第3章 假设检验

第3.1节 基本理论

第3.2节 重要参数检验

第3.3节 非参数检验

第3.4节 统计决策与Bayes 理论

东北大学数学系

3.1.1 假设检验的背景

1. 假设检验(*Test of Hypothesis*) 含义

它是如下的一种统计推断：

对于一个统计模型，我们提出一个假设，根据抽取到的样本，来作出是接受还是拒绝这个假设。

小概率事件在一次试验中不应该发生。

2. 假设检验理论的重要历史事件

(1) K. Pearson 的拟合优度检验，**1900**年

(2) R. A. Fisher 的显著性检验，**1920**年代

(3) J. Neyman 与**E. S. Pearson** 理论，**1928**年开始

(4) A. Wald 的统计决策理论，**1945**年

(5) Bayes 方法，基本观点从 **T. Bayes (1702~1761)** 开始，二战后得到巨大发展。

女士品茶

有一种饮料由 **Tea** 和 **Milk** 混合而成，按照顺序的不同，分为 **TM**、 **MT** 两种，

有位女士声称她有能力品尝出是 **TM** 还是**MT** 。

为了检验她的说法是否可信，准备 **8** 杯饮料，**TM** 和 **MT** 各一半，并且把这一点告诉她。

现在随机的让这位女士品尝，指出哪些是 **TM** ，最终的结果是她全部说对了。

R.A.Fisher 的推理过程如下：

引进一个假设，

H_0：这位女士没有鉴别能力

如果 H_0 是正确的，她只能随机从 **8** 杯饮料中猜测 **4** 杯说是 **TM** 。全部猜对的概率为：

$$\frac{1}{C_8^4}=\frac{1}{70}\approx 0.014$$

现在她正确的说出了全部的 **TM**，要解释这种现象，只能有下面两种可能：

(1) H_0 不成立，即：她的确有鉴别能力；

(2) H_0 成立，意味着一件概率为 0.014 的随机事件在一次试验中发生了。

一个概率不到 2% 的随机事件在一次试验中发生了，这是比较稀奇或者说不太可能的。

Fisher 认为，随机试验的结果(或样本) 构成不利于假设 H_0 的显著性证据，因此应该否定H_0。

这种推理过程就称为：显著性检验

显著性是统计意义上的显著，意思是一个小概率事件是否发生。

思考1　假如这位女士只说对了 3 杯 ？

一个人纯粹靠随机的猜测，能够说对至少 3 杯的概率（即 H_0 成立的情况下，出现这种试验结果的可能性）：

$$\frac{1+C_4^3 C_4^1}{C_8^4} = \frac{17}{70} \approx 0.243$$

显然我们不会对一个概率接近 25% 的随机事件在一次试验中发生而感到惊讶。

试验结果并没有提供不利于H_0的显著性证据，因此不能否定零假设，而应该接受H_0，即应该认为这位女士没有鉴别能力。

3.1.2 假设检验的基本过程

通过一个简单的例题来说明

例**3.1.1** 当包装机器正常工作时，每袋葡萄糖的重量应该是一个服从均值 **0.5 kg**，标准差 **0.015 kg**的随机变量。有一天随机地抽取了 **9** 袋包装好的产品，测量出它们的平均重量是 **0.511 kg**，问这台包装机器是否正常工作？

(假定即使工作异常标准差也不会改变**)**

1. 提出一个统计假设

根据题意每袋产品重量 $X \sim N(\mu, 0.015^2)$，如果机器正常工作，应该是 $\mu = 0.5$，反之应该是 $\mu \neq 0.5$。

因此首先提出统计假设：

$$H_0: \mu = \mu_0\ (= 0.5) \Leftrightarrow H_1: \mu \neq \mu_0\ (\neq 0.5)$$

假设检验的任务就是要根据抽取出的样本，来决定是接受零假设，还是拒绝零假设(接受对立假设)。

2. 选取一个合适的检验统计量

它的分布当零假设成立时应该是已知的，而且一般是从待检验的总体参数的良好的点估计中去寻找。

在例题中需要检验的是总体期望 μ，因此考虑样本均值，

$$\overline{X} \sim N(\mu, \frac{0.015^2}{9})$$

零假设成立时 $(\mu = 0.5)$ 则有：

$$z = \frac{\sqrt{9}(\overline{X} - 0.5)}{0.015} \sim N(0,1)$$

3. 利用零假设成立时检验统计量的分布构造出一个小概率事件

这个小概率就是给定的显著性水平(也称检验水平)，而这个小概率事件就是零假设的拒绝域，并且拒绝域必须和对立假设有关：零假设的拒绝域相当于对立假设的接受域。

在例题中由于样本均值是总体期望 μ 的一个良好的点估计，因此零假设成立($\mu = 0.5$)时，偏差 $|\overline{X} - 0.5|$ 应该比较小，不能够太大。

而如果 $|\overline{X}-0.5|$ 比较大时，自然我们会认为零假设不成立，所以应该接受对立假设。所以零假设 $(\mu = 0.5)$ 的拒绝域的形式就是 ：

$$\text{统计量} |z| = \frac{\sqrt{9}\,|\overline{X}-0.5|}{0.015} > \text{某个常数 } z_0$$

根据检验统计量的分布， 有：

$$\mathrm{P}\{\frac{\sqrt{9}\,|\overline{X}-0.5|}{0.015} > u_{\alpha/2}\} = \alpha$$

这个常数 z_0 就可以取为 $u_{\alpha/2}$

4. 代入样本观察值，如果使得这个小概率事件发生，就否定零假设而去接受对立假设。否则说明样本没有提供否定零假设的显著性证据，因此应该接受零假设。

在这个例题里，检验统计量

$$|z| = \frac{3\times 0.011}{0.015} = 2.2,$$

H_0：$\mu=\mu_0\,(=0.5) \Leftrightarrow H_1$：$\mu\neq\mu_0\,(\neq 0.5)$

的显著水平α的拒绝域就是$\{2.2>u_{\alpha/2}\}$。

假设检验默认的显著水平是$\alpha=0.05$

(1) 如果取$\alpha = 0.05$，则2.2＞常数 $z_0 = 1.96$，说明一个概率为0.05 的随机事件发生了，样本提供了机器异常的显著证据，应该否定零假设；

(2) 如果取 $\alpha = 0.01$，则2.2＜常数 $z_0 = 2.575$，说明一个概率0.01的随机事件没有发生，样本没有提供机器异常的显著证据，应该接受零假设。

□

在不同的显著水平下，可以导致最终得出的检验结论完全不同。这个现象说明了显著水平α对于 H_0的保护：α越小越不容易否定零假设。

3.1.3 Neyman – Pearson 理论

1. 零假设与对立假设

null hypothesis 与 ***alternative hypothesis***

如果有样本 X，取值于样本空间 Λ。我们只知道 X 的分布属于一个分布族$\{F_\theta, \theta \in \Theta\}$。$\Theta_0$ 是参数空间Θ中的一个非空子集，Θ_1 是与Θ_0不相交的参数空间Θ中另一个非空子集。则称：H_0：$\theta \in \Theta_0 \quad \Leftrightarrow \quad H_1$：$\theta \in \Theta_1$
是一个假设检验问题。

H_0 称为零假设，H_1 称为对立假设

Remark

一般来说，零假设总是“受到保护的假设”，没有充分的证据是不能拒绝零假设的。

因此在实际检验问题中检验者总是把自己的倾向性观点作为零假设、再辅助以比较低的检验水平，从而使得没有充分的证据就不能否定它。

2. 否定域(或拒绝域)

有了样本以后需要制定一个法则，根据这个法则去决定究竟是接受零假设，还是拒绝它而接受对立假设，因此一个假设检验的处理，实际上是等价于把样本空间分解成两个不相交的部分：

$$\varLambda_0 \text{ (接受域) 与 } \varLambda_1 \text{ (否定域)}$$

3. 两类检验错误

无论我们采用什么决策，任何一个假设检验都必然可能犯两种错误之一：

第一类错误“ 拒真 ”：零假设本身是正确的却被检验拒绝，接受的是一个错误的对立假设即真实参数 $\theta \in \Theta_0$，但是我们认为它在Θ_1中；

第二类错误“ 采假 ”：零假设本身是错误的而没有被拒绝，最后接受的是一个错误的零假设即真实参数 $\theta \in \Theta_1$，但是我们认为它在Θ_0中。

在理论上可以证明，假设检验的这两类错误的关系如同区间估计中置信度与精度的关系。

如果样本容量固定，则一个检验方法不可能同时使得犯这两类错误的概率都很小。

假设检验和区间估计具有一种内在的联系

实际上，可以把估计未知参数的区间就取成零假设的接受域，反之同理，一个关于参数的假设检验的接受域同样可以作为估计参数的区间。

从发展历史来说先有假设检验的 **N-P** 理论，后有 **Neyman** 的置信区间(**1934**年开始)

4. 功效函数 $\beta_\phi(\theta)$

$$\beta_\phi(\theta) = \mathrm{P}_\theta\{\text{否定零假设 } \mathrm{H}_0\}，\ \theta \in \Theta$$

$$\text{第一类错误的概率} = \begin{cases} \beta_\phi(\theta)，& \text{当 } \theta \in \Theta_0 \\ 0\ ，& \text{当 } \theta \in \Theta_1 \end{cases}$$

$$\text{第二类错误的概率} = \begin{cases} 0\ ，& \text{当 } \theta \in \Theta_0 \\ 1 - \beta_\phi(\theta)\ ，& \text{当 } \theta \in \Theta_1 \end{cases}$$

一个好的检验应该是功效函数 $\beta_\phi(\theta)$在Θ_0中越小越好，而在 Θ_1中越大越好。

Neyman 和 Pearson 提出的解决办法

首先用一个很小的数 $\alpha > 0$ 控制犯第一类错误的概率，在这个基础上再考虑使得犯第二类错误的概率尽量的小。

"犯第一类错误的概率" α 在统计学中称为显著性水平，或者又叫检验水平。

而检验的水平体现了对零假设的保护程度，

显著性水平越小，越不容易否定零假设

例3.1.2　教师给学生出了10 个判断题。为了检验学生是否在猜答案，他采用了如下的决策：

如果答对了7 个或以上的题，就不是凭猜测；

如果答对的题数少于7 个，学生就是在猜答案。

这个决策原则犯第一类错误的概率是多少？

解.　以 p 记答对一个题的概率，则零假设应是：

$$H_0：p = 0.5$$

显然答对的全部题数 $X \sim B(10，p)$。

第一类错误是指：教师认为学生没有猜题而大家的确是靠猜测做题的概率，

因此当 $p = 0.5$ 时，概率 $\mathrm{P}\{X \geq 7\}$ 即为所求，

$$\mathrm{P}\{X \geq 7\} = \mathrm{P}\{X = 7\} + \dots + \mathrm{P}\{X = 10\}$$
$$= 0.1172 + 0.0439 + 0.0098 + 0.0010$$
$$= 0.1719。$$

