

8. 3.2 独立性检验

课标要求	素养要求
了解随机变量 χ^2 的意义，通过对典型案例的分析，了解独立性检验的基本思想和方法.	通过运用列联表进行独立性检验，提升数学抽象及数据分析素养.

新知探究

► 情境引入

山东省教育厅大力推行素质教育，增加了高中生的课外活动时间，某校调查了学生的课外活动方式，结果整理成下表：

	体育	文娱	合计
男生	210	230	440
女生	60	290	350
合计	270	520	790



问题 如何判定“喜欢体育还是文娱与性别是否有联系”？

提示 可通过表格与图形进行直观分析，也可通过统计分析定量判断。

► 知识梳理

1. 临界值

χ^2 统计量也可以用来作相关性的度量. χ^2 越小说明变量之间越独立, χ^2 越大说明变量之间越相关

$$\chi^2 = \frac{n (ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}. \text{ 忽略 } \chi^2 \text{ 的实际分布与该近似分布的误差后, 对}$$

于任何小概率值 α , 可以找到相应的正实数 x_α , 使得 $P(\chi^2 \geq x_\alpha) = \alpha$ 成立. 我们称 x_α 为 α 的临界值, 这个临界值就可作为判断 χ^2 大小的标准.

2. 独立性检验

基于小概率值 α 的检验规则是：

当 $\chi^2 \geq x_\alpha$ 时，我们就推断 H_0 不成立，即认为 X 和 Y 不独立，该推断犯错误的概率不超过 α ；

当 $\chi^2 < x_\alpha$ 时，我们没有充分证据推断 H_0 不成立，可以认为 X 和 Y 独立。

这种利用 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(test of independence).

下表给出了 χ^2 独立性检验中几个常用的小概率值和相应的临界值

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

3.应用独立性检验解决实际问题的大致步骤

- (1)提出零假设 H_0 : X 和 Y 相互独立, 并给出在问题中的解释;
- (2)根据抽样数据整理出 2×2 列联表, 计算 χ^2 的值, 并与临界值 x_α 比较;
- (3)根据检验规则得出推断结论;
- (4)在 X 和 Y 不独立的情况下, 根据需要, 通过比较相应的频率, 分析 X 和 Y 间的影响规律.

拓展深化

[微判断]

1. 概率值 α 越小，临界值 x_α 越大. (✓)
2. 独立性检验的思想类似于反证法. (✓)
3. 独立性检验的结论是有多大的把握认为两个分类变量有关系. (✓)

[微训练]

1. 如果根据小概率 $\alpha=0.01$ 的 χ^2 检测试验, 认为 H_0 成立, 那么具体算出的数据满足()

附表:

α	0.05	0.025	0.010	0.005	0.001
x_α	3.841	5.024	6.635	7.879	10.828

A. $\chi^2 > 6.635$

B. $\chi^2 > 5.024$

C. $\chi^2 > 7.879$

D. $\chi^2 > 3.841$

答案 A

2. 某校为了研究“学生的性别”和“对待某一活动的态度”是否有关，运用 2×2 列联表进行独立性检验，经计算 $\chi^2 = 7.069$ ，则认为“学生性别与支持某项活动有关系”的犯错误的概率不超过()

A. 0.1%

B. 1%

C. 99%

D. 99.9%

解析 $\because \chi^2 = 7.069 > 6.635 = x_{0.01}$, \therefore 认为“学生性别与支持某项活动有关系”的犯错误的概率不超过1%.

答案 B

[微思考]

1. 有人说：“在犯错误的概率不超过0.01的前提下认为吸烟和患肺病有关，是指每100个吸烟者中就会有99个患肺病的。”你认为这种观点正确吗？为什么？

提示 观点不正确。犯错误的概率不超过0.01说明的是吸烟与患肺病有关的程度，不是患肺病的百分数。

2. 应用独立性检验的基本思想对两个变量间的关系作出的推断一定是正确的吗？

提示 不一定。所有的推断只代表一种可能性，不代表具体情况。

题型一 有关“相关的检验”

【例1】 某校对学生课外活动进行调查，结果整理成下表，用你所学过的知识进行分析，能否在犯错误的概率不超过0.005的前提下，认为“喜欢体育还是文娱与性别有关系”？

	体育	文娱	合计
男生	21	23	44
女生	6	29	35
合计	27	52	79

解 零假设为 H_0 : 喜欢体育还是喜欢文娱与性别没有关系

$$\because a=21, b=23, c=6, d=29, n=79,$$

$$\begin{aligned}\therefore \chi^2 &= \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \\ &= \frac{79 \times (21 \times 29 - 23 \times 6)^2}{44 \times 35 \times 27 \times 52} \approx 8.106 > 7.879 = \chi_{0.005}.\end{aligned}$$

根据小概率值 $\alpha=0.005$ 的 χ^2 独立性检验, 我们推断 H_0 不成立, 即认为喜欢体育还是喜欢文娱与性别有关系, 此推断犯错误的概率不大于0.005.

