

第 03 讲 8.3 列联表与独立性检验

(8.3.1 分类变量与列联表+8.3.2 独立性检验)

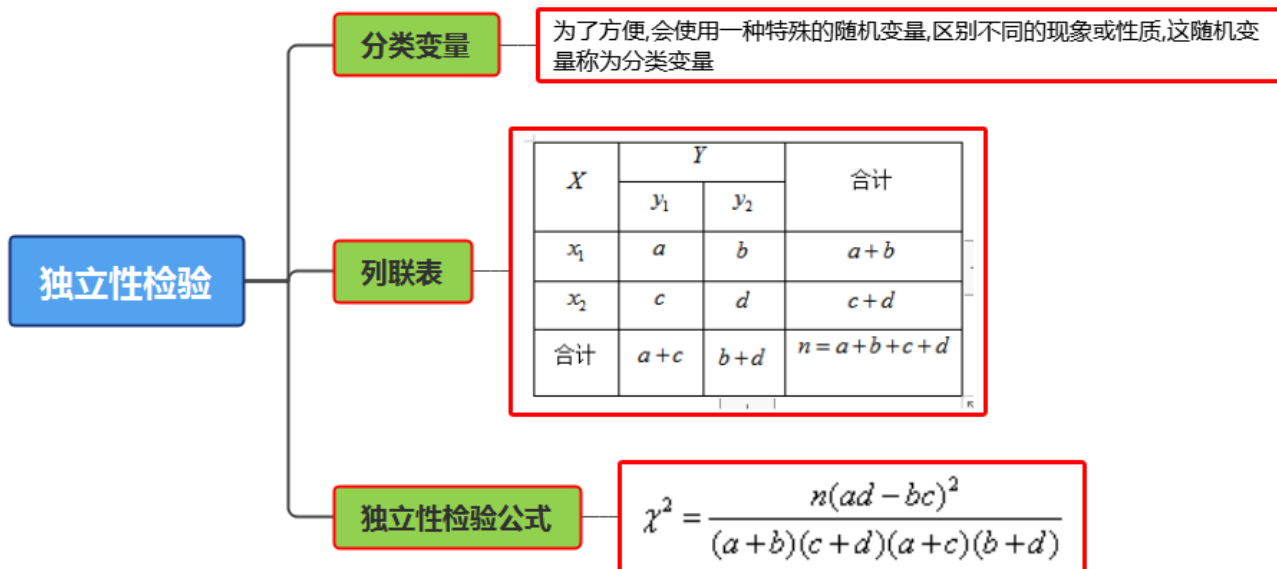
01

学习目标

课程标准	学习目标
①了解分类变量与数值变量的区别,了解回归与相关的区别。 ②通过实例,理解通过比较相关比率,利用2×2列联表或等高图可以初步检验两个随机变量的独立性.理解通过比较相关比率判断随机变量独立性得到的结果有可能会犯错误。 ③理解通过比较相关比率判断随机变量独立性得到的结果有可能会犯错误。	本节课要求会通过比较相关比率,判断两个随机变量的独立性.会对简单的数据分析案例进行初步独立性分析.恰当构造卡方统计量及利用小概率事件原理实现对两个分类变量的是否独立的科学检验.能解决简单的与独立性检验相关的实际问题

02

思维导图



03

知识清单

知识点 1: 分类变量与列联表

(1) 分类变量

为了方便,会使用一种特殊的随机变量,区别不同的现象或性质,这随机变量称为分类变量.

(2) 2×2 列联表

① 2×2 列联表给出了两个分类变量数据的交叉分类频数.

② 定义一对分类变量 X 和 Y , 我们整理数据如下表所示:

X	Y		合计
	y_1	y_2	
x_1	a	b	$a+b$
x_2	c	d	$c+d$
合计	$a+c$	$b+d$	$n=a+b+c+d$

知识点 2: 独立性检验

(1) 独立性检验定义:

利用 χ^2 的取值推断分类变量 X 和 Y 是否独立的方法称为 χ^2 独立性检验, 读作“卡方独立性检验”. 简称独立性检验.

