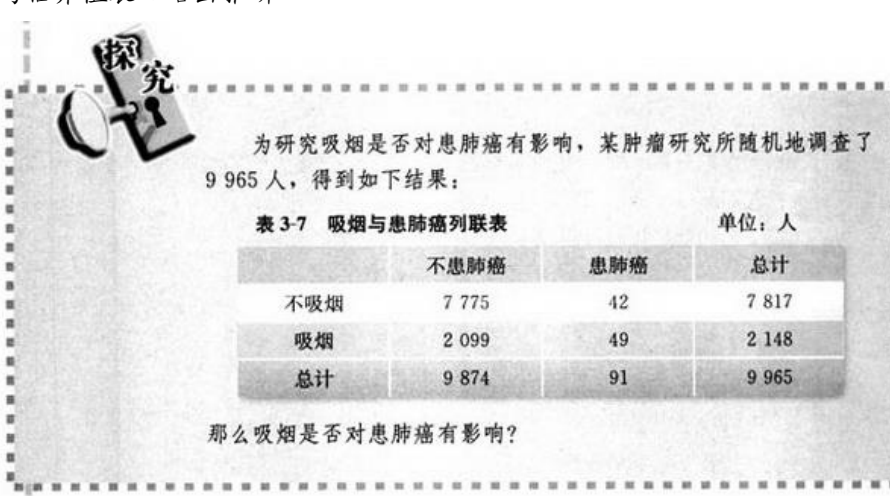
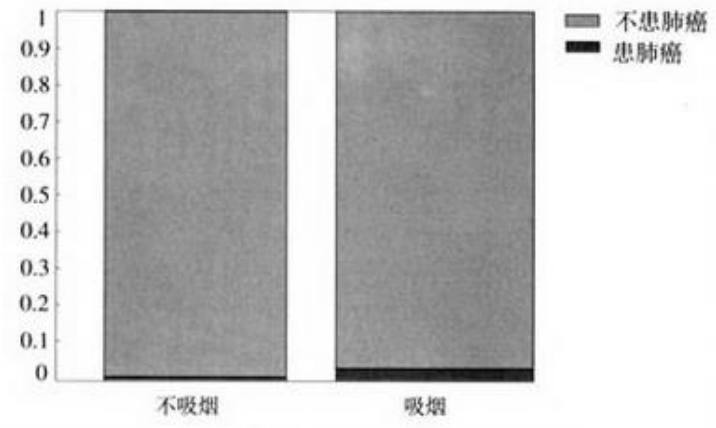


第 145 期
高中教材配套课件创作

课 题	3.2 独立性检验的基本思想及其初步
册别 单元	人教 A 版选修 2—3 第三章 3.2 独立性检验的基本思想及其初步
教材所在页码	选修 2—3 P91~ P96
教材对应截图	<p>2×2 列联表→等高条形图→构造随机变量 K^2 的观测值（检验指标）→比对观测值 K^2 与临界值表→给出推断</p> <div style="border: 1px dashed gray; padding: 10px; margin: 10px 0;">  <p style="text-align: center;">那么吸烟是否对患肺癌有影响？</p> </div> <p>与表格相比，图形更能直观地反映出两个分类变量间是否相互影响，常用等高条形图展示列联表数据的频率特征。图 3.2-1 就是一个等高条形图，其中两个浅色条的高分别表示不吸烟和吸烟样本中不患肺癌的频率；两个深色条的高分别表示不吸烟和吸烟样本中患肺癌的频率。比较图中两个深色条的高可以发现，在吸烟样本中患肺癌的频率要高一些，因此直观上可以认为吸烟更容易引发肺癌。</p> <div style="text-align: center;">  </div>

(1) 根据实际问题的需要确定容许推断“两个分类变量有关系”犯错误概率的上界 α ，然后查表 3-11 确定临界值 k_0 。

表 3-11

$P(K^2 \geq k_0)$	0.50	0.40	0.25	0.15	0.10	0.05	0.025	0.010	0.005	0.001
k_0	0.455	0.708	1.323	2.072	2.706	3.841	5.024	6.635	7.879	10.828

(2) 利用公式 (1)，计算随机变量 K^2 的观测值 k 。

(3) 如果 $k \geq k_0$ ，就推断“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证据支持结论“X 与 Y 有关系”。

例 1 在某医院，因为患心脏病而住院的 665 名男性病人中，有 214 人秃顶，而另外 772 名不是因为患心脏病而住院的男性病人中有 175 人秃顶。

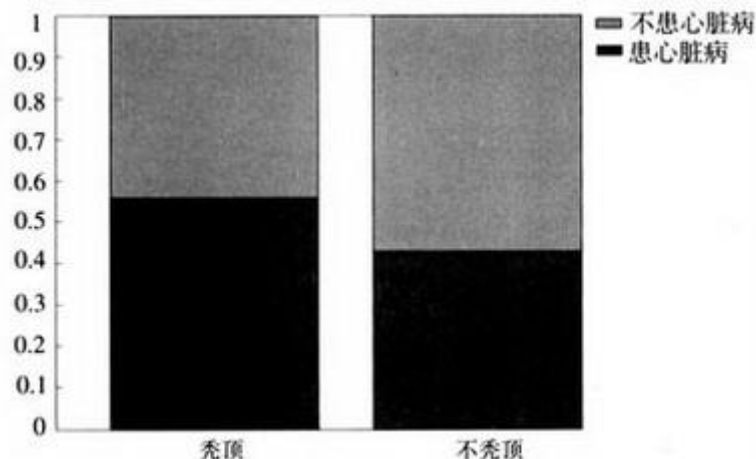
- 利用图形判断秃顶与患心脏病是否有关系；
- 能否在犯错误的概率不超过 0.01 的前提下认为秃顶与患心脏病有关系？