当学生确实在猜答案时，而认为他们不是猜题的概率是 17% (犯第一类错误的概率)。

如果把决策原则改成 " 答对题数少于 8 个才认为在猜答案 "，则第一类错误的概率降为 0.0547。

□

$\mathrm{P}_{p=0.7}\{X \leq 6\}$ 是真实值0.7时犯第二类错误的概率

例3.1.3　如果一组样本X_1，…，X_n 来自总体 $X \sim N(\mu, \sigma_0^2)$，$\sigma_0^2$ 已知。对于假设：

$$H_0: \mu \le \mu_0 \Leftrightarrow H_1: \mu > \mu_0$$

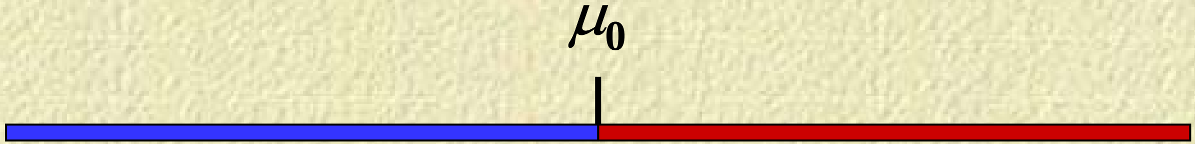
类似例 3.1.1 的讨论，要去构造一个检验：

$\overline{X} >$ 某个常数 C 时拒绝H_0，否则接受。

因为参数空间$\Theta = \mathbf{R}^1$，这种检验的功效函数为：

$$\beta_\phi(\mu) = P_\mu\{\overline{X} > C\} = 1 - \Phi\left(\frac{\sqrt{n}(C-\mu)}{\sigma_0}\right)$$

$\Phi(x)$ 是$N(0,1)$的分布函数，单调非降，故功效函数 $\beta_\phi(\mu) = 1 - \Phi(\frac{n^{1/2}(C-\mu)}{\sigma_0})$ 随着μ的增加而增大，而现在参数空间Θ被分解成两个不相交的部分 H_0 ：$\mu \le \mu_0$ 与 H_1 ：$\mu > \mu_0$



要希望第一类错误概率越小，即在Θ_0中μ应该越向左边偏离 μ_0 来取值；而要希望第二类错误概率越小，在Θ_1中μ应该越向右边远离 μ_0 来取值。

<u>矛盾:两类错误的概率不可能同时都很小</u>

1. $\beta_\phi(\mu)$是μ的单调递增函数，则$\mu \in \Theta_0\ (\mu \le \mu_0)$时 $\beta_\phi(\mu)$在$\mu = \mu_0$达到最大，要使得检验犯第一类错误的概率 $\le \alpha$，只需要取 $\beta_\phi(\mu_0) = \alpha$ 即可。

解出 $C = \mu_0 + \dfrac{u_\alpha \sigma_0}{n^{1/2}}$

因此例题中的一个水平α的检验 ϕ 就是：

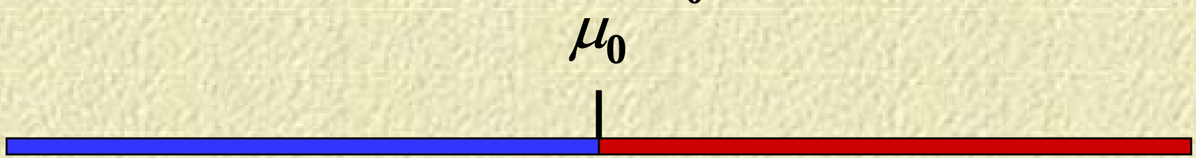
$\overline{X} > \mu_0 + \dfrac{u_\alpha \sigma_0}{\sqrt{n}}$ 时拒绝H_0，否则接受。

而这个检验的功效函数是：

$$\beta_\phi(\mu) = 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + u_\alpha\right) = \Phi\left(\frac{\sqrt{n}(\mu - \mu_0)}{\sigma_0} - u_\alpha\right)$$

2. 详细分析这个检验的功效函数：

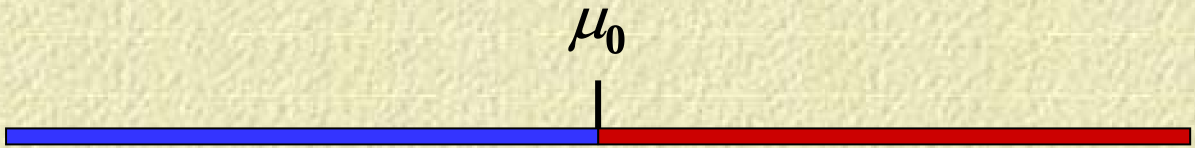
$$\beta_\phi(\mu) = \Phi\left(\frac{\sqrt{n}(\mu - \mu_0)}{\sigma_0} - u_\alpha\right)$$



(1). 真实的μ在$\Theta_1\,(\mu > \mu_0)$中越大，则$\beta_\phi(\mu)$越大，第二类错误的概率就越小，直观理解，这时μ离零假设μ_0越远也就越容易把它们区别开。

如果真实的μ在$\Theta_1\,(\mu > \mu_0)$中越靠近μ_0，则检验犯第二类错误的概率$\approx 1 - \alpha$。

$$\beta_{\phi}(\mu)=\Phi\left(\frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}-u_{\alpha}\right)$$



(2). α 越大 u_{α} 将越小则 $\beta_{\phi}(\mu)$也就越大；反之，α 越小 u_{α} 将越大，则 $\beta_{\phi}(\mu)$也就越小。直观理解 α 大即允许犯第一类错误的概率增加，作为一种补偿，相应地犯第二类错误的概率就应该要降低。

如何控制犯第二类错误的概率？

思想基本上和序贯区间估计的想法一样，

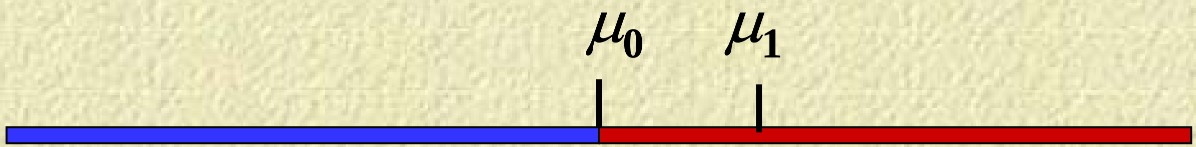
对样本容量的要求

如果在原来的检验问题里，还希望犯第二类错误的概率 $\leq$ 事先指定的一个小的正数 β，即

当 $\mu > \mu_0$ 时 $(\mu \in \Theta_1)$ 有 $\beta_\phi(\mu) \geq 1 - \beta$ ，

按照前面的分析，当 μ 非常接近 μ_0 时这个要求实际上办不到，所以可行做法是放松条件：

$\beta_\phi(\mu) \geq 1 - \beta$，当 $\mu > \mu_1$ (μ_1 比 μ_0 稍大)



同样道理，因为 $\beta_\phi(\mu)$是μ 的单调递增函数，因此只要 $\beta_\phi(\mu_1) \geq 1 - \beta$ ，就可以保证：当 $\mu > \mu_1$ 时，都有 $\beta_\phi(\mu) \geq 1 - \beta$。即，

$$\beta_\phi(\mu_1) = \Phi(\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma_0} - u_\alpha) \geq 1 - \beta$$

所以只要 $n \geq \dfrac{\sigma_0^2 (u_\alpha + u_\beta)^2}{(\mu_1 - \mu_0)^2}$，就可以保证犯第一类错误的概率 $\leq \alpha$，同时当参数真实值大于μ_1 时，犯第二类错误的概率 $\leq \beta$。 □

3.1.4 统计显著并不意味实际也显著

一个统计显著的结论在实际问题中可能并不重要。

例如为了检验一枚硬币是否均匀，把它抛掷100万(10^6)次，其中正面出现了50.1万次。

从实际意义来看正面只比反面多出两千次，占总抛掷次数的0.002，因此这枚硬币应该是均匀的。但如果做关于总体比例的假设检验，将会否定硬币均匀的零假设。(z=2)

第3.2节 重要的参数检验

这一节主要讨论

1. 正态总体均值、方差的检验问题，包括两个独立正态总体之间均值差、方差比的检验；
2. 指数总体参数的检验问题；
3. 两点分布参数的检验，主要是一个总体百分比的检验、两个独立总体百分比差异的检验。

只构造水平为α的检验，不讨论犯第二类错误的概率

3.2.1 正态总体参数的检验

1. 均值的检验

样本X_1，…，X_n 来自总体 $X \sim N(\mu, \sigma_0^2)$

一个总体的均值检验问题

1.1 方差已知 σ_0^2 (u 检验)

$$1.1a.\quad H_0: \mu = \mu_0 \Leftrightarrow H_1: \mu \neq \mu_0$$

$$1.1b.\quad H_0: \mu \leq \mu_0 \Leftrightarrow H_1: \mu > \mu_0$$

$$1.1c.\quad H_0: \mu \geq \mu_0 \Leftrightarrow H_1: \mu < \mu_0$$

利用样本均值来作为检验统计量，分布是：

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma_0} \sim N(0,1)$$

当零假设 H_0 ：$\mu=\mu_0$ 成立时有：

$$\mathrm{P}\{\frac{\sqrt{n}\,|\,\overline{X}-\mu_0\,|}{\sigma_0} > u_{\alpha/2}\} = \alpha$$

因此假设 *1.1a*. H_0：$\mu=\mu_0 \Leftrightarrow \mathrm{H}_1$：$\mu \neq \mu_0$ 一个水平 α 的拒绝域应该是

$$\frac{\sqrt{n}\,|\,\overline{X}-\mu_0\,|}{\sigma_0} > u_{\alpha/2}$$

假设 *1.1b*. $\mathbf{H}_0: \mu \le \mu_0 \Leftrightarrow \mathbf{H}_1: \mu > \mu_0$

例题3.1.3中已经讨论，一个水平α的拒绝域是

$$\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma_0} > u_\alpha$$

假设 *1.1c*. $\mathbf{H}_0: \mu \ge \mu_0 \Leftrightarrow \mathbf{H}_1: \mu < \mu_0$

