

王后雄学案

教材完全解读

选修·专题



6大特色引发学考热潮
推动学习模式全面升级

独具创意

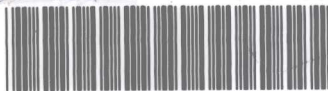
考向指引

考试工具

同步突破

典例导思

核心预测



YZLI0890162204

高中数学 选修1-2

丛书主编：王后雄
本册主编：童亚新 赵志培



全国百佳图书出版单位
Top 100 publishing house in China

王后雄学案

教材完全解读

总策划：熊 辉

选修·专题

高中数学 选修1-2

丛书主编：王后雄
本册主编：童亚新 赵志培
编委：胡建平 丁仁贵
王 涛 汪 飞
林菊芳



YZLI0890152204



全国百佳图书出版单位
Top 100 publishing house in China

图书在版编目(CIP)数据

教材完全解读. 高中数学. 1-2: 选修 / 童亚新主编.—5版.—南宁: 接力出版社, 2011.9

ISBN 978-7-80732-471-3

I. ①教… II. ①童… III. ①中学数学课—高中—教学参考资料 IV. ①G634

中国版本图书馆CIP数据核字(2011)第192199号

总策划: 熊辉

责任编辑: 吴惠娟

责任校对: 陈娟

封面设计: 木头羊

JIAOCAI WANQUAN JIEDU

GAOZHONG SHUXUE

教材完全解读

高中数学 选修1-2

丛书主编: 王后雄 本册主编: 童亚新 赵志培

*

社长: 黄俭 总编辑: 白冰

接力出版社出版发行

广西南宁市园湖南路9号 邮编: 530022

E-mail: jielipub@public.nn.gx.cn

河南省瑞光印务股份有限公司印刷 全国新华书店经销

*

开本: 889毫米×1194毫米 1/16 印张: 10 字数: 264千

2011年10月第5版 2011年10月第6次印刷

ISBN 978-7-80732-471-3

定价: 16.30元

如有印装质量问题,可直接与本社调换。如发现画面模糊,字迹不清,断笔缺画,严重重影等疑似盗版图书,请拨打举报电话。

盗版举报电话: 0771-5849336 5849378

读者服务热线: 4006-980-700

教材完全解读

本书特点

1. 以《课程标准》、《考试大纲》为编写依据，完全解读知识、方法、能力、考试题型，全面提高学习成绩。
2. 采用国际流行的双栏对照案例编写方式，左栏对教材全解全析，在学科层次上力求讲深、讲透、讲出特色；右栏用案例诠释考点，对各个考点各个击破。

明确每课学习要求

以课标为依据，三维目标全解教材学习要求，提供总体的学习策略，提出具体的学习要诀，体现目标控制学习规则。

三层完全解读

从知识、方法、思维三个方面诠释教材知识点和要点，帮您形成答题要点、解题思维，理清解题思路，揭示考点实质和内涵。

教材完全解读 高中数学 选修1-2

第1章 统计案例

1.1 回归分析的基本思想及其初步应用

课标三维目标

1. 知识与技能：通过对具体问题的分析，了解回归分析的必要性，了解回归分析的一般步骤，会求回归直线方程，并能应用所学知识对实际问题进行回归分析，体会回归分析的实际价值和基本思想，并能用回归分析对具体事件进行分析。
2. 过程与方法：本节内容从实际问题中的数量之间的关系入手，求出相应的回归直线方程，从中找出存在的不足，从而有进行回归分析的必要性，进而学习相关系数，用相关系数来刻画回归的效果，进而得出回归分析的一般步骤，并对具体问题进行分析，用于解决实际问题。
3. 情感、态度与价值观：从实际问题中发现自己已有知识的不足之处，激发学生的求知心和求知欲，培养学生不满足于已有知识，勇于求知的好品质，引导学生积极建构，任何事物都是相对的，但又有一定的规律性，我们只要从实际出发，不断探索事物的内在联系，就会找出其中的规律性，形成解决实际问题的方法和能力。

三层完全解读

知识·能力·思维

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

1. 求回归直线方程

>>>解题依据

>>>名题诠释

例1 背景题——2011·江西高考题

变量 x 与 y 相对应的一组数据为 $(10, 1), (11, 3.2), (11.8, 3), (12.5, 4), (13.5, 5)$ ；变量 U 与 V 相对应的一组数据为 $(10, 5), (13, 4), (11.8, 3), (12.5, 2), (13, 1)$ 。表示变量 Y 与 X 之间的线性相关系数为 r_1 ，表示变量 V 与 U 之间的线性相关系数为 r_2 ，则

$A. r_1 < r_2 < 0, B. 0 < r_2 < r_1, C. r_2 < 0 < r_1, D. r_1 < r_2$

教材完全解读 高中数学 选修1-2

4. 相关性检验

(1) 相关性检验的意义

在实际问题中，利用散点图来判断线性相关关系，很快又发现，这时散点图中的各个点应该集中在某一直线的附近，依据求回归直线方程的方法，求出与之相对应的回归直线方程，这条回归直线方程具有实际的应用价值。但是，若散点图中画出的各个点，不是集中在一条直线的附近，而是集中在某一个圆内，依据求回归直线的方法，我们仍然可以求出相应的回归直线方程，但这条回归直线方程已经不能反映出这组数据的变化规律，这时求出的回归直线方程也就失去了意义。那么，到底在数据有什么特征时，才是线性相关的，才能求出回归直线方程呢？这就需要用到 x 与 y 作线性相关性检验，也就是相关性检验。

对于随机抽取的变量 x 与 y 的 n 对数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，检验统计量是样本相关系数 $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$ ，具有以下性质： $|r| \leq 1$ 并且 $|r|$ 越接近于 1，线性相关程度越强； $|r|$ 越接近于 0，线性相关程度越弱。

(2) 相关性检验的步骤

① 作统计假设：假设 x 与 y 不具有线性相关关系；

② 根据小概率 α 与 $n-2$ 在附表中查出 r 的一个临界值 $r_{\alpha, n-2}$ ；

③ 根据样本相关系数的计算公式 $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$ 计算出 r 的值；

④ 作统计判断：如果 $|r| > r_{\alpha, n-2}$ ，表明有 95% 的把握认为 x 与 y 之间具有线性相关关系；如果 $|r| \leq r_{\alpha, n-2}$ ，这时寻找回归直线方程是无意义的。

方法·技巧平台

1. 利用图形分析残差的特性

作图时纵坐标为残差，横坐标可以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。在残差图中，如果有两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存在人为的偏差。如果数据收集有错误，可以先予以纠正，然后重新利用线性回归模型拟合数据。如果数据收集没有错误，则需要寻找其他原因。另外，若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就越高。

