相关.
(D) 样本相关系数 $r\in(-1,1)$.

随堂纠错超级练 数学 1-2 选修 2

发展提高
提供提升知识层次,发展学生解决问题能力的优秀试题。

章测试卷
参照高考题型,提供囊括本章知识要点及考点的试题,供学生自测或教师选用。

出版前言

作为深入贯彻新课程标准精神、全面体现最新教学理念的一个新的尝试，我社精心编辑出版了这套“随堂纠错超级练”丛书，以满足当前高中各科教学的急需。

这是一套涵盖高中各主要学科，包括课堂教学和阶段复习各环节的同步实战型丛书。从书名即反映了其主要特点：随堂，就是基本知识随堂通；纠错，就是出现错误当堂纠；超级练，就是巩固提高分层练。

在设计模块时，我们根据方便、实用的原则，花大力气进行了创新优化：

提炼教材精华，涵盖知识考点　“教材解读”板块，本着“双基”的要求和高考命题的导向，用简练的文字，从识记知识、能力目标与发展提高三个维度归纳整理教材内容，分析学习重点与难点，揭示课标导向，辨疑解惑，为学生指点迷津。

荟萃典例基题，剖析解题方略　“典例剖析”板块，科学选择各类范例“基题”，先通过多角度的详细剖析，给学生示范解题过程，再在分类题型的基础上，总结各类习题的一般解法与规律，以举一反三，提高解题能力。

精选名题范例，循序梯级设置　“同步训练”板块是在分析人民教育出版社A版教科书内容的基础上，根据学科课程标准，并遵照《浙江省普通高中新课程实验学科教学指导意见》的要求，分成“理解巩固”、“发展提高”、“高考链接”三类。其中“理解巩固”是全体学生在模块学习后要达到的要求，重在对学科基本概念、理论以及知识的理解与记忆；而“发展提高”是指部分学生在模块学习后可以达到的较高要求，旨在训练学生对所学知识、概念、原理的应用以及与生产生活实际结合的能力；“高考链接”是根据知识点选择历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。所有这些练习题目，除了荟萃历年来各级各类试卷的名题范例以外，更有许多体现近年高考走向、凝聚名师心得的创新题目。

警示易入歧途，督促随堂自纠　根据心理学认知就是反馈纠错过程的原理和高考状元们都注重自我纠错的成功实践，丛书在各级训练板块的练习和测试部分，预留了一定空间，以方便自我“在线纠错”和归纳、总结、记录纠错心得。

此外，每单元后均附有测试卷，供学生自我测评。

在编排上，为了使各模块条理清晰、方便实用，我们采用了左右分栏、上下切块的版面设计，大致做到了知识体系一目了然，复习翻检信手拈来。

限于水平和时间，本丛书必定存在疏漏和不足，恳切希望得到批评指正，以便我们进一步修订和提高。

浙江教育出版社

2007.9

第一章 统计案例

1.1 回归分析的基本思想及其初步应用

教材解读

1. 随机误差

在线性回归模型 $y=bx+a+e$ 中，b 取估计值$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}$，$a$ 取估计值 $\hat{a}=\overline{y}-\hat{b}\overline{x}$，$e$ 称为随机误差.

2. 线性相关关系

样本相关系数 $r=\dfrac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}$. 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 $|r|>0.75$ 时，认为两个变量有很强的线性相关关系.

3. 总偏差平方和

把每个效应(观测值减去总的平均值)的平方相加，即用 $\sum\limits_{i=1}^{n}(y_i-\overline{y})^2$ 表示总的效应，称为总偏差平方和.

4. 残差、回归平方和

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.

各个残差的平方和 $\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2$ 称为残差平方和，它代表了随机误差的效应.

总偏差平方和减去残差平方和所得的差称为回归平方和，它代表了解释变量的效应.

相关指数 R^2 的计算公式是 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$. 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.

第 1 课时 回归分析(1)

典例剖析

例 1 在画两个变量的散点图时，下列叙述正确的是 ( )

名师引路

MINGSHIYINLU

请对照左栏，仔细阅读教材，思考以下问题：本节有哪些知识要点？具体内容是什么？请尽可能地用自己的话表述出来.