解：根据已知的数据得到如下列联表：

表 3-12 秃顶与患心脏病列联表 单位：人

	患心脏病	患其他病	总计
秃顶	214	175	389
不秃顶	451	597	1 048
总计	665	772	1 437

(1) 等高条形图如图 3.2-2 所示，其中两个深色条的高分别表示秃顶和不秃顶样本中患心脏病的频率。比较图中两个深色条的高可以发现，秃顶样本中患心脏病的频率明显高于不秃顶样本中患心脏病的频率，因此可以认为秃顶与患心脏病有关系。



(2) 根据列联表 3-12 中的数据，得到

$$k = \frac{1\,437 \times (214 \times 597 - 175 \times 451)^2}{389 \times 1\,048 \times 665 \times 772} \approx 16.373 > 6.635.$$

因此，在犯错误的概率不超过 0.01 的前提下认为秃顶与患心脏病有关系。

因为这组数据来自被调查的医院，因此所得结论只适合该医院的住院病人群体。



考察表 3-10, 定义

$$W = \left| \frac{a}{a+b} - \frac{c}{c+d} \right|.$$

根据独立性检验原理, 如何用 W 构造一个判断 X 和 Y 是否有关系的规则, 使得在该规则下把“ X 和 Y 没有关系”错判成“ X 和 Y 有关系”的概率不超过 0.01?



若在“ X 和 Y 没有关系”的情况下有

$$P(K^2 \geq k_0) = 0.01,$$

可以通过 k_0 来确定 u_0 吗?

由 W 的定义可以发现: 它越大, 越有利于结论“ X 和 Y 有关系”; 它越小, 越有利于结论“ X 和 Y 没有关系”. 因此可以建立如下的判断规则: 当 W 的观测值 $w > u_0$ 时, 就判断“ X 和 Y 有关系”; 否则, 判断“ X 和 Y 没有关系”. 这里 u_0 为正实数, 满足如下条件: 在“ X 和 Y 没有关系”的前提下,

$$P(W \geq u_0) = 0.01.$$

事实上,

$$K^2 = W^2 \times \frac{n(a+b)(c+d)}{(a+c)(b+d)},$$

其中 $n = a + b + c + d$. 因此, $K^2 \geq k_0$ 等价于 $W \geq \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}$, 即可取

$$u_0 = \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}.$$

	<div data-bbox="432 163 544 293" data-label="Image"></div> <div data-bbox="619 221 836 253" data-label="Text"> <p>考察表 3-10, 定义</p> </div> <div data-bbox="858 262 1086 324" data-label="Equation-Block"> $W = \left \frac{a}{a+b} - \frac{c}{c+d} \right .$ </div> <div data-bbox="566 329 1385 443" data-label="Text"> <p>根据独立性检验原理, 如何用 W 构造一个判断 X 和 Y 是否有关系的规则, 使得在该规则下把“X 和 Y 没有关系”错判成“X 和 Y 有关系”的概率不超过 0.01?</p> </div> <div data-bbox="456 456 585 607" data-label="Image"></div> <div data-bbox="676 521 1181 562" data-label="Text"> <p>若在“X 和 Y 没有关系”的情况下有</p> </div> <div data-bbox="968 571 1232 607" data-label="Equation-Block"> $P(K^2 \geq k_0) = 0.01,$ </div> <div data-bbox="617 618 976 656" data-label="Text"> <p>可以通过 k_0 来确定 u_0 吗?</p> </div> <div data-bbox="427 707 1385 855" data-label="Text"> <p>由 W 的定义可以发现: 它越大, 越有利于结论“X 和 Y 有关系”; 它越小, 越有利于结论“X 和 Y 没有关系”. 因此可以建立如下的判断规则: 当 W 的观测值 $w > u_0$ 时, 就判断“X 和 Y 有关系”; 否则, 判断“X 和 Y 没有关系”. 这里 u_0 为正实数, 满足如下条件: 在“X 和 Y 没有关系”的前提下,</p> </div> <div data-bbox="801 862 1005 891" data-label="Equation-Block"> $P(W \geq u_0) = 0.01.$ </div> <div data-bbox="434 922 534 958" data-label="Text"> <p>事实上,</p> </div> <div data-bbox="753 969 1080 1028" data-label="Equation-Block"> $K^2 = W^2 \times \frac{n(a+b)(c+d)}{(a+c)(b+d)},$ </div> <div data-bbox="430 1043 1378 1106" data-label="Text"> <p>其中 $n = a + b + c + d$. 因此, $K^2 \geq k_0$ 等价于 $W \geq \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}$, 即可取</p> </div> <div data-bbox="742 1120 1086 1184" data-label="Equation-Block"> $u_0 = \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}.$ </div>
<p>对应的学习目标</p>	<ol style="list-style-type: none"> 1、理解独立性检验的基本思想及实施步骤; 2、制作等高条形图 .
<p>教学/学习难点</p>	<ol style="list-style-type: none"> 1、了解独立性检验的基本思想; 2、理解随机变量 K^2 的含义, K^2 的观测值很大, 就认为两个分类变量是有关系的;
<p>课件设计说明</p>	<ol style="list-style-type: none"> 1、通过列联表按列数据绘制等高条形图; 2、切换行列数据按行数据再次绘制等高条形图; 3、通过列联表数据构造随机变量 K^2 的观测值 (检验指标), 然后把检验指标 K^2 的观测值与临界值表中的数据进行对比, 给出推断结果及其解释.
<p>使用说明</p>	<p>利用课件按钮提示和变量尺进行操作, 按钮切换表格行/列数据 (表格转置) 以及等高条形图的显隐.</p>