$\sum_{i=1}^5 x_i^2 = 84^2 + 64^2 + \dots + 57^2 = 47\ 384$ ，
 $\sum_{i=1}^5 x_i = 120 \times 84 + 108 \times 64 + \dots + 108 \times 71 = 73\ 796$ ，
所以相关系数 $r = \frac{73\ 796 - 10 \times 107.8 \times 68}{\sqrt{(116\ 584 - 10 \times 107.8^2)(47\ 384 - 10 \times 68^2)}} = 0.750\ 6$ ，
查表：显著性水平 $\alpha = 0.05$ ，自由度 $10 - 2 = 8$ 相应的临界值 $r_{\alpha, n-2} = 0.632$ ，由 $r > r_{\alpha, n-2}$ 知两次数学考试成绩有显著的线性相关关系。

【点拨】如果两个变量不具备相关关系，或者相关关系不显著，即使求出回归方程也是无意义的，用统计分析和判定不可行。

背景题

某地区为研究城镇居民家庭人均月生活费支出和人均月收入的相关关系，随机抽取 10 户进行调查，其结果如下：

人均月收入 x /元	人均月生活费 y /元
300	255
390	324
420	335
520	360
570	450
700	520
760	580
800	600
850	630
1 080	750

试预测月人均收入为 1 100 元和月人均收入为 1 800 元的两个家庭的月人均生活费，并进行残差分析。

【解析】未通过数据点画出的两个变量呈线性相关，在用残差图和相关系数法估计模型拟合的效果，进而估计预测值的准确性。

【解】作出散点分布图（如图 1-1-8），由图可知，月人均生活费与月人均收入之间具有线性相关关系。

通过计算可知 $\bar{x} = 659.7, \bar{y} = 480.4, \sum_{i=1}^{10} x_i^2 = 4\ 610\ 300, \sum_{i=1}^{10} y_i^2 = 2\ 540\ 526, \sum_{i=1}^{10} x_i y_i = 3\ 417\ 560, \sum_{i=1}^{10} (x_i - \bar{x})^2 = 106\ 717, \sum_{i=1}^{10} (y_i - \bar{y})^2 = 58\ 724$ ，
线性回归方程为 $\hat{y} = 0.659\ 9x + 58.724$ 。

作残差图如图 1-1-9 所示，由图可知，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。

计算相关指数 $R^2 = 0.986\ 3$ ，说明城镇居民的人均月生活费的差异有 98.63%。

图 1-1-8

图 1-1-9

整体训练方法

针对本节重点、难点、考点及考试能力达标所设计的题目。题目难度适中,是形成能力、考试取得高分的必经阶梯。

整体训练方法

能力·题型设计

题感基础训练

1. 对命题“对顶角相等”的说法中正确的是()。
 - A. 前提是“对顶角”,结论是“相等”
 - B. 前提是“两个角是对顶角”,结论是“这两个角相等”
 - C. 前提是“两个角相等”,结论是“这两个角是对顶角”
 - D. 前提是“两个角相等”,结论是“两个角全等”
2. 已知等式 $1^2 + 2^2 + \dots + n^2 = \frac{1}{3}(5n^2 - 7n + 4)$, 则()。
 - A. n 为任何正整数时都成立
 - B. 仅当 $n=1, 2, 3$ 时成立
 - C. 当 $n=4$ 时成立, $n=5$ 时不成立
 - D. 仅当 $n=4$ 时不成立
3. 下列类比推理恰当的是()。
 - A. 把 $(b+c) \times \log(x+y)$ 类比, 则有 $\log(x+y) = \log x + \log y$
 - B. 把 $(b+c) \times \sin(x+y)$ 类比, 则有 $\sin(x+y) = \sin x + \sin y$
 - C. 把 $(ab)^2$ 与 $(a+b)^2$ 类比, 则有 $(a+b)^2 = a^2 + b^2$
 - D. 把 $(b+c) \times a$ 与 $(b+c)$ 类比, 则有 $a \cdot (b+c) = a \cdot b + a \cdot c$
4. 下列说法正确的是()。
 - A. 合情推理得出的结论一定是正确的
 - B. 合情推理必须有前提有结论
 - C. 合情推理不能猜想
 - D. 合情推理得出的结论无法判定正误

点击考例

- 测试要点 1
【例 1】
- 测试要点 2
【例 2】
- 测试要点 3
- 测试要点 4

教材完全解读 高中数学 选修 1-2

10. 对称性

若 y 与 x 之间有线性相关关系, 设 y 对 x 的回归直线方程为 $\hat{y} = \hat{b}x + \hat{a}$, 则

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

如果我们知道: 相关关系并不是因果关系, 变量 y 与 x 之间有相关关系时, 也可将 y 作为自变量, x 看作因变量, 求 x 对 y 的回归直线方程. 设 x 对 y 的回归直线方程为 $\hat{x} = \hat{d}y + \hat{c}$, 由 x 与 y 的对称性, 即得

$$\hat{d} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n y_i^2 - n \bar{y}^2}$$

值得注意的是: 回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 与 $\hat{x} = \hat{d}y + \hat{c}$ 未必垂直, 比如由于

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

故未必有 $\hat{b} \cdot \hat{d} = -1$, 即两条直线的斜率未必互为负倒数。

整体训练方法

能力·题型设计

题感基础训练

1. 在下列各量之间, 存在相关关系的是()。
 - ① 正方体的体积与棱长之间的关系;
 - 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。
2. 有关回归方程的叙述正确的是()。
 - A. 回归方程只适用于所研究的样本

点击考例

- 测试要点 1.4
- 测试要点 1

解题错因导引

“点击考例”栏目导引每一道试题的“测试要点”。当您解题出错时, 建议您通过“测试要点”的指向, 弄清致错原因, 形成正确答案。

教材完全解读 高中数学 选修 1-2

教材课后习题解答

人数 A 组

1. 若作散点图的目的通过观察散点图判断两个变量之间有什么样的关系, 以决定是否应该用线性回归模型来拟合原始数据. 说明: 在对常用的函数图像了解的情况下, 通过观察散点图可以判断两个变量的关系是“哪种函数”。
2. 答: 分析残差可以帮助我们解决以下两个问题:
 - (1) 寻找异常点, 就是残差特别大的点, 考察异常点的样本数据是否有误;
 - (2) 分析残差图可以发现模型选择的点是否合适. 说明: 残差分析是回归诊断的一部分, 可以帮助我们发现在样本数据中的排列, 分析模型选择是否合适, 是否有其他变量影响模型拟合, 模型是否需要改进等等。
3. 答: (1) 解释变量和预报变量的关系是线性函数关系. (2) $r^2 = 1$. 说明: 如果所有的样本点都在一条直线上, 建立的线性回归模型一定是准确的, 所以每个样本点的残差均为 0, 残差平方和也为 0, 即此处的模型为 $\hat{y} = bx + a$, 没有随机误差项, 是严格的正、负函数关系. 通过计算可以证明解释变量和预报变量之间的相关系数是 1.

习题 1.1 (P9)

解: (1) 散点图中数据制作的散点图如图 1-1 所示, 从散点图中可以看出 GDP 与年份近似呈线性关系.

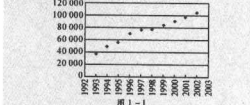


图 1-1

(2) 用 y_t 表示 GDP, t 表示年份, 根据最小二乘法, 得 $\hat{a} = -14 292 537.729$, $\hat{b} = 7 191.969$, 从而得线性回归方程 $\hat{y} = 7 191.969t - 14 292 537.729$. 残差计算结果见下表.

