假 设 检 验

16.1 假设检验问题的提法和标准步骤

假设检验

“那是20 世纪20 年代后期，在英国剑桥一个夏日的午后，一群大学的绅士和他们的夫人们，还有来访者，正围坐在户外的桌旁，享用着下午茶。在品茶过程中，一位女士坚称：先把茶加进奶里，或先把奶加进茶里，不同的做法，会使茶的味道品起来不同。在场的一帮科学精英们，对这位女士的“胡言乱语”嗤之以鼻。这怎么可能呢？他们不能想象，仅仅因为加茶加奶的先后顺序不同，茶就会发生不同的化学反应。然而，在座的一个身材矮小、戴着厚眼镜、下巴上蓄着的短尖髯开始变灰的先生，却不这么看，他对这个问题很感兴趣。他兴奋地说道：“让我们来检验这个命题吧!”并开始策划一个实验…”

****

《女士品茶：20世纪统计怎样变革了科学》

作者：萨尔斯伯格

美国统计学会（the American Statistical Association）的Fellow

该书通俗生动地介绍了二十世纪统计学的发展主线。

“女士品茶”问题是统计学家费舍尔提出的一个有名的统计实验。

如何设计有效的试验，检验这位女士所说的话是否真实？

****

继续引用《女士品茶》书中的描述：在实验中，坚持茶有不同味道的那位女士被奉上一连串的已经调制好的茶，其中，有的是先加茶后加奶制成的，有的则是先加奶后加茶制成的。接下来，在场的许多人都热心地加入到实验中来。几分钟内，他们在那位女士看不见的地方调制出不同类型的茶来。最后，在决战来临的气氛中，蓄短胡须的先生为那位女士奉上第一杯茶，女士品了一小会儿，然后断言这一杯是先倒的茶后加的奶。 这位先生不加评论地记下了女士的说法，然后，又奉上了第二杯……

要辨别女士是否有鉴别力，办法无非就是让她实际的品尝，通过品尝后鉴别正确的多少进行判断。重要的是设计出既方便易行，又能够以清晰的思路给出概率意义解释的方法。

****

试验方案

先加奶后加茶的饮料记为MT，先加茶后加奶的饮料记为TM；

取8个同样的杯子，其中4杯MT，4杯TM，随机排列；

让女士挑选出其中的4杯MT；

根据选中的4杯饮料中，实际的MT数目进行判断。

那么到底这位选对几杯，我们能够相信她有鉴别力。选对3杯是否具有说服力，4杯全选对时能不能够下结论。这就需要做概率分析了，给出对结果的概率意义的理解。

****

我们假设该女士没有鉴别力，并设该女士挑出MT的杯数为随机变量X，

则$X \sim B\left(4,\frac{1}{2}\right)$，即 $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{70} & \frac{16}{70} & \frac{36}{70} & \frac{16}{70} & \frac{1}{70} \end{pmatrix}$

选对3杯时，不差于这一结果的概率为 $P\left(X \geq 3\right)=\frac{17}{70}\approx 0.243$

选对4杯时，不差于这一结果的概率为 $P\left(X \geq 4\right)=\frac{1}{70}\approx 0.014$

如果在实际的测试中，这位女士选对了3杯。计算出现这样的结果，以及比这个结果更好结果的概率，就是X大于等于3的概率，等于70分之17，大约是0.243。也就是说，即使这位女士没有任何的鉴别力，她也能够以接近4分之1的概率猜对3杯或3杯以上。4分之1概率的事情是很容易发生的，所以，当这位女士选对3杯时，若断言她具有对MT和TM的鉴别力，就显得有点太轻率了。

