



湖南省中小学课程资源

成对数据的统计分析单元小结

(第一课时)

年 级：高二年级
主讲人：欧阳博文

学 科：数学(人教A版)
学 校：株洲星雅实验学校



湖南省中小学课程资源

第八章成对数据的统计分析

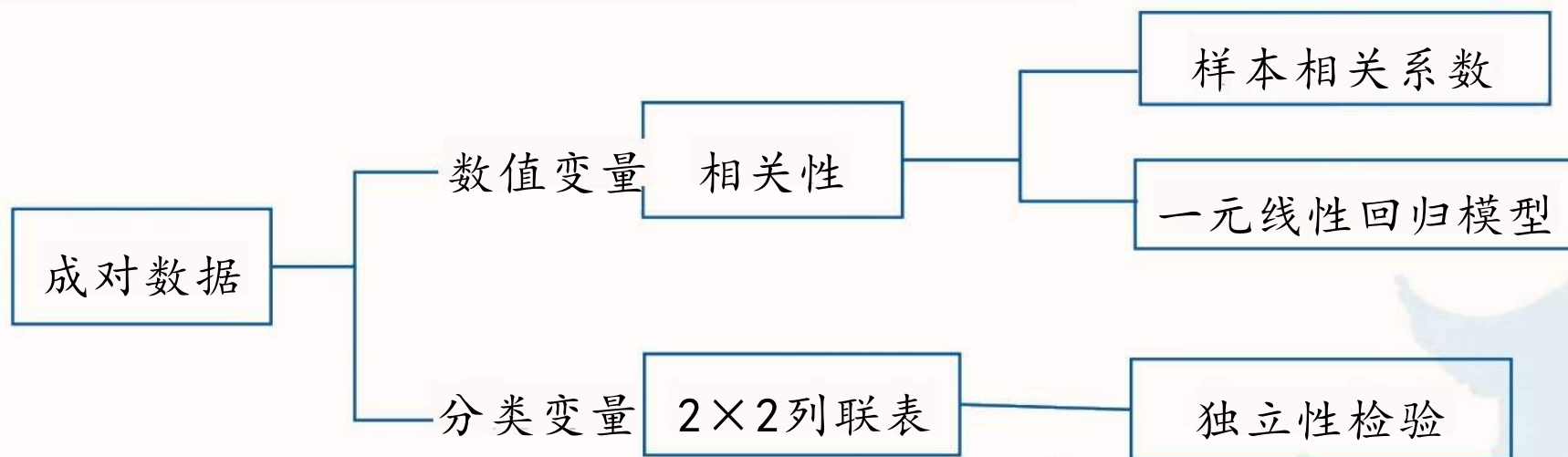
单元小结(第一课时)

年 级：高二年级 学 科：数学(人教A版)

主讲人：欧阳博文 学 校：株洲星雅实验学校

知识梳理、构建网络

问题1：你能画出本章的知识结构图吗？



问题2： 阅读教科书93页到134页的内容，理解相关系数、卡方的统计意义
回答以下三个问题：

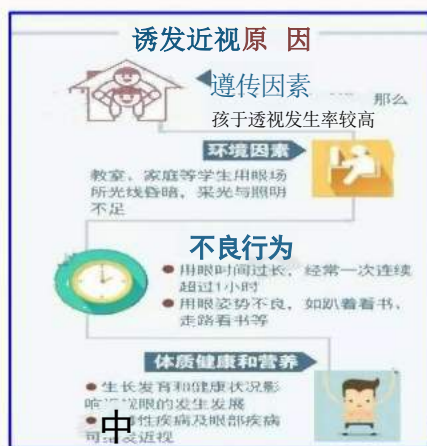
(1) 举例说明成对数据是如何形成的。

(2) 举例说明什么叫相关关系，它与函数关系有什么区别？

(3) 归纳两个变量之间的关系类型。

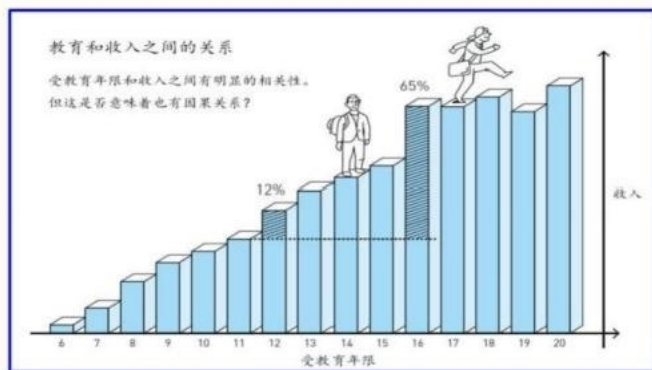
(1) 举例说明成对数据是如何形成的。

例如，为了掌握学生身体健康状况，需要了解身高变量和体重变量之间的关系；医疗卫生部门要制定预防青少年近视的措施，需要了解有哪些因素会影响视力，以及这些因素是如何影响视力的，等等。因为需要了解两个或两个以上变量之间的关系，会对同一个个体的两个维度进行观测，从而形成成对数据。