GDP 与年份线性回归模型					
年份	1993	1994	1995	1997	
残差	-6 422.097	-1 489.066	3 037.665	5 232.196	
GDP 与年份二次回归模型					
年份	1998	1999	2000	2001	2002
残差	1 328.838	-2 140.811	-1 932.118	-1 277.480	-993.618

(3) 2003 年的 GDP 预报值为 112 976.187, 根据国家统计局 2004 年统计, 2003 年实际 GDP 为 117 251.9, 所以预报与实际相差 -4 275.713.

(4) 由散点图可知散点图 $r^2 = 0.974$, 说明年份能够解释 97.4% 的 GDP 变化, 因此建立的模型能够很好地拟合 GDP 和年份的关系.

说明: 关于 2003 年的 GDP 的预报, 不同渠道可能会有所不同.

2. 说明: 去重的站 25 000 站, 与具体的数 20 000 站有关, 所以至少要 15 000 站才唯一.

3. 解: 由表中数据得散点图如图 1-2 所示, 从散点图中可以

看出, 震级 x 与大于或等于该震级的地震数 N 之间不呈线性相关关系. 随着 x 的增加, 所考察的地震数 N 近似地以指数形式增长. 将变换 $y = \lg N$ 得到的数据如下表所示.

x	2.1	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0	
N	4 403	2 309	1 770	1 409	1 049	803	574	426	321	241	181	139	104	79	59	44	33	25	19	14	11	8	6	4	3	2	2	1	1	1

x 和 y 的散点图如图 1-3 所示, 从这个散点图中可以看出 x 和 y 具有很强的线性相关性, 因此可以用线性回归模型拟合它们之间的关系. 根据最小二乘法计算公式, 得 $\hat{a} = 6.704$, $\hat{b} = -0.741$.

线性回归方程为 $\hat{y} = -0.741x + 6.704$.

相关指数 $R^2 = 0.997$, 说明 x 可以解释 y 的 99.7% 的变化. 因此, 可以用回归方程 $N = 10^{-0.741x + 6.704}$ 描述 x 和 N 之间的关系. 说明: 对非线性相关关系, 要选择恰当的变换化为线性相关关系, 再用最小二乘法求线性回归方程.

编写 (P15)

解: 列表的等高条形图如图 1-4. 从图中可以看出成绩与性别有关.

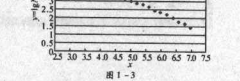


图 1-4

假设成绩与性别无关, 则有 $a = 10$, $b = 35$, $c = 7$, $d = 38$, $n = 90$, $e = 45$, $f = 45$, $g = 17$, $h = 73$, $n = 90$, 代入 A 公式, 得 K^2 的观测值 $k = \frac{90 \times (10 \times 38 - 7 \times 35)}{45 \times 45 \times 17 \times 73} \approx 0.653$.

$\therefore 0.653 < 2.706$, 没有充分证据说明优秀与性别有显著关系, 故与性别与成绩关系无显著性关联的概率为 99%.

习题 1.2 (P16)

1. 解: $10, a = 45$, $20, d = 30$, $a + b = 55$, $c + d = 50$, $n = 30$, $b + d = 75$, $n = 105$.

$k = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

$= \frac{105 \times (10 \times 30 - 45 \times 20)^2}{55 \times 50 \times 75 \times 30} \approx 6.11$.

$\therefore 6.11 > 5.024$, 有 97.5% 的把握认为药物有效.

2. 解: $16, a = 28$, $20, d = 8$, $n = 64$, $c + d = 28$, $n = 36$, $b + d = 36$, $n = 72$.

教辅大师、特级教师王后雄教授科学超前的体例设置，帮您赢在学习起点，成就人生夙愿。

题记

教材完全解读 高中数学 选修1-2

单元知识梳理与能力整合

高考命题趋向

- 了解回归分析的基本思想、方法及常用软件.
 - 了解独立性检验(只要求 2×2 列联表)的基本思想、方法及常用软件.
- 统计作为选修内容,会在近几年高中数学中考有所体现,而且考查量基本的统计知识.

归纳·总结·专题

一、知识结构图



- (3) 若 $K^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 的值:
- (4) 列联表可能性的独立性检验的一般步骤:
- (1) 根据样本数据制成 2×2 列联表.
- (2) 根据公式 $K^2 = \frac{n(ad-bc)^2}{n_1n_2n_3n_4}$ 计算 K^2 的值.
- (3) 比较 K^2 与临界值的大小关系并作出统计推断.
- (三) 基本能力总结
1. 两个变量不一定有线性关系,不能直接利用线性回归方程建立两个变量的关系,可以通过变换的方法转化为线性回归问题.

单元知识整合

单元知识与方法网络化,帮助您将本单元所学教材内容系统化,形成对考点知识的二次提炼与升华,全面提高学习效率.

教材完全解读 高中数学 选修1-2

知识与能力同步测控题

测控时间:90分钟 本卷满分:150分

一、选择题(本大题满分10小题,每小题5分,共50分.在每小题给出的四个选项中,只有一项是符合题目要求的)

- 某校高三年级学生在学习数学的时间(x)与考试成绩(y)之间的线性回归直线方程为 $\hat{y} = a + bx$, 经计算, 方程为 $\hat{y} = 20 - 0.8x$, 该方程中参数().
A. a 值是明显不对的 B. b 值是明显不对的
C. a 值和 b 值都是不对的 D. a 值和 b 值都是正确的
- 对回归变量 y 和 x 进行线性相关性检验, 已知 a 是显著值, 则 r 是相关系数. 已知: ① $a = 7, r = 0.9545$; ② $a = 15, r = 0.3812$; ③ $a = 17, r = 0.4065$; ④ $a = 3, r = 0.9870$. 则变量 y 与 x 具有线性相关关系的题().
A. ①② B. ①③
C. ②④ D. ③④
- 工人月工资(y)依劳动生产率(x , 千元)变化的回归直线方程为 $\hat{y} = 60 + 90x$, 下列判断正确的是().

- C. $1 - P_1$
D. $1 - (1 - P_1)(1 - P_2)$

7. 考察棉花种子经过处理与生病之间的关系得到如下表数据:

	种子未处理	种子经处理	合计
得病	32	101	133
不得病	61	213	274
合计	93	314	407

- 根据以上数据, 则().
- 种子经处理与是否生病有关
 - 种子经处理与是否生病无关
 - 种子是否经处理决定是否生病
 - 以上都是错误的
8. 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为().

考试高分保障

精心选编涵盖本章节或阶段性知识和能力要求的检测试题, 梯度合理、层次分明, 与同步考试接轨, 利于您同步自我测评, 查缺补漏.

答案与提示

第一章 统计案例

1.1 回归分析的基本思想及其初步应用

能力题组设计

1.1.1

2. C 【提示】回归分析的基本思想, 利用最小二乘法, 它不是一种精确的, 也不是唯一的回归模型, 但是样本的取值范围是回归方程的适用范围.

3. D

4. A 【提示】由已知 $\hat{y} = 20 - 0.8x$, $\hat{y} = 20 - 0.8x$, $\hat{y} = 20 - 0.8x$, $\hat{y} = 20 - 0.8x$.

5. (1, 107.2, 2.3925) 【提示】线性回归方程必过样本中心点 (\bar{x}, \bar{y}) , 计算可知 $(1, 107.2, 2.3925)$.