如果这名女士4杯都选对了，又意味着什么呢。我们计算一下概率，在没

有任何鉴别力的条件下，4杯全部蒙对的概率是70分之1，大约百分之1点4。

如果觉得这样小概率的事情发生是非常异常的，那么就倾向于认为没有鉴别力是不太可能。但也有人认为，70分之1概率的事情发生，也不是很异常，还不足以证明这位女士不是蒙的。这时，得出相信与不相信的结论都是有道理的。得出不同的结论只是因为对何种程度属于异常的标准不同而已。有人认为百分之5是异常的，也有人认为小于百分之1才算异常。这个标准因人而异，因情况不同而异。但是，用来进行检验的随机变量的分布，在假设下是确定，实际检验结果以及比它更极端的结果发生的概率是确定的。这就提供了客观的依据，在这个客观依据下，人们可根据自己的标准进行判断。

如果一个人认为概率小于百分之1的事件真实发生才是很异常的，那么即使这位女士选对了全部的4杯饮料，也不足以说明她有鉴别力。这时，需要更设计更为精细的试验，比如对12杯饮料进行试验等。

****

试验方案2

取12个同样的杯子，其中4杯MT，8杯TM，随机排列；

让女士挑选出其中的4杯MT；

根据选中的4杯饮料中，实际的MT数目进行判断。

设挑出MT的杯数为随机变量X，则 $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{70}{495} & \frac{224}{495} & \frac{168}{495} & \frac{32}{495} & \frac{1}{495} \end{pmatrix}$

$$P(X \geq 3) = \frac{33}{495} \approx 6.7\%, \quad P(X \geq 4) = \frac{1}{495} \approx 0.2\%$$

如果选对3杯，则在没有鉴别力的假设下，能得到不差于这一结果的概率为X大于等于3的概率大约是百分之6点7；如果4杯全对时，则，在没有鉴别力的假设下，能得到不差于这一结果的概率为X大于等于4的概率大约是百分之0.2。如果认为10%的极端就很异常了，那么在这个12杯的测试中，只要选对不少于3杯，就认为女士有鉴别力。

如果认为百分之5以下的极端才很异常，那么只选对3杯就不具有充分的说服力了，选对3杯时，还不能认为女士具有鉴别力。发生概率小到什么程度才

被认为是异常，在统计学中被称为显著性水平。按这样思路做的检验称为显著性检验。统计学中要进行检验总是要先做假设，然后选定一个统计量作为检验假设的依据，确定在假设成立的情况下，该统计量所服从的分布，以此作为标准分布。检验的过程就是用观测值与标准分布进行对比，如果观测值在标准分布中处在正常的位置就接受假设，如果观测值在标准分布中处在异常的位置则拒绝假设。

为了对什么是“异常”有更明确的判断，人们还引入了与目标假设相对立的另一个假设。通常人们将这两个假设分别称为原假设和备择假设。

****

假设检验的基本步骤如下：

步骤1. 建立假设 ： H_0： $\theta \in \Theta_0$ vs H_1： $\theta \in \Theta_1$

步骤2. 选择检验统计量，给出拒绝域W的形式。 所谓拒绝域是指使原假设被拒绝的样本观测值所在的区域W，一般将$\overline{W}$称为接受域。

校对1：$\overline{W}$ 称为接受域的$\overline{W}$上面有一横，是W的补集

步骤3. 选择显著性水平α。其定义为$\alpha = P(\text{拒绝}H_0 \mid H_0\text{为真}) = P_\theta(X \in W)$，

$\theta \in \Theta_0$，给出拒绝域的具体范围。

****

例 16.1.1 设某企业员工年收入（单位万元）服从正态分布$N\left(5,1.2^2\right)$，若从某部门随机挑选出的 16 名员工，计算出他们的平均年收入为 4.6 万元，是否可认为这个部门的员工达到了该公司的平均收入。

解 不妨设所考察部门员工的年收入服从正态分布$N\left(\mu,1.2^2\right)$，

步骤 1. 设定原假设和备择假设

H_0： $\mu \in \Theta_0 = \{\mu : \mu \geq 5\}$ vs H_1： $\mu \in \Theta_1 = \{\mu : \mu < 5\}$