在《数学 3》必修模块的"统计"一章中，学习了两个变量之间的相关关系，包括画散点图、求回归直线方程、利用回归直线方程进行预报等内容. 本节在此基础上进一步学习回归分析的基本思想及其初步应用. 从相关系数、相关指数和残差分析等角度探讨回归模型拟合的效果.

本节的重点是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. 难点是解释残差的含义，了解偏差平方和分解的思想.

学好本节内容的关键是理解回归分析的基本思想.

解题方略

JIETIFANGLUE

这里提供的是本节习题的主要题型及一般

解题方略

解法. 阅读后，你理解老师是如何思考并解决问题的吗？你有什么启发？你还有更好的解法吗？

本课时主要有三类题型：

(1) 有关命题的真假性判断. 解这类题的关键是了解有关概念. 如：例 1、例 2.

(2) 根据具有相关关系的两个变量的一组对应数据，画出散点图，建立回归直线方程，并利用回归直线方程进行预报. 如：例 3.

(3) 在尚未断定两个变量之间是否具有线性相关关系的情况下，应先进行相关性检验. 在确定其具有线性相关关系后，再求其回归直线方程. 如：例 4.

(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变量在 x 轴上.

(D) 可以选择两个变量中任意一个变量在 y 轴上.

解 选 B.

例 2 为了考察两个变量 x 和 y 之间的线性相关性，甲、乙两位同学各自独立做了 10 次和 15 次试验，并且利用线性回归方法，求得回归直线分别为 l_1 和 l_2. 已知两人测得的变量 x 和 y 的平均值相等，且分别都是 s 和 t，那么下列说法正确的是 ( )

(A) 直线 l_1 和直线 l_2 重合.

(B) 直线 l_1 和直线 l_2 平行.

(C) 直线 l_1 和直线 l_2 有公共点 (s,t).

(D) 直线 l_1 和直线 l_2 相交，但交点不一定是 (s,t).

解 选 C.

注意 回归直线一定过点 $(\overline{x},\overline{y})$.

例 3 某产品的广告费支出 x(单位：百万元)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；

(3) 估计产品的广告费支出为 1 100 万元时，销售额是多少？

解 (1) 散点图如图 1-1 所示.

图 1-1

(2) 由已知，可得 $\overline{x}=5$，$\overline{y}=50$，$\sum\limits_{i=1}^{5}x_i^2=145$，$\sum\limits_{i=1}^{5}x_iy_i=1\,380$，

则 $\hat{b}=\dfrac{\sum\limits_{i=1}^{5}x_iy_i-5\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{5}x_i^2-5\overline{x}^2}=\dfrac{1\,380-5\times5\times50}{145-5\times5^2}=6.5$，$\hat{a}=\overline{y}-\hat{b}\overline{x}=50-6.5\times5=17.5$，

于是所求的回归直线方程为 $\hat{y}=6.5x+17.5$.

(3) 当 $x=11$ 时，$\hat{y}=6.5\times11+17.5=89$，即销售额的估计值是 8 900 万元.

注意 计算 $\hat{b}$ 的公式可以改写成 $\hat{b}=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$.

例 4 已知某地单位面积菜地年平均使用氮肥量 x(kg)与单位面积蔬菜年平均产量 y(t)之间的关系有如下对应数据：

年份	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
x	70	74	80	78	85	92	90	95	92	108	115	123	130	138	145
y	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) 对变量 y 与 x 进行相关性检验；

(2) 如果 y 与 x 之间具有线性相关关系，求回归直线方程；

(3) 估计单位面积施肥 150 kg 时，单位面积蔬菜的年平均产量.

解题思路　利用样本相关系数计算公式求出样本相关系数 r，如果 $|r|>0.75$. 则线性相关；否则，不线性相关.

解　(1) $n=15, \overline{x}=101, \overline{y}=10.11, \sum_{i=1}^{15}x_i^2=161\ 125, \sum_{i=1}^{15}y_i^2=1\ 628.55, \sum_{i=1}^{15}x_iy_i=16\ 076.8$. 于是得

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

$$=\frac{16\ 076.8-15\times101\times10.11}{\sqrt{(161\ 125-15\times101^2)(1\ 628.55-15\times10.11^2)}}\approx0.864\ 3>0.75,$$

故 y 与 x 具有较强的线性相关关系.

