应用统计

第9讲　　参数假设检验

女士品茶

- 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

- 《女士品茶——20世纪统计怎样变革了科学》是一本20 世纪统计发展史的科普读物。这本书通过皮尔逊、戈赛特、Fisher等统计学家工作和经历的一些生动故事描述了统计学一些重要概念的演进发展。作者萨尔斯伯格（David Salsburg）本人也是一位很好的统计学家，美国统计学会（the American Statistical Association）的Fellow。

- “女士品茶” 是一个统计史上非常有名的统计实验，由现代统计学得奠基人之一费歇尔 (R·A· Fisher)提出。

女士品茶：检验一

- 先放奶记为MT，先加茶记为TM
- 设计试验：取8个一样的杯子，其中4杯MT，4杯TM，随机排列，让该女士挑出4杯MT
- 如果4杯全部挑对 $C(8,4)=70$，$1/70\approx 0.0143$
- 如果挑对3杯　$C(4,3)*C(4,1)=16$，$17/70\approx 0.243$
- 挑对2杯，　$53/70\approx 0.757$

女士品茶：检验二

- 设计试验：取12个一样的杯子，其中4杯MT，8杯TM，随机排列，让该女士挑出4杯MT
- 如果4杯全部挑对 C(12,4)=495, 0.002
- 如果挑对3杯 C(4,3)*C(8,1)=32, 0.0667
- 挑对2杯， C(4,2)*C(8,2)=168, 0.4061

判断产品是否为优质

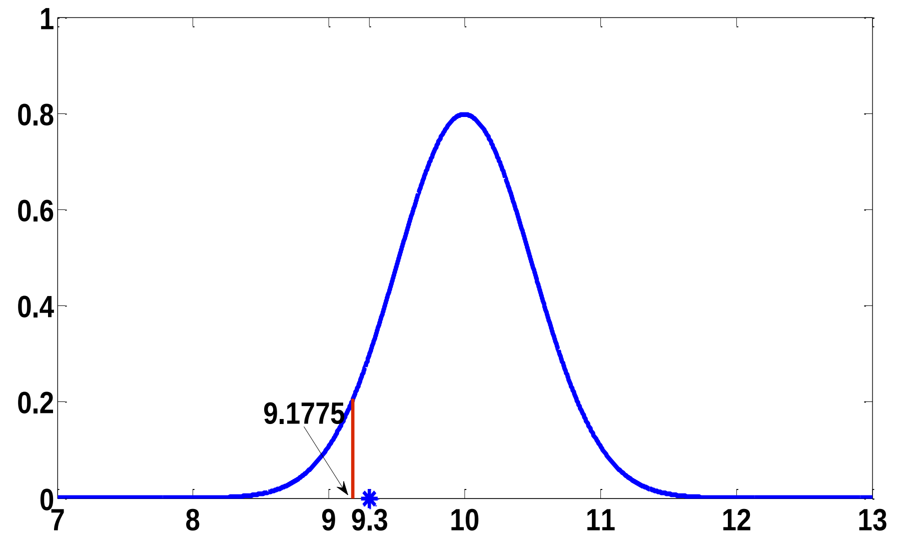
设某工厂生产一种产品，其质量指标服从正态分布 $N\left(\mu, 2^2\right)$，为平均质量标准，其值越大则质量越好，$\mu = 10$ 是达到优质的标准。进货商从一批产品中抽取 16 个样本测得指标分别为:

9.6	9.2	9.3	9.8	10.1	8.5	9.8	8.4
9.2	9.1	8.7	9.3	8.6	9.8	9.2	10.2,

则进货商是否接受这批产品为优质

如果这品产品来自于$\mu = 10$,

则 $\overline{X} \sim N\left(10, \frac{1}{2}^2\right)$



假设检验问题的提法

假设检验的标准提法是提出一个原假设 H_0 和一个备择假设 H_1。

$$H_0: \quad \theta \in \Theta_0 = \{\theta : \theta \geq 10\} \quad \text{vs} \quad H_1: \quad \theta \in \Theta_1 = \{\theta : \theta < 10\}$$

或简写为 $H_0: \quad \theta \geq 10 \quad \text{vs} \quad H_1: \quad \theta < 10$

检验借助一个统计量完成，该统计量称为检验统计量。
检验：在原假设条件下，该统计量的取值是否正常。上例取的检验统计量是样本均值。直观上，样本均值越大，意味着总体期望可能越大；样本均值小则很可能总体期望也较小。