同理讨论，一个水平α的拒绝域如下

$$\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma_0} < -u_\alpha$$

p-值 (*p*-value)

p-值是零假设成立时得到所观测数据或者更为极端数据的概率。

p-值越小越应该否定零假设。有时候 *p*-值又被理解成拒绝零假设的最小显著性水平。

p-值的计算依赖于所使用的检验统计量以及对立假设是单边假设还是双边假设。

拟合优度检验中的 *p*-值又称为是拟合优度，即数据与总体分布拟合的程度。

例3.2.1　从正态总体 $X \sim N(\mu, 10.1^2)$中随机抽取 100 个样本，样本均值是 59.22。检验以下假设：

$$H_0：\mu = 60 \Leftrightarrow H_1：\mu < 60$$

解. 从对立假设的形式知道零假设的水平α的拒绝域应该是样本均值偏小，即

$$z = \frac{10 \times (59.22 - 60)}{10.1} < -u_\alpha$$

根据样本数据计算出的 $z = -0.77$，

$$u_{0.05} = 1.64，\quad u_{0.20} = 0.84，\quad u_{0.25} = 0.68$$

甚至在0.2的水平下也不会拒绝零假设，但是在水平0.25下将会拒绝零假设。

如果采用 *EXCEL*或 *SPSS* 等软件处理这个问题，则将输出一个检验统计量 z 对应的 p-值 ：

$$p = P\{ z < -0.77 \} = 0.22065$$

只要这个p-值小于事先给出的水平α，就否定零假设

□

Remark 如果对立假设是 H_1 ： $\mu \neq 60$ ，则检验的拒绝域的形式应该为 $|z| > u_{\alpha/2}$。

因此在0.4的水平下也不会拒绝零假设，但是在水平0.50下将会拒绝零假设。

相应的 p-值将是 $p = P\{ |z| > 0.77 \} = 0.4413$

1.2 方差σ^2未知 (t 检验)

样本X_1，…，X_n 来自总体 $X \sim N(\mu, \sigma_0^2)$

$$1.2a.\quad H_0: \mu = \mu_0 \Leftrightarrow H_1: \mu \neq \mu_0$$
$$1.2b.\quad H_0: \mu \leq \mu_0 \Leftrightarrow H_1: \mu > \mu_0$$
$$1.2c.\quad H_0: \mu \geq \mu_0 \Leftrightarrow H_1: \mu < \mu_0$$

检验的构造与方差已知时基本相同，需要使用抽样分布中的定理1.3.1

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t(n-1)$$

假设 *1.2a*. $\mathrm{H}_0：\mu = \mu_0 \Leftrightarrow \mathrm{H}_1：\mu \neq \mu_0$

一个水平α的拒绝域是

$$\frac{\sqrt{n}\,|\,\overline{X} - \mu_0\,|}{S} > t_{\alpha/2}(n-1)$$

假设 *1.2b*. $\mathrm{H}_0：\mu \leq \mu_0 \Leftrightarrow \mathrm{H}_1：\mu > \mu_0$

一个水平α的拒绝域是

$$\frac{\sqrt{n}(\overline{X} - \mu_0)}{S} > t_{\alpha}(n-1)$$

假设 *1.2c*. $\mathrm{H}_0：\mu \geq \mu_0 \Leftrightarrow \mathrm{H}_1：\mu < \mu_0$

一个水平α的拒绝域是

$$\frac{\sqrt{n}(\overline{X} - \mu_0)}{S} < -t_{\alpha}(n-1)$$

两个总体均值差的检验

假定两组简单随机样本 X_1，…，X_{n_1} 与 Y_1，…，Y_{n_2} 分别来自两个独立的正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 与 $Y \sim N(\mu_2, \sigma_2^2)$，

1.3 方差 σ_1^2、σ_2^2已知 (u 检验)

1.3a. H_0：$\mu_1 - \mu_2 = \delta \Leftrightarrow H_1$：$\mu_1 - \mu_2 \neq \delta$

1.3b. H_0：$\mu_1 - \mu_2 \leq \delta \Leftrightarrow H_1$：$\mu_1 - \mu_2 > \delta$

1.3c. H_0：$\mu_1 - \mu_2 \geq \delta \Leftrightarrow H_1$：$\mu_1 - \mu_2 < \delta$

因为两个总体的样本均值分别服从正态分布，

$$\overline{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1}), \quad \overline{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

所以有

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

则当H_0：$\mu_1 - \mu_2 = \delta$成立时，统计量 $z \sim N(0, 1)$

$$z = \frac{(\overline{X} - \overline{Y}) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

假设 *1.3a*. $H_0：\mu_1 - \mu_2 = \delta \Leftrightarrow H_1：\mu_1 - \mu_2 \neq \delta$ 的一个水平α的拒绝域是

$$\frac{|\bar{X}-\bar{Y}-\delta|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} > u_{\alpha/2}$$

假设 *1.3b*. $H_0：\mu_1 - \mu_2 \leq \delta \Leftrightarrow H_1：\mu_1 - \mu_2 > \delta$ 的一个水平α的拒绝域是

$$\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} > u_{\alpha}$$

假设 *1.3c*. $H_0：\mu_1 - \mu_2 \geq \delta \Leftrightarrow H_1：\mu_1 - \mu_2 < \delta$ 的一个水平α的拒绝域是

$$\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < -u_{\alpha}$$

1.4 方差 σ_1^2、σ_2^2未知但相等(t 检验)

$\sigma_1^2 \neq \sigma_2^2$ 时称为Behrens-Fisher问题

使用抽样分布中的定理1.3.2，定义：

$$S_W^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

则有 $$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

则当H_0：$\mu_1 - \mu_2 = \delta$ 成立时，统计量：

$$\frac{\overline{X}-\overline{Y}-\delta}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

假设 1.4a. $H_0: \mu_1 - \mu_2 = \delta \Leftrightarrow H_1: \mu_1 - \mu_2 \neq \delta$

的一个水平α的拒绝域是 $$\frac{|\bar{X}-\bar{Y}-\delta|}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha/2}(n_1+n_2-2)$$

假设 1.4b. $H_0: \mu_1 - \mu_2 \leq \delta \Leftrightarrow H_1: \mu_1 - \mu_2 > \delta$

的一个水平α的拒绝域是 $$\frac{\bar{X}-\bar{Y}-\delta}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha}(n_1+n_2-2)$$

假设 1.4c. $H_0: \mu_1 - \mu_2 \geq \delta \Leftrightarrow H_1: \mu_1 - \mu_2 < \delta$

的一个水平α的拒绝域是 $$\frac{\bar{X}-\bar{Y}-\delta}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < -t_{\alpha}(n_1+n_2-2)$$

2. 方差的检验

一个总体的方差检验问题

2.1 均值 μ_0已知(χ^2 检验)

$$2.1a.\quad H_0：\sigma^2 = \sigma_0^2 \Leftrightarrow H_1：\sigma^2 \neq \sigma_0^2$$

$$2.1b.\quad H_0：\sigma^2 \leq \sigma_0^2 \Leftrightarrow H_1：\sigma^2 > \sigma_0^2$$

$$2.1c.\quad H_0：\sigma^2 \geq \sigma_0^2 \Leftrightarrow H_1：\sigma^2 < \sigma_0^2$$

$$\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \sim N(0,1) \quad or \quad \sum_{k=1}^{n} \frac{(X_k - \mu_0)^2}{\sigma^2} \sim \chi^2(n)?$$

当零假设 H_0：$\sigma^2 = \sigma_0^2$ 成立时有：

$$\sum_{k=1}^{n} \frac{(X_k - \mu_0)^2}{\sigma_0^2} \sim \chi^2(n)$$

因此假设 *2.1a*. H_0：$\sigma^2 = \sigma_0^2 \Leftrightarrow H_1$：$\sigma^2 \neq \sigma_0^2$ 一个水平 α 的拒绝域应该是

$$\sum_{k=1}^{n} \frac{(X_k - \mu_0)^2}{\sigma_0^2} > \chi^2_{\alpha/2}(n) \quad or \quad \sum_{k=1}^{n} \frac{(X_k - \mu_0)^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(n)$$

同理<u>假设 *2.1b*</u>. $\mathbf{H_0}$：$\sigma^2 \le \sigma_0^2 \Leftrightarrow \mathbf{H_1}$：$\sigma^2 > \sigma_0^2$
一个水平α的拒绝域是

$$\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2} > \chi_\alpha^2(n)$$

<u>假设 *2.1c*</u>. $\mathbf{H_0}$：$\sigma^2 \ge \sigma_0^2 \Leftrightarrow \mathbf{H_1}$：$\sigma^2 < \sigma_0^2$

一个水平α的拒绝域是

$$\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2} < \chi_{1-\alpha}^2(n)$$

2.2 均值 μ 未知(χ^2 检验)

$$2.2a.\quad H_0: \sigma^2 = \sigma_0^2 \Leftrightarrow H_1: \sigma^2 \neq \sigma_0^2$$
$$2.2b.\quad H_0: \sigma^2 \leq \sigma_0^2 \Leftrightarrow H_1: \sigma^2 > \sigma_0^2$$
$$2.2c.\quad H_0: \sigma^2 \geq \sigma_0^2 \Leftrightarrow H_1: \sigma^2 < \sigma_0^2$$

根据抽样分布中定理1.3.1中关于样本方差的分布，有

$$\sum_{k=1}^{n} \frac{(X_k - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

当零假设 H_0：$\sigma^2 = \sigma_0^2$ 成立时有：

$$\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{\sigma_0^2}\sim\chi^2(n-1)$$

因此假设 <u>2.2*a*</u>. H_0：$\sigma^2 = \sigma_0^2 \Leftrightarrow H_1$：$\sigma^2 \neq \sigma_0^2$ 一个水平α的拒绝域应该是

$$\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{\sigma_0^2}>\chi^2_{\alpha/2}(n-1)\quad or\quad \sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{\sigma_0^2}<\chi^2_{1-\alpha/2}(n-1)$$

同理<u>假设 *2.2b*</u>. $\mathbf{H_0}$：$\sigma^2 \le \sigma_0{}^2 \Leftrightarrow \mathbf{H_1}$：$\sigma^2 > \sigma_0{}^2$

一个水平α的拒绝域是

$$\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{\sigma_0^2} > \chi_{\alpha}^{2}(n-1)$$

<u>假设 *2.2c*</u>. $\mathbf{H_0}$：$\sigma^2 \ge \sigma_0{}^2 \Leftrightarrow \mathbf{H_1}$：$\sigma^2 < \sigma_0{}^2$

一个水平α的拒绝域是

$$\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{\sigma_0^2} < \chi_{1-\alpha}^{2}(n-1)$$

两个总体方差比的检验

假定两组简单 随机样本 X_1，…，X_{n_1} 与 Y_1，…，Y_{n_2} 分别来自两个独立的正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 与 $Y \sim N(\mu_2, \sigma_2^2)$，

2.3 均值 μ_1、μ_2 都未知 (F 检验)

$$2.3a.\quad H_0: \sigma_1^2 = \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$$

$$2.3b.\quad H_0: \sigma_1^2 \leq \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 > \sigma_2^2$$

$$2.3c.\quad H_0: \sigma_1^2 \geq \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 < \sigma_2^2$$