6. 1 【提示】若 a, b 互为相反数, 即 $a = -b$, 则 $\hat{y} = a + bx = a - bx = a - b(x - 1) = a - b(x - 1) + b = a + b - b(x - 1) = 1 - b(x - 1)$.

7. D 【提示】根据最小二乘法的基本思想, 回归方程的系数应使残差平方和最小.

8. 11.69 【提示】当 $x = 25$ 时, $\hat{y} = 0.5 \times 25 - 0.1 = 11.69$.

9. 0 【提示】解相关系数

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

$$\bar{x} = 101, \bar{y} = 10, 113, \sum_{i=1}^n x_i^2 = 161, 125, \sum_{i=1}^n y_i^2 = 628, 55, \sum_{i=1}^n x_i y_i = 16, 076, 8$$

$$r = \frac{16, 076, 8 - 10 \times 113}{\sqrt{(161, 125 - 10 \times 113)(628, 55 - 10 \times 113)}} = \frac{16, 076, 8 - 1, 130}{\sqrt{(151, 125)(415, 55)}} = \frac{14, 946, 8}{\sqrt{62, 812, 875}} \approx 0, 865, 2$$

这说明蔬菜产量与氮肥量之间存在较强的线性关系.

(2) 设所求的线性回归方程为 $\hat{y} = a + bx$, 则

$$\begin{cases} a + 10b = 10 \\ a + 113b = 101 \end{cases}$$

$$\Rightarrow \begin{cases} a = 10 - 10b \\ 10 - 10b + 113b = 101 \end{cases}$$

$$\Rightarrow \begin{cases} a = 10 - 10b \\ 103b = 91 \end{cases}$$

$$\Rightarrow \begin{cases} a = 10 - 10 \times \frac{91}{103} \\ b = \frac{91}{103} \end{cases}$$

$$\Rightarrow \begin{cases} a = \frac{100 - 910}{103} \\ b = \frac{91}{103} \end{cases}$$

$$\Rightarrow \begin{cases} a = -\frac{810}{103} \\ b = \frac{91}{103} \end{cases}$$

$$\Rightarrow \begin{cases} a \approx -7, 864 \\ b \approx 0, 8835 \end{cases}$$

$$\Rightarrow \hat{y} = -7, 864 + 0, 8835x$$

$$\Rightarrow \hat{y} = -7, 864 + 0, 8835 \times 100 = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

$$\Rightarrow \hat{y} = 12, 136$$

点拨解题思路

试题皆提供详细的解题步骤和思路点拨, 鼓励一题多解. 不但知其然, 且知其所以然, 帮助您养成良好规范的答题习惯.

目录

模块学习指南	1
全书知识结构图解·名师学法指津	2
第1章 统计案例	4
1.1 回归分析的基本思想及其初步应用	4
1.2 独立性检验的基本思想及其初步应用	13
教材课后习题解答	20
单元知识梳理与能力整合	22
知识与能力同步测控题	26
第2章 推理与证明	28
2.1 合情推理与演绎推理	28
2.1.1 合情推理	28
2.1.2 演绎推理	39
2.2 直接证明与间接证明	50
2.2.1 综合法和分析法	50
2.2.2 反证法	62
教材课后习题解答	70
单元知识梳理与能力整合	73
知识与能力同步测控题	78

目录

第3章 数系的扩充与复数的引入	79
3.1 数系的扩充和复数的概念	79
3.1.1 数系的扩充和复数的概念	79
3.1.2 复数的几何意义	84
3.2 复数代数形式的四则运算	90
3.2.1 复数代数形式的加减运算及其几何意义	90
3.2.2 复数代数形式的乘除运算	97
教材课后习题解答	105
单元知识梳理与能力整合	107
知识与能力同步测控题	112
第4章 框图	113
4.1 流程图	113
4.2 结构图	120
教材课后习题解答	126
单元知识梳理与能力整合	129
知识与能力同步测控题	132
教材学业水平考试试题	134
答案与提示	136

知识与方法 阅读索引

第1章 统计案例

1.1 回归分析的基本思想及其初步应用

- 1. 求回归直线方程 4
- 2. 随机误差 5
- 3. 回归分析 5
- 4. 相关性检验 6
- 5. 利用图形分析残差的特性 6
- 6. 建立回归模型的基本步骤 8
- 7. 非线性回归问题的处理方法 8
- 8. 回归系数 \hat{a} 、 \hat{b} 的简单推导方法 9
- 9. 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 9
- 10. 对称性 10

1.2 独立性检验的基本思想及其初步应用

- 1. 相互独立的含义 13
- 2. 与列联表相关的概念 13
- 3. 独立性检验 14
- 4. 独立性检验的基本思想 15
- 5. 两个分类变量相关性的检验方法 15
- 6. 卡方 χ^2 的表达式的推导 17
- 7. K^2 大小划分的由来 17

第2章 推理与证明

2.1 合情推理与演绎推理

2.1.1 合情推理

- 1. 推理 28
- 2. 归纳推理 29

- 3. 类比推理 30
- 4. 合情推理 32
- 5. 关于类比的几种题型 32
- 6. 归纳法分类 35

2.1.2 演绎推理

- 1. 演绎推理 39
- 2. 三段论法 39
- 3. 合情推理与演绎推理的区别 40
- 4. 应用三段论证明数学问题 41
- 5. 推理证明的基本规则 45
- 6. 演绎推理的常见推理形式 46

2.2 直接证明与间接证明

2.2.1 综合法和分析法

- 1. 综合法 50
- 2. 分析法 51
- 3. 综合法与分析法的联系与区别 52
- 4. 综合法与分析法应用举例 54
- 5. 几种常见的证明方法 55
- 6. 证明不等式要注意的几个问题 58
- 7. 直接证明方法在高考中的体现 59

2.2.2 反证法

- 1. 反证法 62
- 2. 反证法的分类 62
- 3. 用反证法证明命题“若 p 则 q ” 62
- 4. 反证法的一般步骤 62
- 5. 常见的主要矛盾 63
- 6. 反证法所能证明的问题类型 64
- 7. 应用反证法证明问题时应注意的问题 64
- 8. 反证法应用举例 65

9. 反证法思想的应用	67
-------------	----

第3章 数系的扩充与复数的引入

3.1 数系的扩充和复数的概念

3.1.1 数系的扩充和复数的概念

1. 数系扩充的原则与虚数单位	79
2. 复数的代数形式	79
3. 复数的分类	80
4. 复数相等	80
5. 怎样判定一个复数是什么类型的数	80
6. 复数相等充要条件的适用	81
7. 两个复数只能说相等或不相等,不能比较大小	81

3.1.2 复数的几何意义

1. 复平面	84
2. 复数的几何意义	84
3. 复数的模	85
4. 复数几何意义应用举例	85
5. 复数模的几何意义的拓展	86
6. 关于模的最值问题	87

