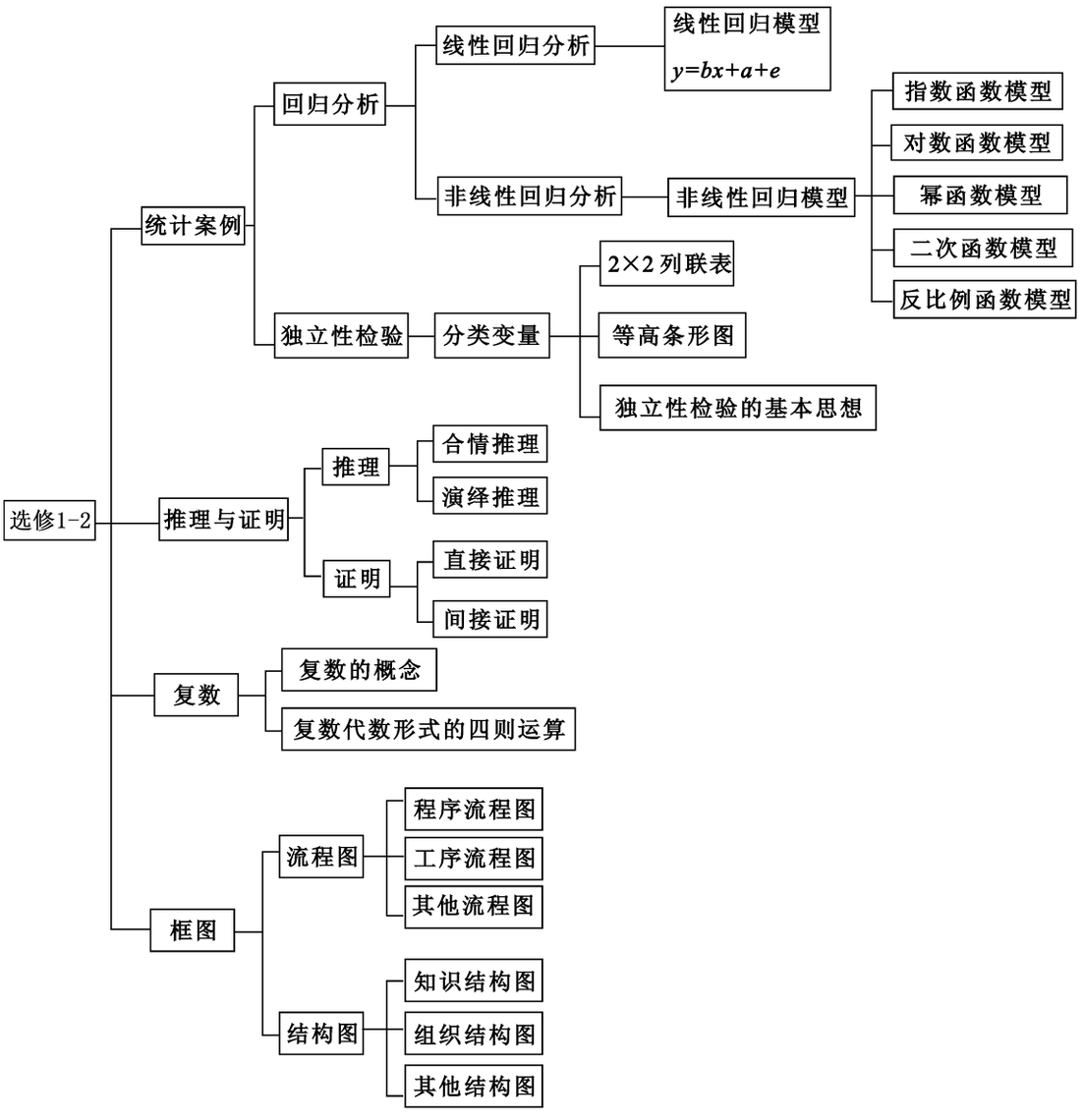
模块整体感悟

模块知识梳理



学法点拨

统计案例是在《数学3》学习统计的基础上，通过对典型案例的讨论，了解和使用一些常用的统计方法，进一步体会运用统计方法解决实际问题的基本思想，认识统计方法在决策中的作用.本章的学习重点是了解两种统计方法的基本思想及其初步应用，对其理论基础不作要求，应避免单纯记忆和机械套用公式；鼓励同学们使用计算器、计算机等现代技术手段来处理数据，建议大家学会使用计算器中的“回归计算”功能进行有关计算.

推理与证明是数学的基本思维过程，也是人们学习和生活中经常使用的思维方式.这些内容其实在以前的初中数学中也不知不觉地应用到，只是没有作为专门的章节单独来进行学习.而这种推理与证明的思想实际上贯穿了整个数学思想体系，它是学习新课标教材的一个亮点，是对以前所学数学思维方法的总结、归纳，并对后续学习起到引领的作用.通过本章的学习，同学们要更加明确这些方法，并能在今后的学习中有意识地使用，养成言之有理、论证有据的习惯.

数系扩充的过程体现了数学的发现和创造过程，也体现了数学发生、发展的客观需求.学习时应注意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（数的运算规则、方程求根）在数系扩充过程中的作用，感受人类理性思维的作用以及数与现实世界的联系.由复数在整个高中数学所处的地位看，复数的考查从分值上、难度上在逐渐下降，这也是目前教学内容改革的趋势.学习时要注意复数问题实数化处理的化归思想、方程思想和数形结合的思想方法.

框图的学习基础是《数学3》中的算法流程图.教材一开始就要求画“二分法求方程近似根”的程序框图，对多数同学来说可能是比较困难的，因此建议在学习本章知识时重点从以下4个层次去把握：

（1）从分析实例入手，先认识流程图与结构图的一般形式、特征和作用；

（2）学会读流程图和结构图；

（3）学会画流程图和结构图；

（4）用流程图写一些数学问题的解题思路与步骤，把学过的某些章节的知识结构图写出来，这对理清知识联系、形成知识体系、完善认知结构是很有帮助的.

第一章　统计案例

本章学习导航

先行一步，步步先行

内容概要

本章我们将在《数学 3》学习了概率统计内容的基础上，通过典型统计案例的学习，进一步体会运用统计方法解决实际问题的基本思想，认识统计方法在决策中的作用. 本章主要包括两部分内容：

1. 回归分析的基本思想及其初步应用，其中回归分析包括：线性回归分析与非线性回归分析两大部分内容，线性回归分析是基础和重点.

我们需要掌握以下知识点：

(1)作散点图，初步判断两个变量是否具有线性相关关系；

(2)求相关系数 r，进行相关性检验；

(3)用最小二乘法求回归直线方程，通过解释变量估计预报变量；

(4)了解随机误差产生的原因，以及它对预报变量产生的影响；

(5)求样本点的残差、残差平方和、总偏差平方和、相关指数 R^2，作残差图进行残差分析，发现原始数据中的可疑数据，衡量模型的拟合效果，建立适当的线性回归模型；

(6)利用散点图判断并建立非线性回归模型，通过适当的线性变换转化成线性回归模型，利用线性回归模型来建立 y 关于 x 的非线性回归方程.

2. 独立性检验的基本思想及其初步应用. 我们需要掌握以下知识点：

(1)认识什么是分类变量，能根据题意列出 2×2 列联表；

(2)用等高条形图直观展示列联表数据的频数特征，粗略判断两个变量之间是否有关系；

(3)利用列联表数据，通过比较 ad 与 bc，$\frac{a}{a+b}$ 与 $\frac{c}{c+d}$ 的大小，粗略判断两个变量之间是否有关系；

(4)利用构造出的随机变量 $K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 进行独立性检验，精确刻画两个变量有关系的可靠程度；

(5)利用独立性检验的基本思想和方法，解决和指导生产、生活、学习中判断两个变量是否有关系的问题.

《课程标准》要求

内容标准	学习要求
回归分析的基本思想及其初步应用，包括：线性回归分析与非线性回归分析两部分.	1. 通过对典型案例（如“人的体重与身高的关系”）的探究，进一步了解线性回归分析的基本思想、方法及其初步应用. 2. 通过对典型案例（如“红铃虫的产卵数和温度的关系”）的探究，初步了解非线性回归分析向线性回归分析转换的基本思想、方法及其简单应用.
独立性检验的基本思想及其初步应用.	通过对典型案例（如“吸烟是否对患肺癌有影响”）的探究，了解独立性检验（只要求2×2列联表）的基本思想、方法及其初步应用.

学法指导

1. 本章我们主要通过探究典型统计案例的方式，亲身经历、体验数据处理的过程，学习线性回归分析、非线性回归分析、独立性检验的基本思想和方法，初步体会统计方法的应用特点. 因此，学习中应尽量涉及更多的案例讨论，最好是选择一个案例，单独完成数据处理过程，亲历对这些统计方法的体验过程.

2. 本章内容由于涉及对大量数据的处理，学习中可在老师指导下，学会借助计算器、计算机等现代技术手段（运用一些常见的统计软件如Excel、几何画板）来处理数据.

3. 对本章典型的统计案例，学习时应侧重理解和掌握这几种统计方法的基本思想和初步应用，可以适当避开其繁琐的理论证明.

4. 对本章出现的计算公式，要明确各公式的结构特点、意义以及它们之间的内在联系，避免单纯记忆和机械套用公式进行计算. 由于公式相对复杂，计算量较大，计算时一定要做到代入数字细心谨慎，保证计算结果准确无误.

1.1　回归分析的基本思想及其初步应用

第一学时　回归直线方程

板块一　创设问题，引领目标

导入问题，直指学时重点

问题呈现

2007年5月，中共中央、国务院发布《关于加强青少年体育增强青少年体质的意见》(简称《意见》).该《意见》指出，由于片面追求升学率的影响，社会和学校存在重智育、轻体育的倾向，学生课业负担过重，休息和锻炼时间严重不足.通过体质健康监测表明，青少年耐力、力量、速度等体能指标持续下降，视力不良率居高不下，城市超重和肥胖青少年的比例明显增加，部分农村青少年营养状况亟待改善.

实际上，我们可以通过调查得到一些有关“身高与体重”的数据，从数学的角度来看，我们怎么来分析这些数据呢？健康的“身高与体重”指标应该是什么样的呢？什么是相关关系？什么是回归分析？什么是线性回归分析？什么是独立性检验？从本学时开始，结合《数学3》的概率统计的知识，我们进一步学习、了解运用统计方法解决实际问题的基本思想方法.

材料链接

1.《数学3》p84～p93的2.3“变量间的相关关系”及“阅读与思考”中关于“相关关系的强弱”.