由于两个总体的样本方差 S_1^2、S_2^2 分别是它们各自方差 σ_1^2、σ_2^2的无偏估计，因此，

对于假设 *2.3a*. $H_0: \sigma_1^2 = \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$ 拒绝域的形式应该是 S_1^2/S_2^2 偏大或偏小；

同理假设 *2.3b*. $H_0: \sigma_1^2 \leq \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 > \sigma_2^2$ 拒绝域的形式应该是 S_1^2/S_2^2 偏大；

而假设 *2.3c*. $H_0: \sigma_1^2 \geq \sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 < \sigma_2^2$ 拒绝域的形式应该是 S_1^2/S_2^2 偏小。

利用抽样分布中定理1.3.2，

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

假设 *2.3a*. $H_0：\sigma_1^2 = \sigma_2^2 \Leftrightarrow H_1：\sigma_1^2 \neq \sigma_2^2$ 的一个水平α的拒绝域是：

$$S_1^2/S_2^2 > F_{\alpha/2}(n_1 - 1, n_2 - 1) \text{ 或 }$$
$$S_1^2/S_2^2 < F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$$

同理假设 *2.3b*. $H_0：\sigma_1^2 \leq \sigma_2^2 \Leftrightarrow H_1：\sigma_1^2 > \sigma_2^2$ 一个水平α的拒绝域是

$$S_1^2/S_2^2 > F_{\alpha}(n_1 - 1, n_2 - 1)$$

假设 *2.3c*. $H_0：\sigma_1^2 \geq \sigma_2^2 \Leftrightarrow H_1：\sigma_1^2 < \sigma_2^2$ 一个水平α的拒绝域是

$$S_1^2/S_2^2 < F_{1-\alpha}(n_1 - 1, n_2 - 1)$$

例3.2.2　　从一批矿砂中随机抽取了5 个样品来测量镍的百分比含量，数据如下：

3.25，3.27，3.24，3.26，3.24

设测量值总体是正态分布，在显著性水平**0.01**下能不能认为这批矿砂的镍含量均值为 **3.25** ？

解.　这是一个总体方差未知，关于均值的检验：

$$H_0：\mu = 3.25 \Leftrightarrow H_1：\mu \neq 3.25$$

采用t 检验，拒绝域为：

$$\frac{\sqrt{5}\,|\,\overline{X} - 3.25\,|}{S} > t_{0.01/2}(4)$$

代入数据，计算出

样本均值为 3.252，样本标准差为0.01304；

查表得到 t 分布的上侧 0.005 分位点

$$t_{0.005}(4) = 4.6041$$

因此有统计量 0.343 ＜ 4.6041

即样本没有提供否定零假设的显著证据。 □

$t_{0.25}(4) = 0.7407$ 说明水平 0.5 下也不会否定零假设

相应的 p-值是 $p = 0.74887$

例3.2.3　从马克·吐温 (**Mark Twain**) 和斯诺特格拉斯(**Snodgrass**) 作品中各选了 **8** 篇和 **10** 篇小品文。得到**3** 个字母组成单词的比例如下：

Mark Twain	Snodgrass
0.225, 0.262, 0.217, 0.240, 0.230, 0.229, 0.235, 0.217。	0.209, 0.205, 0.196, 0.210, 0.202, 0.207, 0.224, 0.223, 0.220, 0.201。

设这两组数据都来自正态总体而且独立的，问在检验水平 **0.05** 下，它们的方差是否相同？

解. 这是关于两个独立总体方差是否相同的检验，两个总体的期望都未知，采用**2.3a.**的 F 检验：

$$H_0：\sigma_1^2 = \sigma_2^2 \Leftrightarrow H_1：\sigma_1^2 \neq \sigma_2^2$$

取检验统计量

$$F = S_1^2/S_2^2$$

$$= \frac{2.12125 \times 10^{-4}}{9.3344 \times 10^{-5}} = 2.2725$$

拒绝域的形式为：

$$F > F_{0.025}(7,9) \text{ 或者 } F < F_{0.975}(7,9)$$

查表得到 $F_{0.025}(7,9)=4.20$,

但是一般查不到 $F_{0.975}(7,9)$的数值

必须利用 F 分布的性质,

$$F_{1-\alpha}(m,n) = 1/F_{\alpha}(n,m)$$

所以有

$$F_{0.975}(7,9)=1/F_{0.025}(9,7)=1/4.82=0.2075$$

只有当检验统计量 F (= 2.2725) 大于 4.20 或小于0.2075 时才否定零假设，因此可以认为两个总体方差相同。

相应的 p-值是 $p = 0.125$ □

例3.2.4　　在刚才例题里已经通过检验认为方差相同。即马克·吐温与斯诺特格拉斯 的作品中 3 个字母的单词占的比例来自相同方差的正态总体。 那么在水平 0.05 下，能不能认为这两个总体均值有显著差异？

解. 采用*1.4a.* 的 t 检验($\delta = 0$)，要检验的是：

$$H_0：\mu_1 = \mu_2 \Leftrightarrow H_1：\mu_1 \neq \mu_2$$

取检验统计量为

$$T = \frac{\overline{X} - \overline{Y}}{S_W\sqrt{\frac{1}{8} + \frac{1}{10}}}$$

拒绝域是 $|T| > t_{0.025}(16) = 2.1199$

根据例题3.2.3 提供的数据，算出：

$$\bar{x} = 0.2319\,,\quad \bar{y} = 0.2097$$

$$S_w = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{1.4531\times 10^{-4}} = 0.0121$$

最后得到检验统计量 T 的观察值

$$|T| = 3.8679 > t_{0.025}(16) = 2.1199$$

否定零假设，认为这两位作家写作习惯显著不同。

注：$t_{0.005}(16) = 2.9208$，即水平0.01下也否定。

相应的 p-值是 $p = 0.0014$ □

Remark

假如现在新发现了一部佚名的文集，怀疑可能是马克·吐温或者是斯诺特格拉斯的作品。那么从统计学的角度又应该如何来讨论 ？

这里需要检验的是：一组数据是否来自于一个已知的总体。统计理论中称为<u>非参数检验</u>。

下一节的拟合优度检验能够处理这类问题，“拟合优度”的含义是：一组样本数据和某个总体拟合的程度。

3. 成对数据的检验问题

这也是一种 t 检验，常常用来检验性能、及方法或者属性是否有显著差异。

例如要比较不同的药品的疗效，有两个办法：

A. 把患者分成两组，一组服用甲另一组服用乙；

B. 同一组患者各服用甲、乙。

A 方法要检验的就是两个独立正态总体均值是否相同的问题 (应该事先检验方差是否相同)

B 方法中的数据来自正态总体，但不应该认为是独立的，因此要检验的是这些成对数据的差来自的正态总体的均值是否为 0 。

例3.2.5　**10** 个失眠患者分别服用了甲、乙两种安眠药，延长睡眠的时间如下(单位：小时)：

甲	1.9	0.8	1.1	0.1	– 0.1	4.4	5.5	1.6	4.6	3.4
乙	0.7	– 1.6	– 0.2	– 1.2	– 0.1	3.4	3.7	0.8	0.0	2.0

解.　这十组成对数据的差(甲-乙) 是：

1.2　2.4　1.3　1.3　0　1　1.8　0.8　4.6　1.4

可以认为它们来自正态总体 $X \sim N(\mu, \sigma^2)$

需要检验的是*1.2a*.　$\mathbf{H_0}: \mu = \mu_0 \Leftrightarrow \mathbf{H_1}: \mu \neq \mu_0$

由于总体方差未知，采用 t 检验，

否定域为：$$|T|=\frac{\sqrt{10}\,|\overline{X}|}{S}>t_{0.025}(9)$$

计算出数据的

样本均值为 1.58，样本标准差 1.23

查表得到 $t_{0.025}(9)=2.2622$

因为 $|T|=4.06>2.2622$，在水平 0.05 下否定零假设，认为两种药的疗效有显著差异。

注：$t_{0.005}(9)=3.2498$，即水平0.01下也否定。

相应的 p-值是 $p=0.0028$ □

3.2.2 指数总体参数的检验

如果一组样本X_1，…，X_n 来自参数为 θ 的指数总体，密度函数是：

$$f(x) = \theta e^{-\theta x}, \quad x > 0$$

可以证明，对于样本均值，有 $2n\theta\overline{X} \sim \chi^2(2n)$

需要对参数θ(平均寿命的倒数) 作检验：

$$a.\quad H_0: \theta = \theta_0 \Leftrightarrow H_1: \theta \neq \theta_0$$

$$b.\quad H_0: \theta \leq \theta_0 \Leftrightarrow H_1: \theta > \theta_0$$

$$c.\quad H_0: \theta \geq \theta_0 \Leftrightarrow H_1: \theta < \theta_0$$

假设 a. H_0：$\theta=\theta_0 \Leftrightarrow H_1$：$\theta\neq\theta_0$
的一个水平α检验的拒绝域为：

$$2n\theta_0\overline{X} < \chi^2_{1-\alpha/2}(2n) \quad or \quad 2n\theta_0\overline{X} > \chi^2_{\alpha/2}(2n)$$

假设 b. H_0：$\theta\leq\theta_0 \Leftrightarrow H_1$：$\theta>\theta_0$
的一个水平α的检验的拒绝域为：

$$2n\theta_0\overline{X} < \chi^2_{1-\alpha}(2n)$$

假设 c. H_0：$\theta\geq\theta_0 \Leftrightarrow H_1$：$\theta<\theta_0$
的一个水平α的检验的拒绝域为：

$$2n\theta_0\overline{X} > \chi^2_{\alpha}(2n)$$

例3.2.6　一种电子元件要求寿命不得低于1000小时，现在随机抽取15个，测量出样本平均寿命为900小时，不妨假定这种元件寿命是指数分布，在0.05的水平下能否认为这批元件合乎要求？

解. 需要检验 $H_0: \theta \le 0.001 \Leftrightarrow H_1: \theta > 0.001$

拒绝域为 $K^2 < \chi_{0.95}{}^2(30)\ (= 18.493)$，

这里统计量 $K^2 = \dfrac{30 \times 900}{1000} = 27$。

样本没有提供元件不合格的显著证据，即应该接收这批元件。

相应的 p-值是 $p = 0.3767$ □

练习3.2.7

分别计算例题3.2.4 - 例题3.2.6 中的 p-值。

练习3.2.8

利用*Gamma* 分布性质证明，如果 X_1，…，X_n 独立同分布于参数为 θ 的指数分布，密度函数为 $p(x) = \theta e^{-\theta x}$，$x > 0$，则有

$$2\theta(X_1 + \ldots + X_n) \sim \chi^2(2n)。$$

3.2.3 两点分布参数的检验

总体属性比例的检验

假定总体中具有某种属性的比例是 p ，从总体抽取了n 个观察值，p_s 记样本中这种属性的百分比，

$$H_0：p = p_0 \Leftrightarrow H_1：p \neq p_0$$

的水平α 检验的拒绝域近似为：

$$|z| = \frac{|p_s - p_0|}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} > u_{\alpha/2}$$