规律方法 独立性检验的具体做法

①根据实际问题的需要确定允许推断“两个分类变量有关系”犯错误的概率的上界 α ，然后查表确定临界值 x_α 。

②利用公式 $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 计算 χ^2 。

③如果 $\chi^2 > x_\alpha$ ，则“ X 与 Y 有关系”这种推断犯错误的概率不超过 α ；否则，就认为在犯错误的概率不超过 α 的前提下不能推断“ X 与 Y 有关系”，或者在样本数据中没有发现足够的证据支持结论“ X 与 Y 有关系”。

【训练1】 打鼾不仅影响别人休息，而且可能与患某种疾病有关。下表是一次调查所得的数据：

	患心脏病	未患心脏病	合计
每一晚都打鼾	30	224	254
不打鼾	24	1 355	1 379
合计	54	1 579	1 633

根据独立性检验，能否在犯错误的概率不超过0.001的前提下认为每一晚都打鼾与患心脏病有关系？

解 零假设为 H_0 : 打鼾与患心脏病无关系

由列联表中的数据, 得

$$\chi^2 = \frac{1633 \times (30 \times 1355 - 224 \times 24)^2}{254 \times 1379 \times 54 \times 1579}$$

$$\approx 68.033 > 10.828 = \chi_{0.001}.$$

根据小概率值 $\alpha=0.001$ 的 χ^2 独立性检验, 我们推断 H_0 不成立, 即认为打鼾与患心脏病有关系, 此推断犯错误的概率不大于0.001.

题型二 有关“无关的检验”

【例2】 为了探究学生选报文、理科是否与对外语的兴趣有关，某同学调查了361名高二在校学生，调查结果如下：理科生对外语有兴趣的有138人，无兴趣的有98人，文科生对外语有兴趣的有73人，无兴趣的有52人。试分析学生选报文、理科与对外语的兴趣是否有关？

解 零假设为 H_0 ：选报文、理科与对外语的兴趣无关。

列出 2×2 列联表

	理	文	合计
有兴趣	138	73	211
无兴趣	98	52	150
合计	236	125	361

代入公式得 χ^2 的观测值

$$\chi^2 = \frac{361 \times (138 \times 52 - 73 \times 98)^2}{236 \times 125 \times 211 \times 150} \approx 1.871 \times 10^{-4}.$$

$$\because 1.871 \times 10^{-4} < 2.706 = x_{0.1},$$

根据小概率值 $\alpha=0.1$ 的 χ^2 独立性检验，没有充分证据推断 H_0 不成立，即选报文、理科与对外语的兴趣无关.

规律方法 独立性检验的关注点

在 2×2 列联表中，如果两个分类变量没有关系，则应满足 $ad - bc \approx 0$ ，因此 $|ad - bc|$ 越小，关系越弱； $|ad - bc|$ 越大，关系越强。

【训练2】 某教育机构为了研究成年人具有大学专科以上学历(包括大学专科)和对待教育改革态度的关系，随机抽取了392名成年人进行调查，所得数据如下表所示：

	积极支持教 育改革	不太赞成教育 改革	合计
大学专科以上学历	39	157	196
大学专科以下学历	29	167	196
合计	68	324	392

对于教育机构的研究项目，根据上述数据能得出什么结论？

解 零假设为 H_0 : 成年人具有大学专科以上学历(包括大学专科)和对待教育改革态度无关.

根据表中数据, 计算得

$$\chi^2 = \frac{392 \times (39 \times 167 - 157 \times 29)^2}{196 \times 196 \times 68 \times 324} \approx 1.78.$$

因为 $1.78 < 2.706 = \chi_{0.1}$,

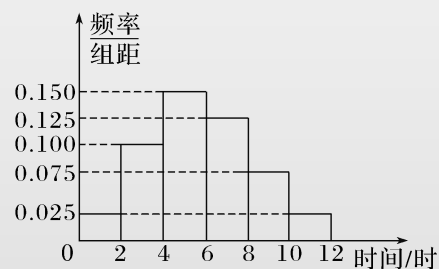
根据小概率值 $\alpha=0.1$ 的 χ^2 独立性检验, 没有充分证据推断 H_0 不成立, 所以我们没有理由说成年人具有大学专科以上学历(包括大学专科)和对待教育改革态度有关.