3.2 复数代数形式的四则运算

3.2.1 复数代数形式的加减运算及其几何意义

1. 复数的加法	90
2. 复数加法的运算律	90
3. 复数加法的几何意义	91
4. 复数的减法	91
5. 复数减法的几何意义	92
6. 复数加(减)法运算公式	92
7. 复数加(减)法的几何意义的应用	92
8. 复平面内两点间距离公式的复数表示式	93

9. 复平面的轨迹问题	93
-------------	----

3.2.2 复数代数形式的乘除运算

1. 复数的乘法	97
2. 复数乘法的运算律	97
3. 复数的乘方	97
4. 共轭复数	98
5. 复数的除法	98
6. 虚数单位 i 的乘方	98
7. 三次虚根 ω 的性质	99
8. 复数运算与技巧	99
9. 证明 $z \in \mathbf{R}$ 的方法	100
10. 证明 z 为纯虚数的方法	101
11. 重要等式 $z \cdot \bar{z} = z ^2 = \bar{z} ^2$ 的定义及应用	101
12. 复数集上的方程	102

第4章 框图

4.1 流程图

1. 程序框图	113
2. 工序流程图	114
3. 流程图	114
4. 流程图的构成	114
5. 工序流程图的画法	114
6. 通过设计算法编写程序	115

4.2 结构图

1. 结构图的概念	120
2. 结构图的画法	121
3. 结构图的具体画法	121
4. 结构图与流程图的异同	121
5. 体验用结构图表示所学知识的优越性	122

模块学习指南

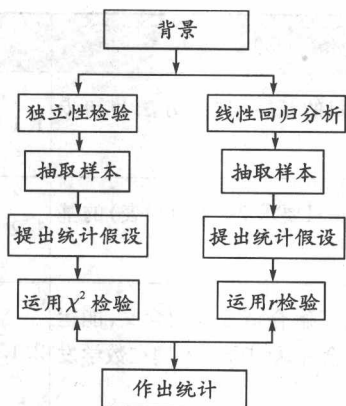
“课程标准”与“完全解读”内容对照表

序号	课程标准(高中数学选修1-2)	完全解读内容*页码
1	通过对典型案例的探究,进一步了解回归的基本思想、方法及初步应用	1.1 回归分析的基本思想及其初步应用 * P ₄
2	通过对典型案例的探究,了解独立性检验(只要求 2×2 列联表)的基本思想、方法及初步应用	1.2 独立性检验的基本思想及其初步应用 * P ₁₃
3	结合已学过的数学实例和生活中的实例,了解合情推理的含义,能运用归纳和类比等方法进行简单的推理,体会并认识合情推理在数学发展中的作用	2.1.1 合情推理 * P ₂₈
4	结合已学过的数学实例和生活中的实例,体会演绎推理的重要性,掌握演绎推理的基本方法,并能运用它们进行一些简单推理.通过具体实例,了解合情推理和演绎推理之间的联系和差异	2.1.2 演绎推理 * P ₃₉
5	结合已经学过的数学实例,了解直接证明的两种基本方法:分析法和综合法,并了解分析法和综合法的思考过程 and 特点	2.2.1 综合法和分析法 * P ₅₀
6	结合已经学过的数学实例,了解间接证明的一种基本方法——反证法,并了解反证法的思考过程 and 特点	2.2.2 反证法 * P ₆₂
7	在问题情境中了解数系的扩充过程,体会实际需求与数学内部的矛盾在数系扩充过程中的作用,感受人类理性思维的作用及数与现实世界的联系.理解复数的基本概念以及复数相等的充要条件	3.1.1 数系的扩充和复数的概念 * P ₇₉
8	了解复数的几何意义	3.1.2 复数的几何意义 * P ₈₄
9	(1)了解复数的代数表示法及其几何意义 (2)能进行复数代数形式的四则运算,了解复数代数形式的加、减运算的几何意义	3.2.1 复数代数形式的加减运算及其几何意义 * P ₉₀ 3.2.2 复数代数形式的乘除运算 * P ₉₇
10	(1)通过具体实例,进一步认识程序框图 (2)通过具体实例,了解工序流程图 (3)能绘制简单实际问题的流程图,体会流程图在解决实际问题中的作用	4.1 流程图 * P ₁₁₃
11	(1)通过实例,了解结构图,运用结构图梳理已学过的知识、收集整理资料信息 (2)结合作出的结构图与他人进行交流,体会结构图在揭示事物联系中的作用	4.2 结构图 * P ₁₂₀

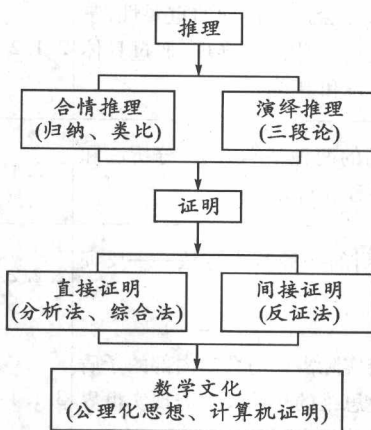
全书知识结构图解·名师学法指津

一 全书知识结构图解

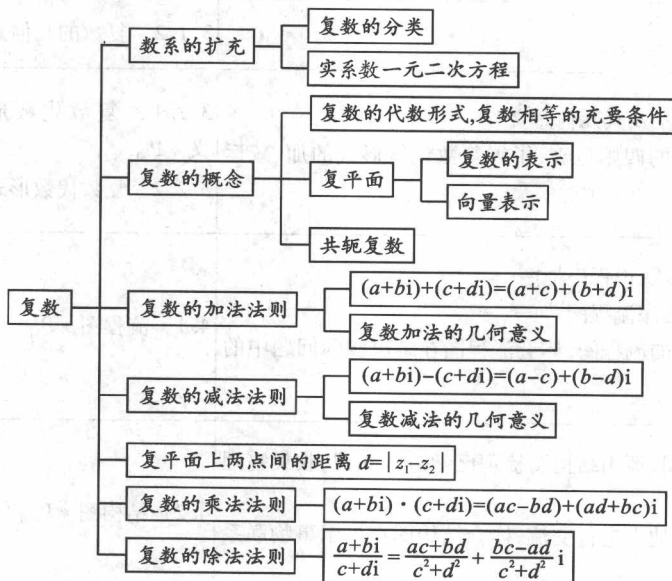
第一章 统计案例



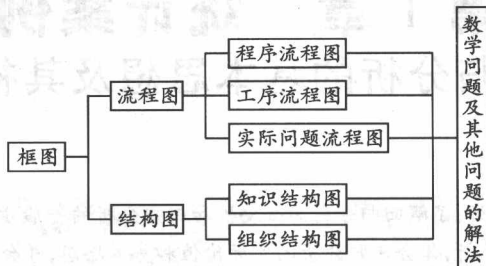
第二章 推理与证明



第三章 数系的扩充与复数的引入



第四章 框图



二 名师学法指津

第一章“统计案例”中学习的统计方法是在前面学习统计和回归直线的基础上,通过分析诸多典型案例来进行的,在学习时要注意以下三点:

1. 同学们需要亲自实践,参与解决具体的实际问题,学习处理数据的一些常用方法,会用必要的现代技术手段来处理有关数据,如计算器、计算机等.
2. 学习回归分析要在复习、回顾两个变量的相关性的基础上进行,做到前后联系,逐步推进.
3. 在独立性检验中要特别关注方法的应用,不要过多探究方法的理论依据;要注意不能单纯记忆公式和套用公式,要多探寻规律,合理建模,形成方法.