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

则 $\hat{b}=\dfrac{\sum_{i=1}^{15}x_iy_i-15\overline{x}\,\overline{y}}{\sum_{i=1}^{15}x_i^2-15\overline{x}^2}=\dfrac{16\ 076.8-15\times101\times10.11}{161\ 125-15\times101^2}\approx0.093\ 7$，

$\hat{a}=\overline{y}-\hat{b}\overline{x}=10.11-0.093\ 7\times101\approx0.646\ 3$，

所以回归直线方程为 $\hat{y}=0.093\ 7x+0.646\ 3$.

当 $x=150$(kg)时，$\hat{y}=0.093\ 7\times150+0.646\ 3\approx14.701$(t).

同步训练

理解巩固

1. 散点图在回归分析过程中的作用是　(　　)
   (A) 查找个体个数.　(B) 比较个体数据大小关系.
   (C) 探究个体分类.　(D) 粗略判断变量是否线性相关.
2. 设有一个回归直线方程为 $\hat{y}=2-1.5x$，则变量 x 增加一个单位时　(　　)
   (A) y 平均增加 1.5 个单位.　(B) y 平均增加 2 个单位.
   (C) y 平均减少 1.5 个单位.　(D) y 平均减少 2 个单位.
3. 下列变量关系是线性相关的是　(　　)
   (A) 人的身高与视力.　(B) 角的大小与所对的圆弧长.
   (C) 收入水平与纳税水平.　(D) 人的年龄与身高.
4. 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 y(cm)与年龄 x(岁)的回归直线方程为 $\hat{y}=7.19x+73.93$，由此预测这个孩子 10 岁时的身高　(　　)
   (A) 一定是 145.83 cm.　(B) 在 145.83 cm 以上.
   (C) 在 145.83 cm 左右.　(D) 在 145.83 cm 以下.
5. 作一个关于两个变量的散点图的主要目的是____________.

纠错在线

JIUCUOZAIXIAN

做题的目的是评估自己的学习效果，提高解题的准确率与速度. 每次做题时，你都应该认真、仔细. 题目做错是正常的，但作业完成后，务必将做错的那些习题标出来，分析出错的原因，这样你就可以在纠错中不断进步.

做对________题；

做错________题；

原因分析：________.

纠错在线
JIUCUOZAIXIAN

6. 回归模型的基本步骤,不仅适用于线性回归模型,也适用于________,比如用________曲线和________曲线来拟合.

7. 已知回归直线的斜率的估计值是 2.21,样本点的中心是(3,5),则回归直线方程是____________.

8. 一项调查表对 9 个不同的 x 值,测得 y 的 9 个对应值如下表:

x	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
y	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13.6	15.3

试作出该数据的散点图,并由此判断是否存在回归直线.若有,求出直线方程.

9. 要分析初中升学考试中的数学成绩对高一年级学生的数学学习有什么影响,在高一年级中随机抽取 10 名学生,统计他们入学时的数学成绩 x(分)和高一年级期末数学考试成绩 y(分)如下:

学生编号	1	2	3	4	5	6	7	8	9	10
x	75	45	52	82	90	71	63	58	97	67
y	76	52	70	81	82	84	65	56	95	77

(1) 画出散点图;
(2) 判断 y 与 x 是否具有线性相关关系;
(3) 如果 y 与 x 具有线性相关关系,求回归直线方程;
(4) 若某学生的入学数学成绩为 70 分,试估计他高一期末数学考试成绩.

发展提高

10. 已知 x,y 之间的对应数据如下表所示:

x	1.08	1.12	1.19	1.28
y	2.25	2.37	2.40	2.55

则 y 与 x 之间的线性回归直线过点 ( )
(A) (0,0). (B) $(\bar{x},0)$. (C) $(0,\bar{y})$. (D) $(\bar{x},\bar{y})$.

11. 在某校高三年级学生学习数学的时间 x(h)与考试成绩 y(分)之间建立线性回归方程,经计算方程为$\hat{y}=20-0.8x$,该方程中的参数计算 ( )
(A) $\hat{a}$的值是明显错误的. (B) $\hat{b}$的值是明显错误的.
(C) $\hat{a}$和$\hat{b}$的值都是错误的. (D) $\hat{a}$和$\hat{b}$的值都是正确的.