(2) 独立性检验公式:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \quad \text{其中 } n = a+b+c+d \quad (\text{注意使用公式时分子的平方不要忽略了})$$

【即学即练 1】 (2024 上·江西九江·高二统考期末) 某校随机调查了 100 名高中生是否喜欢篮球, 按照男女区分得到列联表, 经计算得 $\chi^2 = 8.133$. 根据独立性检验的相关知识, 对照下表, 可以认为有 () 把握喜欢篮球与性别有关.

p	0.05	0.01	0.005	0.001
χ^2	3.841	6.635	7.879	10.828

A. 95%

B. 99.5%

C. 99%

D. 99.9%

【答案】 B

【详解】 $Q \chi^2 = 8.133 \in (7.879, 10.828)$,

$\therefore$ 有 99.5% 把握认为与性别有关,

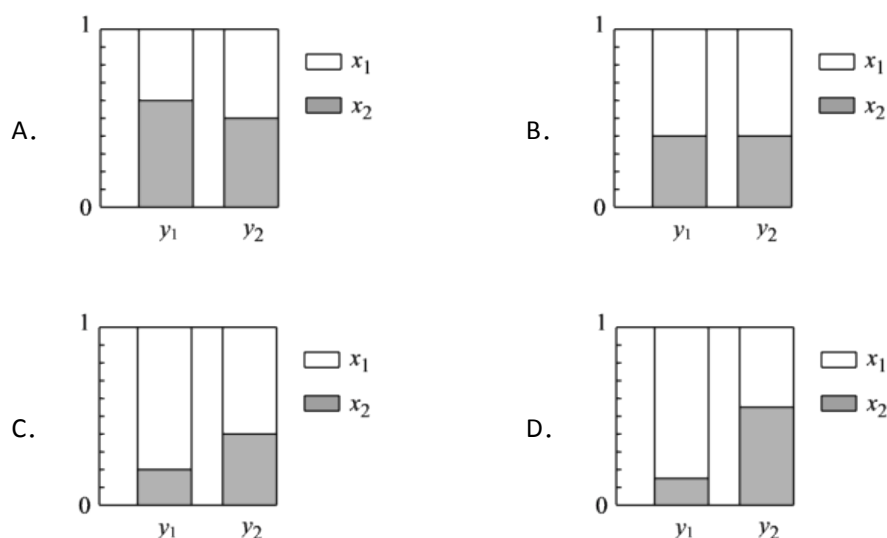
故选: B.