2.用WPS表格(或Excel)作散点图的步骤示例：

(1)列表输入数据；

(2)选择上述表格数据，然后点击：插入—图表—XY散点图，选择需要的散点图类型(一般可选择第一个类型)；

(3)点击下一步，自动选择图表源数据；

(4)点击下一步，按需要修改图表名称以及XY轴标记；

(5)点击完成按钮就得到我们通常看到的散点图.

3.百度搜索“认识最小二乘法”.

板块二　自学思疑，初探问题

尊重认知规律，亲历感悟知识生成

教材导读

阅读教材p1～p3的有关内容，思考并回答下列问题：

问题一

什么是回归分析？什么是线性回归分析？

1.两个变量的相关关系与函数关系有什么不同吗？如下列一些关系：

(1)人的年龄与他(她)拥有的财富之间的关系；

(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.

你能判别其中哪些是相关关系,哪些是函数关系吗?

❀2. 如何作散点图(见“材料链接”)? 两个变量线性相关时,其散点图呈现什么样的形态分布?

❀3. 什么是回归分析? 什么是线性回归分析? 回归分析与线性回归分析之间有什么区别和联系?

问题二

如何求回归直线方程?

❀对于线性相关的两个变量,用什么方法来刻画它们之间的关系呢?

自主测评

1. 下列两个变量具有相关关系的是 (　　)

A. 正方体的体积与棱长

B. 匀速行驶的车辆的行驶距离与行驶时间

C. 一块稻田的产量与施肥量

D. 人的身高与视力

2. 函数关系中两个变量之间是一种确定性关系. 而当一个变量取值一定时,另一个变量的取值带有一定_____性的两个变量之间的关系叫做相关关系. 如果散点图中,点的分布从整体上看大致在一条直线附近,我们就称这两个变量之间具有_____关系.

3. 样本相关系数$|r|\leqslant 1$. 当$r>0$时,表明两个变量_____相关;当$r<0$时,表明两个变量_____相关;r的绝对值越接近_____时,表明两个变量的线性相关性越强;r的绝对值越接近_____时,表明两个变量之间几乎不存在线性相关关系. 通常,当$|r|\geqslant$_____时,我们就认为两个变量有很强的线性相关关系.

4. 已知一组观测值(x_i,y_i),$i=1,2,\cdots,n$,作出散点图后确定具有线性相关关系. 若对于$\hat{y}=\hat{b}x+\hat{a}$,求得$\hat{b}=0.8$,$\bar{x}=12.5$,$\bar{y}=17.6$,则回归直线方程为_____.

5. 调查了某地若干户家庭的年收入x(单位:万元)和年饮食支出y(单位:万元). 调查显示年收入x与年饮食支出y具有线性相关关系,并由调查数据得到y对x的回归直线方程:$\hat{y}=0.254x+0.321$. 由回归直线方程可知,家庭年收入每增加1万元,年饮食支出平均增加_____万元.

疑点归纳

通过上面的学习,你了解线性回归分析的基本过程了吗? 会求回归直线方程了吗? 此外,你还有什么疑惑,赶紧记录下来,和你的同学、老师交流解答吧.

板块三　合作互助,共析问题

发展创新思维,形成主动探究与合作的意识和能力

问题三

求回归直线方程的主要步骤是什么?

❀1. 如何直观地判断变量x,y之间的线

性相关关系？

❀2. 求回归直线方程的主要步骤是什么？

板块四　展示交流，探究问题

提高分析问题、解决问题的能力

问题四

你了解回归分析的基本过程吗？

❀1. 多个模型，怎样知道哪个模型效果更好呢？

❀2. “回归直线过样本点的中心$(\overline{x},\overline{y})$”，这个结论对吗？

❀3. 函数模型与线性回归模型之间有什么异同呢？

展题设计

❀**展题 1**　研究某灌溉渠道水的流速 y 与水深 x 之间的关系，测得一组数据如下：

水深 x/m	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10
流速 y/(m/s)	1.70	1.79	1.88	1.95	2.03	2.10	2.16	2.21

(1)作出散点图，你能从中发现水深 x 与水的流速 y 之间的一般规律吗？

(2)求 y 对 x 的回归直线方程；

(3)预测水深为 1.95 m 时，水的流速是多少？

指导要求：要了解线性回归分析的基本过程，应注意以下几点：

(1)一定要做相关关系检验吗？通常如何进行相关关系检验？

(2)会用最小二乘法公式求回归系数 $\hat{a}$ 与 $\hat{b}$ 吗？

(3)会根据回归直线方程进行预测吗？

❀**展题 2**　某地最近十年粮食需求量逐年上升，下表是部分统计数据：

年份	2010	2012	2014	2016	2018
需求量/万吨	236	246	257	276	286

(1)利用所给数据求年需求量与年份之间的回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$；

(2)利用(1)中所求出的直线方程估计该地 2020 年的粮食需求量.

指导要求：同“展题 1”.

变式 一般来说，一个人脚掌越长，他的身高就越高.现对10名成年人的脚掌长 x 与身高 y 进行测量，所得的数据(单位：cm)作为样本如下表所示.

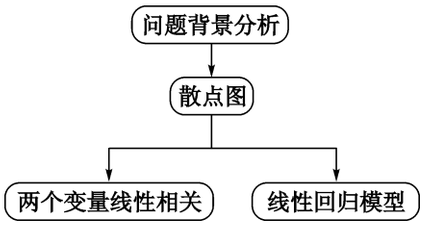
脚掌长 x	20	21	22	23	24	25	26	27	28	29
身高 y	141	146	154	160	169	176	181	188	197	203

(1)在上表数据中，以“脚掌长”为横坐标，“身高”为纵坐标，作出散点图后，发现散点在一条直线附近，试求“身高”与“脚掌长”之间的线性回归方程 $\hat{y}=\hat{b}x+\hat{a}$；

(2)若某人的脚掌长为26.5 cm，试估计此人的身高.

参考数据：$\sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})=577.5$，$\sum_{i=1}^{10}(x_i-\overline{x})^2=82.5$，$\overline{x}=24.5$，$\overline{y}=171.5$.

归纳总结



板块五 应用演练，再生新疑

分级设题，精选精练

分层演练

A. 基础巩固

1. 下列各变量之间存在相关关系的是 (　　)

①人的年龄与身高

②角的弧度数与它的正弦值

③日照时间与水稻的亩产量

④球的表面积与半径

A. ①②　　B. ③④　　C. ①③　　D. ②④

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的任意一个变量在 x 轴上

D. 可以选择两个变量中的任意一个变量在 y 轴上

3. 已知变量 x 和 y 满足关系 $y=-0.1x+1$，变量 y 与 z 正相关，则下列结论中正确的是 (　　)

A. x 与 y 负相关，x 与 z 负相关

B. x 与 y 正相关，x 与 z 正相关

C. x 与 y 正相关，x 与 z 负相关

D. x 与 y 负相关，x 与 z 正相关

4. r 是相关系数，给出下列命题：

①当 $r\in[-1,-0.75]$ 时，两个变量负相关很强；

②当 $r\in[0.75,1]$ 时，两个变量正相关很强；

③当 $r\in(-0.75,-0.3]$ 或 $r\in[0.3,0.75)$ 时，两个变量相关性一般；

④当 $r=0.1$ 时，两个变量相关性很弱.

其中正确的是_____.

B. 能力提升

5. 人的年龄 x 与人体脂肪含量 y 的回归直线方程为 $\hat{y}=0.577x-0.448$. 如果某人 36 岁，那么这个人的脂肪含量　　(　　)

A. 一定是 20.3%　　B. 大于 20.3%

C. 小于 20.3%　　D. 大约为 20.3%

6. 设某大学女生的体重 y(单位：kg)与身高 x(单位：cm)具有线性相关关系. 根据一组样本数据 $(x_i,y_i)(i=1,2,\cdots,n)$，用最小二乘法建立的回归方程为 $\hat{y}=0.85x-85.71$，则下列结论中不正确的是　　(　　)

A. y 与 x 具有正的线性相关关系

B. 回归直线过样本点的中心 $(\overline{x},\overline{y})$

C. 若该大学某女生身高增加 1 cm，则其体重约增加 0.85 kg

D. 若该大学某女生身高为 170 cm，则可断定其体重必为 58.79 kg

7. 根据如下样本数据得到的回归方程为 $\hat{y}=bx+a$，则　　(　　)

x	3	4	5	6	7	8
y	4.0	2.5	−0.5	0.5	−2.0	−3.0

A. $a>0,b<0$

B. $a>0,b>0$

C. $a<0,b<0$

D. $a<0,b>0$

8. 为了了解儿子身高与其父亲身高的关系，随机抽取 5 对父子的身高数据如下：

父亲身高 x/cm	174	176	176	176	178
儿子身高 y/cm	175	175	176	177	177

则 y 对 x 的线性回归方程为　　(　　)

A. $\hat{y}=x-1$　　B. $\hat{y}=x+1$

C. $\hat{y}=88+\frac{1}{2}x$　　D. $\hat{y}=176$

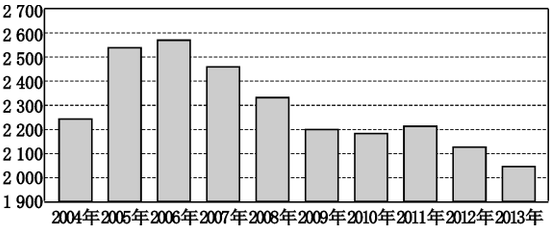
C. 拓展创新

9. 在一组样本数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)(n\geqslant 2,x_1,x_2,\cdots,x_n$ 不全相等)的散点图中，若所有样本点 $(x_i,y_i)(i=1,2,\cdots,n)$ 都在直线 $y=\frac{1}{2}x+1$ 上，则这组样本数据的样本相关系数为　　(　　)

A. −1　　B. 0

C. $\frac{1}{2}$　　D. 1

10. 根据下面给出的 2004 年至 2013 年我国二氧化碳年排放量(单位：万吨)的柱形图判断，以下结论中不正确的是　　(　　)



A. 逐年比较，2008 年减少二氧化碳排放量的效果最显著

B. 2007 年我国治理二氧化碳排放显现成效

C. 2006 年以来我国二氧化碳年排放量呈逐渐减少趋势

D. 2006 年以来我国二氧化碳年排放量与年份正相关

11. 某医院用光电比色计检验尿汞时，得到尿汞含量 x(mg/L)与消化系数 y 的数据如下表：

尿汞含量 x	2	4	6	8	10
消化系数 y	64	138	205	285	260

(1)对变量 y 与 x 作相关性检验；

(2)如果 y 与 x 之间具有线性相关关系，求回归直线方程.