或简写为 ： $H_0 : \mu \geq 5$ vs $H_1 : \mu < 5$

步骤2．选取样本均值为检验统计量，$\mu=5$ 时，$\bar{X}\sim N\left(5,\dfrac{1.2^2}{16}\right)$，即

$\bar{X}\sim N\left(5,0.3^2\right)$。原假设下，$\bar{X}$ 倾向于比较大，而备择假设下，$\bar{X}$ 倾向于比较小。所以 $\bar{X}$ 小意味着异常。拒绝域为 $\bar{X}$ 小于某个值的区域，即 $\bar{X}\le c$；

步骤3．选择显著性水平 $\alpha=0.05$，由 $P\left(\bar{X}\le c\right)=0.05$，得

$$P\left(\frac{\bar{X}-5}{0.3}\le\frac{c-5}{0.3}\right)=0.05,$$

$\dfrac{c-5}{0.3}=u_{0.05}=-1.645$，$c=5-0.3\times1.645=4.507$，所以拒绝域

$W=\left\{\bar{X}\le 4.507\right\}$。

样本均值的观测值为 $\bar{x}=4.6$。$\bar{x}=4.6\notin W$，所以该检验的结果是接受原假设，可以认为这个部门的员工达到了该公司的平均收入。

****

16.2 假设检验问题的两类错误和 p 值

假设检验两类错误

	原假设成立	原假设不成立
接受	√	第二类错误（受伪）
拒绝	第一类错误（拒真）	√

第一类错误即为显著性水平 $\alpha=P(\text{拒绝}H_0\mid H_0\text{为真})=P_\theta(X\in W)$，

第二类错误的概率表达为 $\beta=P(\text{接受}H_0\mid H_1\text{为真})=P_\theta(X\in\overline{W})$，$\theta\in\Theta_1$。

****

假设检验中，两类错误的概率不能同时减小，二者相互制约。

犯第一类错误的概率越小，则犯第二类错误的概率越大，

犯第二类错误的概率越小，则犯第一类错误的概率越大。

原假设和备择假设不能随意互换位置，原假设是人们经验上认为正常的假设。

理想的检验应该是在控制犯第一类错误的基础上，尽量少犯第二类错误。

显著性检验具有“保护原假设”的特点，显著性水平α也不是越小越好。

固定第一类错误的概率，可通过增加样本量降低犯第二类错误的概率。

****

例 16.2.1 某厂生产一种标准长度35mm的螺钉，实际生产的产品长度服从正态分布 $N(\mu,3^2)$。做假设检验，样本容量 $n=36$， $H_0:\mu=35$， $H_1:\mu\neq35$，拒绝域为 $W=\{\bar{x}:|\bar{x}-35|>1\}$。

(1) 犯第一类错误的概率。(2) μ =36时，犯第二类错误的概率。

解 (1) 检验统计量 $\bar{X}$ 的分布为 $\bar{X}\sim N\left(\mu,\left(\frac{1}{2}\right)^2\right)$，第一类错误的概率为

$$\alpha=P\{|\bar{X}-35|>1|\mu=35\}=1-P\{|\bar{X}-35|\le1|\mu=35\}$$

$$=1-P\left\{-2<\frac{\bar{X}-35}{1/2}>2|\mu=35\right\}$$

$$=1-\Phi(2)+\Phi(-2)=2-2\Phi(2)=0.0455\text{。}$$

(2) 第二类错误的概率为

$$\beta=P\{|\bar{X}-35|\le1|\mu=36\}=P(-1\le\bar{X}-35\le1\,|\,\mu=36)$$

$$=P\left(-4\le\frac{\bar{X}-36}{\frac{1}{2}}\le0\,|\,\mu=36\right)$$

$$=\Phi(0)-\Phi(-4)=\Phi(0)+\Phi(4)-1=0.5\text{。}$$

****

***p* 值：**在一个假设检验问题中，利用观测值能够做出拒绝原假设的最小显著性水平，称为检验的 p 值。或者说是在原假设下，出现比观测值更极端情况的概率。 p 值值时相对于某一次具体的观测结果而言的。

例 16.2.2 总体服从 $N\left(\mu,3^2\right)$。做假设检验，样本容量 $n=36$， $H_0:\mu=35$，$H_1:\mu\neq 35$。现得到样本均值的观测值为36.5mm，求其 p 值。