12. 某化工厂为预测某产品的回收率 y,需研究 y 和原料有效成分含量 x 之间的相关关系,现抽取了 8 对观察值,计算得 $\sum_{i=1}^{8}x_i=52$, $\sum_{i=1}^{8}y_i=228$, $\sum_{i=1}^{8}x_i^2=478$, $\sum_{i=1}^{8}x_iy_i=$

1 849，则 y 与 x 的回归直线方程为____________.

13. 根据下列给出的条件，回答各题中的两个变量的关系.
   (1) 某学生期中考试数学成绩与复习时间投入量的关系；
   (2) 目前各家商场都进行促销活动，商品的销售额与广告费的关系.

14. 下面是两个变量之间的一组数据：

x	1	2	3	4	5	6	7	8
y	1	4	9	16	25	36	49	64

试求出这两个变量之间的回归直线方程，并与 $y=x^2$ 进行比较，考察当 x 越来越大时，此回归直线方程有没有意义？为什么？

第 2 课时　回归分析(2)

典例剖析

例 1　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.

解　利用公式 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$，故选 A.

例 2　炼钢是一个氧化降碳的过程，钢水含碳量的多少直接影响冶炼时间的长短，必须掌握钢水含碳量和冶炼时间的关系. 已测得炉料熔化完毕时钢水的含碳量 x(0.01%) 与冶炼时间 y(min)(从炉料熔化完毕到出钢的时间)的一列数据，如下表所示：

x	104	180	190	177	147	134	150	191	204	121
y	100	200	210	185	155	135	170	205	235	125

(1) 作出 y 与 x 的散点图，根据该图猜想它们之间的关系；
(2) 建立以 x 为解释变量，y 为预报变量的回归模型，并计算残差.
(3) 你认为这个模型能较好地刻画 y 与 x 的关系吗？请说明理由.

解　(1) 画出散点图如图 1-2 所示：

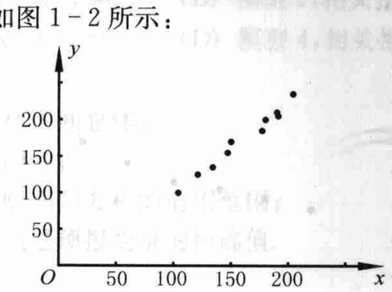


图 1-2

纠错在线
JIUCUOZAIXIAN

解题方略
JIETIFANGLUE

本课时主要有三类题型：

(1) 有关相关指数 R^2、残差平方和、回归平方和的关系问题. 只需利用公式 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$ 求解即可. 如：例 1、例 4.

(2) 在具有相关关系的两个变量的散点图中，样本点在一条直线附近，利用线性回归模型. 通过计算相关指数判断线性回归模型是否较好地刻画两个变量的关系. 如：例 2.

(3) 在具有相关关系的两个变量的散点图中，样本点不在一条直线附近，利用非线性回归模型，如二次回归模型、指数回归模型等，通过计算相关指数判断非线性回归模型是否较好

解题方略

地刻画两个变量的关系.如:例3.

从散点图中可以看出 y 与 x 近似呈线性关系.

(2) $\overline{x}=159.8$, $\overline{y}=172$, $\sum_{i=1}^{10} x_i^2=265\ 448$, $\sum_{i=1}^{10} y_i^2=312\ 350$, $\sum_{i=1}^{10} x_iy_i=287\ 640$.

$$\therefore\ \hat{b}=\frac{\sum_{i=1}^{10} x_iy_i-10\overline{x}\ \overline{y}}{\sum_{i=1}^{10} x_i^2-10\overline{x}^2}\approx 1.267.$$

$\hat{a}=\overline{y}-\hat{b}\overline{x}=-30.47$.

故所求的回归直线方程为 $\hat{y}=1.267x-30.47$. 残差如下表所示:

x	104	180	190	177	147	134	150	191	204	121
y	100	200	210	185	155	135	170	205	235	125
$\hat{y}_i$	101.298	197.59	210.26	193.789	155.779	139.308	159.58	211.527	227.998	122.837
$\hat{e}_i=y_i-\hat{y}_i$ (残差)	-1.298	2.41	-0.26	-8.789	-0.779	-4.308	10.42	-6.527	7.002	2.163

(3) 上面建立的回归方程的相关指数 $R^2=0.962\ 1$,说明 x 能解释约 96.21% 的 y 值变化,因此所建立的模型能够很好地刻画 y 与 x 的关系.