12. 从某居民区随机抽取10个家庭，获得第i个家庭的月收入x_i（单位：千元）与月储蓄y_i（单位：千元）的数据资料，算得$\sum_{i=1}^{10}x_i=80$，$\sum_{i=1}^{10}y_i=20$，$\sum_{i=1}^{10}x_iy_i=184$，$\sum_{i=1}^{10}x_i^2=720$.

（1）求家庭的月储蓄y对月收入x的线性回归方程$y=bx+a$；

（2）判断变量x与y之间是正相关还是负相关；

（3）若该居民区某家庭月收入为7千元，预测该家庭的月储蓄.

附：线性回归方程$y=bx+a$中，$a=\overline{y}-b\overline{x}$，$b=\dfrac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}$，

其中$\overline{x}$，$\overline{y}$为样本平均值，线性回归方程也可写为$\hat{y}=\hat{b}x+\hat{a}$.

再生新疑

通过本学时的学习，我们了解了线性回归分析的基本过程，学会了怎样求回归直线方程.

此外，学完本学时，请思考如下问题：

1. 有同学说，从本学时的线性回归过程来看，在我们分析的任何两个变量之间都可以写出其回归直线方程，你怎么理解这个“结论”？

2. 你意识到线性回归分析的“误差”了吗？

第二学时　随机误差与线性回归模型

板块一　创设问题，引领目标

导入问题，直指学时重点

问题呈现

上一学时，我们通过“女大学生的身高与体重”的实例再现了求回归直线方程的过程，并利用“变量间的相关关系、散点图、最小二乘法”这些已有的知识完成了回归预报. 那么，你知道这样的预报准确吗？要建立变量间的线性回归模型还应该注意什么问题呢？从这一学时开始，我们继续通过实例来进一步学习随机误差与线性回归模型的有关知识.

材料链接

随机误差（又称偶然误差）是指测量结果与同一待测量的大量重复测量的平均结果之差.

“同一待测量的大量重复测量的平均结果”指在重复条件下得到待测量的期望值或所有可能测得值的平均值.

它的特点：大小和方向都不固定，也无法测量或校正. 随机误差的性质是：随着测定次数的增加，正负误差可以相互抵消，误差的平

均值将逐渐趋向于零.

随机误差的特征：

①只要测试系统的灵敏度足够高，在相同的测量条件下，对同一量值进行多次等精度测量时，仍会有各种偶然的、无法预测的不确定因素干扰而产生测量误差，其绝对值和符号均不可预知.

②虽然单次测量的随机误差没有规律，但多次测量的总体却服从统计规律，通过对测量数据的统计处理，能在理论上估计其对测量结果的影响.

③随机误差不能用修正或采取某种技术措施的办法来消除.

产生随机误差的原因：产生随机误差的原因很多、很复杂，观测时目的物对得不准，读数不准确，周围环境的偶然变化，如电磁场的微变，零件的摩擦和配合间隙，热起伏，空气扰动，气压及湿度的变化，测量人员的感觉器官的生理变化等，以及它们的综合影响都可以成为产生随机误差的因素.

板块二　自学思疑，初探问题

尊重认知规律，亲历感悟知识生成

教材导读

阅读教材 p3～p6 的有关内容，思考并回答下列问题：

问题一

什么是随机误差？

结合教材例 1 谈谈什么是随机误差？产生随机误差的原因是什么？

问题二

什么是残差？如何进行残差分析？

1. 什么是残差？什么是残差图？

2. 如何进行残差分析？

(1)在残差图中，如果某个样本点的残差比较大，我们应该做怎样的数据处理？如果样本点的残差比较小，残差点比较均匀地落在水平的带状区域中，那么你选用的模型合适吗？带状区域的宽窄表明了什么？

(2)在线性回归模型中，相关指数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}$ 刻画的是什么？表达式中的分子 $\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2$ 与分母 $\sum\limits_{i=1}^{n}(y_i-\overline{y})^2$ 的含义分别是什么？其中哪个是确定的数？哪个是不确定的数？R^2 是如何受其中不确定数值的影响而变化的？R^2 的大小表明了什么？

自主测评

1. 通过下列哪个数据可以发现原始数据中的可疑数据，从而判断所建立模型的拟合效果 (　　)

A. 回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 的系数 $\hat{b}$

B. 样本点的中心 $(\overline{x},\overline{y})$

C. 相关系数 r

D. 残差 $\hat{e}_i$

2. 回归分析中，相关指数 R^2 的值越大，说明残差平方和 (　　)

A. 越大　　B. 越小

C. 可能大也可能小　　D. 以上都不对

3. 在散点图中，如果样本点散布在某条直线附近，而不是在一条直线上，我们就用线性回归模型 $y=bx+a+e$ 来描述两个变量之间的关系，我们把 e 称为_____.

4. e 是一个不可观测的量. 在线性回归模型 $y=bx+a+e$ 中，把随机误差 e 写成 $e=y-(bx+a)$，我们用回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 中的 $\hat{y}$ 估计 $y=bx+a+e$ 中的 $bx+a$，那么 $\hat{e}=y-\hat{y}$ 就是 e 的_____. 相应于样本点 (x_i, y_i)，它的随机误差 $e_i=y_i-bx_i-a$，e_i 的估计值 $\hat{e}_i=y_i-\hat{y}_i=y_i-\hat{b}x_i-\hat{a}$ 叫做该样本点的_____.

5. 相关指数 $R^2=$_____，其中把_____叫做残差平方和，_____叫做总偏差平方和；R^2 表示解释变量对预报变量变化的_____. R^2 越接近_____，表示模型回归的效果越好.

疑点归纳

通过上面的学习，相信多数同学已经学会如何进行残差分析，你学会了吗？此外，你还有什么疑惑，赶紧记录下来，和你的同学、老师交流解答吧.

板块三　合作互助，共析问题

发展创新思维，形成主动探究与合作的意识和能力

问题三

在本学时的实例中，用身高预报体重时应注意什么问题呢？

1. 利用回归方程得到的预报值就是预报变量的精确值吗？

2. “女大学生”的身高与体重的回归方程能用来描述“女运动员”的身高与体重的关系吗？还有哪些因素会影响回归方程的适用范围呢？

3. 如何发现数据中的错误？如何衡量回归模型的拟合效果？

板块四　展示交流，探究问题

提高分析问题、解决问题的能力

问题四

建立回归模型需要注意什么？

1. 建立回归模型的基本步骤是什么？

2. 画散点图是必要的吗？

3. 在建立回归模型时，如何确定回归方程的类型？

4. 得出回归模型后如何分析它的拟合效果？

展题设计

展题1　一个车间为了规定工时定额，需要确定加工零件所花费的时间，为此进行了10次试验，测得的数据如下：

零件数 x/个	10	20	30	40	50	60	70	80	90	100

加工时间 y/min	62	68	75	81	89	95	102	108	115	122

把零件数 x 作为解释变量，加工时间 y 作为预报变量.

(1)y 与 x 是否具有线性相关关系？如果 y 与 x 具有线性相关关系，求回归直线方程；

(2)计算总偏差平方和、残差平方和及相关指数；

(3)作出残差图；

(4)进行残差分析.

指导要求：(1)建立回归模型的基本(主要)步骤是什么？

(2)在回归分析的过程中涉及的概念有：随机误差 e、残差 $\hat{e}_i$、相关指数 R^2 以及残差分析，你能通过本题把这些概念“串”起来吗？

展题 2　某运动员训练次数 x 与运动成绩 y 之间的数据关系如下：

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

画出散点图初步判断 y 与 x 具有线性相关关系，请参考如下数据完成下列多选判断(把你认为正确的答案都选上)：

$\sum\limits_{i=1}^{8}x_i^2=12\ 656$，$\sum\limits_{i=1}^{8}y_i^2=13\ 731$，$\sum\limits_{i=1}^{8}x_iy_i=13\ 180$，$\overline{x}=39.25$，$\overline{y}=40.875$，$r\approx0.992\ 704$，$R^2\approx0.985\ 5$.

(1)由 $r\approx0.992\ 704$ 可知 y 与 x 之间存在很强的线性相关关系；

(2)由 $r\approx0.992\ 704$ 可知 y 与 x 之间线性相关关系不明显；

(3)相关指数 $R^2\approx0.985\ 5$，说明了该运动员的成绩的差异有 98.55% 是由训练次数引起的；

(4)相关指数 $R^2\approx0.985\ 5$ 接近 1，说明线性回归模型拟合数据效果很好；

(5)回归直线方程为 $\hat{y}=1.041\ 5x-0.003\ 875$；

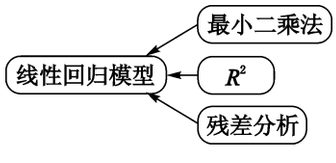
(6)回归直线方程为 $\hat{y}=-0.003\ 875x+1.041\ 5$；

(7)预测该运动员训练 47 次的成绩为 49；

(8)预测该运动员训练 57 次的成绩为 55.

指导要求：你能完成一个完整的回归分析过程吗？你会借助计算器(或电脑软件)进行计算或分析吗？

归纳总结



板块五　应用演练，再生新疑

分级设题，精选精练

分层演练

A. 基础巩固

1. 甲、乙、丙、丁四位同学各自对 A，B 两变量的线性相关性做试验，并用回归分析的方法分别求得相关系数 r 与残差平方和 m 如下表：

	甲	乙	丙	丁
r	0.86	0.78	0.69	0.82
m	x	115	124	106

则关于 x 的取值以及试验结果体现 A，B 两变量有更强的线性相关性的描述正确的是（　　）

A. $x>124$ 且丙最强

B. $x>124$ 且甲最强

C. $x<106$ 且丙最强

D. $x<106$ 且甲最强

2. 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 如下表：

模型	1	2	3	4
R^2	0.98	0.80	0.50	0.25

其中拟合效果最好的是（　　）

A. 模型 1　　B. 模型 2

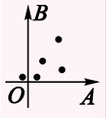
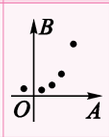
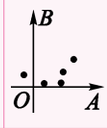
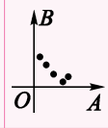
C. 模型 3　　D. 模型 4

3. 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. 如果 $R^2\approx0.81$，表示自变量 x 解释了因变量 y _____的变化. 同时，随机误差对预报变量的影响效应为_____.

4. 在残差图中，残差点比较均匀地落在_____区域中，说明选用的模型比较合适，这个区域的宽度越_____，说明模型拟合精度越高.