两个总体比例差的检验

假定两个独立总体某种属性比例分别是 p_1 和 p_2，分别抽取n_1、n_2个观察值，样本比例是 p_{s1}、p_{s2}。则

$$H_0: p_1 - p_2 = 0 \Leftrightarrow H_1: p_1 - p_2 \neq 0$$

的水平α 检验的拒绝域近似为：

$$|z| = \frac{|p_{s1} - p_{s2}|}{\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1} + \dfrac{p_{s2}(1-p_{s2})}{n_2}}} > u_{\alpha/2}$$

例3.2.9　1987 年Diane 承诺为费城的Belmont小学毕业的112名学生中考上大学的人提供学费。6年后这些学生中的45%高中毕业，而相应的1986年Belmont 小学毕业生中只有28%的人在 6 年以后高中毕业。新闻报道由此下结论说这两个总体的百分比差异显著。

解.　不妨假定1986年的毕业生人数也是 112 人，

$$|z| = \frac{|0.45 - 0.28|}{\sqrt{\dfrac{0.45 \times 0.55}{112} + \dfrac{0.28 \times 0.72}{112}}} \approx 2.68 > 1.96$$

相应的 p-值为 $p = P\{\,|z| > 2.68\,\} = 0.007$　□

3.2.4 似然比检验(*Likelihood ratio test*)

这是基于 **Fisher** 的"似然原理"建立的构造参数检验的一种比较统一的方法

由于似然函数 $L(x, \theta)$ 的大小体现了在样本给定以后参数 θ 的"似然性"，考虑如下假设：

$$H_0：\theta \in \Theta_0 \quad \Leftrightarrow \quad H_1：\theta \in \Theta_1$$

这里 Θ_0 与 Θ_1 构成 Θ 的一个划分。

"似然比" (统计量)定义成

$$LR = \frac{\sup\{L(x, \theta),\ \theta \in \Theta_0\}}{\sup\{L(x, \theta),\ \theta \in \Theta\}}$$

显然有 $0 < \mathrm{LR} \le 1$，根据极大似然估计的性质，当样本容量n 足够大时，极大似然估计应该以很大的概率非常接近于参数 θ 的真实值，因此如果零假设为真，则极大似然估计应该很接近Θ_0 ，甚至就落在Θ_0里，因此LR应该以相当大的概率接近甚至等于 1。反之如果零假设不真，则LR应该远离 1 而接近于 0 。

只要事先给出一个检验水平α，计算常数C ：

$$\sup\{P_\theta(\mathrm{LR} < C),\ \theta \in \Theta_0\} = \alpha$$

从而零假设 H_0：$\theta \in \Theta_0$的拒绝域就是 $\{\mathrm{LR} < C\}$

练习3.2.10

样本X_1，…，X_n 来自总体 $X \sim N(\mu_0, \sigma^2)$，$\mu_0$ 已知，采用似然比检验的方法给出假设

$$H_0：\sigma^2 = \sigma_0^2 \Leftrightarrow H_1：\sigma^2 \neq \sigma_0^2$$

的一个水平α的拒绝域。

练习3.2.11

讨论区间估计中置信区间与假设检验中拒绝域的关系。

第3.3节　非参数检验

参数检验是在假定总体分布类型已知，对于总体的未知参数或者数字特征的检验。而非参数检验则不同，它是对于总体分布类型的检验。

其中最重要的是拟合优度检验 (***Goodness-of-fit test***)，即一组样本数据是否来自某种分布的问题。

此外，还有一些重要的非参数检验，包括判断独立性的列联表检验(卡方分析)，基于顺序统计量的秩检验、符号检验等等。

3.3.1 卡方 检验

1. 卡方检验的想法

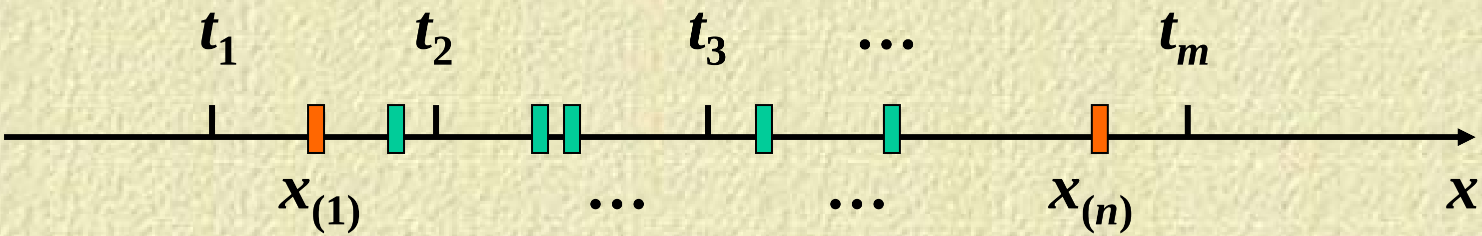
如果一组样本 X_1，…，X_n 来自分布 F，需要检验是如下问题：

$$\mathrm{H}_0: F = F_0 \Leftrightarrow \mathrm{H}_1: F \neq F_0$$

从理论上来说无论 F 是离散还是连续分布，卡方检验都可以处理；不过它更适用于离散的总体，对于连续的总体 F，采用 **Kolmogrov** 检验更好。

K.Pearson 的拟合优度检验思想

在实数轴上取 m 个点把 R^1 分成 $m+1$ 个部分，以 v_i 表示落在第i 个区间里的样本个数，p_i 是总体随机变量 X 在这个区间中的概率：



Remark

F_0 是离散分布时，m 及分点 t_i 的选取很自然；而F_0 是连续分布时，一般建议取 $m \le n/5$ ，再取一个稍大的区间$[a，b] \supset (x_{(1)}，x_{(n)})$，把$[a，b]$ m 等分得到分点t_i。

当零假设 H_0：$F = F_0$ 成立时p_i 可以计算出：

$p_i = F_0(t_i) - F_0(t_{i-1})$，$1 \le i \le m+1$；

这里 $F_0(t_0) = 0$，$F_0(t_{m+1}) = 1$

n 充分大时，频率 v_i/n 与概率 p_i 应该相当接近，因此如果零假设成立则统计量：

$$K^2 = \sum_{i=1}^{m+1} \frac{n}{p_i}\left(\frac{v_i}{n} - p_i\right)^2$$

应该偏小，反之则可以否定零假设 H_0：$F = F_0$。

1900年K.Pearson 证明了极限分布 $K^2 \to \chi^2(m)$，因此 H_0 的一个水平α 拒绝域近似为 $K^2 > \chi_\alpha^2(m)$。

2. 完全已知离散分布的卡方检验

总体 X 只可能取有限个值a_i，$1 \le i \le k$。相应地，样本 X_1，…，X_n中取值为a_i 的个数为v_i，$1 \le i \le k$。需要检验：

$$H_0：P\{X = a_i\} = p_i，\ 1 \le i \le k$$

取检验统计量：

$$K^2 = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{v_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{(v_i - np_i)^2}{np_i} = \frac{1}{n}\sum_{i=1}^{k} \frac{v_i^2}{p_i} - n$$

则H_0 的一个水平α 检验的拒绝域为 $K^2 > \chi_\alpha^2(k-1)$

3. 含未知参数离散分布的卡方检验

如果总体 X 的分布含有 r 个未知参数， **Fisher** 证明了在 $\mathrm{H_0}$ 成立的条件下用极大似然估计估计出这 r 个参数，计算出相应的概率 $\hat{p}_i$， $1 \le i \le k$ ，则修改自由度后**Pearson** 的结论仍然成立，即：

取检验统计量：

$$K^2 = \sum_{i=1}^{k} \frac{n}{\hat{p}_i} \left(\frac{v_i}{n} - \hat{p}_i\right)^2$$

则$\mathrm{H_0}$ 的一个水平α的检验的拒绝域为 $K^2 > \chi_\alpha^2(k - r - 1)$

例3.3.1　**Mendel** 的遗传学例子

Mendel 研究豌豆时发现豌豆有两种特性：圆与皱、黄与绿，他观察了 **556** 颗豌豆：

圆黄	皱黄	圆绿	皱绿	(总数)
315	101	108	32	(556)

而根据他的遗传学理论，**Mendel** 认为这些组合关系应该有理论上的概率：

圆黄	皱黄	圆绿	皱绿	(概率)
9/16	3/16	3/16	1/16	(1)

解. 总体分布的 $k = 4$，对应 K^2 统计量为：

$$K^2 = \frac{1}{n}\sum_{i=1}^{k}\frac{v_i^2}{p_i} - n$$

$$= \frac{1}{556}\{\frac{315^2}{9/16} + \frac{101^2}{3/16} + \frac{108^2}{3/16} + \frac{32^2}{1/16}\} - 556 = 0.47$$

$\chi_{0.05}{}^2(3)=7.815$，$\chi_{0.90}{}^2(3)=0.584$，$\chi_{0.95}{}^2(3)=0.352$

甚至在水平0.90下都可以接受零假设，即认为Mendel 的遗传学理论是正确的。

从p-值的角度拟合优度 $p = P\{\chi^2(3) > 0.47\}$这个值是0.9254，理论分布与实际数据相当吻合。

□

例3.3.2　**Rutherford** 与**Geiger** 的放射性衰变

观测一个放射源，每7.5秒记录一次 α 粒子个数，一共进行了 **2608** 次观测，数据如下：

放射出的粒子数	观测到的次数	放射出的粒子数	观测到的次数
0	57	6	273
1	203	7	139
2	383	8	45
3	525	9	27
4	532	≥10	16
5	408		

是否能认为衰变产生粒子个数服从泊松分布？

解. 以 X 记每 7.5 秒衰变产生的 α 粒子个数，则如果 X 服从泊松分布，应该有：

$$H_0: P\{X=i\}=\frac{\lambda^i}{i!}e^{-\lambda},\ i\geq 0$$

为了方便，实际检验的是截尾泊松分布：

$$H_0: p_i=P\{X=i\}=\frac{\lambda^i}{i!}e^{-\lambda},\ 0\leq i\leq 9$$

$$p_{10}=P\{X=10\}=1-(p_0+..+p_9)$$

λ 的极大似然估计为

$$=\frac{0\times 57+1\times 203+\ldots+10\times 16}{2608}=3.87$$

从而算出 p_0，…，p_{10}，得到检验统计量 K^2

$$K^2=\sum_{i=0}^{10}\frac{n}{\hat{p}_i}\left(\frac{v_i}{n}-\hat{p}_i\right)^2$$

$n = 2608$，v_i 是 2608 次观测中粒子个数恰好是 i 个的观测次数：$v_0 = 57$, $v_1 = 203$, …, $v_{10} = 16$。

$$\hat{p}_i=\frac{3.87^i}{i\,!}e^{-3.87}\ ,\quad 0\le i\le 9 \qquad\qquad \hat{p}_{10}=1-\sum_{i=0}^{9}\hat{p}_i$$