标准应该既能区分差别，又能较为简单地区分差别

假设检验的基本步骤

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

2. 选择检验统计量，给出拒绝域形式。

使原假设被拒绝的样本观测值所在的区域 W 称为**拒绝域**。一般将 $\overline{W}$ 称为接受域。

3. 选择显著性水平。 $\alpha = P(\text{拒绝}H_0 \mid H_0\text{为真}) = P_\theta(X \in W)$， $\theta \in \Theta_0$

分布情况是客观的，但是选择标准却是较为主观的

倾向于选α比较小，也即不会轻易拒绝原假设，因此较为主观

女士品茶检验：建立假设

- 设这位女士对一杯饮料能够鉴别正确的概率为 p

- 原假设（零假设）：女士没有鉴别力

  H_0： $p \leq 1/2$

- 备择假设：相反的结果

  H_1： $p > 1/2$

女士品茶检验：检验统计量

- 该女士挑出MT的杯数 X，它的分布
- 试验一：$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{70} & \frac{16}{70} & \frac{36}{70} & \frac{16}{70} & \frac{1}{70} \end{pmatrix}$
- 试验二：$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{70}{495} & \frac{224}{495} & \frac{168}{495} & \frac{32}{495} & \frac{1}{495} \end{pmatrix}$
- 拒绝域的形式：　检验统计量大于等于某个值

女士品茶：检验标准

- 选择显著性水平 $\alpha = 0.01$

试验一：当检验统计量 $x>4$ 时，拒绝原假设

试验二：当检验统计量 $x\geq 4$ 时，拒绝原假设

- 选择显著性水平 $\alpha = 0.05$

试验一：当检验统计量 $x\geq 4$ 时，拒绝原假设

试验二：当检验统计量 $x\geq 4$ 时，拒绝原假设

女士品茶：检验标准

- 如果选择显著性水平 $\alpha = 0.1$

试验一： 当检验统计量x ?时，拒绝原假设

试验二： 当检验统计量x ?时，拒绝原假设

假设检验的两类错误

上述 α 称为第一类错误（拒真）概率，

第二类错误（受伪）概率 $\beta = P\left(\text{接受}H_0 \mid H_1\text{为真}\right) = P_\theta\left(X \in \overline{W}\right)$， $\theta \in \Theta_1$ 。

	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）	√

例　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 $N\left(\mu,3^2\right)$。做假设检验，样本容量 $n=36$， $H_0:\mu=35$， $H_1:\mu\neq 35$，拒绝域为 $W=\left\{\bar{x}:\left|\bar{x}-35\right|>1\right\}$。

(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

$$\alpha=P\left\{\left|\bar{X}-35\right|>1\middle|\mu=35\right\}=1-P\left\{\left|\bar{X}-35\right|\le 1\middle|\mu=35\right\}$$

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

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

例　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 $N\left(\mu,3^2\right)$。做假设检验，样本容量 $n=36$， $H_0:\mu=35$， $H_1:\mu\neq35$，拒绝域为 $W=\left\{\bar{x}:\left|\bar{x}-35\right|>1\right\}$。

(1) 犯第一类错误的概率。(2) μ =36时，犯第二类错误的概率。

(2) 第二类错误的概率为

$$\beta=P\left\{\left|\bar{X}-35\right|\le1\middle|\mu=36\right\}=P\left(-1\le\bar{X}-35\le1\mid\mu=36\right)$$

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

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

显著性水平，保护原假设

限制显著性水平α为一个较小的值就是限制犯第一类错误的概率，这体现了“保护零假设”的思想。

另一方面，两类错误相互制约，不能同时减小。犯第一类错误的概率越小，则犯第二型错误的概率越大。因此，一般来说，显著性水平α也不是越小越好。首先保证第一类错误较小，即保证参数为真时，以尽可能小的概率拒绝原假设。

判断产品是否为优质

设某工厂生产一种产品 其质量指标服从正态分布 $N\left(\mu,2^2\right)$, 为平均质量标准其值越大则质量越好，$\mu=10$ 是达到优质的标准。进货商从一批产品中抽取16个样本测得指标分别为