B. 能力提升

5. 甲、乙、丙、丁四位同学各自对 A，B 两个变量做回归分析，分别得到散点图与残差平方和 $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ 如下表：

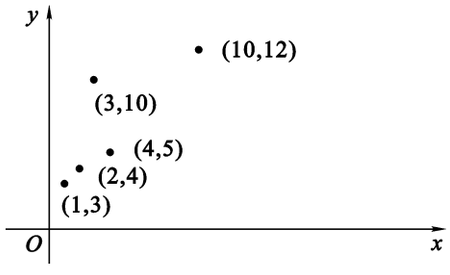
	甲	乙	丙	丁
散点图				
残差平方和	115	106	124	103

对 A，B 两个变量的关系，所建立模型的拟合精度高的同学是（　　）

A. 甲　　B. 乙

C. 丙　　D. 丁

6. 如图所示的 5 组数据中，去掉哪组数据后，剩下的 4 组数据的线性相关性最强（　　）



A. (1,3)　　B. (2,4)

C. (3,10)　　D. (10,12)

7. 某产品的广告费用 x（单位：万元）与销售额 y（单位：万元）的统计数据如下表：

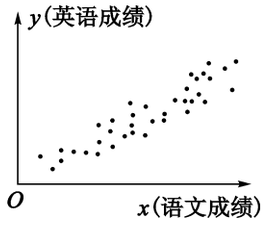
广告费用 x/万元	4	2	3	5
销售额 y/万元	49	26	39	54

根据上表，可得回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 中的 $\hat{b}$ 为 9.4. 据此模型预报广告费用为 6 万元时销售额为（　　）

A. 63.6 万元　　B. 65.5 万元

C. 67.7 万元　　D. 72.0 万元

8. 为了研究语文成绩和英语成绩之间是否具有线性相关关系，统计两科成绩得到如图所示的散点图（两坐标轴单位长度相同），用回归直线 $y=bx+a$ 近似地刻画其相关关系. 根据图形，以下结论最有可能成立的是（　　）



A. 线性相关关系较强，b 的值为 1.25

B. 线性相关关系较强，b 的值为 0.83

C. 线性相关关系较强，b 的值为 −0.87

D. 线性相关关系太弱，无研究价值

C. 拓展创新

9. 四名同学根据各自的样本数据研究变量 x，y 之间的相关关系，并求得回归直线方程，分别得到以下四个结论：

① y 与 x 负相关，且 $y=2.347x-6.423$；

② y 与 x 负相关，且 $y=-3.476x+5.648$；

③ y 与 x 正相关，且 $y=5.437x+8.493$；

④ y 与 x 正相关，且 $y=-4.326x-4.578$.

其中一定不正确的结论的序号是（　　）

A. ①②　　B. ②③

C. ③④　　D. ①④

10. 已知 x 与 y 之间的几组数据如下表：

x	1	2	3	4	5	6
y	0	2	1	3	3	4

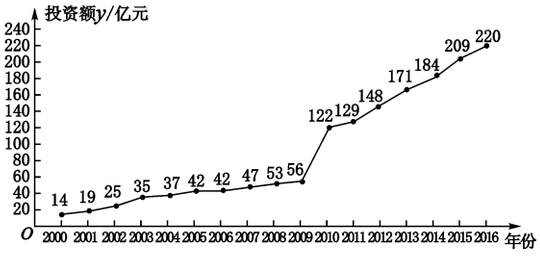
假设根据上表数据所得的线性回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$. 若某同学根据上表中前两组数据（1，0）和（2，2）求得的直线方程为 $y=b'x+a'$，则以下结论正确的是（　　）

A. $\hat{b}>b'$，$\hat{a}>a'$　　B. $\hat{b}>b'$，$\hat{a}<a'$

C. $\hat{b}<b'$，$\hat{a}>a'$　　D. $\hat{b}<b'$，$\hat{a}<a'$

11. 若对于样本点 (x_1,y_1)，(x_2,y_2)，…，(x_n,y_n)，它们的随机误差 $e_i=y_i-bx_i-a$ 的估计值残差 $\hat{e}_i=y_i-\hat{y}_i$ 越趋近于 0，则 R^2 越趋近于_____.

12.（2018 · 全国卷Ⅱ）下图是某地区 2000 年至 2016 年环境基础设施投资额 y（单位：亿元）的折线图.



为了预测该地区 2018 年的环境基础设施投资额，建立了 y 与时间变量 t 的两个线性回归模型. 根据 2000 年至 2016 年的数据（时间变量 t 的值依次为 1，2，…，17）建立模型①：$\hat{y}=-30.4+13.5t$；根据 2010 年至 2016 年的数据（时间变量 t 的值依次为 1，2，…，7）建立模型②：$\hat{y}=99+17.5t$.

（1）分别利用这两个模型，求该地区 2018 年的环境基础设施投资额的预测值；

（2）你认为利用哪个模型得到的预测值更可靠？并说明理由.

再生新疑

通过本学时的学习，我们比较完整地了解了一个回归分析的过程，初步学会了以建立回归直线方程为主的线性回归分析的方法

与基本思想.如果我们判断变量之间不适合选用线性回归方程,那么我们该如何对其进行分析呢?

第三学时　非线性回归模型

板块一　创设问题,引领目标

导入问题,直指学时重点

问题呈现

在前面的学习中,我们比较完整地了解了一个回归分析的过程.那么,如果我们判断变量之间不适合选用线性回归方程时又该如何进行分析呢?如问题:

为了研究3月下旬的平均气温x(℃)与4月20日前棉花害虫化蛹高峰日y的关系,某地区观察了2000年～2005年的情况,得到下面的数据:

年份	2000	2001	2002	2003	2004	2005
x/℃	24.4	29.6	32.9	28.7	30.3	28.9
y/日	19	6	1	10	1	8

据气象预报,该地区2006年3月下旬的平均气温为27 ℃,你能根据所学的数学知识,估计一下2006年4月化蛹高峰日为哪天吗?

你觉得能用线性回归方程来预报结果吗?

板块二　自学思疑,初探问题

尊重认知规律,亲历感悟知识生成

教材导读

阅读教材p6～p8的有关内容,思考并回答下列问题:

问题一

什么是非线性回归?

若两个变量呈现非线性关系,你将如何建立这两个变量之间的关系?

问题二

如何建立非线性回归模型?

1.常用的非线性拟合函数有哪些?

2.建立非线性回归模型与建立线性回归模型的主要步骤有什么不同?

自主测评

1.如果样本点分布在某一条指数函数曲线$y=ae^{bx}$的周围,其中a和b是参数,通过两边取自然对数的方法,把指数关系式变成对数关系式后,下列变换结果正确的是(　　)

A. $\ln y=bx\ln a$　　B. $\ln y=bx+\ln a$

C. $\ln y=\ln a\cdot\ln bx$　　D. $\ln y=\ln a+\ln bx$

2.下列回归方程中,_____是线性回归方程,_____是非线性回归方程.

①$\hat{y}=0.688x+3.27$；②$\hat{y}=0.25x^2-12.8$；③$\hat{y}=2.6e^{1.3x}$；④$\hat{y}=4-1.5x$；⑤$\hat{y}=1.38e^{-\frac{0.185}{x}}$.

3. 非线性回归模型的建立，一般可以通过变量置换（如两边取对数的变换），转化成_____模型而实现.

4. 对于同一个样本数据，在已经建立的几个不同的回归模型中，我们可以通过分别考察它们的_____和_____来判断各个模型拟合效果的好坏.

疑点归纳

通过上面的学习，有的同学还是不太清楚非线性回归与线性回归的关系，你弄清楚了吗？此外，你还有什么疑惑，赶紧记录下来，和你的同学、老师交流解答吧.

板块三　合作互助，共析问题

发展创新思维，形成主动探究与合作的意识和能力

问题三

非线性回归问题与线性回归问题之间有什么联系？

1. 如何把非线性回归问题转化为线性回归问题？（如$y=ae^{bx}$，你能通过什么方式把指数关系转化为线性关系？）

2. 对于相同的样本数据，可建立的回归模型是唯一的吗？

3. 如何分析、比较不同回归模型的拟合效果？

板块四　展示交流，探究问题

提高分析问题、解决问题的能力

问题四

你了解非线性回归的基本思想吗？

1. 在解决非线性回归问题的过程中涉及的转换思想是什么？

2. 如何判断、选择不同的回归模型？

展题设计

展题 1　混凝土的抗压强度x较易测定，其抗剪强度y不易测定. 工程中希望由x估算出y，以便应用. 现测得一批对应数据如下表：

x	141	152	168	182	195	204	223	254	277
y	23.1	25.3	27.9	29.8	31.1	31.8	32.5	34.8	35.2

其中：x——抗压强度（kg/cm^2），y——抗剪强度（kg/cm^2）.

试建立y与x的回归方程.

指导要求：你能大致完成建立非线性回归模型的过程吗？

（1）如何理解用“线性回归模型”解决“非线性回归问题”的转换思想？

（2）你会根据散点图选择相应的函数类型吗？你会采用适当的变量置换吗？

（3）你会比较、分析不同模型的拟合效果吗？为什么不同的函数模型中x，y与$(y_i-\bar{y})^2$这三组的数据均相同，但残差平方和$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$却因为选择了不同的回归方程而有所变化？如何通过残差与相关指数来分析、比较不同模型的拟合效果？

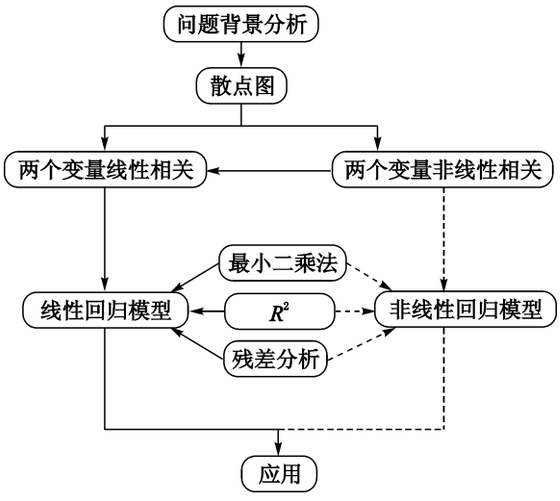
展题2 有一个测量水流量的实验装置，测得实验数据如下表：

i	1	2	3	4	5	6	7
水高 h /cm	0.7	1.1	2.5	4.9	8.1	10.2	13.5
流量 Q /(L/min)	0.082	0.25	1.8	11.2	37.5	66.5	134

根据表中的数据，建立 Q 与 h 之间的回归方程.