题型三 独立性检验的综合应用

【例3】 某高校共有学生15 000人，其中男生10 500人，女生4 500人．为调查该校学生每周平均体育运动时间的情况，采用分层随机抽样的方法，收集300位学生每周平均体育运动时间(单位：时)的样本数据．

(1)应收集多少位女生的样本数据？

(2)根据这300个样本数据，得到学生每周平均体育运动时间的频率分布直方图(如图)，其中样本数据的分组区间为 $[0, 2]$ ， $(2, 4]$ ， $(4, 6]$ ， $(6, 8]$ ， $(8, 10]$ ， $(10, 12]$ ．估计该校学生每周平均体育运动时间超过4小时的概率．



(3)在样本数据中，有60位女生的每周平均体育运动时间超过4小时，请完成每周平均体育运动时间与性别的列联表，并判断是否认为“该校学生的每周平均体育运动时间与性别有关”。

附：

α	0.100	0.050	0.010	0.005
x_α	2.706	3.841	6.635	7.879

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

解 (1)由分层随机抽样可得 $300 \times \frac{4\ 500}{15\ 000} = 90$, 所以应收集 90 位女生的样本数据.

(2)由频率分布直方图得学生每周平均体育运动时间超过4小时的频率为 $1 - 2 \times (0.100 + 0.025) = 0.75$, 所以该校学生每周平均体育运动时间超过4小时的概率的估计值为0.75.

(3)由(2)知，300位学生中有 $300 \times 0.75 = 225$ (人)的每周平均体育运动时间超过4小时，75人的每周平均体育运动时间不超过4小时。样本数据中有210份是关于男生的，90份是关于女生的，可得每周平均体育运动时间与性别的列联表如下：

	男生	女生	合计
每周平均体育运动时间不超过4小时	45	30	75
每周平均体育运动时间超过4小时	165	60	225
合计	210	90	300

零假设为 H_0 ：该校学生的每周平均体育运动时间与性别无关。

结合列联表可算得

$$\chi^2 = \frac{300 \times (45 \times 60 - 30 \times 165)^2}{75 \times 225 \times 210 \times 90} \approx 4.762 > 3.841 = \chi_{0.05}.$$

根据小概率值 $\alpha=0.1$ 的 χ^2 独立性检验，我们推断 H_0 不成立，即认为“该校学生的每周平均体育运动时间与性别有关”，此推断犯错误的概率不大于0.05。

规律方法 (1) 解答此类题目的关键在于正确利用 $\chi^2 =$

$\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 计算 χ^2 的值, 再用它与临界值 x_α 的大小作比较来

判断假设检验是否成立, 从而使问题得到解决.

(2) 此类题目规律性强, 解题比较格式化, 填表计算分析比较即可, 要熟悉其计算流程, 不难理解掌握.

【训练3】 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、总分成绩优秀的人数如下表所示，能否在犯错误的概率不超过0.001的前提下认为数学成绩优秀与物理、化学、总分成绩优秀有关系？

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

注：该年级在此次考试中数学成绩优秀的有 360 人，非优秀的有 880 人。

解 零假设为 H_0 : 数学成绩优秀与物理、化学、总分成绩优秀都无关系.

列出数学成绩与物理成绩的 2×2 列联表如下:

	物理优秀	物理非优秀	合计
数学优秀	228	132	360
数学非优秀	143	737	880
合计	371	869	1 240

将表中数据代入公式, 得

$$\chi_1^2 = \frac{1\,240 \times (228 \times 737 - 132 \times 143)^2}{360 \times 880 \times 371 \times 869} \approx 270.1 > 10.828 = \chi_{0.001}.$$

列出数学成绩与化学成绩的 2×2 列联表如下：

	化学优秀	化学非优秀	合计
数学优秀	225	135	360
数学非优秀	156	724	880
合计	381	859	1 240

将表中数据代入公式，得

$$\chi_2^2 = \frac{1\,240 \times (225 \times 724 - 156 \times 135)^2}{360 \times 880 \times 381 \times 859} \approx 240.6 > 10.828 = x_{0.001}.$$