第二章“推理与证明”中内容的实用性较强,同时又是数学逻辑性的重要体现.在学习中要做到以下两点:

1. 多联系生活实际和其他学科的结论,体会推理的应用性.
2. 通过数学实例去理解合情推理和演绎推理,不必刻意追求概念的抽象表述.

第三章“数系的扩充与复数的引入”,新教材降低了对内容的要求,只要求学习复数的概念、复数的代数形式及几何意义、加减乘除运算及加减法的几何意义.学习时要做到以下三点:

1. 在学习复数的概念和运算时,要注意避免繁琐的计算,要学会灵活运用一些基本技巧.
2. 在学习复数的几何意义时,要学会运用数形结合的思想解决有关问题.
3. 多阅读一些引申的内容,扩大自己的知识面及数学视野.

第四章“框图”,本章是体现数学应用性的重要内容,学习时要做到以下三点:

1. 学习框图,要从分析实际入手,学会运用框图表示数学计算和证明过程中的主要思路与步骤、实际问题中的工序流程、某一知识系统的结构关系等.
2. 在运用框图的过程中理解流程图和结构图的特征.
3. 积极实践,掌握框图的用法,体验用框图解决问题的优越性.

第1章 统计案例

1.1 回归分析的基本思想及其初步应用

► 课标三维目标

1. 知识与技能:通过对具体问题的分析,了解回归分析的必要性和回归分析的一般步骤,会求回归直线方程,作散点图,并会运用所学习的知识对实际问题进行回归分析,体会回归分析的实际价值和基本思想,并会用回归分析对具体事件进行分析。

2. 过程与方法:本节内容先从实际问题中的数量之间的关系入手,求出相应的回归直线方程,从中也找出存在的不足,从而有进行回归分析的必要性,进而学习相关指数,用相关指数来刻画回归的效果,进而得出回归分析的一般步骤,并对具体问题进行分析,用于解决实际问题。

3. 情感、态度与价值观:从实际问题中发现自己已有知识的不足之处,激发学生的好奇心和求知欲,培养学生不满足于已有知识、勇于求知的良好个性品质,引导学生积极进取。任何事物都是相对的,但又有一定的规律性,我们只要从实际出发,不断探求事物的内在联系,就会找出其中的规律性,形成解决实际问题的方法和能力。

► 三层完全解读

>>> 解题依据

1 知识·能力聚焦

1. 求回归直线方程

求回归直线方程的一般方法是:

(1) 作出散点图,将问题所给的数据在平面直角坐标系中描点,这样表示出的具有相关关系的两个变量的一组数据的图形就是散点图。从散点图中我们可以看出样本点是否呈条状分布,从而判断两个量是否具有线性相关关系。

(2) 求回归系数 \hat{a}, \hat{b} , 其中

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

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

将所给的数据 x, y 列成相应的表格,如下表所示:

序号	x	y	x^2	y^2	xy
1	x_1	y_1	x_1^2	y_1^2	$x_1 y_1$
2	x_2	y_2	x_2^2	y_2^2	$x_2 y_2$
...
n	x_n	y_n	x_n^2	y_n^2	$x_n y_n$
Σ	Σx_i	Σy_i	Σx_i^2	Σy_i^2	$\Sigma x_i y_i$

由此可知 $\bar{x} = \frac{\Sigma x_i}{n}, \bar{y} = \frac{\Sigma y_i}{n}$, 其中 (\bar{x}, \bar{y}) 称为样本点的中心。

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

(3) 写出回归直线方程 $\hat{y} = \hat{b}x + \hat{a}$, 并用回归直线方程进行预测说明:当 x 取 x_0 时,由线性回归方程可得 \hat{y}_0 的值,从而可进行相应的判断。

>>> 名题诠释

例题1 容易题——2011·江西高考题

变量 X 与 Y 相对应的一组数据为 $(10, 1), (11.3, 2), (11.8, 3), (12.5, 4), (13, 5)$; 变量 U 与 V 相对应的一组数据为 $(10, 5), (11.3, 4), (11.8, 3), (12.5, 2), (13, 1)$. r_1 表示变量 Y 与 X 之间的线性相关系数, r_2 表示变量 V 与 U 之间的线性相关系数, 则()。

A. $r_2 < r_1 < 0$ B. $0 < r_2 < r_1$ C. $r_2 < 0 < r_1$ D. $r_2 = r_1$

【解析】 $\because r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \therefore$ 通过计算得第一组变量正相关, 第二组变量负相关。故选 C。

【答案】C

例题2 容易题

某工厂1~8月份某种产品的产量与成本的统计数据见下表:

月份	1	2	3	4	5	6	7	8
产量(吨)	5.6	6.0	6.1	6.4	7.0	7.5	8.0	8.2
成本(万元)	130	136	143	149	157	172	183	188

以产量为 x , 成本为 y 。

(1) 画出散点图;

(2) y 与 x 是否具有线性相关关系? 若有, 求出其回归方程。

【解析】本题主要考查对两个变量的回归分析, 画出散点图, 代入回归系数公式即可得。

【解】(1) 由表画出散点图, 如图 1-1-7 所示。

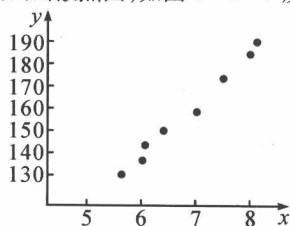


图 1-1-7



2. 随机误差

人教A版

对于某一个 x_i , 由回归直线方程: $\hat{y} = \hat{b}x + \hat{a}$ 可以确定一个 \hat{y}_i 的值, 但是由于测量本身存在误差, 或者测量受其他因素的影响, 或者回归直线方程本身存在误差, 或者方程受某些随机因素的影响, 使得 \hat{y}_i 与测得的实际数据 y_i 之间存在误差, 并不相等, 出现 $y_i = \hat{y}_i + \Sigma_i$, 其中 Σ_i 为随机误差.

从散点图中我们可以看到, 样本点分布在某一条直线附近, 而不是一条直线上, 所以不能用一次函数 $y = bx + a$ 来描述它们之间的关系, 我们用下面的线性回归模型来表示: $y = bx + a + e$, 其中 a, b 为模型的未知参数, e 称为随机误差. 我们选用的线性模型往往只是一种近似的模型, 除解释变量外, 还有其他因素也会导致随机误差 e 的产生.

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 的因素不只是变量 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 (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) 的原因, 导致 y 的观测值产生误差, 这样的误差也包含在 e 中.