指导要求：同“展题1”.

归纳总结



板块五 应用演练，再生新疑

分级设题，精选精练

分层演练

A. 基础巩固

1. 下列回归方程中，属于线性回归方程的是 （ ）

A. $\hat{y}=\frac{2}{x}+1.4$　　B. $\hat{y}=0.13e^{0.9x}$

C. $\hat{y}=1.8-3.5x$　　D. $\hat{y}=0.6x^2+0.45$

2. 当我们建立多个模型来拟合某个数据时，为了比较各个模型的拟合效果，我们可以通过计算下列哪些量来确定 （ ）

①残差平方和；②样本点的中心；③相关指数 R^2；④相关系数 r.

A. ①　　B. ①③

C. ①②③　　D. ③④

3. 若一个样本的总偏差平方和为256，残差平方和为32，则相关指数为 （ ）

A. 0.125　　B. 1

C. 1.125　　D. 0.875

4. 已知样本点散落在某一条曲线 $y=e^{a+bx}$ 附近，作变换 $z=\ln y$，利用线性回归模

型来求其中的参数 a，b，则拟合其变换后的样本点的直线方程为_____.

B. 能力提升

5. 下列说法正确的有　　(　　)

①回归方程适用于一切样本的总体；

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.

A. ①②　　B. ①③

C. ②③　　D. ③④

6. 若一个样本的残差平方和是 325，总偏差平方和是 923，则解释变量与随机误差对预报变量的贡献率与影响效应分别约为　　(　　)

A. 64.8% 和 35.2%

B. 60% 和 40%

C. 68.9% 和 31.1%

D. 79.5% 和 10.5%

7. 在某种新型材料的研制中，试验人员获得了下列一组试验数据. 现准备用下列四个函数中的一个近似地表示这些数据的规律，其中最接近的一个是　　(　　)

x	1.95	3.00	3.94	5.10	6.12
y	0.97	1.59	1.98	2.35	2.61

A. $y=2^x$　　B. $y=\log_2 x$

C. $y=\frac{1}{2}(x^2-1)$　　D. $y=2.6\cos x$

8. 已知因变量 y 与自变量 x 的倒数 u $\left(u=\frac{1}{x}\right)$之间具有线性相关关系，如果求得 y 与 u 的回归直线方程是 $\hat{y}=1.25+5.3u$，那么 y 关于 x 的非线性回归方程是　　(　　)

A. $\hat{y}=1.25x+5.3$

B. $\hat{y}=1.25+5.3x$

C. $\hat{y}=1.25+\frac{x}{5.3}$

D. $\hat{y}=1.25+\frac{5.3}{x}$

C. 拓展创新

9. 在一个非线性回归过程中，如果我们选用指数型函数曲线 $y=a\cdot e^{\frac{b}{x}}$ 来进行模拟，并通过变量置换得到线性回归模型 $y'=a'+b'x'$，那么所用变量置换为 $x'=$____，$y'=$____，$a'=$____，$b'=$____.

10. 某种图书每册的成本费 y(元)与印刷册数 x(千册)有关. 经统计，得到如下数据：

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

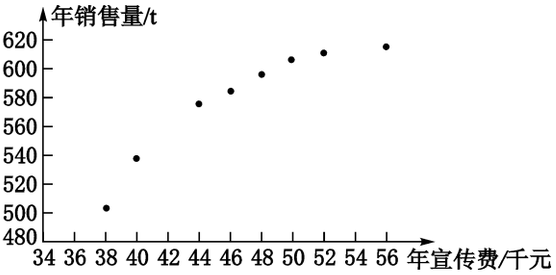
检验每册书的成本费 y 与印刷册数的倒数 $\frac{1}{x}$ 之间是否具有线性相关关系，如果有，求出 y 对 x 的回归方程.

11. 在彩色显像中，由经验知：形成染料光学密度 y 与析出银的光学密度 x 由公式 $y=Ae^{\frac{b}{x}}(b<0)$ 表示. 现测得实验数据如下：

x_i	0.05	0.06	0.25	0.31	0.07	0.10
y_i	0.10	0.14	1.00	1.12	0.23	0.37
x_i	0.38	0.43	0.14	0.20	0.47	
y_i	1.19	1.25	0.59	0.79	1.29	

试求 y 对 x 的回归方程.

12. 某公司为确定下一年度投入某种产品的宣传费，需了解年宣传费 x（单位：千元）对年销售量 y（单位：t）和年利润 z（单位：千元）的影响，对近 8 年的年宣传费 x_i 和年销售量 $y_i(i=1,2,\cdots,8)$ 数据作了初步处理，得到下面的散点图及一些统计量的值.



$\bar{x}$	$\bar{y}$	$\bar{w}$	$\sum_{i=1}^{8}(x_i-\bar{x})^2$	$\sum_{i=1}^{8}(w_i-\bar{w})^2$	$\sum_{i=1}^{8}(x_i-\bar{x})(y_i-\bar{y})$	$\sum_{i=1}^{8}(w_i-\bar{w})(y_i-\bar{y})$
46.6	563	6.8	289.8	1.6	1 469	108.8

表中 $w_i=\sqrt{x_i}$，$\bar{w}=\frac{1}{8}\sum_{i=1}^{8}w_i$.

(1) 根据散点图判断，$y=a+bx$ 与 $y=c+d\sqrt{x}$ 哪一个适宜作为年销售量 y 关于年宣传费 x 的回归方程类型？（给出判断即可，不必说明理由）

(2) 根据(1)的判断结果及表中数据，建立 y 关于 x 的回归方程.

(3) 已知这种产品的年利率 z 与 x，y 的关系为 $z=0.2y-x$. 根据(2)的结果回答下列问题：

①年宣传费 $x=49$ 时，年销售量及年利润的预报值是多少？

②年宣传费 x 为何值时，年利率的预报值最大？

再生新疑

1.2　独立性检验的基本思想及其初步应用

板块一　创设问题，引领目标

导入问题，直指学时重点

问题呈现

在生活与学习中，我们会遇到如下一些类似的问题：

某校高三年级在一次全年级的大型考试中，数学成绩优秀和非优秀的学生中，物理、化学、总分也为优秀的人数如下表所示．那么你能从表中给出的数据，判断一下数学成绩优秀分别与物理、化学、总分优秀哪个关系较大吗？

	物理	化学	总分
数学优秀	228	225	267
数学非优秀	143	156	99

注：该年级此次考试中数学成绩优秀的有360人，非优秀的有880人．

那么，你能判断数学成绩优秀与其他学科成绩优秀之间有关系吗？有关系的话又有“多大程度”的关系呢？你知道什么是独立性检验吗？从这一学时开始，我们通过实例来学习独立性检验的有关知识．

材料链接

百度搜索：如何用 Excel 生成等高条形图、二维条形图、三维柱形图．

板块二　自学思疑，初探问题

尊重认知规律，亲历感悟知识生成

教材导读

阅读教材 p10～p15 的有关内容，思考并回答下列问题：

问题一

什么是分类变量？什么是列联表？

1. 分类变量是本节给出的概念，你能指

出分类变量和我们以前学习的变量在本质上有什么差别吗？你能否再举出几个分类变量？你能否举出一对分类变量，使它们之间有着某种关联？

❀ 2. 为了判别吸烟与患肺癌两件事是否相关，教材上采用了两种方法：等高条形图法和独立性检验法. 你能说出等高条形图法的优缺点吗？

❀ 3. 列联表是我们进行独立性检验不可或缺的重要工具，表格中的 a,b,c,d 无疑是最重要的数据. 教材在一个不太起眼的角落提到了我们进行独立性检验的时候对这四个基本数据的基本要求，你找到了吗？

问题二

什么是独立性检验？

❀ 1. 有的书上是这样描述独立性检验的原理的：在假设 H_0 下，如果出现一个与 H_0 相矛盾的小概率事件，就推断 H_0 不成立，且该推断犯错误的概率不超过这个小概率. 请仔细阅读教材，结合例1(秃顶与患心脏病)，说说在该问题中，假设 H_0 是什么，小概率是多少，与 H_0 相矛盾的小概率事件又是什么.

❀ 2. 假设有两个分类变量，我们计算得到的 K^2 的值是 9.264，那么我们应该说这两个分类变量是相关还是无关呢？这个时候，我们应该确定容许推断“两个分类变量有关系”犯错误的最大概率，只有明确这一标准，我们才能得到最终的结论.

假如我们确定这种推断犯错误的概率不超过 0.05，那么 9.264 大于临界值 3.841，这个时候我们应该认为这两个分类变量是相关还是无关？为什么？

自主测评

1. 下表是一个 2×2 列联表：

	y_1	y_2	总计
x_1	a	21	73
x_2	2	25	27
总计	b	46	100

则表中 a,b 的值分别为 （ ）

A. 94，96　　B. 52，54

C. 52，50　　D. 54，52

2. 可以粗略地判断两个分类变量是否有关系的是（　　）

A. 散点图

B. 残差图

C. 等高条形图

D. 以上都不对

3. 在等高条形图中，下列哪两个比值相差越大，要推断的论述成立的可能性就越大（　　）

A. $\frac{a}{a+b}$与$\frac{c}{c+d}$

B. $\frac{a}{c+d}$与$\frac{c}{a+d}$

C. $\frac{a}{a+d}$与$\frac{c}{b+c}$

D. $\frac{a}{b+d}$与$\frac{c}{a+c}$

4. 在一项“打鼾与患心脏病是否有关”的调查中，共调查了 1 978 人，经过计算，得 $K^2=28.63$. 根据这一数据分析，我们有理由认为打鼾与患心脏病是_____（填“有关”或“无关”）的.

5. 下面是一个 2×2 列联表：

	y_1	y_2	总计
x_1	8	26	34
x_2	24	31	55
总计	32	57	89

根据上述数据，我们计算得出的 $K^2=$_____.

疑点归纳

通过上面的学习，相信多数同学已经理解了独立性检验的基本思想，你清楚吗？此外，你还有什么疑惑，赶紧记录下来，和你的同学、老师交流解答吧.