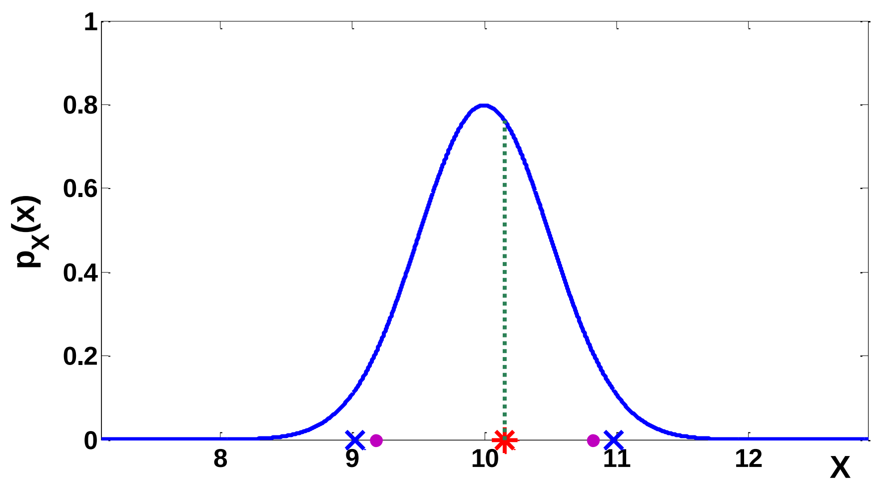
10.6	11.2	9.2	9.8	10.1	10.5	8.8	12.4
9.2	10.1	9.7	9.3	10.6	9.8	11.2	9.9

，

则进货商是否接受这批产品为优质

$H_0:\ \theta \ge 10 \quad \text{vs} \quad H_1:\ \theta < 10$

$H_0:\ \theta < 10 \quad \text{vs} \quad H_1:\ \theta \ge 10$



两种情况的讨论

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，当然这也不排除偶尔出现一批质量较差的产品。于是以 $\mu \geq \mu_0 = 10$ 作为原假设，取 $\alpha = 0.05$。此时的拒绝域为

$$W = \left\{ u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \leq u_\alpha \right\} = \left\{ \bar{x} \leq 10 - 1.645 \frac{2}{\sqrt{n}} \right\}$$

。只要

$\bar{x} > 10 - 1.645 \dfrac{2}{\sqrt{n}}$，就接受产品为优级。

两种情况的讨论

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。这时，商店就可能坚持以 $\mu \le \mu_0$ 作为原假设，并选定较低水平的 α，例如 $\alpha = 0.05$。此时的拒绝域为

$$W = \left\{ u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \ge u_{1-\alpha} = \left\{ \bar{x} \ge 10 + 1.645 \frac{2}{\sqrt{n}} \right\} \right\}$$ 。只有