解： $p=P\left(\left|\bar{X}-35\right|\geq 1.5\middle|\mu=35\right)=1-P\left(\left|\bar{X}-35\right|<1.5\middle|\mu=35\right)$

$$=1-P\left\{-3<\frac{\bar{X}-35}{1/2}<3\middle|\mu=35\right\}=1-\left(\Phi(3)-\Phi(-3)\right)=0.0027。$$

****

16.3 单个正态总体参数的假设检验

设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的样本，考虑如下三种关于μ的检验问题

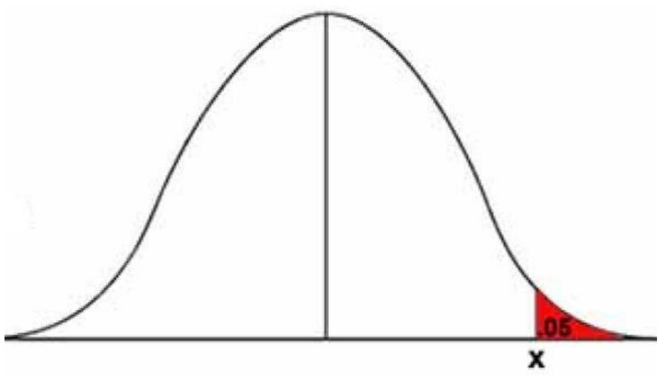
(1) H_0： $\mu\leq\mu_0$ vs H_1： $\mu>\mu_0$ 单侧检验

(2) H_0： $\mu\geq\mu_0$ vs H_1： $\mu<\mu_0$ 单侧检验

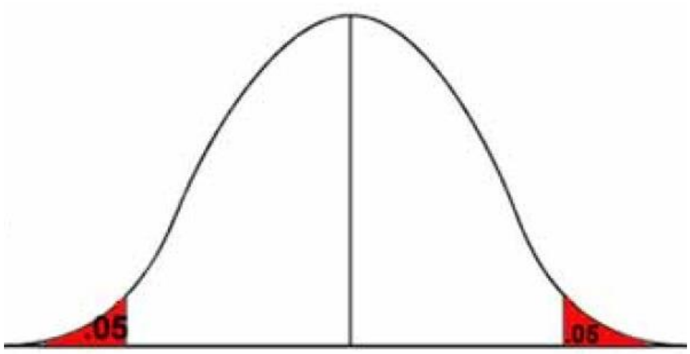
(3) H_0： $\mu=\mu_0$ vs H_1： $\mu\neq\mu_0$ 双侧检验

****

(1) H_0： $\mu\leq\mu_0$ vs H_1： $\mu>\mu_0$ 单侧检验



(3) H_0： $\mu=\mu_0$ vs H_1： $\mu\neq\mu_0$ 双侧检验



****

下面给出σ已知时，上述三种检验情况的具体实现。

σ已知时的，对于单侧检验问题(1) H_0： $\mu \le \mu_0$ vs H_1： $\mu > \mu_0$，

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$，故选用服从标准正态分布的检验统计量$u = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$，

通常称此检验为u检验。

拒绝域选为$W = \left\{ (x_1, \cdots, x_n) : u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \ge c \right\}$，$c$为临界值，简记为

$\{u \ge c\}$。若显著性水平要求为α，则可确定$c = u_{1-\alpha}$。

同理对

问题(2)，H_0： $\mu \ge \mu_0$ vs H_1： $\mu < \mu_0$，水平为α的检验的拒绝域为

$W = \left\{ (x_1, \cdots, x_n) : u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \le u_\alpha \right\}$。

问题(3)，H_0： $\mu = \mu_0$ vs H_1： $\mu \ne \mu_0$，水平为α的检验的拒绝域为

$W = \left\{ (x_1, \cdots, x_n) : u = \left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| \le u_{1-\frac{\alpha}{2}} \right\}$。