例3 某城市 2001 年~2006 年人口总数 y(万)与年份 x 的关系如下表所示:

x	2001	2002	2003	2004	2005	2006
y	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 2007 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 1-3 所示.

图 1-3

由图知,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 $z=\hat{b}x+\hat{a}$ ($\hat{a}=\ln c_1$, $\hat{b}=c_2$),得到变换后的数据如下表所示:

x	2001	2002	2003	2004	2005	2006
z	3.912	4.234	4.477	4.700	5.247	5.858

作散点图如图 1-4 所示.

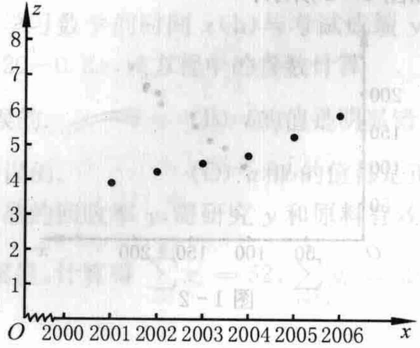


图 1-4

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.

计算得$\overline{x}=2\ 003.5$，$\overline{z}=4.738$，$\sum\limits_{i=1}^{6}x_i^2=24\ 084\ 091$，$\sum\limits_{i=1}^{6}z_i^2=137.211\ 202$，$\sum\limits_{i=1}^{6}x_iz_i=56\ 961.994$，于是$\hat{b}=\dfrac{\sum\limits_{i=1}^{6}x_iz_i-6\overline{x}\,\overline{z}}{\sum\limits_{i=1}^{6}x_i^2-6\overline{x}^2}=\dfrac{56\ 961.994-6\times2\ 003.5\times4.738}{24\ 084\ 091-6\times2\ 003.5^2}=0.371\ 2$，$\hat{a}=\overline{z}-\hat{b}\overline{x}=-738.96$. 故得线性回归方程为$\hat{z}=0.371\ 2x-738.96$，因此年份$x$与人口数$y$(万)之间的非线性回归方程为$\hat{y}=e^{0.371\ 2x-738.96}$.

(2) 令$x=2\ 007$，则$\hat{y}=e^{0.371\ 2\times2\ 007-738.96}\approx419.22$，于是估计2007年人口总数为419.22万.

(3) 列表如下：

x	2001	2002	2003	2004	2005	2006
y	50	69	88	110	190	350
$\hat{y}_i$	45.205	65.523	94.974	137.662	199.537	289.224
$\hat{e}_i=y_i-\hat{y}_i$	4.795	3.477	−6.974	−27.662	−9.537	60.776

由表中数据得残差平方和为$\sum\limits_{i=1}^{6}(y_i-\hat{y}_i)^2\approx4\ 633.581$，又$\sum\limits_{i=1}^{6}(y_i-\overline{y})^2=63\ 302.547$. 故相关指数$R^2=1-\dfrac{\sum\limits_{i=1}^{6}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{6}(y_i-\overline{y})^2}=1-\dfrac{4\ 633.581}{63\ 302.547}\approx0.926\ 8$.

例4 若有一组数据的总偏差平方和为100，相关指数为0.5，求残差平方和与回归平方和.

解 由题意知，$\sum\limits_{i=1}^{n}(y_i-\overline{y})^2=100$，$R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{100}=0.5$，则$\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2=50$，即残差平方和为50，于是回归平方和为$100-50=50$.

同步训练

理解巩固

1. 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，它们的相关指数R^2如下，其中拟合效果最好的模型是 ( )

   (A) 模型1，相关指数R^2为0.98. (B) 模型2，相关指数R^2为0.80.

   (C) 模型3，相关指数R^2为0.50. (D) 模型4，相关指数R^2为0.25.

2. 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.

   其中正确的有 ( )

   (A) ①②. (B) ②③. (C) ③④. (D) ①③.