板块三　合作互助，共析问题

发展创新思维，形成主动探究与合作的意识和能力

问题三

独立性检验的基本思想是什么？

1. 独立性检验的基本思想是什么？

2. 独立性检验和反证法有什么异同？如何通过教材 p12 表 1－9 更好地理解独立性检验的原理呢？

板块四　展示交流，探究问题

提高分析问题、解决问题的能力

问题四

你了解判断“两个分类变量有关系”的主要方式或步骤了吗？

我们已经知道，若两件事 A，B 相互独立（完全无关），则概率上满足 $P(AB)=P(A)\cdot P(B)$. 而在我们的统计结果中，可以用频率近似代替概率，于是在 2×2 列联表（表中 $\overline{A}$，$\overline{B}$ 分别表示事件 A，B 的对立事件）中，

	A	$\overline{A}$	合计
B	a	b	$a+b$
$\overline{B}$	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d$

A 与 B 两件事越“相关”，我们用到的统计量 $K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$（其中 $n=a+b+c+d$）应该是越大还是越小？

展题设计

展题1 在一次对人们的休闲方式的调查中，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) 粗略判断休闲方式与性别是否有关系；

(3) 如果休闲方式与性别有关系，则求它们有关系的程度.

指导要求：(1)能根据所给数据建立一个2×2列联表吗？

(2)用什么方式可粗略地判断分类变量X与Y间是否有关？$|ad-bc|$值的大小与“X与Y之间有关系”的可能性大小之间有什么规律？$\left|\frac{a}{a+b}-\frac{c}{c+d}\right|$值的大小与“$X$与$Y$之间有关系”的可能性大小之间有什么规律？

(3)你熟悉独立性检验的具体做法或步骤吗？k的大小与判断“X与Y有关系”的可信度的大小(或判断出错的可能性)之间的规律是什么呢？

展题2 通过随机询问110名性别不同的大学生是否爱好某项运动，得到如下的列联表：

	男	女	总计
爱好	40	20	60
不爱好	20	30	50
总计	60	50	110

由$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，得

$K^2=\frac{110\times(40\times30-20\times20)^2}{60\times50\times60\times50}\approx7.8$.

附表：

$P(K^2\geqslant k)$	0.050	0.010	0.001
k	3.841	6.635	10.828

参照附表，得到的正确结论是 (　　)

A. 在犯错误的概率不超过0.010的前提下，认为“爱好该项运动与性别有关”

B. 在犯错误的概率不超过0.010的前提下，认为“爱好该项运动与性别无关”

C. 在犯错误的概率不超过0.1%的前提下，认为“爱好该项运动与性别有关”

D. 在犯错误的概率不超过0.1%的前提下，认为“爱好该项运动与性别无关”

指导要求：略.

展题3 有2 000名网购者在11月11日当天于某购物网站进行网购消费(消费金额不超过1 000元)，其中有女士1 100名，男士900名. 该购物网站为优化营销策略，根据性别采用分层抽样的方法从这2 000名网购者中抽取200名进行分析，如下表(消费金额

的单位：元）：

女士消费情况：

消费金额	(0, 200)	[200, 400)	[400, 600)	[600, 800)	[800, 1 000]
人数	10	25	35	30	x

男士消费情况：

消费金额	(0, 200)	[200, 400)	[400, 600)	[600, 800)	[800, 1 000]
人数	15	30	25	y	5

(1)计算 x，y 的值；在抽出的 200 名且消费金额在[800，1 000](单位：元)的网购者中随机选出 2 名发放网购红包，求选出的 2 名网购者都是男士的概率.

(2)若消费金额不低于 600 元的网购者为“网购达人”，低于 600 元的网购者为“非网购达人”，根据以上统计数据填写下面 2×2 列联表，并回答能否在犯错误的概率不超过 0.05 的前提下认为“是否为‘网购达人’与性别有关”？

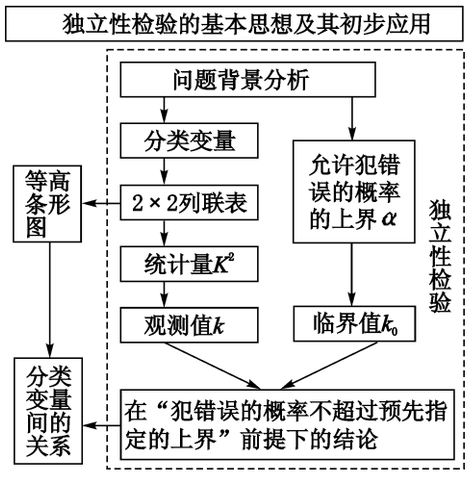
	女士	男士	总计
网购达人			
非网购达人			
总计			

附：

$P(K^2\geqslant k_0)$	0.10	0.05	0.025	0.010	0.005
k_0	2.706	3.841	5.024	6.635	7.879

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

归纳总结



板块五　应用演练，再生新疑

分级设题，精选精练

分层演练

A. 基础巩固

1. 在判断两个分类变量是否有关系的常用方法中，最为精确的方法是　　(　　)

A. 考察随机误差 e

B. 考察线性相关系数 r

C. 考察相关指数 R^2

D. 考察独立性检验中的 K^2

2. 假设有两个分类变量 X 与 Y，其 2×2 列联表如下：

	y_1	y_2	总计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

以下各组数据中，对于同一样本，能说明

X 与 Y 有关系的可能性最大的一组是 (　　)

A. $a=5,b=4,c=3,d=2$

B. $a=5,b=3,c=4,d=2$

C. $a=2,b=3,c=4,d=5$

D. $a=2,b=3,c=5,d=4$

3. 当某矿石粉厂生产一种矿石粉时，在数天内即有部分工人患职业性皮肤炎. 在生产季节开始时，随机抽取 75 名车间工人穿上新防护服，其余工人仍穿原来的防护服，生产进行一个月后，检查两组工人的皮肤炎患病人数如下：

	患病	不患病	合计
新防护服	5	70	75
旧防护服	10	18	28
合计	15	88	103

则以下说法正确的是 (　　)

A. 新防护服对工人预防职业性皮肤炎没有效果

B. 新防护服与旧防护服对预防这种皮肤炎的效果一样

C. 穿上新防护服就有 99.9% 的工人不患这种皮肤炎

D. 在犯错误的概率不超过 0.001 的前提下，认为新防护服比旧防护服对预防这种皮肤炎有效果

4. 调查者通过随机询问 72 名男女中学生喜欢文科还是喜欢理科，得到如下 2×2 列联表：

	喜欢文科	喜欢理科	总计
男生	8	28	36
女生	20	16	36
总计	28	44	72

则在犯错误的概率不超过 0.005 的前提下，认为中学生的性别和喜欢文科还是理科_____(填“有”或“没有”)关系.

B. 能力提升

5. 某大型企业人力资源部为了研究企业员工工作积极性和对待企业改革的关系，随机抽取了 189 名员工进行调查. 其中支持企业改革的调查者中，工作积极的 54 人，工作一般的 32 人；而不太赞成企业改革的调查者中，工作积极的 40 人，工作一般的 63 人. 若由此推断企业的全体员工对待企业改革的态度与其工作的积极性有关系，则这种推断犯错误的概率不超过 (　　)

A. 0.001　　B. 0.005

C. 0.01　　D. 0.025

6. 某人研究中学生的性别与成绩、视力、智商、阅读量这 4 个变量之间的关系，随机抽查 52 名中学生，得到统计数据如表 1 至表 4，则与性别有关联的可能性最大的变量是 (　　)

表 1

成绩 / 性别	不及格	及格	总计
男	6	14	20
女	10	22	32
总计	16	36	52

表 2

视力 / 性别	好	差	总计
男	4	16	20
女	12	20	32
总计	16	36	52

表 3

智商 / 性别	偏高	正常	总计
男	8	12	20
女	8	24	32
总计	16	36	52

表 4

阅读量 / 性别	丰富	不丰富	总计
男	14	6	20
女	2	30	32
总计	16	36	52

A. 成绩　　B. 视力

C. 智商　　D. 阅读量

7. 考察棉花种子经过处理跟生病之间的关系，得到下表数据：

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

根据以上数据，知　　（　　）

A. 种子经过处理跟是否生病有关

B. 种子经过处理跟是否生病无关

C. 种子是否经过处理决定是否生病

D. 以上说法都是错误的

8. 两个分类变量 X 和 Y 可能的取值分别为 $\{x_1,x_2\}$ 和 $\{y_1,y_2\}$，其样本频数满足 $a=10,b=21,c+d=35$. 若在犯错误的概率不超过 0.1 的前提下认为 X 与 Y 有关系，则 c 的值可能等于　　（　　）

A. 4　　B. 5　　C. 6　　D. 7

C. 拓展创新

9. 经过对 K^2 的统计量的研究，得到了若干个临界值，当 $K^2\leqslant 2.706$ 时，我们认为两个分类变量 X 与 Y 之间　　（　　）

A. 在犯错误的概率不超过 0.05 的前提下认为 X 与 Y 有关系

B. 在犯错误的概率不超过 0.01 的前提下认为 X 与 Y 有关系

C. 没有充分的理由说明 X 与 Y 有关系

D. 在犯错误的概率不超过 0.05 的前提下认为 X 与 Y 没有关系

10. 某医疗研究所为了检验某种血清预防感冒的作用，把 500 名使用血清的人与另外 500 名未用血清的人一年中的感冒记录作比较，提出假设 H_0：这种血清不能起到预防感冒的作用. 利用 2×2 列联表计算，得 $K^2\approx 3.918$，经查对临界值表知，$P(K^2\geqslant 3.841)\approx 0.05$. 对此，四名同学作出了以下的判断：

p：在犯错误的概率不超过 0.05 的前提下认为“能起到预防感冒的作用”；

q：如果某人未使用该血清，那么他在一年中有 95% 的可能性得感冒；

r：这种血清预防感冒的有效率为 95%；

s：这种血清预防感冒的有效率为 5%.

则下列结论中，正确的是_____（填序号）.

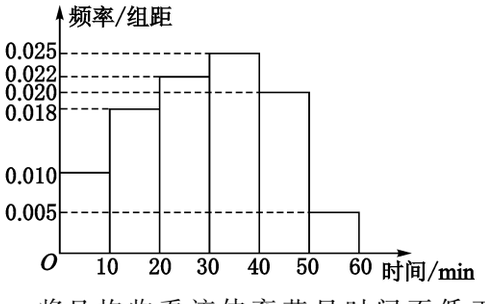
① $p\wedge(\neg q)$；

② $(\neg p)\wedge q$；

③ $(\neg p\wedge\neg q)\wedge(r\vee s)$；

④ $(p\vee\neg r)\wedge(\neg q\vee s)$.

11. 电视传媒公司为了了解某地区电视观众对某体育节目的收视情况，随机抽取了 100 名观众进行调查，其中女性 55 名. 下面是根据调查结果绘制的观众日均收看该体育节目时间的频率分布直方图：



将日均收看该体育节目时间不低于 40 min的观众称为“体育迷”. 已知“体育迷”中有 10 名女性.