****

例16.3.1 设某工厂生产一种产品，其质量指标服从正态分布$N(\mu,2^2)$，μ为平均质量指标，其值越大则质量越好，$\mu=10$是达到优级的标准。进货商店从一批产品抽取样本$X_1,X_2,\cdots,X_n$，$n=16$，取显著性水平为$\alpha=0.05$，如何检验这一批产品是否达到优秀。

分析： 根据工厂产品社会声誉可能的不同，分以下两种情况讨论。

情形一，按照过去长时间的记录，商店的检验人员相信该厂的产品质量很好。

情形二，按照过去长时间的记录，该厂的产品质量一直不够好。

情形一求解，按照过去长时间的记录，商店的检验人员相信该厂的产品质量很好。H_0： $\mu\geq 10$ vs H_1： $\mu<10$

当$\mu=10$时，样本均值$\bar{X}\sim N\left(10,\dfrac{2^2}{16}\right)$，即$\dfrac{\bar{X}-10}{1/2}\sim N(0,1)$。

此时的拒绝域为$W=\left\{\dfrac{\bar{x}-10}{1/2}\leq u_{0.05}\right\}=\left\{\bar{x}\leq 10-1.645\cdot\dfrac{1}{2}=9.18\right\}$。

当样本均值的观测值$\bar{x}>9.18$时，就接受产品为优级。

情形二求解，按照过去长时间的记录，该厂的产品质量一直不够好。

这时，商店就可能坚持以$\mu\leq\mu_0$作为原假设，

H_0： $\mu<10$ vs H_1： $\mu\geq 10$

此时的拒绝域为

$$W=\left\{\frac{\bar{x}-10}{1/2}\geq u_{1-0.05}\right\}=\left\{\bar{x}\geq 10+1.645\frac{1}{2}\right\},$$

当样本均值的观测值$\bar{x}\geq 10+1.645\dfrac{1}{2}=10.82$时，接受产品为优级。

这个例子反映了假设检验“保护零假设”的特性。情形一的做法对接受产品为优级有利；情形二的做法则对接受产品为优级比较苛刻，要求有强有力的证据证明产品质量优秀，才能接受产品为优级。假设检验中，都是以常识性的、经验性的假设作为原假设，也就是以通常认为正常的情况作为原假设，除非数据表明特别的异常，才拒绝原假设。

****

σ 未知时，正态总体 $N\left(\mu,\sigma^2\right)$ 关于 μ 的假设检验，可利用服从 $n-1$ 自由度的 t 分布统计量 $t=\dfrac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S}$ 进行检验，通常称为t检验。

三种关于 μ 的检验问题的显著性水平为 α 的拒绝域分别为

(1) $W=\left\{\left(x_1,\cdots,x_n\right): t=\dfrac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{s} \geq t_{1-\alpha}(n-1)\right\}$

(2) $W=\left\{\left(x_1,\cdots,x_n\right): t=\dfrac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{s} \leq t_{\alpha}(n-1)\right\}$

(3) $W=\left\{\left(x_1,\cdots,x_n\right): t=\left|\dfrac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{s}\right| \leq t_{1-\frac{\alpha}{2}}(n-1)\right\}$。

****

假设检验与置信区间存在密切的关联。

(1) H_0： $\mu \leq \mu_0$ vs H_1： $\mu > \mu_0$ 接受域为 $1-\alpha$ 置信水平的上侧置信限的置信区间，

(2) H_0： $\mu \geq \mu_0$ vs H_1： $\mu < \mu_0$ 接受域为 $1-\alpha$ 置信水平的下侧置信限的置信区间，

(3) H_0： $\mu = \mu_0$ vs H_1： $\mu \neq \mu_0$ 接受域为 $1-\alpha$ 置信水平的双边置

信区间。

****

16. 4 拟合优度检验

卡方检验

设总体服从离散分布 $X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$，进行 n 次独立的观测， k 个取值出现的频次分别为 $n_i\left(i=1,\cdots,k\right)$，则 $X=\sum_{i=1}^{k}\frac{\left(n_i-np_i\right)^2}{np_i}$ 近似服从 $\chi^2\left(k-1\right)$。