纠错在线 JIUCUOZAIXIAN

纠错在线
JIUCUOZAIXIAN

3. 若散点图中所有样本点都在一条直线上，则解释变量和预报变量之间的残差平方和是 ( )
(A) 1. (B) 0.75. (C) 0.5. (D) 0.

4. 若某回归模型相对一组数据的残差平方和为 90，相关指数为 0.9，则其总偏差平方和为 ( )
(A) 9. (B) 100. (C) 810. (D) 900.

5. 对于一组数据的两个回归模型，其残差平方和分别为 153.4 和 200，若从中选取一个拟合程度较好的回归模型，应选残差平方和为________的那个.

6. 在回归分析中，若残差图中存在异常点(个别数据对应残差过大，或残差呈现不随机的规律性等)，则检查数据____________，或模型____________等.

7. 若某回归模型相对一组数据的总偏差平方和为 1 780，其相关指数为 0.95，则其残差平方和为________.

8. 某工厂的原料中含有两种有效成分 A 和 B，随机抽取 10 袋，测得原料中 A 和 B 的含量 $x_i, y_i(\%)$ 如下表所示：

i	1	2	3	4	5	6	7	8	9	10
x_i	24	15	23	19	16	11	20	16	17	13
y_i	67	54	72	64	39	22	58	43	46	34

(1) 作出散点图；
(2) 求回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$；
(3) 求残差平方和、总偏差平方和、相关指数 R^2.

9. 为了研究随时间 x(天)变化，某种细菌的繁殖个数 y 的情况，收集数据如下：

x	1	2	3	4	5	6
y	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) 建立解释变量与预报变量之间的回归模型；
(3) 计算残差和相关指数；
(4) 你认为这个模型能较好地刻画解释变量与预报变量的关系吗？说明理由.

纠错在线

发展提高

10. 在研究身高和体重的关系时，如果身高解释了64%的体重变化，而随机误差贡献了剩余的36%，那么求得的相关指数 R^2 约为　　　　（　　）

(A) 0.　　(B) 0.36.　　(C) 0.64.　　(D) 1.

11. 已知方程 $\hat{y}=0.849x-85.712$ 是根据女大学生的身高预报她的体重的回归方程，其中 x，$\hat{y}$ 的单位分别是 cm，kg，则在样本(165，61)处的残差是________.

12. 一台机器使用的时间较长，但还可以使用，它按不同的转速生产出来的某机械零件有些会有缺陷，每小时生产有缺陷的零件数 y 随机器的运转速度 x(转/秒)而变化.下表为抽样试验的结果：

x	16	14	12	8
y	11	9	8	5

(1) 画出散点图；

(2) 若 y 与 x 有线性相关关系，求回归直线方程；

(3) 计算相关指数 R^2；

(4) 若实际生产中，允许每小时生产的产品中有缺陷的零件最多为10件，那么机器的运转速度应控制在什么范围？

1.2　独立性检验的基本思想及其初步应用

教材解读

1. 分类变量与定量变量

分类变量也称为定性变量，它的取值一定是离散的，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.

定量变量的取值一定是实数，它们的取值大小有特定的含义，不同取值之间的运算也有特定的含义，如身高、体重等.

2. 列联表

列联表一般为两个或两个以上分类变量的汇总统计表，教科书中仅限于研究两个分类变量的列联表，并且每个分类变量只取两个值，这样的列联表称为2×2列联表.

3. 独立性检验

假设有两个分类变量 X 和 Y，它们的值域分别为 $\{x_1,x_2\}$ 和 $\{y_1,y_2\}$，其样本频数列联表(称为2×2列联表)为：

	y_1	y_2	总　计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总　计	$a+c$	$b+d$	$a+b+c+d$

若要推断的论述为 H_1："X 与 Y 有关系"，可以按如下步骤判断 H_1 成立的可能性：

(1) 先假设 H_0："X 与 Y 没有关系".

名师引路

在日常生活中，我们常常关心两个分类变量之间是否有关系.例如，吸烟是否与患肺癌有关系？性别是否对于喜欢数学课程有影响？等等.通过三维柱形图和二维条形图，可以粗略地判断两个分类变量是否有关系，但是这种判断无法精确地给出所得结论的可靠程度.本节介绍利用独立性检验来考察两个分类变量是否有关系，它能较精确地给出这种判断的可靠程度.