(2) 举例说明什么叫相关关系，它与函数关系有什么区别？

例如，身高与脚长的关系，一般来说，身高较高的人脚长也会较长，但身高相同的人脚长未必相同；受教育程度和收入水平的关系，一般来说，受教育程度高的人收入也较高，但受教育程度相同的人收入未必相同。

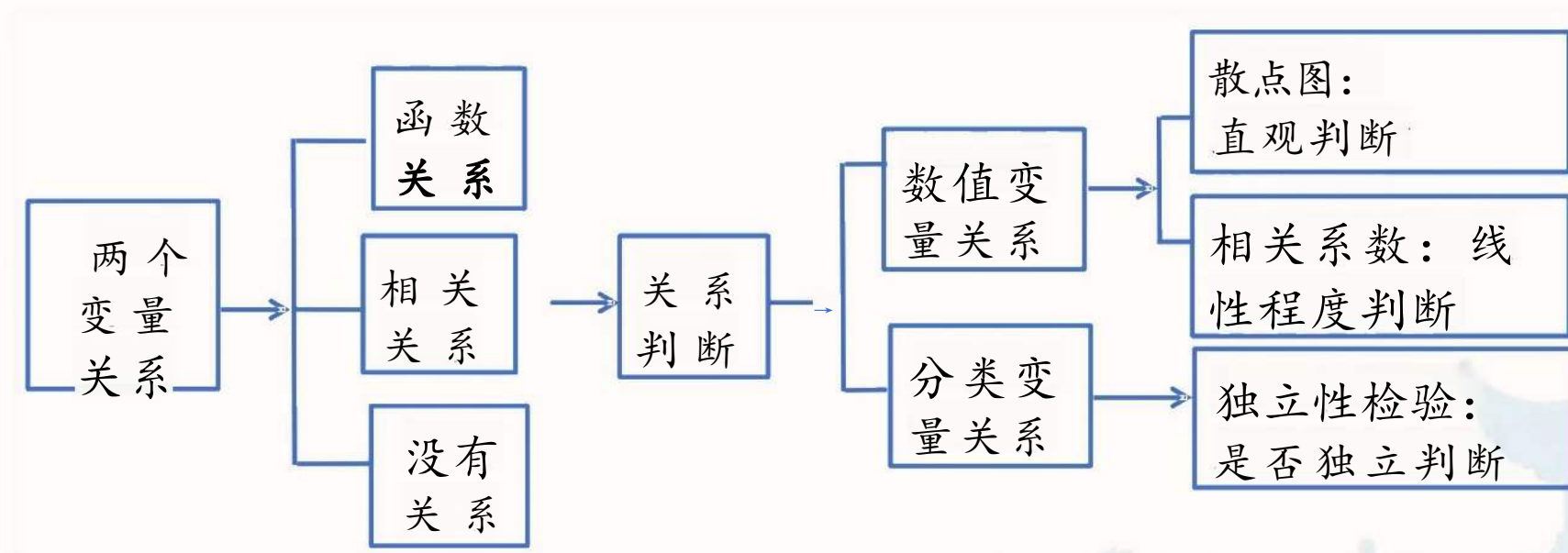


相关关系是指两个变量有关系，但又没有确切到可由其中的一个去精确地决定另一个的程度。

与函数关系的区别：

- > 函数关系中两个变量间是一种确定性关系，函数值 y 由 x 唯一确定。
- > 相关关系中两个变量间是一种不确定性关系（是一种伴随关系）。

(3) 归纳两个变量之间的关系类型。



问题3: 回顾回归分析形成过程，回答下面的问题：

(1) 借助样本相关系数和回归模型，可以刻画两个变量的非线性相关程度的高低吗？

(2) 一元线性回归模型中，模型参数 a 和 b 的统计意义是什么？

(3) 最小二乘原理是什么？你能说一说它的基本思想吗？

(4) 你能画出回归分析流程图吗？

(1) 借助样本相关系数和回归模型，可以刻画两个变量的非线性相关程度的高低吗？

样本相关系数 r 的取值范围 $[-1, 1]$ ， $|r|$ 刻画了样本点集中于某条直线的程度。

当 $r > 0$ 时，称成对样本数据正相关；

当 $r < 0$ 时，称成对样本数据负相关。

当 $|r|$ 越接近1时，成对样本数据的线性相关程度越强；当 $|r|$ 越接近0时，成对样本数据的线性相关程度越弱。

利用样本相关系数 r 判定相关程度的高低。样本相关系数和回归模型，不能刻画两个变量的非线性相关程度的高低。

(2) 一元线性回归模型中，模型参数 a 和 b 的统计意义是什么？

$$\begin{cases} Y = bx + a + e, \\ E(e) = 0, D(e) = \sigma^2. \end{cases}$$

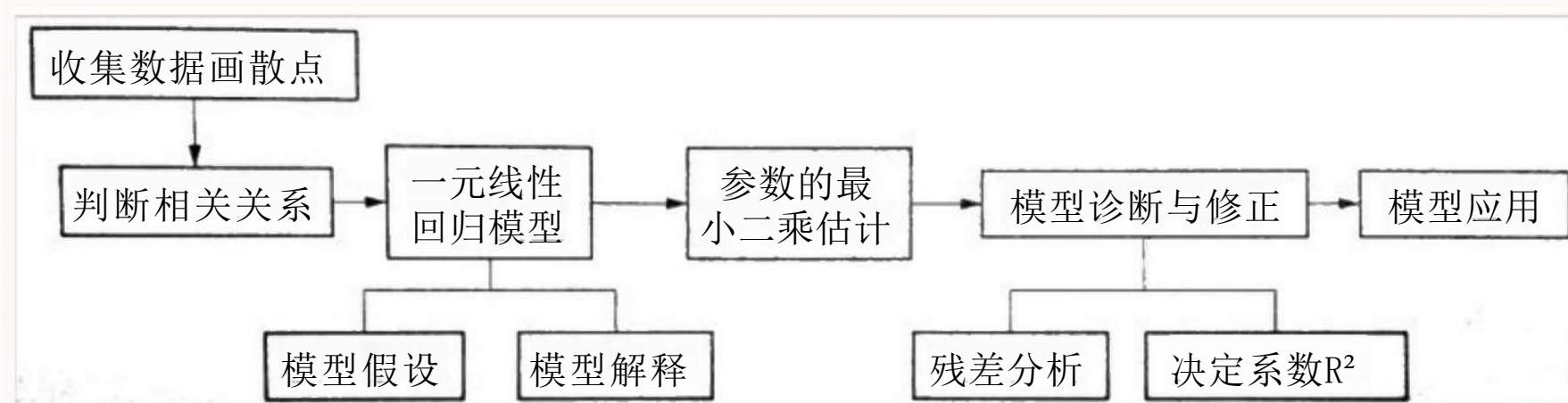