检验统计量 $K^2 = 13.064$，自由度是 $11 - 1 - 1 = 9$ ；而分位点 $\chi_{0.05}{}^2(9) = 16.9$， $\chi_{0.15}{}^2(9) = 13.288$。

水平 0.05 下不能拒绝零假设，即认为粒子数服从泊松分布(拟合优度为0.1597)。

练习3.3.3　抛掷一枚骰子63次，结果如下

点数	1	2	3	4	5	6
次数	10	9	11	8	12	13

这颗骰子能否被认为是均匀的？

练习3.3.4

收集某种公开发行的彩票开奖结果，讨论中奖号码是否具有随机性。

练习3.3.5

用计算机生成 100 个随机数，或者查找一个随机数表取 50 个随机数，讨论所得结果。

3.3.2 Kolmogrov 检验

如果一组样本 X_1，…，X_n 来自分布 F，考虑检验问题：

$$\mathrm{H}_0：F(x)=F_0(x) \Leftrightarrow \mathrm{H}_1：F(x)\neq F_0(x)$$

首先计算经验分布函数：
$$F_n(x)=\begin{cases} 0, & x\le x_{(1)} \\ \dfrac{k}{n}, & x_{(k)}<x\le x_{(k+1)} \\ 1, & x>x_{(n)} \end{cases}$$

F_n 与 F_0 的一致距离定义为：

$$D_n=\sup|F_n(x)-F_0(x)|，对所有 x\in\mathrm{R}^1$$

Glivenko-Cantelli 定理

如果 F_n 是分布函数 F_0 的经验分布函数，则 一致距离D_n几乎处处收敛到零，即

$$\mathbf{P}\{\lim D_n = 0\} = 1。$$

Kolmogrov 定理

当 F_0 是一个连续分布时， D_n 的极限分布是

$$\mathbf{P}\{\, n^{1/2} D_n \le x \,\} \to Q(x)，对所有 x \in \mathbf{R}^1$$

$$Q(x) = \sum_{k=-\infty}^{k=+\infty} (-1)^k \exp(-2k^2 x^2)\,, \quad x > 0$$

D_n 的计算公式

对于 $1 \le k \le n$，定义 $d_k = \max\{d_{k1}, \ d_{k2}\}$：

$$d_{k1} = \left|\frac{k}{n} - F_0(x_{(k)})\right|, \quad d_{k2} = \left|\frac{k-1}{n} - F_0(x_{(k)})\right|。$$

则有 $D_n = \max\{d_1, \ d_2, \ \ldots, \ d_n\}$；

并且当 F_0 是连续分布时 D_n 的分布与 F_0 无关。

H_0：$F(x) = F_0(x)$ 的拒绝域应该是 $\{D_n \ge C\}$，而一般取拒绝域形式为 $\{n^{1/2}D_n \ge d\}$。当给定检验水平 α 时即 $P\{n^{1/2}D_n \ge d\} = \alpha$，

具体的临界值 d 需要查表。

Kolmogrov 检验临界值($n \geq 40$)

$$P\{ n^{1/2} D_n \geq d \} = \alpha$$

α	0.9	0.75	0.50	0.25	0.10	0.05	0.01
d	0.575	0.678	0.830	1.02	1.23	1.36	1.63

在各种分布检验里**Kolmogrov** 检验效果比较好，灵敏度较高。但同时这种检验也有两点不足：

(1) 只适用于对连续分布的检验；

(2) 不能用来检验分布族(零假设分布不能含参数)。

练习3.3.6

从 1902 年 12月16日 开始到 1977 年 3月4日一共记录了强烈地震 63 次 (*Richter Scale* ≥7.5 或者*Fatalities* ≥1000)，62 个间隔如下(天)：

840	736	76	1336	304	129	638	157	584	710	335
375	9	937	145	887	46	1354	567	209	735	44
263	402	454	139	599	38	33	1901	194	36	780
83	365	121	695	759	667	203	832	92	150	294
319	40	436	328	82	280	562	460	556	30	246
220	434	721	40	99	384	1617				

检验相继两次强震的时间间隔是否来自参数为1/430 (即平均间隔为430 天) 的指数分布？

3.3.3 卡方分析(列联表检验)

设 X 的可能值是 1，…，s，Y 的可能值是 1，…，t。现在进行了 n 次独立的观察，其中“X 取 i，Y 取 j”的次数是 n_{ij}，需要检验：

H_0：X、Y 相互独立。

$X \backslash Y$	1	…	j	…	t	
1	n_{11}	…	n_{1j}	…	n_{1t}	$n_{1\bullet}$
i	n_{i1}	…	n_{ij}	…	n_{it}	$n_{i\bullet}$
s	n_{s1}	…	n_{sj}	…	n_{st}	$n_{s\bullet}$
	$n_{\bullet 1}$	…	$n_{\bullet j}$	…	$n_{\bullet t}$	n

记 $p_{ij} = \mathrm{P}(X = i, Y = j)$，

$p_i = \mathrm{P}(X = i)$，$q_j = \mathrm{P}(Y = j)$ ；需要检验的是

H_0：$p_{ij} = p_i \times q_j$ 对一切 i、j 成立

1. p_i、q_j 的极大似然估计

似然函数显然是 $L = \prod_{i=1}^{s} \prod_{j=1}^{t} (p_i q_j)^{n_{ij}}$

因此对数似然函数为

$$\ln L = \sum_{i=1}^{s} \sum_{j=1}^{t} (n_{ij} \ln p_i + n_{ij} \ln q_j)$$

下面分别对 p_i、q_j 求导建立似然方程组，

似然方程组

$$\frac{\partial \ln L}{\partial p_i} = 0,\quad \frac{\partial \ln L}{\partial q_j} = 0,\quad 1 \le i \le s-1,\quad 1 \le j \le t-1$$

而 $p_s = 1 - (p_1 + \ldots p_{s-1})$，$q_t = 1 - (q_1 + \ldots q_{t-1})$

得到 p_i、q_j 的极大似然估计为

$$\hat{p}_i = \frac{n_{i\bullet}}{n} \qquad \hat{q}_j = \frac{n_{\bullet j}}{n} \qquad 1 \le i \le s,\quad 1 \le j \le t$$

注意这里只估计了 $(s-1)+(t-1)$ 个未知参数

2. (分类变量) 独立性的检验

Pearson 的卡方统计量：

$$K^2 = \sum_{i=1}^{s}\sum_{j=1}^{t}\frac{(n_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

的极限分布是自由度为

$$st - (s - 1) - (t - 1) - 1 = (s - 1)(t - 1)$$

因此零假设(X、Y两个变量独立) 的水平 α 检验的拒绝域为 $K^2 > \chi_\alpha^2((s - 1)(t - 1))$。

2.1 列联表检验统计量的计算公式

$$K^2 = n\left[\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1\right]$$

2.2 四格表检验统计量的计算公式

$$K^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}}$$

四格表的零假设(X、Y两个变量独立) 的水平α检验的拒绝域为 $K^2 > \chi_\alpha^2(1)$。

$\chi_{0.01}^2(1) = 6.63$， $\chi_{0.05}^2(1) = 3.84$

例3.3.7 **P.G.Norton** 与**E.V.Dunn 1985**年研究关于“打鼾”与“心脏病” 的关系，

	从来不打鼾	每晚都打鼾
有心脏病	**24**	**30**
没有心脏病	**1355**	**224**

解. 计算有关的K^2统计量，

$$K^2 = \frac{1633\times(24\times224-30\times1355)^2}{54\times1579\times1379\times254} = 68$$

认为睡觉打鼾与心脏病有显著的统计关系。 □

分类变量中两个变量的相关程度

对于一个$s \times t$的列联表，Cramer 提出利用：

$$V=\sqrt{\frac{K^2}{n(L-1)}}$$

来刻画这两个分类变量的相关性；K^2是 Pearson 的卡方检验统计量，$L=\min(s,t)$。

V越小相关程度越弱，反之V越接近 1 则越强

对于 2×2 的四格表，V习惯上被记成ϕ。

例3.3.8　1969年美国征兵法案是否公平？

一年内的每一天被指定对应于一个1~366 的号码，按照抽取出的顺序作为服兵役的顺序。

显然被抽到一个小号码的概率不应该依赖于一个人的出生月份。但是最终公布的结果是

	1~6月	7~12月
1~183	73	110
183~366	109	74

是否应该认为下半年出生更可能先服兵役？

解. 计算有关的K^2统计量，

$$K^2=\frac{366\times(73\times74-110\times109)^2}{183\times183\times182\times184}=14.16$$

p-值是 p = P($\chi^2(1)$ ＞14.16) = 0.00017

$$\phi=\frac{|73\times74-110\times109|}{\sqrt{183\times183\times182\times184}}=0.197$$