$\bar{x} \ge 10 + 1.645 \frac{2}{\sqrt{n}}$ 时，接受产品为优级。。

情形一的做法对接受产品为优级有利；情形二的做法则比较苛刻地接受产品为优级，要求有较强的证据证明产品质量优秀。

情形3:如果不了解厂家，不知道该怎么建立原假设，因此要增加样本容量，使得两组方法得到的结果类似

正态总体均值的假设检验

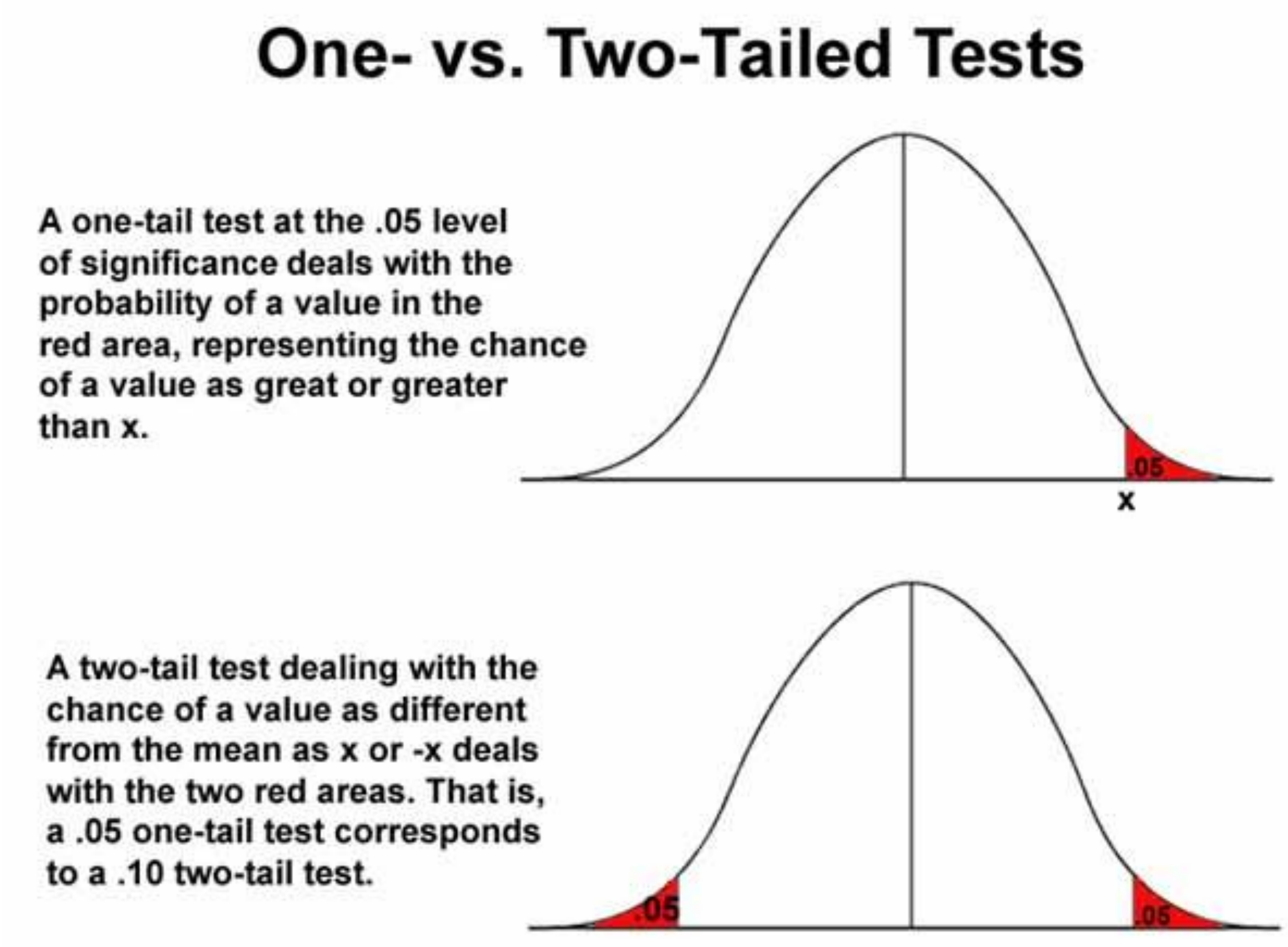
设 $x_1, x_2, \cdots, x_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，考虑如下三种关于 μ 的检验问题

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

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

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

单边与双边检验


One- vs. Two-Tailed Tests
A one-tail test at the .05 level of significance deals with the probability of a value in the red area, representing the chance of a value as great or greater than x.
.05
x
A two-tail test dealing with the chance of a value as different from the mean as x or -x deals with the two red areas. That is, a .05 one-tail test corresponds to a .10 two-tail test.
.05
.05

U 检验（方差已知）

σ 已知时的 u 检验

由于 μ 的点估计是 $\bar{x}$，且 $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$，故选用检验统计量

$$u = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$$ 。一般称此检验为 u 检验。

若显著性水平要求为 α，双边检验的拒绝域为

$$W = \{(x_1, \cdots, x_n) : u \le u_{\alpha/2} \text{ 或 } u \ge u_{1-\alpha/2}\}$$

显著性水平为α的单边检验

对于单侧检验问题， H_0： $\mu \le \mu_0$ vs H_1： $\mu > \mu_0$，

拒绝域选为 $W = \left\{ (x_1, \cdots, x_n) : u = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \ge u_{1-\alpha} \right\}$。

对于单侧检验问题， H_0： $\mu \ge \mu_0$ vs H_1： $\mu < \mu_0$，

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

σ 未知时均值的t检验

对于检验问题， H_0： $\mu \le \mu_0$ vs H_1： $\mu > \mu_0$， σ 未知时，可利用 t 统计量 $t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s}$ 进行检验。