本节的重点是理解独立性检验的基本思想及实施步骤.难点是了解独立性检验的基本思想及随机变量 K^2 的含义.

(2) 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 $n=a+b+c+d$ 为样本容量. 计算 K^2 的观测值 k.

(3) 统计推断：当 $K^2>2.706$ 时，有 90%的把握认为“X 与 Y 有关系”；当 $K^2>6.635$ 时，有 99%的把握认为“X 与 Y 有关系”.

4. 通过三维柱形图和二维条形图，可以粗略地判断两个分类变量是否有关系.

在二维条形图中，可以估计满足条件 $X=x_1$ 的个体中具有 $Y=y_1$ 的个体所占的比例 $\dfrac{a}{a+b}$，也可以估计满足条件 $X=x_2$ 的个体中具有 $Y=y_1$ 的个体所占的比例 $\dfrac{c}{c+d}$. 两个比例的值相差越大，“X 与 Y 有关系”的可能就越大.

在三维柱形图中，主对角线上两个柱形高度的乘积 ad 与副对角线上两个柱形高度的乘积 bc 相差越大，“X 与 Y 有关系”的可能性越大.

名师引路 MINGSHIYINLU

独立性检验的思想来自于统计上的假设检验思想，它与反证法类似. 假设检验和反证法都是先假设结论不成立，然后根据是否能够推出“矛盾”来断定结论是否成立. 但两者“矛盾”的含义不同，反证法中的“矛盾”是指不符合逻辑的事件的发生；而假设检验中的“矛盾”是指不符合逻辑的小概率事件的发生，即在结论不成立的假设下推出有利于结论成立的小概率事件的发生.

第 1 课时 独立性检验(1)

典例剖析

例 1 通过三维柱形图和二维条形图 ( )

(A) 可以粗略地判断两个分类变量是否有关系，精确地给出结论的可靠程度.

(B) 可以粗略地判断两个分类变量是否有关系，不能给出结论的可靠程度.

(C) 可以精确地判断两个分类变量是否有关系.

(D) 可以给出结论的可靠程度.

解 选 B.

例 2 在独立性检验中，选用随机变量 K^2，用其取值大小来推断独立性是否成立. 当 K^2 满足条件________时，我们有 99%的把握认为两个变量有关系.

解题思路 利用教科书中的临界值表：

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

解 $K^2>6.635$.

例 3 为了研究色盲与性别的关系，调查了 1 000 人. 结果男性中 442 人正常，38 人色盲；女性中 514 人正常，6 人色盲.

(1) 根据以上数据建立一个 2×2 列联表；

(2) 在多大程度上可以认为色盲与性别之间有关系？为什么？

解题思路 第(2)题先计算随机变量 K^2 的观测值，然后利用临界值表.

解 (1) 列联表如下：

	男	女	总 计
正 常	442	514	956
色 盲	38	6	44
总 计	480	520	1 000

(2) 由列联表，得 K^2 的观测值为

$$k=\frac{1\,000(442\times6-38\times514)^2}{956\times44\times480\times520}\approx27.139>10.828.$$

由临界值表知 $P(K^2\geqslant10.828)=0.001$，所以有 99.9%的把握认为色盲与性别有关.

解题方略 JIETIFANGLUE

本课时主要有三类题型：

(1) 利用三维柱形图和二维条形图来粗略分析两个分类变量否有关系. 在三维柱形图中，主对角线上两个柱形高度的乘积 ad 与副对角线上两个柱形高度的乘积 bc 相差越大，两分类变量有关系的可能性也越大. 在二维条形图中，满足条件 $X=x_1$ 的个体中具有 $Y=y_1$ 的个体所占比例 $\dfrac{a}{a+b}$，满足条件 $X=x_2$ 的个体中具有 $Y=y_1$ 的个体所占的比例 $\dfrac{c}{c+d}$，两个比例的值相差越大，两分类变量有关系的可能性也越大. 如：例 1.

(2) 有关独立性检验的概念题. 已知随机变量 K^2 的观测值，求有多少把握认为两分类变量有关. 或者已知两分类变量有关的把握程度，求随机变量 K^2 的观测值 k 的范围，关键是熟悉临界值表. 如：例 2.