04

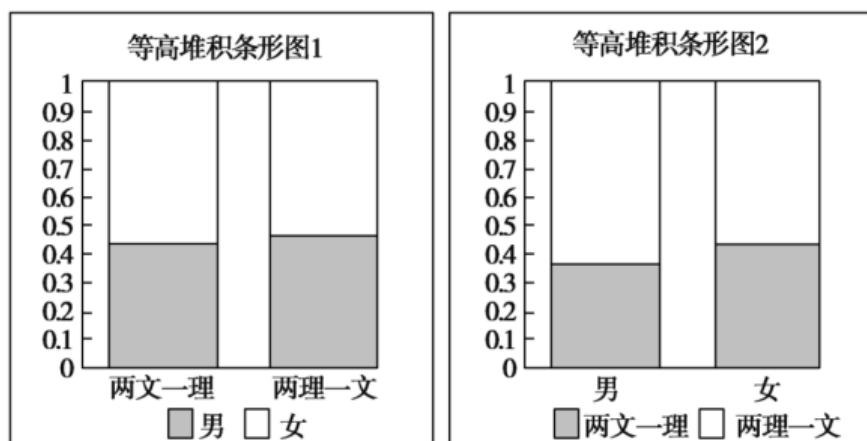
题型精讲

题型 01 通过等高堆积条形图判断两个分类变量是否存在差异

【典例 1】（2014·吉林长春·统考三模）观察下列各图，其中两个分类变量 x, y 之间关系最强的是（ ）



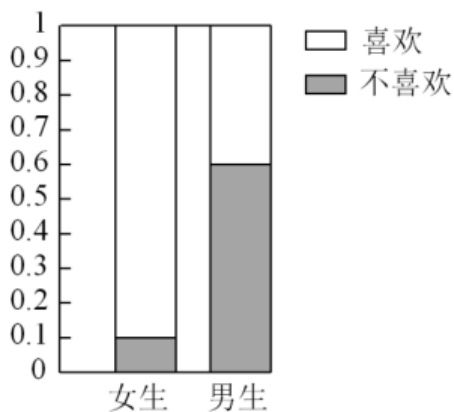
【典例 2】（2017·广东佛山·统考一模）现行普通高中学生在高一时面临着选科的问题，学校抽取了部分男、女学生意愿的一份样本，制作出如下两个等高堆积条形图：



根据这两幅图中的信息，下列哪个统计结论是不正确的（ ）

- A. 样本中的女生数量多于男生数量
- B. 样本中有两理一文意愿的学生数量多于有两文一理意愿的学生数量
- C. 样本中的男生偏爱两理一文
- D. 样本中的女生偏爱两文一理

【典例 3】（2012 下·山东青岛·高二山东省青岛第一中学校考期中）为了了解某高校学生喜欢使用手机支付是否与性别有关，抽取了部分学生作为样本，统计后作出如图所示的等高条形图，则下列说法正确的是（ ）

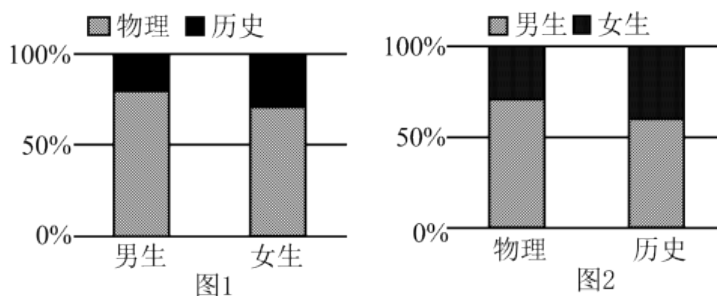


- A. 喜欢使用手机支付与性别无关
- B. 样本中男生喜欢使用手机支付的约 60%
- C. 样本中女生喜欢使用手机支付的人数比男生多
- D. 女生比男生喜欢使用手机支付的可能性大些

【变式 1】（2021·高二课时练习）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 $\frac{d}{c+d}$
- B. $\frac{c}{a+b}$ 与 $\frac{a}{c+d}$
- C. $\frac{a}{a+b}$ 与 $\frac{c}{c+d}$
- D. $\frac{a}{a+b}$ 与 $\frac{c}{b+c}$

【变式 2】（2023·四川达州·统考一模）四川省将从 2022 年秋季入学的高一年级学生开始实行高考综合改革，高考采用“3+1+2”模式，其中“1”为首选科目，即物理与历史二选一.某校为了解学生的首选意愿，对部分高一学生进行了抽样调查，制作出如下两个等高条形图，根据条形图信息，下列结论正确的是（ ）



- A. 样本中选择物理意愿的男生人数少于选择历史意愿的女生人数
- B. 样本中女生选择历史意愿的人数多于男生选择历史意愿的人数
- C. 样本中选择物理学科的人数较多
- D. 样本中男生人数少于女生人数

题型 02 独立性检验的概念及辨析

【典例 1】（2024 上·全国·高三专题练习）根据分类变量 x 与 y 的观测数据，计算得到 $\chi^2 = 3.974$. 依据 $\alpha = 0.05$ 的独立性检验，结论为（ ）