a 为截距参数， b 为斜率参数。它们刻画了变量 Y 与变量 x 的线性关系，参数 a 的意义是指不受 x 影响的部分。参数 b 的含义可以解释为自变量 x 对响应变量 Y 的均值的影响，变量 x 每增加1个单位，响应变量 Y 的均值将增加 b 个单位。

(3) 最小二乘原理是什么？你能说一说它的基本思想吗？

最小二乘原理是根据散点到直线的竖直距离平方之和 $Q(a,b) = \sum_{i=1}^n [y_i - (bx_i + a)]^2$ 最小时，得到一元线性回归模型中参数 a 和 b 的最小二乘估计 \hat{a} 和 \hat{b} ，从而得到最小经验回归方程 $\hat{y} = bx + \hat{a}$ 。

最小二乘法形成过程体现了函数的思想，运用它求解参数的估计值体现的基本思想是样本估计总体的思想，因此得出的经验回归方程未必准确，还要通过残差分析、决定系数等进行拟合效果的评价和改进。

(4) 你能画出回归分析流程图吗?



问题4： 回顾独立性检验学习过程，回答下面的问题：

- (1) 举例说明用条件概率的语言刻画两个分类变量关联性的关键点是什么？
- (2) 分类变量与数值变量有何本质不同？
- (3) 独立性检验的基本思想是什么？它和反证法有何异同？
- (4) 若依据 $\alpha = 0.05$ 的独立性检验，结论是两个变量之间有关联，如何解释这个结论？如果是两个变量之间没有关联，又该如何解释这个结论？
- (5) 你能画出独立性检验的流程图吗？

(1) 举例说明用条件概率的语言刻画两个分类变量关联性的关键点是什么？

教科书第126页例1在分析两校学生中数学成绩优秀之间是否存在差异中，用条件概率的语言刻画两个分类变量关联性的关键是：根据概率运算关系得到零假设 H_0 ： $P(Y=1|X=0)=P(Y=1|X=1)$ 的四个等价形式，如：

$P(X=0, Y=1)=P(X=0)P(Y=1)$ 等，

即事件 $\{Y=0\}$ 、 $\{Y=1\}$ 、 $\{X=0\}$ 、 $\{X=1\}$ 相互独立。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/726114155223010141>