可以认为一个人的出生日期和他在1969年美国国防部公布的被征召入伍的顺序具有统计上的显著关系，但是是一种比较弱的关系。 □

EXCEL 处理卡方分析

CHITEST(观察值域,理论值域)直接输出*p*-值

$$\text{其中观察值域}(s\times t)\text{为：}\begin{matrix} & \vdots & \\ \cdots & n_{ij} & \cdots \\ & \vdots & \end{matrix}$$

$$\text{其中理论值域}(s\times t)\text{为：}\begin{matrix} & \vdots & \\ \cdots & C_{ij} & \cdots \\ & \vdots & \end{matrix}$$

$$C_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$$

练习3.3.9

下表是美国人口普查局1992年高等教育年鉴的一组样本数据。分析这两组变量是否有关。

	亚裔	西班牙裔	白人
中学或更低	24	98	419
上过大学	27	34	310
专业人员或硕士	9	6	61

3.3.4 秩检验

对于一组来自连续总体 F 的样本 X_1，…，X_n，相应顺序统计量为 $X_{(1)} \le \ldots \le X_{(n)}$，每个样本$X_i$ 的秩记为 R_i (即它在顺序统计量中的排序),

$$R_i = \min\{ k : 1 \le k \le n,\ X_i = X_{(k)} \}$$

不难证明随机向量 $R = (R_1, \ldots, R_n)$ 在集合 $A = \{ a : a$ 是 1，2，…，n 的一个排列 $\}$ 上均匀分布，即对于任意一个固定的排列 a，都有：

$$P\{R = a\} = \frac{1}{n!}$$

因此依赖于 R 的统计量(秩统计量) $T(R)$ 具有与X 的总体 F 无关、固定不变的分布

F. Wilcoxon 秩和检验

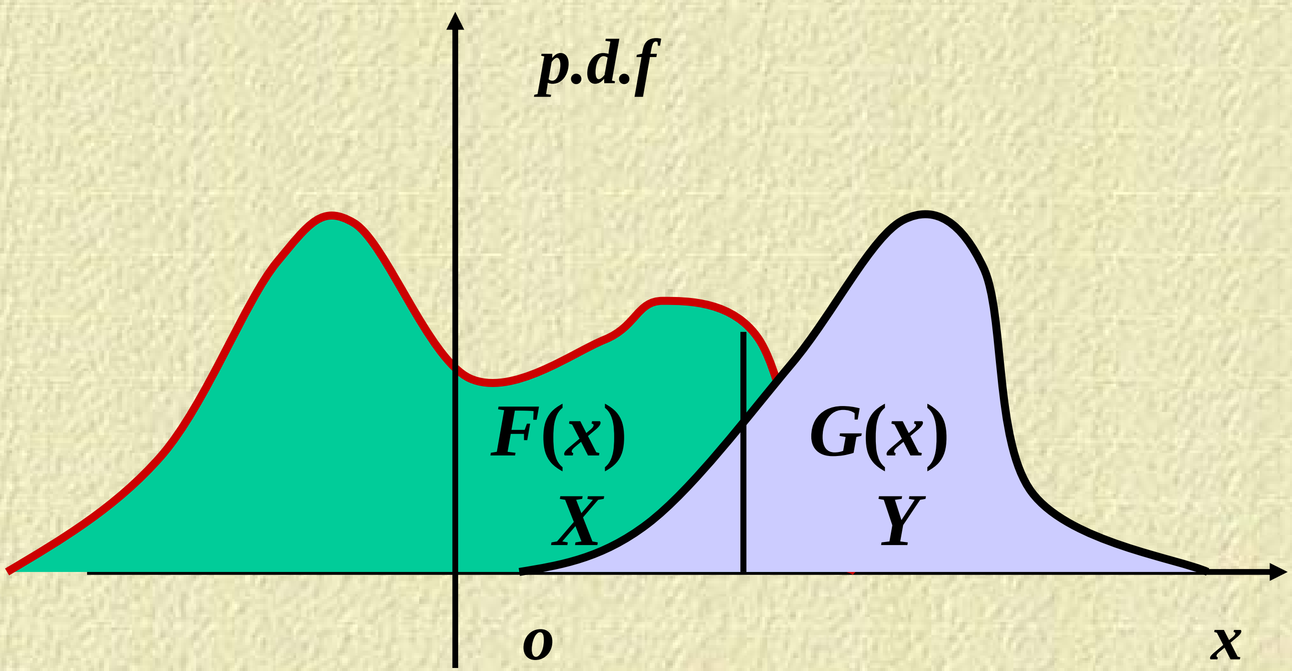
1945年 Wilcoxon 提出用来比较两个总体分布函数的非参数检验方法。

假定 $F(x)$ 与 $G(x)$ 是两个独立的连续总体，相应样本分别是 $X_1, \ldots, X_m$ 与 $Y_1, \ldots, Y_n$。把混合后的样本($X_1, \ldots, X_m, Y_1, \ldots, Y_n$)排序，记 $R_1, \ldots, R_n$ 为 $Y_1, \ldots, Y_n$ 的秩。

Wilcoxon 提出，利用秩和：

$$W = R_1 + \ldots + R_n$$

作为检验统计量来比较这两个分布的大小。



由于 W 是 Y 样本(即$G(x)$) 在全体样本中的秩和，如果 $F(x) > G(x)$ 则 W 应该偏大；反之W应该偏小

1. 两个总体分布函数的秩和检验方法

H_0	H_1	拒绝域
$F(x) \le G(x) \Leftrightarrow$	$F(x) > G(x)$	$W \ge d$
$F(x) \ge G(x) \Leftrightarrow$	$F(x) < G(x)$	$W \le c$
$F(x) = G(x) \Leftrightarrow$	$F(x) \ne G(x)$	$W \le c \cup W \ge d$

秩和统计量 $W = R_1 + \ldots + R_n$ 的计算公式

$$W = \sum_{i=1}^{m}\sum_{j=1}^{n} I\,[\,Y_j > X_i\,] + \frac{1}{2}n(n+1)$$

2. 非参数的均值检验

Wilcoxon 秩和检验也可以如下方式提出：假定 μ_1 与 μ_2 分别是两个非参数总体 $F(x)$ 与$G(x)$ 的均值，则

H_0	H_1	拒绝域
$\mu_1 \ge \mu_2$	$\Leftrightarrow \mu_1 < \mu_2$	$W \ge d$
$\mu_1 \le \mu_2$	$\Leftrightarrow \mu_1 > \mu_2$	$W \le c$
$\mu_1 = \mu_2$	$\Leftrightarrow \mu_1 \ne \mu_2$	$W \le c \cup W \ge d$

例3.3.10　一种产品可以用 A、B两种材料制造，随机抽取了12个产品按照性能从低到高排序：

B，B，A，B，B，A，A，B，A，A，A，A

解. 如果检验这两种材料产品性能是否有差异，

$$\mathrm{H_0}: \mu_1 = \mu_2 \Leftrightarrow \mathrm{H_1}: \mu_1 \neq \mu_2$$

0.05水平下的拒绝域为：$W \le 21$ 或$W \ge 44$；

如果需要检验 A 材料的产品性能是否更好，

$$\mathrm{H_0}: \mu_1 \ge \mu_2 \Leftrightarrow \mathrm{H_1}: \mu_1 < \mu_2$$

0.05水平下的拒绝域为：$W \ge 43$。

而 B 材料产品的秩和为 $W = 1+2+4+5+8 = 20$。

□

3.3.5 符号检验

使用简单，如果在重复观察时总体没有变化，则n 个样本中应该是正、负号基本上各占一半；

可以用二项分布计算出正号的 p-值：

$$p\text{-值}= BINOMDIST(k, n, p, 1)$$

在处理配对的数据时，符号检验的显著优点是相对比较安全

例3.3.11

Darwin 收集的异花与自花授粉植物高度

数据	1	2	3	4	5	6	7	8
异花	23.5	12.0	21.0	22.0	19.1	21.5	22.1	20.4
自花	17.4	20.4	20.0	20.0	18.4	18.6	18.6	15.3

数据	9	10	11	12	13	14	15
异花	18.3	21.6	23.3	21.0	22.1	23.0	12.0
自花	16.5	18.0	16.3	18.0	12.8	15.5	18.0

三种不同的统计分析方法

① 假定这些数据配对，需要检验它们的差值是否来自均值0的正态，*p*-值是**0.0478**；

② 使用符号检验，*p*-值是**0.0037**；

③ 假定这些数据来自不同总体，需要检验的是它们的均值是否相同，*p*-值是**0.0203**。

都认为自花与异花授粉植株高度显著不同

□

练习3.3.12 是否某些国家对陌生人更为友好？

	国家			
	丹麦	法国	荷兰	西德
信任	625	206	468	393
怀疑	360	763	463	513

练习3.3.13

把一枚一元硬币抛掷100 次，利用符号检验判断它是否均匀。

第3.4节 统计决策与Bayes理论

统计决策(又称统计判决)理论，通过引进各种优良性标准，把统计问题转化为数学最优化问题的求解。

Bayes 统计从 **1763** 年英国学者 **T.Bayes** 的论文《论有关机遇问题的求解》发展而来，实际上是一种归纳推理的理论。最核心(也是争议最大)的原则是其主观概率的思想，即先验分布的引入。

3.4.1 统计决策(*Statistical decision*)的思想

1945年**A. Wald** 提出统计学应该是在不确定情况下作出决策的一门科学，包括参数估计、假设检验等都可以归结到“统计决策”的理论中。

1. 统计决策理论的三个要素

假定总体分布函数是$F(x, \theta)$，未知参数θ全部可能值构成参数空间Θ；如果把我们在实际工作中可能采取的各种行为看作一个集合D，称为决策空间D；定义在$\Theta \times D$上的非负函数就称为损失函数$L(\theta, d)$，表示参数是θ时采取行为d而引起的损失。

2. 常见的两种损失函数

平方损失： $L(\theta, d) = (\theta - d)^2$,

绝对损失： $L(\theta, d) = |\theta - d|$

3. 风险函数 (*Risk fuction*)

由于我们采取的行为 d 一般都建立在一组样本的基础上即 $d = \delta(X_1, .., X_n)$ 。有了样本以后，样本的函数 δ 就是一个决策，它的平均损失

$$R(\theta, \delta) = E\,[L(\theta, \delta(X_1, .., X_n))]$$

就称为是决策 δ 的风险函数。

统计决策观点下的估计与检验

(1). 在参数点估计问题中常常取平方损失函数：

$$L(\theta, d) = (\theta - d)^2$$

此时风险函数就是一个估计量的均方误差

UMVUE 就是在无偏估计类中，
平方损失下风险最小

练习3.4.1

假定X_1，..，X_n 来自参数 p 的两点分布，分别在平方损失与绝对损失下求 p 风险最小的点估计。

(2). 在假设检验问题中，取决策空间 $D = \{0, 1\}$，其中k 表示接受假设 H_k，损失函数为 0-1 损失：

$$L(\theta, k) = \begin{cases} 0，如果\ \theta \in \Theta_k \\ 1，如果\ \theta \in \Theta_{1-k} \end{cases}$$

则一个检验的风险函数就是：

$$R(\theta, \phi) = \begin{cases} \beta_\phi(\theta)，当\ \theta \in \Theta_0 \\ 1 - \beta_\phi(\theta)，当\ \theta \in \Theta_1 \end{cases}$$

Neyman-Pearson 的假设检验理论，即指定一个 $0 < \alpha < 1$，把风险函数 $R(\theta, \phi)$在Θ_0 上的取值控制为不超过α，同时使得 $R(\theta, \phi)$ 在 Θ_1 上尽量小。

(3). 在区间估计理论中，取决策空间D为全部区间(a, b) 的集合，损失函数一共有两个：

$$L_1(\theta, (a, b)) \text{ 与 } L_2(\theta, (a, b))$$

L_1 是 0-1 损失函数：

$$L_1(\theta, (a, b)) = \begin{cases} 0\text{，当 } a < \theta < b \\ 1\text{，否则} \end{cases}$$