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

- A. 变量 x 与 y 不独立，这个结论犯错误的概率不超过 0.01
 B. 变量 x 与 y 不独立，这个结论犯错误的概率不超过 0.05
 C. 变量 x 与 y 独立，这个结论犯错误的概率不超过 0.05
 D. 变量 x 与 y 独立，这个结论犯错误的概率不超过 0.01

【典例 2】（2024·全国·高三专题练习）针对时下的“短视频热”，某校团委对学生性别和喜欢短视频是否有关联进行了一次调查，其中被调查的男生、女生人数均为 $5m$ ($m \in \mathbf{N}^*$) 人，男生中喜欢短视频的人数占男生人数的 $\frac{4}{5}$ ，女生中喜欢短视频的人数占女生人数的 $\frac{3}{5}$. 零假设为 H_0 : 喜欢短视频和性别相互独立. 若依据 $\alpha = 0.05$ 的独立性检验认为喜欢短视频和性别不独立，则 m 的最小值为（ ）

附: $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$, 附表:

α	0.05	0.01
x_α	3.841	6.635

- A. 7 B. 8 C. 9 D. 10

【变式 1】（2024 上·全国·高三专题练习）北京冬奥会的举办掀起了一阵冰雪运动的热潮. 某高校在本校学生中对“喜欢滑冰是否与性别有关”做了一次调查，参与调查的学生中，男生人数是女生人数的 3 倍，有 $\frac{2}{3}$ 的男生喜欢滑冰，有 $\frac{1}{3}$ 的女生喜欢滑冰. 若根据独立性检验的方法，有 95% 的把握认为是否喜欢滑冰和性别有关，则参与调查的男生人数可能为（ ）

参考公式: $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$, 其中 $n = a+b+c+d$.

参考数据:

$P(\chi^2 \geq k_0)$	0.10	0.05	0.025	0.010
k_0	2.706	3.841	5.024	6.635

- A. 12 B. 18 C. 36 D. 48

【变式 2】（多选）

(2024 上·全国·高三期末)为考察一种新型药物预防疾病的效果,某科研小组进行动物实验,收集整理数据后将所得结果填入相应的 2×2 列联表中,由列联表中的数据计算得 $\chi^2 \approx 9.616$.参照附表,下列结论正确的是 ()

附表:

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

- A. 根据小概率值 $\alpha = 0.001$ 的独立性检验,分析认为“药物有效”
- B. 根据小概率值 $\alpha = 0.001$ 的独立性检验,分析认为“药物无效”
- C. 根据小概率值 $\alpha = 0.005$ 的独立性检验,分析认为“药物有效”
- D. 根据小概率值 $\alpha = 0.005$ 的独立性检验,分析认为“药物无效”

题型 03 卡方的计算

【典例 1】(2024 上·湖南长沙·高三湖南师大附中校考阶段练习)为检验预防某种疾病的 A, B 两种疫苗的免疫效果,随机抽取接种 A, B 疫苗的志愿者各 100 名,化验其血液中某项医学指标(该医学指标范围为 $[0, 100]$),统计如下:

该项医学指标	$[0, 25)$	$[25, 50)$	$[50, 75)$	$[75, 100]$
接种 A 疫苗人数	$10 + m$	10	50	$30 - m$
接种 B 疫苗人数	$10 - m$	30	40	$20 + m$

个别数据模糊不清,用含字母 $m(m \in \mathbf{N})$ 的代数式表示.

- (1)为检验该项医学指标在 $[0, 50)$ 内的是否需要接种加强针,先从医学指标在 $[25, 50)$ 的志愿者中,按接种 A, B 疫苗分层抽取 8 人,再次抽血化验进行判断.从这 8 人中随机抽取 4 人调研医学指标低的原因,记这 4 人中接种 B 疫苗的人数为 X ,求 X 的分布列与数学期望;
- (2)根据 (1) 化验研判结果,医学认为该项医学指标低于 50,产生抗体较弱,需接种加强针,该项医学指标不低于 50,产生抗体较强,不需接种加强针.请先完成下面的 2×2 列联表,若根据小概率 $\alpha = 0.025$ 的独立性检验,认为接种 A, B 疫苗与志愿者产生抗体的强弱有关联,求 m 的最大值.