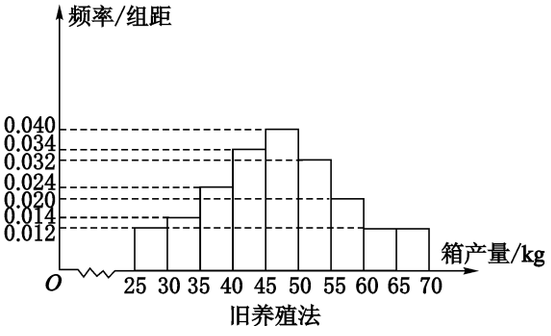
(1)根据已知条件完成下面 2×2 列联表，并据此资料判断能否在犯错误的概率不超过 0.05 的前提下认为“体育迷”与性别有关？

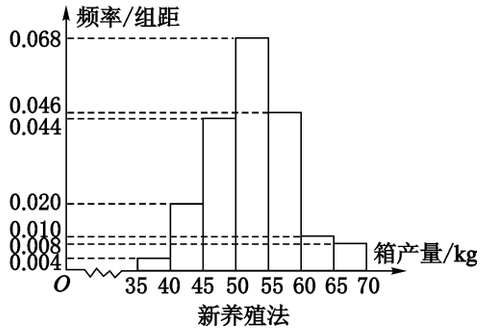
	非体育迷	体育迷	合计
男			
女			
合计			

(2)将日均收看该体育节目不低于 50 min

的观众称为“超级体育迷”，已知“超级体育迷”中有2名女性.若从“超级体育迷”中任意选取2人，求至少有1名女性观众的概率.

12.(2017·全国卷Ⅱ)海水养殖场进行某水产品的新、旧网箱养殖方法的产量对比，收获时各随机抽取了100个网箱，测量各箱水产品的产量(单位：kg)，其频率分布直方图如下：





(1)记A表示事件“旧养殖法的箱产量低于50 kg”，估计A的概率；

(2)填写下面列联表，并根据列联表判断能否在犯错误的概率不超过0.010的前提下，认为箱产量与养殖方法有关：

	箱产量<50 kg	箱产量≥50 kg
旧养殖法		
新养殖法		

(3)根据箱产量的频率分布直方图，对两种养殖方法的优劣进行比较.

附：

$P(K^2\geqslant k)$	0.050	0.010	0.001
k	3.841	6.635	10.828

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

再生新疑

第一章学习报告

一、完成本章知识结构图.

二、编写本章问题集.

参考以下问题：

1.通过本章的学习，你学会判别两个变量之间的关系了吗？函数关系是一种什么关系？相关关系是一种什么关系？回归分析是对具有哪种关系的两个变量进行统计分析的一种常用方法？

❀ 2. 我们用什么来衡量两个变量之间线性关系的强弱？它的计算公式是什么？你能说出 r 的取值范围吗？什么情况下表明两个变量正相关？什么情况下表明两个变量负相关？r 的绝对值越接近 1，说明了什么？r 的绝对值越接近 0，又说明了什么？通常在什么情况下，我们就认为两个变量有很强的线性相关关系？

❀ 3. 我们用什么方法求回归直线方程 $\hat{y}=bx+a$ 中未知参数 b 和 a 的估计值 $\hat{b}$ 和 $\hat{a}$？对于解释变量的某个值，根据回归直线方程求出的预报变量的值是精确值吗？解释变量能完全解释预报变量的变化吗？预报变量还会受什么因素的影响？

❀ 4. 你能解释一下相关指数 $R^2=1-\dfrac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}$ 的含义吗？你弄清楚相关指数 R^2 与残差平方和之间的变化规律了吗？如何用相关指数 R^2 与残差平方和来判断模型的拟合效果呢？

❀ 5. 如果我们发现散点图中，样本点的分布不在某条直线附近，我们还能用线性回归方程来建立两个变量之间的关系吗？

❀ 6. 建立线性回归模型和非线性回归模型的基本步骤是什么？

❀ 7. 你能在生活中搜集一些关于两个分类变量的实例吗？就你搜集到的材料，你能作出有关数据的 2×2 列联表和等高条形图吗？能否简单判断出这两个分类变量之间是否有关系呢？你的判断的可靠程度有多高？

❀ 8. 独立性检验的基本思想是什么？我们是如何构造随机变量 K^2 的？如何根据 K^2 的观测值 k 来说明有多大把握认为“X 与 Y 有关系”？

三、尝试编拟一份测试题.

数学探究活动

体验数学研究的过程和创造的激情

“回归”名称的由来

“回归”(英文“regression”)一词是由英国著名生物学家兼统计学家高尔顿(Galton)在研究人类遗传问题时提出来的. 高尔顿是优生学的创立者，他主要研究体质和智力的遗传.

1889 年，高尔顿在研究人类身高的遗传时，为了研究父代与子代身高的关系，搜集了 1 078 对父亲及其儿子的身高数据. 他发现这些数据的散点图大致呈直线状态，也就是说，总的趋势是父亲的身高增加时，儿子的身高也倾向于增加.

但是，高尔顿对试验数据进行了深入的分析，发现了一个很有趣的现象——回归效应. 因为当父亲的身高高于平均水平时，他们儿子的身高比他高的概率要小于比他矮的概率；父亲矮于平均身高时，他们儿子的身高比他矮的概率要小于比他高的概率. 这种现象

反映出了一个规律，即这两类儿子（父亲的身高高于平均水平与父亲的身高低于平均水平）的身高，有向他们父辈的平均身高回归的趋势. 这就是统计学上最初出现“回归”一词的涵义. 对这个一般结论的解释是：大自然具有一种约束力，使人类身高的分布相对稳定而不至于产生两极分化（高的更高，矮的更矮），这就是所谓的回归效应.

高尔顿根据试验数据还推算出儿子的身高（Y）与父亲的身高（X）的关系式是 $Y=a+bX$，它代表的是一条直线，这条直线称为回归直线，并把相应的统计分析称为回归分析.

后来，人们把“回归”一词沿用到其他具有相关关系的回归效应现象的分析中. 尽管关系式 $Y=a+bX$ 仅反映了变量相关关系的一种特殊情况（线性相关），对于更多的相关关系（非线性相关），特别是涉及多个变量的情况，该关系式并不适用或成立，但是我们依然把相应的相关分析统称为回归分析，“回归”一词便这样一直沿用下来了.

生活中有趣的相关关系

其实，只要我们细心观察就可以发现，在我们的生活中也蕴含着许多有趣的回归现象，例如：

1. 人体脂肪含量与年龄之间具有相关关系.

在一次对人体脂肪含量 y（%）和年龄 x 关系的研究中，研究人员获得了一组样本数据：

年龄	23	27	39	41	45	49	50
脂肪含量	9.5	17.8	21.2	25.9	27.5	26.3	28.2
年龄	53	54	56	57	58	60	61
脂肪含量	29.6	30.2	31.4	30.8	33.5	35.2	34.6

根据上述数据所提供的信息，从散点图分析发现，人体的脂肪含量与年龄之间具有线性相关关系，而且正相关关系很强.

2. 足迹长度与身高的相关关系.

人的身高与足长存在一定的线性相关关系，回归系数约为 7 或 6.875 1. 在实际工作中，侦查、技术人员经常运用现场足迹分析犯罪嫌疑人的身高. 但是足长与身高之间的关系受性别、年龄以及地区等因素的影响较大. 为了完善这一研究工作，研究人员对全国范围内不同地区、不同年龄、不同性别的人群进行了大量的抽样调查，逐一测量样本足迹长度，建立数据库，利用计算机进行综合统计、回归分析，对足迹长度与身高之间的相关关系及其变化规律作了全面的分析研究，得出结论：足迹长度与身高之间的相关关系明显，可以根据足迹长度分析留痕人的身高，推算公式为：

身高（cm）＝脚印长度（cm）×6.876.

因此，根据脚印的大小，能够推测出罪犯的身高，这是符合科学的.

3. 学习中的相关关系.

在学习过程中，我们已经发现了各门学科之间也有一定的相关关系，比如数学成绩对物理成绩也会产生一定的影响等.

4. 俗语中的相关关系.

生活中我们还有一些俗语，比如说“龙生龙，凤生凤”“名师出高徒”“强将手下无弱兵”“虎父无犬子”“差之毫厘，谬以千里”“一着不慎，满盘皆输”等等. 这些俗语是不是或多或少都体现了两个事物之间的一种相关关系呢？

课后发现探究

同学们课后可以进一步开动脑筋，仔细观察一下周围的各种事物，尽可能多地运用我们本章所学习的统计思想和方法，就你观察到的一些相关现象，作出初步的判断和分析.

特别地，可以完成下列活动内容：

搜集本班同学中 10 对父子（或母女）的身高数据，按照线性回归分析的方法和步骤，把你获得的数据进行处理，求得回归直线方程，并初步检验回归效果是否和经典回归效应相类似.

第一章综合测试

（时间：90 分钟　满分：100 分）

一、选择题（本大题共 10 小题，每小题 3 分，共 30 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）

1. 下列结论正确的是 （　　）

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方法.

A. ①②③　　B. ①②④

C. ①②　　D. ①②③④

2. 一位母亲记录儿子 3～9 岁的身高，由此建立的身高与年龄的回归直线方程为 $\hat{y}=7.19x+73.93$，由此预测这个孩子 10 岁时的身高，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 以下

D. 身高在 145.83 cm 左右

3. 由一组数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 得到的回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$，则下列说法不正确的是 （　　）

A. 直线 $\hat{y}=\hat{b}x+\hat{a}$ 至少经过点 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 中的一个点

B. 直线 $\hat{y}=\hat{b}x+\hat{a}$ 必过点 $(\overline{x},\overline{y})$

C. 直线 $\hat{y}=\hat{b}x+\hat{a}$ 的斜率为 $\dfrac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\ \overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2}$

D. 直线 $\hat{y}=\hat{b}x+\hat{a}$ 和各点 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 的偏差是该坐标平面上所有直线与这些点的偏差中最小的直线

4. 在某种产品表面进行腐蚀刻线试验，得到腐蚀深度 y 与腐蚀时间 x 之间的一组观测值如下表：

x/s	5	10	15	20	30	40	50	60	70	80	90
y/μm	6	10	10	13	16	17	19	23	25	29	46

根据上表我们可以得到腐蚀深度 y 与腐蚀时间 x 的回归直线方程为 $\hat{y}=0.304x+5.36$，则以下对回归系数 $\hat{b}=0.304$ 的含义叙述正确的是 （　　）

①腐蚀时间 x 每增加 1 个单位，腐蚀深度 y 平均减少 0.304 个单位；

②腐蚀时间 x 每减少 1 个单位，腐蚀深度 y 平均增加 0.304 个单位；

③腐蚀时间 x 每增加 1 个单位，腐蚀深度 y 平均增加 0.304 个单位；

④腐蚀时间 x 每减少 1 个单位，腐蚀深度 y 平均减少 0.304 个单位.