3. 回归分析

人教A版

(1) 线性相关关系强弱的描述, 样本相关系数的计算公式:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \text{ 当 } r > 0 \text{ 时, 表明两个变量正相}$$

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.

(2) $\sum_{i=1}^n (y_i - \bar{y})^2$ 表示总的效应, 称为总偏差平方和; 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; $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 称为残差平方和, $\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 为两者之差, 它代表了随机误差的效应, 解释变量效应的这个值称为回归平方和. 我们用相关指数 R^2 来刻画回归的效果, 其计算公式是 $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$, 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 的值来作出选择, 即选择 R^2 大的模型作为这组数据的模型.

(2) 从图上可看出, 这些点基本上散布在一条直线附近, 可以认为 x 和 y 线性相关关系显著, 下面求其回归方程, 首先列出下表.

序号	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	5.6	130	31.36	16 900	728.0
2	6.0	136	36.00	18 496	816.0
3	6.1	143	37.21	20 449	872.3
4	6.4	149	40.96	22 201	953.6
5	7.0	157	49.00	24 649	1 099.0
6	7.5	172	56.25	29 584	1 290.0
7	8.0	183	64.00	33 489	1 464.0
8	8.2	188	67.24	35 344	1 541.6
Σ	54.8	1 258	382.02	201 112	8 764.5

$$\bar{x} = 6.85, \bar{y} = 157.25.$$

$$\therefore \hat{b} = \frac{\sum_{i=1}^8 x_i y_i - 8 \bar{x} \bar{y}}{\sum_{i=1}^8 x_i^2 - 8 \bar{x}^2} = \frac{8 764.5 - 8 \times 6.85 \times 157.25}{382.02 - 8 \times 6.85^2} \approx 22.17,$$

$$\hat{a} = \bar{y} - \hat{b} \bar{x} = 157.25 - 22.17 \times 6.85 \approx 5.39,$$

故线性回归方程为 $\hat{y} = 22.17x + 5.39$.

【点拨】求回归直线方程时, 首先应注意到, 只有在散点图大致呈线性分布时, 求出的回归直线方程才有实际意义, 否则, 求出的回归直线方程毫无意义.

例题 3 容易题——2011·山东青岛模拟

为了考察两个变量 y 与 x 的线性相关性, 测得 x, y 的 13 对数据, 若 y 与 x 具有线性相关关系, 则相关系数 $|r|$ 的取值范围是

【解析】查表得显著性水平 0.05, 自由度 $13 - 2 = 11$ 相应的相关系数临界值 $r_{0.05} = 0.553$. 所以 y 与 x 若具有线性相关关系, 则相关系数 $|r|$ 的取值范围是 $(0.553, 1]$.

【答案】 $(0.553, 1]$

例题 4 容易题

现随机抽取了我校 10 名学生在入学考试中的数学成绩 (x) 与入学后的第一次考试数学成绩 (y), 数据如下表:

学生号	1	2	3	4	5	6	7	8	9	10
x_i	120	108	117	104	103	110	104	105	99	108
y_i	84	64	84	68	69	68	69	46	57	71

请问: 这 10 名学生的两次数学考试成绩是否具有显著的线性相关关系?

【解析】若 x 与 y 呈线性相关关系, 就无需进行相关性检验, 否则需进行相关性检验, 可由公式 $r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}}$ 判断.

$$\text{【解】} \bar{x} = \frac{1}{10} (120 + 108 + \dots + 99 + 108) = 107.8,$$

$$\bar{y} = \frac{1}{10} (84 + 64 + \dots + 57 + 71) = 68,$$

$$\sum_{i=1}^{10} x_i^2 = 120^2 + 108^2 + \dots + 99^2 + 108^2 = 116 584,$$



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 $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n$ 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.

4. 相关性检验

人教B版·苏教版

(1) 相关性检验的必要性.

在实际问题中,利用散点图来判断线性相关关系,既快又方便,但这时散点图中的各个点应该集中在某一直线的附近,依据求回归直线方程的方法,求出与之相对应的回归直线方程,这条回归直线方程具有实际的应用价值.但是,若散点图中画出的各个点,不是集中在一条直线的附近,而是集中在某一个圆内,依据求回归直线的方法,我们仍然可以求出相应的回归直线方程,但这条回归直线方程已经不能反映出这组数据的变化规律,这时求回归直线方程也就失去了意义.那么,到底在数据有什么特征时,才是线性相关的,才能求出回归直线方程呢?这就需要与 x 与 y 作线性相关性检验,也就是相关性检验.

对于随机抽取到的变量 x 与 y 的 n 对数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$,检验统计量是样本相关系数 $r =$

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

r 具有以下性质: $|r| \leq 1$ 并且 $|r|$ 越接近于1,线性相关程度越强; $|r|$ 越接近于0,线性相关程度越弱.

(2) 相关性检验的步骤.

- ①作统计假设.假设 x 与 y 不具有线性相关关系;
- ②根据小概率0.05与 $n-2$ 在附表中查出 r 的一个临界值 $r_{0.05}$;

- ③根据样本相关系数的计算公式 $r =$

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}} \text{计算出 } r \text{ 的值;}$$

- ④作统计推断:如果 $|r| > r_{0.05}$,表明有95%的把握认为 x 与 y 之间具有线性相关关系;如果 $|r| \leq r_{0.05}$,这时寻找回归直线方程是毫无意义的.

2 方法·技巧平台

5. 利用图形分析残差的特性

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.在残差图中,如果有两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,就先予以纠正,然后重新利用线性回归模型拟合数据;如果数据采集没有错误,则需要寻找其他原因.另外,若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就越高.

$$\begin{aligned} \sum y_i^2 &= 84^2 + 64^2 + \dots + 57^2 + 71^2 = 47\,384, \\ \sum x_i y_i &= 120 \times 84 + 108 \times 64 + \dots + 108 \times 71 = 73\,796, \end{aligned}$$

所以相关系数

$$r = \frac{73\,796 - 10 \times 107.8 \times 68}{\sqrt{(116\,584 - 10 \times 107.8^2)(47\,384 - 10 \times 68^2)}} \approx 0.7506,$$

查表:显著性水平0.05,自由度 $10-2=8$ 相应的相关关系临界值 $r_{0.05}=0.632$,由 $r > r_{0.05}$ 知两次数学考试成绩有显著的线性相关关系.

【点拨】如果两个变量不具备相关关系,或者相关关系不显著,即使求回归直线方程也是无意义的,用其估计和预测是不可信的.

例题5 容易题

某城区为研究城镇居民家庭月人均生活费支出和月人均收入的相关关系,随机抽取10户进行调查,其结果如下:

月人均收入 x /元	月人均生活费 y /元
300	255
390	324
420	335
520	360
570	450
700	520
760	580
800	600
850	630
1 080	750

试预测月人均收入为1 100元和月人均收入为1 200元的两个家庭的月人均生活费,并进行残差分析.

【解析】先通过散点图确定两个变量是否线性相关,再用残差图和相关指数 R^2 来分析模型拟合的效果,进而估计预测值的准确性.