疫苗	抗体	合计
----	----	----

	抗体弱	抗体强	
A 疫苗			
B 疫苗			
合计			

附： $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ ，其中 $n = a + b + c + d$.

α	0.25	0.025	0.005
x_α	1.323	5.024	7.879

【典例 2】（2024·云南楚雄·云南省楚雄彝族自治州民族中学模拟预测）全国“村 BA”篮球赛点燃了全民的运动激情，深受广大球迷的喜爱.每支球队都有一个或几个主力队员，现有一支“村 BA”球队，其中甲球员是其主力队员，经统计该球队在某个赛季的所有比赛中，甲球员是否上场时该球队的胜负情况如表.

甲球员是否上场	球队的胜负情况		合计
	胜	负	
上场	40		45
未上场		3	
合计	42		

- (1)完成 2×2 列联表，并判断依据小概率值 $\alpha = 0.01$ 的独立性检验，能否认为球队的胜负与甲球员是否上场有关；
- (2)由于队员的不同，甲球员主打的位置会进行调整，根据以往的数据统计，甲球员上场时，打前锋、中锋、后卫的概率分别为 0.3，0.5，0.2，相应球队赢球的概率分别为 0.7，0.8，0.6.
- (i) 当甲球员上场参加比赛时，求球队赢球的概率；

(ii) 当甲球员上场参加比赛时，在球队赢了某场比赛的条件下，求甲球员打中锋的概率. (精确到 0.01)

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

α	0.15	0.10	0.05	0.025	0.010	0.001
χ_α	2.072	2.706	3.841	5.024	6.635	10.828

【典例 3】（2024 上·广西北海·高二统考期末）在对人们休闲方式的一次调查中，共调查了 120 人，其中女性 70 人，男性 50 人. 女性中有 40 人主要的休闲方式是看电视，另外 30 人主要的休闲方式是运动；男性中有 20 人主要的休闲方式是看电视，另外 30 人主要的休闲方式是运动.

(1)根据以上数据完成以下列联表；

休闲方式 性别	看电视	运动	合计
女			
男			
合计			

(2)能否有 95%把握认为性别与休闲方式有关系？

附： $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$, 其中 $n = a+b+c+d$.

α	0.15	0.1	0.05	0.025	0.01	0.001
χ_α	2.072	2.706	3.841	5.024	6.635	10.828

【变式 1】

（2024 上·四川成都·高三成都七中校考期末）在某病毒疫苗的研发过程中，需要利用基因编辑小鼠进行动物实验.现随机抽取 100 只基因编辑小鼠对该病毒疫苗进行实验，得到如下 2×2 列联表（部分数据缺失）：

		被某病毒感染		未被某病毒感染	合计
注射疫苗		10			50
未注射疫苗				30	50
合计		30			100
α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828

计算可知，根据小概率值 $\alpha = \underline{\hspace{1cm}}$ 的独立性检验，分析“给基因编辑小鼠注射该种疫苗能起到预防该病毒感染的效果”（ ）

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

- A. 0.001 B. 0.05 C. 0.01 D. 0.005

【变式 2】（2024 上·江西鹰潭·高二统考期末）积化和差的重要应用在于求解傅里叶级数. 为了解学生掌握该组公式的情况，在高一、高三两个年级中随机抽取了 100 名学生进行考查，其中高三年级的学生占 $\frac{3}{5}$ ，其他相关数据如下表：

	合格	不合格	总计
高三年级学生	54		
高一年级学生		16	
总计			100

(1)请完成 2×2 列联表，依据小概率值 $\alpha = 0.001$ 的独立性检验，分析“对公式的掌握情况”与“学生所在年级”是否有关？

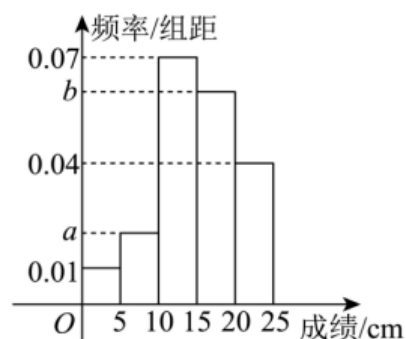
(2)以频率估计概率，从该校高一年级学生中抽取 3 名学生，记合格的人数为 X ，求 X 的分布列和数学期望.