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

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

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

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

判断产品是否为优质

设某工厂生产一种产品，其质量指标服从正态分布 $N\left(\mu,\sigma^2\right)$，为平均质量标准，其值越大则质量越好，$\mu = 10$ 是达到优质的标准。进货商从一批产品中抽取 16 个样本测得指标分别为:

10.6	11.2	9.2	9.8	10.1	10.5	8.8	12.4
9.2	10.1	9.7	9.3	10.6	9.8	11.2	9.9

，则进货商是否接受这批产品为优质

$\bar{x} = 10.15,\ s = 0.92$

$H_0:\ \theta \geq 10 \quad \text{vs} \quad H_1:\ \theta < 10$

类似地还可以做方差是否为4的假设检验

$$\frac{4\left(\bar{x} - 10\right)}{s} < t_\alpha\left(15\right),\quad \text{拒绝域：}\quad W = \left\{ \bar{x} < 10 + \frac{t_\alpha}{4} \cdot 0.92 \right\}$$

选定 $\alpha = 0.05$ 时， $W = \left\{ \bar{x} < 10 - 1.753 * 0.23 = 9.5968 \right\}$

p值较大的时候不能讨论什么更正常，要在较小的时候才有意义

P 值

针对特定检验结果，在原假设下异常的程度

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；即原假设成立条件下，样本量出现在观测值以外的概率的最大值，称为检验的 p 值。利用 p 值做检验比较方便。

在判断产品是否为优质的例题中，以 $\theta \geq 10$ 为原假设，$\bar{x}=10.15$ 对应的 p 值是 $P(\bar{x} \leq 10.15)=0.6179$。以 $\theta<10$ 为原假设，$\bar{x}=10.15$ 对应的 p 值是 $P(\bar{x}>10.15)=0.3821$。结果都是不拒绝原假设。

可以先算p值然后再判断原假设是否满足

例　总体服从 $N(\mu,3^2)$。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。

这个时候是双边检验，因此要计算偏离的绝对值

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

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

K.Pearson 的χ^2检验 （拟合优度检验）

设总体服从离散分布
$$\begin{array}{cccc} X & x_1 & \cdots & x_k \\ P & p_1 & \cdots & p_k \end{array}$$

进行n次独立地观测，k个取值出现的频次分别为$N_i\left(i=1,\cdots,k\right)$，则

$$X=\sum_{i=1}^{k}\frac{\left(N_i-np_i\right)^2}{np_i}$$ 近似服从自由度为$k-1$的χ^2分布。

每一个被求和项近似于正态分布

χ^2 检验实例

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。

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 $\dfrac{1}{2608}$，共记录下10094个粒子， $X\sim B\left(10094,\dfrac{1}{2608}\right)$

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 $\dfrac{1}{2608}$，共记录下10094个粒子，$X \sim B\left(10094, \dfrac{1}{2608}\right)$

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

泊松分布随机变量不同取值的概率

$$\hat{p}_i = \frac{\hat{\lambda}^i}{i!} e^{-\hat{\lambda}} \left(i = 0, 1, \cdots, 9\right), \quad \hat{p}_{10} = 1 - \sum_{i=0}^{9} \hat{p}_i$$

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

$Y \dot{\sim} \chi^2\left(10\right)$， 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。

检验色子的均匀性

例 为检验骰子的均匀性，甲乙两人分别进行试验。

甲投掷 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；

乙投掷 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

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

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

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

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

$Y_1 \dot\sim \chi^2(5)$，　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

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

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

乙投掷 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。

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 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？

列联表检验（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

$$Y=\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{\left(n_{ij}-n\cdot\frac{c_i}{n}\cdot\frac{d_j}{n}\right)^2}{n\cdot\frac{c_i}{n}\cdot\frac{d_j}{n}}=\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{\left(nn_{ij}-c_id_j\right)^2}{nc_id_j}$$ 近似服从自由度为$(s-1)(t-1)$的χ^2分布

从左边这个式子上看，也即检验是否满足服从
这样的分布，其实和上一例很类似

左右手习惯是否与性别相关

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

$Y = 5.65,\qquad \chi_1^2(0.05) = 3.841,\qquad \chi_1^2(0.01) = 6.635$

女性左右率0.08，男性左右率0.10

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

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

作业

习题三（ 175-178页）

1，2，11，15，20，21

并计算2(2)，11和20题检验的p值