A. ①③　　B. ②④

C. ①②　　D. ③④

5. 设商品的销售额与利润之间的样本相关系数 $r=0.78$，说明两者之间存在着 （　　）

A. 高度相关关系

B. 中度相关关系

C. 弱度相关关系

D. 极弱相关关系

6. （2017 · 山东）为了研究某班学生的脚

长 x(单位:cm)和身高 y(单位:cm)的关系,从该班随机抽取 10 名学生,根据测量数据的散点图可以看出 y 与 x 之间有线性相关关系,设其回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$. 已知 $\sum_{i=1}^{10}x_i=225$, $\sum_{i=1}^{10}y_i=1\ 600$, $\hat{b}=4$. 该班某学生的脚长为 24 cm,据此估计其身高为 (　　)

A. 160 cm　　B. 163 cm

C. 166 cm　　D. 170 cm

7. 为研究人的肥胖程度(胖瘦)与家庭富裕水平(贫富)之间是否相关,随机调查了 50 000 人,其中有 5 000 名胖人. 下列独立性检验的方案中,较合理有效的是 (　　)

A. 随机抽取 100 名胖人和 100 名瘦人

B. 随机抽取 0.08% 的胖人和瘦人

C. 随机抽取 900 名瘦人和 100 名胖人

D. 随机抽取 0.1% 的瘦人和 1% 的胖人

8. 为了了解某社区居民的家庭年收入与年支出的关系,随机调查了该社区 5 户家庭,得到如下统计数据表:

收入 x/万元	8.2	8.6	10.0	11.3	11.9
支出 y/万元	6.2	7.5	8.0	8.5	9.8

根据上表,可得回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$,其中 $\hat{b}=0.76$, $\hat{a}=\overline{y}-\hat{b}\,\overline{x}$. 据此估计,该社区一户收入为 15 万元的家庭年支出为 (　　)

A. 11.4 万元　　B. 11.8 万元

C. 12.0 万元　　D. 12.2 万元

9. 某学校为考察数学成绩与物理成绩的关系,在高二随机抽取了 300 名学生,得到如下列联表:

	(数学) 85～100 分	(数学) 85 分以下	合计
(物理)85～100 分	37	85	122
(物理)85 分以下	35	143	178
合计	72	228	300

现判断数学成绩与物理成绩有关,则该判断的出错率不超过 (　　)

A. 0.5%　　B. 1%

C. 2%　　D. 5%

10. 已知变量 x 与 y 正相关,且由观测数据算得样本的平均数 $\overline{x}=3$, $\overline{y}=3.5$,则由该观测数据算得的线性回归方程可能为 (　　)

A. $\hat{y}=0.4x+2.3$

B. $\hat{y}=2x-2.4$

C. $\hat{y}=-2x+9.5$

D. $\hat{y}=-0.3x+4.4$

二、填空题(本大题共 8 小题,每小题 3 分,共 24 分)

11. 若施肥量 x 与水稻产量 y 的回归直线方程为 $\hat{y}=5x+250$,当施肥量为 80 kg 时,预报水稻产量为_____.

12. 已知样本数为 11,计算得 $\sum_{i=1}^{11}x_i=510$, $\sum_{i=1}^{11}y_i=214$,回归直线方程为 $\hat{y}=0.3x+\hat{a}$,则 $\overline{x}\approx$_____, $\hat{a}\approx$_____.(精确到 0.01)

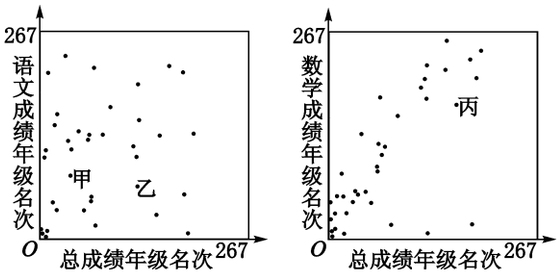
13. 在研究硝酸钠的可溶性程度时,在不同的温度下观测它在水中的溶解度,得观测结果如下表所示:

温度(x)	0	10	20	50	70
溶解度(y)	66.7	76.0	85.0	112.3	128.0

由此得到的回归直线的斜率是_____.

14. 已知一组观测值 (x_i,y_i),作出散点图后确定具有线性相关关系. 若对于回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$,求得 $\hat{b}=0.51$, $\overline{x}=61.75$, $\overline{y}=38.14$,则回归直线方程为_____.

15. 高三年级 267 位学生参加期末考试，某班 37 位学生的语文成绩、数学成绩与总成绩在全年级中的排名情况如下，甲、乙、丙为该班三位学生.



从这次考试成绩看：

(1)在甲、乙两人中，其语文成绩名次比其总成绩名次靠前的学生是_____；

(2)在语文和数学两个科目中，丙同学的成绩名次更靠前的科目是_____.

16. 变量 x 与 y 的列联表如下：

	y_1	y_2	合计
x_1	a	b	$a+b$
x_2	c	d	$c+d$

在 $a+b=c+d$ 的情况下，随机变量 $\xi=\left|\frac{a}{a+b}-\frac{c}{c+d}\right|$ 越大，表明 x 与 y 之间的关系越_____(填“强”或“弱”).

17. 某人研究中学生的性别与三种心理障碍——焦虑、说谎、懒惰这 3 个变量的关系，随机抽查了 110 名中学生，得到统计数据，如表 1 至表 3，则最有利于判断与性别有关联的变量是_____.

表 1

	焦虑	不焦虑	合计
女生	5	25	30
男生	20	60	80
合计	25	85	110

表 2

	说谎	不说谎	合计
女生	10	20	30
男生	10	70	80
合计	20	90	110

表 3

	懒惰	不懒惰	合计
女生	15	15	30
男生	50	30	80
合计	65	45	110

18. 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了 3 年的跟踪研究，调查他们是否又发作过心脏病. 调查结果如下表所示：

	又发作过心脏病	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

试根据上述数据判断：在犯错误的概率不超过 0.15 的前提下_____(填“能”或“不能”)说做过心脏搭桥手术与又发作心脏病有关系.

三、解答题(本大题共 4 小题，共 46 分. 解答时应写出必要的文字说明、证明过程或推演步骤)

19.(本小题满分 10 分)

某地区 2007 年至 2013 年农村居民家庭纯收入 y(单位：千元)的数据如下表：

年份	2007	2008	2009	2010	2011	2012	2013
年份代号 t	1	2	3	4	5	6	7
人均纯收入 y	2.9	3.3	3.6	4.4	4.8	5.2	5.9

(1)求 y 关于 t 的线性回归方程；

(2)利用(1)中的回归方程，分析 2007 年

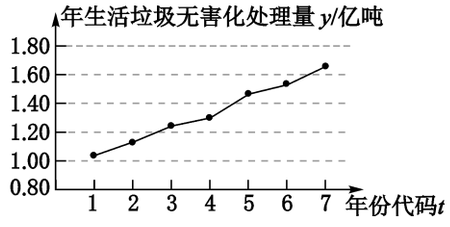
至2013年该地区农村居民家庭人均纯收入的变化情况，并预测该地区2015年农村居民家庭人均纯收入.

附：回归直线的斜率和截距的最小二乘法估计公式分别为：

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}(t_i-\overline{t})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(t_i-\overline{t})^2},\hat{a}=\overline{y}-\hat{b}\,\overline{t}.$$

20.（本小题满分12分）

（2016·全国卷Ⅲ）下图是我国2008～2014年生活垃圾无害化处理量（单位：亿吨）的折线图：



注：年份代码1～7分别对应年份2008～2014.

（1）由折线图看出，可用线性回归模型拟合y与t的关系，请用相关系数加以说明；

（2）建立y关于t的回归方程（系数精确到0.01），预测2016年我国生活垃圾无害化处理量.

附注：

参考数据：$\sum\limits_{i=1}^{7}y_i=9.32$，$\sum\limits_{i=1}^{7}t_iy_i=40.17$，$\sqrt{\sum\limits_{i=1}^{7}(y_i-\overline{y})^2}=0.55$，$\sqrt{7}\approx2.646$.

参考公式：

$$r=\frac{\sum\limits_{i=1}^{n}(t_i-\overline{t})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(t_i-\overline{t})^2\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}}.$$

回归方程$\hat{y}=\hat{a}+\hat{b}t$中斜率和截距的最小二乘估计公式分别为：

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}(t_i-\overline{t})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(t_i-\overline{t})^2},\hat{a}=\overline{y}-\hat{b}\overline{t}.$$

21.(本小题满分 12 分)

(2018·全国卷Ⅲ)某工厂为提高生产效率,开展技术创新活动,提出了完成某项生产任务的两种新的生产方式.为比较两种生产方式的效率,选取 40 名工人,将他们随机分成两组,每组 20 人,第一组工人用第一种生产方式,第二组工人用第二种生产方式.根据工人完成生产任务的工作时间(单位:min)绘制了如下茎叶图:

第一种生产方式		第二种生产方式
8	6	5 5 6 8 9
9 7 6 2	7	0 1 2 2 3 4 5 6 6 8
9 8 7 7 6 5 4 3 3 2	8	1 4 4 5
2 1 1 0 0	9	0

(1)根据茎叶图判断哪种生产方式的效率更高?并说明理由;

(2)求 40 名工人完成生产任务所需时间的中位数 m,并将完成生产任务所需时间超过 m 和不超过 m 的工人数填入下面的列联表;

	超过 m	不超过 m	合计
第一种生产方式			
第二种生产方式			
合计			

(3)根据(2)中的列联表判断,能否在犯错误的概率不超过 0.01 的前提下认为两种生产方式的效率有差异?

22.(本小题满分 12 分)

(2019·全国卷Ⅰ)某商场为提高服务质量,随机调查了 50 名男顾客和 50 名女顾客,每位顾客对该商场的服务给出满意或不满意的评价,得到下面列联表:

	满意	不满意
男顾客	40	10
女顾客	30	20

(1)分别估计男、女顾客对该商场服务满意的概率;

(2)在犯错误的概率不超过 0.05 的前提下能否认为男、女顾客对该商场服务的评价有差异?

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

$P(K^2\geqslant k)$	0.05	0.01	0.001
k	3.841	6.635	10.828