附表及公式：

α	0.1	0.05	0.01	0.005	0.001
χ^2	2.706	3.841	6.635	7.879	10.828

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

【变式 3】（2024 上·福建泉州·高三统考期末）教育部印发的《国家学生体质健康标准》，要求学校每学年开展全校学生的体质健康测试工作.某中学为提高学生的体质健康水平，组织了“坐位体前屈”专项训练.现随机抽取高一男生和高二男生共 60 人进行“坐位体前屈”专项测试.高一男生成绩的频率分布直方图如图所示，其中成绩在 $[5,10)$ 的男生有 4 人.



高二男生成绩（单位：cm）如下：

10.2	12.8	6.4	6.6	14.3	8.3	16.8	15.9	9.7	17.5
18.6	18.3	19.4	23.0	19.7	20.5	24.9	20.5	25.1	17.5

(1)估计高一男生成绩的平均数和高二男生成绩的第 40 百分位数；

(2)《国家学生体质健康标准》规定，高一男生“坐位体前屈”成绩良好等级线为15cm，高二男生为16.1cm.已知该校高一年男生有 600 人，高二年男生有 500 人，完成下列 2×2 列联表，依据小概率值 $\alpha = 0.005$ 的独立性检验，能否认为该校男生“坐位体前屈”成绩优良等级与年级有关？

等级 年级	良好及以上	良好以下	合计
高一			
高二			
合计			

附： $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ ，其中 $n = a+b+c+d$.

α	0.05	0.010	0.005	0.001
χ_α	3.841	6.635	7.879	10.828

题型 04 独立性检验的基本思想

【典例 1】（2024·四川绵阳·绵阳南山中学实验学校校考模拟预测）在某病毒疫苗的研发过程中，需要利用基因编辑小鼠进行动物实验。现随机抽取 100 只基因编辑小鼠对该病毒疫苗进行实验，得到如下 2×2 列联表（部分数据缺失）：

	被某病毒感染	未被某病毒感染	合计
注射疫苗	10		50
未注射疫苗		30	50
合计	30		100

计算可知，根据小概率值 $\alpha = \underline{\hspace{2cm}}$ 的独立性检验，分析“给基因编辑小鼠注射该种疫苗能起到预防该病毒感染的效果”（ ）

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

α	0.1	0.05	0.01	0.005	0.001
$\chi\alpha$	2.706	3.841	6.635	7.879	10.828

A. 0.001

B. 0.05

C. 0.01

D. 0.005

【典例 2】（2024 上·全国·高三专题练习）为了有针对性地提高学生体育锻炼的积极性，某中学需要了解性别因素是否对学生体育锻炼的经常性有影响，为此随机抽取了 40 名学生，按照性别和体育锻炼情况整理得到如下的列联表：

性别	锻炼		合计
	不经常	经常	
女生	5	10	15
男生	5	20	25
合计	10	30	40

(1) 根据上表数据，依据小概率值 $\alpha = 0.01$ 的独立性检验，能否认为性别因素会影响学生体育锻炼的经常性？

(2) 如果将表中的数据都扩大为原来的 n ($n \in \mathbf{N}^*$) 倍，在相同的检验标准下，得到与 (1) 中不一样的结论。

(i) 求 n 的最小值；

(ii) 如果抽样方式不变，你认为 (1) 和 (2) 的结论哪个更可靠，并说明理由。

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/006041234135010134>