

第四章 概率与统计

4.3 统计模型

4.3.2 独立性检验

知识梳理

一. 独立性检验

1.如果随机事件 A 与 B 的样本数据的 2×2 列联表如下:

记 $n=a+b+c+d$.统计学中有一个非常有用的统计量 χ^2 (读作“卡方”).

	A		总计
B	a	b	a+b
	c	d	c+d
总计	a+c	b+d	a+b+c+d

它的表达式是 $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$.

2.任意给定一个 α (称为显著性水平, 通常取为 0.05, 0.01 等), 可以找到满足条件 $P(\chi^2 \geq k) = \alpha$ 的数 k (称为显著性水平 α 对应的分位数). χ^2 是一个随机变量, 其分布能够求出, 上面的概率是可以计算的. 因此, 如果根据样本数据算出 χ^2 的值后, 发现 $\chi^2 \geq k$ 成立, 就称在犯错误的概率不超过 α 的前提下, 可以认为 A 与 B 不独立(也称为 A 与 B 有关); 或说有 $1-\alpha$ 的把握认为 A 与 B 有关. 若 $\chi^2 < k$ 成立, 就称不能得到前述结论. 这一过程通常称为独立性检验.

3. A 与 B 独立时, 也称为 A 与 B 无关. 当 $\chi^2 < k$ 成立时, 一般不直接说 A 与 B 无关, 也就是说, 独立性检验通常得到的结果, 或者是有 $1-\alpha$ 的把握认为 A 与 B 有关, 或者没有 $1-\alpha$ 的把握认为 A 与 B 有关.

4.独立性的判断方法

- (1)当 $\chi^2 < 2.706$ 时, 没有充分的证据判定变量 A, B 有关联, 可以认为变量 A, B 是没有关联的;
- (2)当 $\chi^2 \geq 2.706$ 时, 有 90% 的把握判定变量 A, B 有关联;
- (3)当 $\chi^2 \geq 3.841$ 时, 有 95% 的把握判定变量 A, B 有关联;
- (4)当 $\chi^2 \geq 6.635$ 时, 有 99% 的把握判定变量 A, B 有关联.

常见考点

考点一 完善列联表

典例 1. 假设有两个变量 X 和 Y ，他们的取值分别为， x_2 和， y_2 ，其列联表为：

		y_2	总计
	a	21	73
x_2	8	25	33
总计	b	46	106

则表中 a ， b 的值分别是（ ）

- A. 94，96 B. 54，52 C. 52，50 D. 52，60

变式 1-1. 下表是一个 2×2 列联表，则表中 a ， b 的值分别为（ ）

	y_1	y_2	总计
	b	21	e
x_2	c	25	33
总计	a	d	100

- A. 46，54 B. 54，46 C. 52，54 D. 50，52

变式 1-2. 为了解某大学的学生是否爱好体育锻炼，用简单随机抽样方法在校园内调查了 120 位学生，得到如下 2×2 列联表：

	男	女	总计
爱好	a	b	73
不爱好	c	25	
总计	74		

则 $a-b-c$ 等于（ ）

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

变式 1-3. 某村庄对该村内 50 名老年人、年轻人每年是否体检的情况进行了调查，统计数据如表所示：

把握说明两个事件有关，当 $\chi^2 > 6.635$ 时，有 99% 的把握说明两个事件相关，当 $\chi^2 < 3.841$ 时，认为两个事件无关. 在一项调查某种药是否对心脏病有治疗作用时，共调查了 3000 人，经计算的 $\chi^2 = 9.56$ ，根据这一数据分析，认为此药物与心脏病之间（ ）

- A. 有 95% 的把握认为两者相关
- B. 约有 95% 的心脏病患者使用药物有作用
- C. 有 99% 的把握认为两者相关
- D. 约有 99% 的心脏病患者使用药物有作用

变式 2-3. 独立性检验中，假设变量与变量 Y 没有关系，则在假设成立的情况下，估算概率

$P(K^2 \geq 10.828) \approx 0.001$ 表示的意义是（ ）

- A. 在犯错误的概率不超过 0.01 的前提下，认为“变量与变量 Y 有关”
- B. 在犯错误的概率不超过 0.01 的前提下，认为“变量与变量 Y 无关”
- C. 有 99.9% 以上的把握认为“变量与变量 Y 无关”
- D. 有 99.9% 以上的把握认为“变量与变量 Y 有关”

考点三 卡方的计算与应用

典例 3. 为考察某种药物预防疾病的效果，进行动物试验，得到如下 2×2 列联表：

	患病	未患病	总计
服用药	10	40	50
没服用药	20	30	50
总计	30	70	100

附：，其中 $n = a + b + c + d$

$P(K^2 \geq k_0)$	0.05	0.010	0.001
k_0	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 0.1% 的前提下，认为“服药与患病无关”
- B. 在犯错误的概率不超过 0.1% 的前提下，认为“服药与患病有关”
- C. 有 95% 以上的把握认为“服药与患病有关”

D. 有 95%以上的把握认为“服药与患病无关”

变式 3-1. 为了丰富教职工业余文化生活, 某校计划在假期组织 70 名老师外出旅游, 并给出了两种方案(方案一和方案二), 每位老师均选择且只选择一种方案, 其中有 50%的男老师选择方案一, 有 75%的女老师选择方案二, 且选择方案一的老师中女老师占 40%, 则参照附表, 得到的正确结论是 ()

附:

$P(K^2 \geq k_0)$	0.10	0.05	0.025
k_0	2.706	3.841	5.024

, $n = a + b + c + d$.