这一结果是 20 世纪初著名统计学家皮尔森（Pearson）发现的结果，用这一结果可以构造观测数据与假设离散分布的拟合程度，该方法称为皮尔森的卡方（χ^2）检验。

****

例 16. 4. 1 卢瑟福和盖革在 1910 年观察了放射性物质放出 α 粒子的个数的情况，共观察 $n=2608$ 次，每次观察间隔 7. 5 秒，记录到达指定区域的 α 粒子数，共记录下 10094 个粒子， n_k 表示恰好记录到 k 个 α 粒子的观察次数。

k	0	1	2	3	4	5	6	7	8	9	≥ 10
n_k	57	203	383	525	532	408	273	139	45	27	16
$n\cdot\hat{p}_k$	54	211	407	525	508	394	254	140	68	29	17

现在希望检验这组数据是否来自于泊松分布。

解： 设1次观察中出现的 α 粒子数为随机变量 X，共有2608次观测，所以1个粒子落入该次观察的概率是 $\frac{1}{2608}$，共记录下10094个粒子，

$$X \sim B\left(10094,\frac{1}{2608}\right)$$

根据泊松定理， $X \sim P\left(\hat{\lambda}\right)$，其中 $\hat{\lambda}=\frac{10094}{2608}=3.87$

泊松分布随机变量不同取值的概率 $\hat{p}_i=\frac{\hat{\lambda}^i}{i!}e^{-\hat{\lambda}}\left(i=0,1,\cdots,9\right)$， $\hat{p}_{10}=1-\sum_{i=0}^{9}\hat{p}_i$

计算这组观测数据下卡方检验统计量的取值，$Y=\sum_{i=0}^{10}\frac{\left(N_i-N\hat{p}_i\right)^2}{N\hat{p}_i}=12.88$，

$Y\sim\chi^2\left(10\right)$， p 值 $p=P\left(Y>12.88\right)=0.236$，因此可以接受这组数据来自于参数 $\hat{\lambda}$ 的分布。

****

例16.4.2　为检验骰子的均匀性，甲乙两人分别进行试验。

甲掷60次，结果出现1—6点的次数分别为：7，6，12，14，5，16；

相应的频率依次为：0.117 ， 0.100 ， 0.200 ， 0.233 ， 0.083 ， 0.267；

乙掷了 9，000，000次，结果出现1—6点的次数分别为：

1500300，1502100，1503000，1498500，1496700，1499400；

相应的频率依次为：0.1667 ， 0.1669 ， 0.1670 ， 0.1665 ， 0.1663 ， 0.1666。

试判断甲、乙所用骰子是否均匀。

解　在骰子均匀的假设下，设掷一次所的点数为随机变量 X，其概率分布为

$$P\left(X=i\right)=p_i=\frac{1}{6}，\ i=1,2,\cdots,6。$$

甲掷骰子的试验进行了60次，所以此时 $n=60$，$n\cdot p_i=10$，

$i=1,2,3,4,5,6$，投掷结果的卡方统计量取值为

$$Y_1=\sum_{i=1}^{6}\frac{\left(n_i-np_i\right)^2}{np_i}$$

$$=\frac{\left(7-10\right)^2+\left(6-10\right)^2+\left(12-10\right)^2+\left(14-10\right)^2+\left(5-10\right)^2+\left(16-10\right)^2}{10}=8.6$$

$Y_1 \sim \chi^2(5)$， p 值 $p=P(Y_1>8.6)>0.1$，可以接受均匀假设。

乙掷筛子的试验进行了9，000，000次，所以此时 $n=9000000$，$n \cdot p_i = 1500000$，$k=1,\cdots,6$，投掷结果的卡方统计量取值为

$$Y_2=\sum_{i=1}^{6}\frac{(n_i-np_i)^2}{np_i}=16.07$$

$Y_2 \sim \chi^2(5)$， p 值 $p=P(\chi^2>16.07)<0.01$，拒绝均匀假设。

思考：甲掷筛子所得各点数的频率与理想值 $\frac{1}{6}$ 有较明显的差异，而乙掷筛子所得各点数的频率都非常接近理想值 $\frac{1}{6}$，为什么甲的结果能够通过均匀假设，而乙的结果却反而不能？