【解】作出散点分布图(如图1-1-8),由图可知,月人均生活费与月人均收入之间具有线性相关关系.

通过计算可知 $\bar{x} = 639, \bar{y} = 480.4, \sum_{i=1}^{10} x_i^2 = 4\,610\,300, \sum_{i=1}^{10} y_i^2 = 2\,540\,526, \sum_{i=1}^{10} x_i y_i = 3\,417\,560, \therefore \hat{b} = \frac{\sum_{i=1}^{10} x_i y_i - 10 \bar{x} \bar{y}}{\sum_{i=1}^{10} x_i^2 - 10 \bar{x}^2} \approx 0.6599, \hat{a} = \bar{y} - \hat{b} \bar{x} \approx 58.724.$

\therefore 线性回归方程为 $\hat{y} = 0.6599x + 58.724$.

作残差图如图1-1-9所示,由图可知,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.

计算相关指数得 $R^2 \approx 0.9863$,说明城镇居民的月人均生活费的差异有98.63%

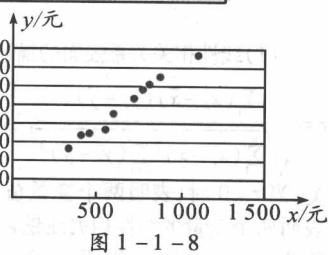


图1-1-8

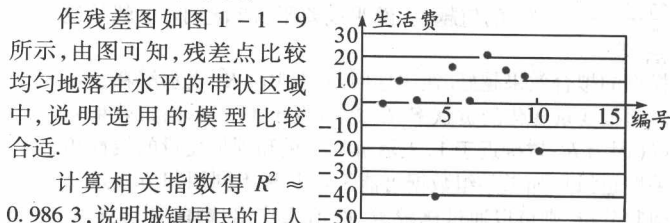


图1-1-9



几种常见的残差图:

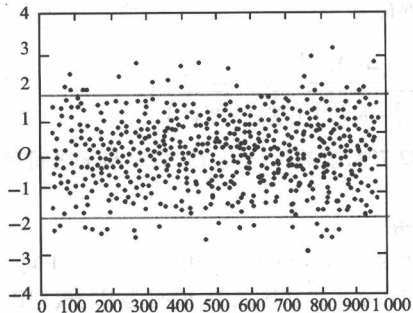


图 1-1-1

图 1-1-1: 残差图中的点分布在以原点为中心的水平和带状区域上, 并且沿水平方向散点的分布规律相同, 说明残差是随机的, 所选择的回归模型是合理的。

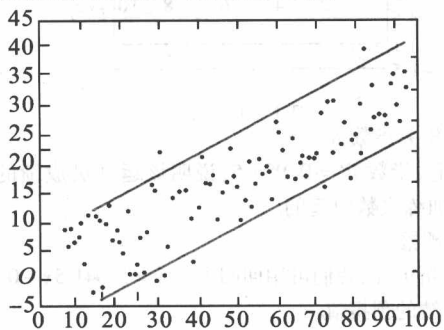


图 1-1-2

图 1-1-2: 残差图中的点分布在一条倾斜的带状区域上, 并且沿带状区域方向散点的分布规律相同, 说明残差与横坐标变量间有线性关系, 此时所选择的回归模型的效果不是最好的, 有改进的余地。

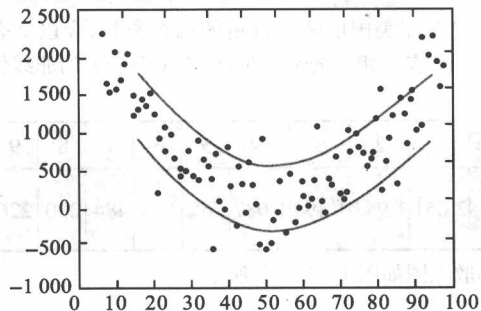


图 1-1-3

图 1-1-3: 残差图中的点分布在一条二次曲线的弯曲带状区域上, 说明残差与横坐标变量间有二次关系, 此时所选用的回归模型的效果不是最好的, 有改进的余地。

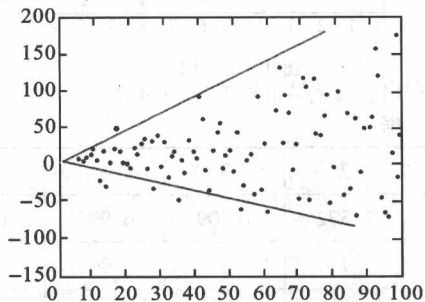


图 1-1-4

是由月人均收入引起的。

由以上分析可知, 我们可以利用回归方程 $\hat{y} = 0.6599x + 58.724$ 来作为月生活费的预报值。

将 $x = 1100$ 代入回归方程得 $y = 784.61$ (元);

将 $x = 1200$ 代入回归方程得 $y = 850.60$ (元)。

故预测月人均收入分别为 1100 元和 1200 元的两个家庭的月人均生活费分别为 784.61 元和 850.60 元。

例题 6 中难题

某运动员训练次数与运动成绩的数据如下表:

次数(x)	30	33	35	37	39	44	46	50
成绩(y)	30	34	37	39	42	46	48	51

- (1) 作出散点图;
- (2) 求出回归方程;
- (3) 计算相关系数, 并利用其检验两变量间线性相关关系的显著性;
- (4) 作出残差图;
- (5) 计算相关指数 R^2 ;
- (6) 试预测该运动员训练 47 次及 55 次的成绩。

【解】(1) 作出该运动员训练次数(x)与成绩(y)的散点图, 如图 1-1-10 所示, 由散点图可知, 它们之间具有线性相关关系。

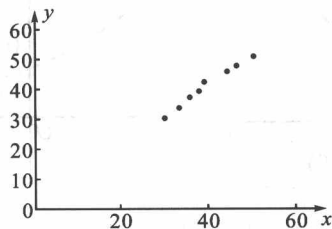


图 1-1-10

(2) 列表计算:

次数 x_i	成绩 y_i	x_i^2	y_i^2	$x_i y_i$
30	30	900	900	900
33	34	1 089	1 156	1 122
35	37	1 225	1 369	1 295
37	39	1 369	1 521	1 443
39	42	1 521	1 764	1 638
44	46	1 936	2 116	2 024
46	48	2 116	2 304	2 208
50	51	2 500	2 601	2 550

由上表可求得 $\bar{x} = 39.25$, $\bar{y} = 40.875$, $\sum_{i=1}^8 x_i^2 = 12 656$,

$\sum_{i=1}^8 y_i^2 = 13 731$, $\sum_{i=1}^8 x_i y_i = 13 180$,

$$\therefore \hat{b} = \frac{\sum_{i=1}^8 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^8 (x_i - \bar{x})^2} = \frac{\sum_{i=1}^8 x_i y_i - 8 \bar{x} \bar{y}}{\sum_{i=1}^8 x_i^2 - 8 \bar{x}^2} \approx 1.041 5,$$

$$\hat{a} = \bar{y} - \hat{b} \bar{x} = -0.003 875,$$

\therefore 回归方程为 $\hat{y} = 1.041 5x - 0.003 875$ 。