L_2 反映了区间的精度，

例如 $L_2(\theta, (a, b)) = b - a$。

对于任意一个区间估计 $\delta(x) = (\varphi_1(x), \varphi_2(x))$，损失函数$L_1$ 导致风险函数：

$$R_1(\theta, \delta) = 1 - P\{\varphi_1(x) < \theta < \varphi_2(x)\}$$；

而损失 L_2 的风险函数是区间平均长度：

$$R_2(\theta, \delta) = E[\varphi_2(x) - \varphi_1(x)]$$

指定置信度为$1 - \alpha$ 就理解为 $R_1(\theta, \delta) \le \alpha$，然后在这个条件下要求 $R_2(\theta, \delta)$ 尽可能地小。

4. 两个常用的决策原则

(1). Minimax 决策 **(**极小化极大方法**)**

对每一个决策都考虑最坏的可能状态，然后选择那个对于全部最坏状态而言是最佳的决策。

(2). Bayes 决策

对于参数θ定义一个概率分布**(**称为先验分布**)**，把风险函数 $R(\theta, d)$关于这个先验分布取数学期望**(**称为 **Bayes** 风险**)**，在决策空间D 中取使得**Bayes**风险最小的那一个作为问题的解。

例3.4.2　汽车决策问题

一位统计学家准备买一辆汽车，计划使用两年行驶 40,000 英里。他把选择的范围缩小到两部：汽车*A* 价格 ＄5,000，平均每加仑汽油行驶 20 英里， 汽车*B* 价格 ＄6,700，平均每加仑汽油行驶 40 英里；　假定汽油平均价格有三种，每加仑＄1，＄2，＄3，问这位统计学家应该选择哪一辆车？

解.　如果选择汽车 *A* 需要 2,000 加仑汽油；
　　如果选择汽车 *B* 需要 1,000 加仑汽油；

油价	＄1	＄2	＄3
汽车*A* 总费用	＄7,000	＄9,000	＄11,000
汽车*B* 总费用	＄7,700	＄8,700	＄9,700

如果采用 Minimax 决策，*A* 车的最大总费用＄11,000，*B*车的最大总费用＄9,700 ；他应该选择汽车*B* 。

如果采用 Bayes 决策，需要给出一组概率：假定油价(主观概率)：

＄1	＄2	＄3
3/4	1/8	1/8

则*A* 车的平均总费用＄7,750，*B*车的平均总费用＄8,075；他应该选择汽车*A* 。 □

练习3.4.3　收藏家问题

一位收藏家考虑购买一幅据说是著名画家的画，标价＄5,000 。如果是真品可值＄10,000 ；如果赝品则一文不值。并且买了假画或者是没有买下真画都会损害她的名誉。损失表如下：

	真品	赝品
买	-＄5,000	＄6,000
不买	＄3,000	0

她跑去征询一位鉴赏家的意见，而这位鉴赏家能够以 0.95 的概率判断一幅真画，同时以 0.70 的概率识别一幅假画。

现在有四种方案可供考虑：

决策1：一定要买；

决策2：坚决不买；

决策3：抛硬币决定是否买；

决策4：如果鉴赏家说是真品就买。

(1). 这位收藏家的 **Minimax** 决策是什么？

(2). 假定这幅画有**0.75**的先验概率是真画，**0.25** 的概率为假画，她的 **Bayes** 决策又应该是什么？

3.4.2 Bayes 统计理论

<u>频率学派的观点</u>

对于总体的未知参数θ，无论做点估计、区间估计或者假设检验，θ始终仅仅是一个常数。

我们抽取样本之前，对于 θ无任何了解，所有关于θ的信息都全部包含在样本之中。

<u>**Bayes** 学派则认为</u>，

在抽取样本之前，对θ已经有一些认识(先验知识)，这种知识可通过一个概率分布体现出来(先验分布$h(\theta)$)，它综合了我们在抽样之前对θ的全部信息。

1. Bayes 统计的基本思想

在随机向量(X, θ) 中，我们只能通过对 X 的观察去间接地推导出 θ。概率函数 $f(x, \theta)$ 被看成样本 x 关于θ的条件分布 $f(x|\theta)$ ，

因此抽取一组样本 x，也就意味着对θ得到了新的信息，它将修正原来的观点(θ的先验分布)，这种对θ的新认识通过θ关于样本 x 的条件分布(后验分布) 体现出来；

对θ的估计或者检验都依据后验分布来进行

总体分布律或密度

$p(x,\theta)$

根据样本

概率函数 $f(x|\theta)$

$=\prod p(x_i,\theta)$

先验分布 $h(\theta)$

X 与 θ 的联合分布

$=h(\theta)\times\prod p(x_i,\theta)$

样本 X 的边缘分布

$=\int_{\Theta} h(\theta)\times\prod p(x_i,\theta)\,d\theta$

对 θ 积分或求和

θ 关于样本 X 的后验分布律或者后验密度函数：

$$h(\theta|x)=\frac{h(\theta)\times\prod p(x_i,\theta)}{\int_{\Theta} h(\theta)\times\prod p(x_i,\theta)\,d\theta}$$

2. Bayes 统计推断理论

得到θ关于样本X的后验分布$h(\theta|X)$以后，其余的如样本观察值、样本分布、先验分布等都不再有用，对θ的处理全部依赖于后验分布。

(1) **Bayes 点估计**：根据损失函数取成后验分布的中位数或者数学期望；

(2) **Bayes 区间估计**：直接由后验分布构造；

(3) **Bayes 检验**：由后验分布计算θ落在零假设或者对立假设中的概率，哪个大就接受哪个。

3. 如何确定先验分布

(1) 过去的经验或资料

(2) 同等无知原则 **(Bayes 假定)**：

θ以相同的概率可以取参数空间Θ中任何一个。

Remark

采用同等无知原则时，先验分布可以不再是一个分布律或密度函数。如 $h(\theta)\equiv 1$，$\theta\in \mathbf{R}^1$，只要能保证样本X 的边缘分布 $\int_{\Theta} h(\theta)\times\Pi\, p(x_i, \theta)\, d\theta$有限从而 $h(\theta|X)$存在即可，此时称为广义先验分布。

例3.4.4　假定随机事件 A 发生的概率是 p，做了 n 次独立重复试验，根据试验结果估计 p。

解. 总体是参数 p 的两点分布，n 个样本显然是：

$$X_i = \begin{cases} 0\text{，第 } i \text{ 次试验时}A\text{不发生；} \\ 1\text{，第 } i \text{ 次试验时}A\text{ 发生。} \end{cases}$$

考虑样本的概率函数：

$$f(x, p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

根据“同等无知”观点，取先验分布 $p \sim U(0, 1)$

因此 X 与θ的联合分布，即联合密度为：

$$h(\theta)\times f(x|\theta) = p^{\sum x_i}(1-p)^{n-\sum x_i},\ 0 < p < 1$$

所以样本x 的边缘分布为一个 Beta 积分：

$$f(x) = \int_0^1 p^{\sum x_i}(1-p)^{n-\sum x_i}\,dp = \mathrm{B}\left(\sum x_i + 1, n - \sum x_i + 1\right)$$

$$= \frac{\Gamma(\sum x_i + 1)\Gamma(n - \sum x_i + 1)}{\Gamma(n+2)}$$

则参数 θ 关于样本 x 的后验分布为：

$$h(\theta \mid x) = \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\mathrm{B}\left(\sum x_i + 1, n - \sum x_i + 1\right)}, \quad 0 < p < 1$$

如果取损失为平方损失函数，则参数 p 的 **Bayes** 估计应该是后验分布的均值，

$$\hat{p} = \int_0^1 p \times \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\mathrm{B}\left(\sum x_i + 1, n - \sum x_i + 1\right)} dp$$

$$= \frac{\mathrm{B}\left(\sum x_i + 2, n - \sum x_i + 1\right)}{\mathrm{B}\left(\sum x_i + 1, n - \sum x_i + 1\right)} = \frac{\sum x_i + 1}{n+2}$$

n 较小时，**Bayes** 估计要比样本均值更合理

□

例3.4.5

假定总体 $X \sim B(1, p)$，只有三个样本 X_1, X_2, X_3，考虑：H_0: p=0.4 $\Leftrightarrow$ H1: p=0.5。记决策 d_k为"接受假设 H_k"，损失函数是：$L(d_0,0.4) = L(d_1,0.5) = 0$; $L(d_0,0.5) = 1$, $L(d_1,0.4) = 2$ ；

再假定参数 p 的先验分布是：

p	0.4	0.5
	1/3	2/3

求这个假设的Bayes 检验。

Bayes 检验为：

样本和等于0 或者1 时接受零假设 $p = 0.4$，当样本和等于2 或者3 时接受零假设 $p = 0.5$。 □

例3.4.6　样本X_1，…，X_n 来自正态总体$N(\theta, 1)$，并且已知θ的先验分布是 $N(\mu, \sigma^2)$，这里 μ, σ^2 已知。讨论θ的Bayes 估计与检验：

$$H_0：\theta \le 0 \Leftrightarrow H_1：\theta > 0$$

解. 可以计算出参数 θ的后验分布是正态分布

$$N\left(\frac{\mu + (\sum x_i)\sigma^2}{n\sigma^2 + 1}, \frac{\sigma^2}{n\sigma^2 + 1}\right)$$

在平方损失(或绝对损失)函数下，θ的**Bayes** 估计为：

$$\hat{\theta} = (\frac{n\sigma^2}{n\sigma^2+1})\overline{X} + (\frac{1}{n\sigma^2+1})\mu$$

Bayes 估计实际上是样本均值与先验均值的加权估计，而且这种权的比例为：

$$\overline{X} : \mu \quad = \quad n : (\frac{1}{\sigma^2})$$

要检验假设 H_0：$\theta \le 0 \Leftrightarrow H_1$：$\theta > 0$，由于$\theta$ 的后验分布为：

$$N\left(\frac{\mu + (\sum x_i)\sigma^2}{n\sigma^2 + 1}, \frac{\sigma^2}{n\sigma^2 + 1}\right)$$

所以只需要比较P $\{\theta \in \Theta_0 \mid X\}$ 与0.5 的大小，即后验分布的均值是否不小于 0 。

Bayes 检验为：

$$\overline{X} \le -\frac{\mu}{n\sigma^2},$$ 接受零假设，认为 $\theta \le 0$ ；

$$\overline{X} > -\frac{\mu}{n\sigma^2},$$ 否定零假设，认为 $\theta > 0$ 。

□

练习3.4.7

给出例题3.4.6 中参数θ的置信度 $1-\alpha$ 的**Bayes**区间估计。

练习3.4.8

把例题3.4.6 中参数θ的**Bayes**估计、检验结果与传统的估计、检验结果加以比较和讨论。

练习3.4.9

证明如果取参数θ的先验分布为均匀分布，则后验分布的概率函数的众数就是θ的极大似然估计。