****

独立性检验

拟合优度的 χ^2 检验还可以用来判断不同属性的相关性。

例 16.4.3 曾经有人统计了 6672 名学生使用左、右手的习惯，

	男	女	合
右	2780	3281	6061
左	311	300	611
合	3091	3581	6672

其中男性左手率为 0.1，女性左手率为 0.08，

试问使用左、右手的习惯是否与性别相关？

****

双向列联表的独立性检验

$A\backslash B$	1	2	$\cdots$	j	$\cdots$	t	行合计
1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	c_1
2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	c_2
$\vdots$	.	.	.	.	.	.	$\vdots$
i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	c_i
$\vdots$	.	.	.	.	.	.	$\vdots$
s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	c_s
列合计	d_1	d_2	$\cdots$	d_j	$\cdots$	d_t	n

$$\sum_{i=1}^{s}\sum_{j=1}^{t} n_{ij} = n$$

$$c_i = \sum_{j=1}^{t} n_{ij}, \quad d_j = \sum_{i=1}^{s} n_{ij}$$

$$\frac{c_i}{n}\cdot\frac{d_j}{n} \approx \frac{n_{ij}}{n}$$

$$Y = \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{\left(n_{ij} - n\cdot\frac{c_i}{n}\cdot\frac{d_j}{n}\right)^2}{n\cdot\frac{c_i}{n}\cdot\frac{d_j}{n}} = \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{\left(nn_{ij} - c_i d_j\right)^2}{nc_i d_j} \ \dot\sim\ \chi^2\left((s-1)(t-1)\right)$$

****

例 16.4.3（续） 男女生使用左右手的数据

	男	女	合
右	2780	3281	6061
左	311	300	611
合	3091	3581	6672

$$Y = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(nn_{ij} - c_i d_j\right)^2}{nc_i d_j} = \frac{(6672\cdot 2780 - 6061\cdot 3091)^2}{6672\cdot 6061\cdot 3091} + \frac{(6672\cdot 3281 - 6061\cdot 3581)^2}{6672\cdot 6061\cdot 3581}$$
$$+\frac{(6672\cdot 311 - 611\cdot 3091)^2}{6672\cdot 611\cdot 3091} + \frac{(6672\cdot 300 - 611\cdot 3581)^2}{6672\cdot 611\cdot 3581} = 5.65,$$

Y 近似服从 1 个自由度的χ^2分布，

Y 的 0.95 和 0.99 分位数分别为 $\chi^2_{0.95}(1) = 3.841 \quad \chi^2_{0.99}(1) = 6.635$。

$\alpha = 0.05$时，拒绝原假设；$\alpha = 0.01$时，接受原假设

事实上，这组数据确实表现出男女生的左右率有一定的差别，但这种差别又不是特别的显著。而检验统计量的取值也和直观的感觉的差不多，p 值在 0.01 和 0.05 之间，处于比较边缘状态，结论也就是仁者见仁，智者见智了。虽然无法得到非常确切的结论，但检验统计量的概率意义是完全清晰的，使用者根据自己的尺度得到相应结论。

****

例16.4.4 下面是1936年瑞典对25263个家庭的小孩数与收入的调查表

小孩数\收入	0−1	1−2	2−3	≥3	行合计
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
4	39	98	31	14	182
列合计	6116	10928	5173	3046	25263

试问家庭的小孩数与收入水平是否相关。

解： 计算 $Y=\sum_{i=1}^{5}\sum_{j=1}^{4}\frac{\left(nn_{ij}-c_i d_j\right)^2}{nc_i d_j}=75.173$，

检验统计量 Y 近似服从 $4\times3=12$ 自由度的 χ^2 分布，

查表可知 $\chi^2_{0.999}\left(12\right)=32.909$， p 值小于 0.001，拒绝原假设。

****