- A. 在犯错误的概率不超过 2.5%的前提下, 认为“选择方案与性别有关”
- B. 在犯错误的概率不超过 2.5%的前提下, 认为“选择方案与性别无关”
- C. 有 95%以上的把握认为“选择方案与性别有关”
- D. 有 95%以上的把握认为“选择方案与性别无关”

变式 3-2. 某机构为了解某地区中学生在校月消费情况, 随机抽取了 100 名中学生进行调查, 将月消费金额不低于 550 元的学生成为“高消费群”, 调查结果如表所示. 参照公式, 正确的结论是 ()

	高消费群	非高消费群	合计
男	15	35	50
女	10	40	50
合计	25	75	100

参考公式:, 其中 $n = a + b + c + d$.

$P(K^2 \geq k_0)$	0.10	0.05	0.025	0.010	0.005	0.001
k_0	2.706	3.841	5.024	6.635	7.879	10.828

- A. 有 90%以上的把握认为“高消费群与性别有关”
- B. 没有 90%以上的把握认为“高消费群与性别有关”

- C. 在犯错误的概率不超过1%的前提下，认为“高消费群与性别无关”
- D. 在犯错误的概率不超过1%的前提下，认为“高消费群与性别有关”

变式 3-3. 通过随机询问 110 名性别不同的大学生是否爱好体育，

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

得到上表：参照附表，得到的正确结论是（ ）

附：由公式算得： $K^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \approx 7.8$

附表：

$P(k^2 > k_0)$	0.25	0.15	0.10	0.05	0.025	0.010	0.005
k_0	1.323	2.702	2.706	3.841	5.024	6.635	7.879

- A. 有 99% 以上的把握认为“爱好体育运动与性别有关”
- B. 有 99% 以上的把握认为“爱好体育运动与性别无关”
- C. 在犯错误的概率不超过 0.1% 的前提下，认为“爱好体育运动与性别有关”
- D. 在犯错误的概率不超过 0.1% 的前提下，认为“爱好体育运动与性别无关”

考点四 独立性检验解决实际问题

典例 4. 为了调查某校高二学生是否需要学校提供学法指导，用简单随机抽样的方法从该校高二年级调查了 55 名学生，结果如下：

	男	女
需要	20	10
不需要	10	15

）估计该校高二年级学生中，需要学校提供学法指导的学生的比例；（用百分数表示，保留两位有效数字）

（2）能否有 95%的把握认为该校高二年级学生是否需要学校提供学法指导与性别有关？

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

参考数据：

$P(\chi^2 \geq 0)$	0.10	0.05	0.025	0.010	0.005	0.001
x_0	2.706	3.481	5.024	6.635	7.879	10.828

变式 4-1. 网购是现在比较流行的一种购物方式，现随机调查 50 名收入不同的消费者是否喜欢网购，调查结果表明：在喜欢网购的 30 人中有 20 人是低收入的人，在不喜欢网购的 20 人中有 10 人是低收入的人.

（1）试根据以上数据完成下面的列联表；

	喜欢网购	不喜欢网购	总计
低收入的人			
高收入的人			

总计			
----	--	--	--

(2) 判断能否有 90% 的把握认为“是否喜欢网购与个人收入高低有关”?

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

$P(K^2 \geq k_0)$	0.10	0.05	0.025	0.010	0.005	0.001
k_0	2.706	3.841	5.024	6.635	7.879	10.828

变式 4-2. 为了推动智慧课堂的普及和应用, A 市现对全市中小学智慧课堂的应用情况进行抽样调查, 统计数据如下表:

	经常应用	偶尔应用或者不应用	总计
农村学校	40		
城市学校			80
总计	100		160

(1) 补全上面的列联表;

(2) 通过计算判断能否有 99.5%的把握认为智慧课堂的应用与区域有关.

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

$P(K^2 \geq k_0)$	0.500	0.050	0.005
k_0	0.445	3.841	7.879

变式 4-3. 某手机生产企业为了解消费者对某款手机的认同情况, 通过销售部随机抽取 50 名购买该款手机的消费者, 并发出问卷调查(满分 50 分), 该问卷只有 20 份给予回复, 这 20 份的评分如下:

男	47,36,28,48,48,44,50,46,50,37,35,49
女	38,37,50,36,38,45,29,39

(1) 完成下面的茎叶图;

(2) 若大于 40 分为“满意”, 否则为“不满意”, 完成上面的 2×2 列联表, 并判断是否有 95%

的把握认为消费者对该款手机的“满意度”与性别有关.

	满意	不满意	合计
男			
女			
合计			

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

参考数据:

$P(K^2 \geq k_0)$	0.050	0.010	0.001
k_0	3.841	6.635	10.828

巩固练习

练习一 完善列联表

1. 如表是 2×2 列联表, 则表中的 a 、 b 的值分别为 ()

		y	合计
	a	8	35
	11	34	45
合计	b	42	80

- A. 27、38 B. 28、38 C. 27、37 D. 28、37

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/977123030010010025>