列出数学成绩与总分成绩的 2×2 列联表如下：

	总分优秀	总分非优秀	合计
数学优秀	267	93	360
数学非优秀	99	781	880
合计	366	874	1 240

将表中数据代入公式，得

$$\chi_3^2 = \frac{1240 \times (267 \times 781 - 93 \times 99)^2}{360 \times 880 \times 366 \times 874} \approx 486.1 > 10.828 = x_{0.001}.$$

根据小概率值 $\alpha=0.001$ 的 χ^2 独立性检验，我们推断 H_0 不成立，即认为数学成绩优秀与物理、化学、总分成绩优秀都有关系，此推断犯错误的概率不大于0.001.

一、素养落地

1. 通过本节课的学习，提升数学抽象及数据分析素养.
2. 对独立性检验思想的理解

独立性检验的基本思想类似于数学中的反证法. 先假设“两个分类变量没有关系”成立, 计算 χ^2 的值, 如果 χ^2 值很大, 说明假设不合理, χ^2 越大, 两个分类变量有关系的可能性越大.

二、素养训练

1. 对两个分类变量 A , B 的下列说法中正确的个数为()

- ① A 与 B 无关, 即 A 与 B 互不影响;
- ② A 与 B 关系越密切, 则 χ^2 的值就越大;
- ③ χ^2 的大小是判定 A 与 B 是否相关的唯一依据

A. 0

B. 1

C. 2

D. 3

解析 ①正确, A 与 B 无关即 A 与 B 相互独立; ②不正确, χ^2 的值的的大小只是用来检验 A 与 B 是否相互独立; ③不正确, 例如借助三维柱形图、二维条形图等. 故选B.

答案 B

2. 高二第二学期期中考试，按照甲、乙两个班学生的数学成绩优秀和及格统计人数后，得到如下列联表：

	优秀	及格	合计
甲班	11	34	45
乙班	8	37	45
合计	19	71	90

则 χ^2 的观测值约为()

A. 0.600

B. 0.828

C. 2.712

D. 6.004

解析 根据列联表中的数据, 可得 $\chi^2 = \frac{90 \times (11 \times 37 - 34 \times 8)^2}{45 \times 45 \times 19 \times 71} \approx 0.600$. 故选 A.

答案 A

3. 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, 可得出()

- A. 种子是否经过处理跟是否生病有关
- B. 种子是否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. 以上都是错误的

解析 由 $\chi^2 = \frac{407 \times (32 \times 213 - 61 \times 101)^2}{93 \times 314 \times 133 \times 274} \approx 0.164 < 2.706 = \chi_{0.1}$, 故没有把握认为种子是否经过处理跟是否生病有关.

答案 B

4. (多选题)对于分类变量 X 与 Y 的随机变量 χ^2 的值, 下列说法正确的是()

A. χ^2 越大, “ X 与 Y 有关系”的可信程度越小

B. χ^2 越小, “ X 与 Y 有关系”的可信程度越小

C. χ^2 越接近于0, “ X 与 Y 没有关系”的可信程度越小

D. χ^2 越大, “ X 与 Y 没有关系”的可信程度越小

解析 χ^2 越大，“ X 与 Y 没有关系”的可信程度越小，则“ X 与 Y 有关系”的可信程度越大， χ^2 越小，“ X 与 Y 有关系”的可信程度越小。

答案 BD

5. 高中流行这样一句话“文科就怕数学不好，理科就怕英语不好”。下表是一次针对高三文科学生的调查所得的数据。

	总成绩好	总成绩不好	合计
数学成绩好	478	a	490
数学成绩不好	399	24	423
合计	b	c	913

(1) 计算 a , b , c 的值;

(2) 文科学生总成绩不好与数学成绩不好有关系吗?

解 (1)由 $478+a=490$, 得 $a=12$.

由 $a+24=c$, 得 $c=12+24=36$.

$b=478+399=877$.

(2)零假设为 H_0 : 文科学生总成绩不好与数学成绩不好没有关系. 计算得

$$\chi^2 = \frac{913 \times (478 \times 24 - 399 \times 12)^2}{490 \times 423 \times 877 \times 36} \approx 6.233 > 5.024 = x_{0.05},$$

根据小概率值 $\alpha=0.05$ 的 χ^2 独立性检验, 我们推断 H_0 不成立, 即认为文科学生总成绩不好与数学成绩不好有关系, 此推断犯错误的概率不大于0.05.