第七章　假设检验

上一章是对总体参数作出估计，本章是检验总体参数为某给定值是否合理。

§7.1　假设检验的基本思想与概念

7.1.1　假设检验问题

例　某袋装食品正常情况下每袋重量（克）服从 $N(500,100)$，现抽取 25 袋测得平均重量为 495.3 克。问该袋装食品重量是正常？

通常假设需检验的参数是给定的值，再检验是否合理。本题中，假设 $H_0:\mu=500$ vs $H_1:\mu\neq 500$。称 H_0 为原假设，而 H_1 为备择假设，此时备择假设 H_1 是双侧的。此外，备择假设 H_1 有时可以是单侧的。如上例中，若问题改为问该袋装食品重量是否不足，则假设 $H_0:\mu\geq 500$ vs $H_1:\mu<500$。

这里是对总体的一个参数进行检验，称为单参数检验。一般地，设 θ 是需检验的总体参数，参数空间为 Θ，而 Θ_0 与 Θ_1 是其两个互不相容的子集，则统计假设一般形式为假设 $H_0:\theta\in\Theta_0$ vs $H_1:\theta\in\Theta_1$。通常当原假设中 Θ_0 只含一个点时，称为简单原假设；否则，称为复合原假设。

例　掷一枚骰子 100 次，结果如下：

点数	1	2	3	4	5	6
次数	14	25	17	10	23	11

可否认为这枚骰子不均匀？

假设 $H_0:p_i=\dfrac{1}{6},\quad i=1,2,3,4,5,6$。

这里是对总体的多个参数进行检验，称为多参数检验。本章最后一节还将介绍非参数检验。一般单参数检验需要写出备择假设，而多参数检验和非参数检验不必写出备择假设。

对于作出的原假设 H_0，需检验是否合理。假设检验的基本原理是小概率原理：即认为小概率事件在一次试验中实际上不会发生。假设检验在方法上类似于反证法，通常是根据在假设 H_0 成立的条件下，取得问题中样本观测值概率的大小而定。发生的概率大，则认为 H_0 合理，接受 H_0；发生的概率小（相当于反证法中推出了矛盾），则认为 H_0 不合理，拒绝 H_0。直观上，原假设 H_0 通常是问题中不轻易拒绝的结果；数学上，原假设 H_0 所对应的参数 θ 取值范围 Θ_0 通常是闭集（一个点或闭区间）。

7.1.2　假设检验的基本步骤

1．　提出统计假设 H_0 与 H_1

如检验期望值 μ，可采用双侧假设：$H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$，单侧假设 $H_0:\mu\geq\mu_0$ vs $H_1:\mu<\mu_0$ 或 $H_0:\mu\leq\mu_0$ vs $H_1:\mu>\mu_0$。

2． 选取检验统计量

如正态总体 $X \sim N(\mu, \sigma^2)$，已知 σ^2，检验 μ，建立统计量

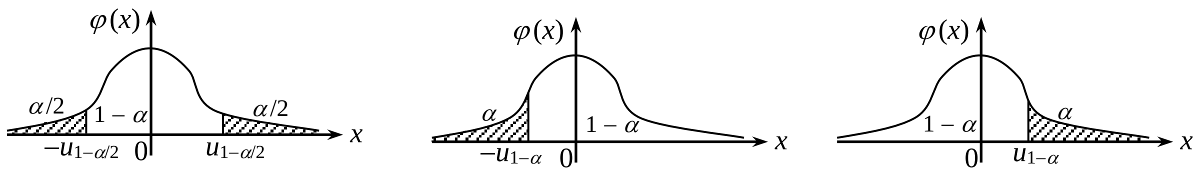
$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

在原假设 $H_0: \mu = \mu_0$ 成立的条件下，检验统计量为 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$。

3． 确定拒绝域

根据显著水平 α 及检验统计量的分布，确定拒绝 H_0 的检验统计量观测值的取值范围。显著水平 α 为小概率，一般 $0.01 \le \alpha \le 0.1$。选择拒绝域 W 使得检验统计量观测值落入 W 的概率不超过显著水平 α。

如选取 U 检验统计量，应使得 $P\{U \in W\} \le \alpha$。双侧假设 $H_1: \mu \ne \mu_0$，拒绝域为 $W = \{|u| \ge u_{1-\alpha/2}\}$；而单侧假设 $H_1: \mu < \mu_0$，拒绝域为 $W = \{u \le -u_{1-\alpha}\}$；单侧假设 $H_1: \mu > \mu_0$，拒绝域为 $W = \{u \ge u_{1-\alpha}\}$。



一般情况下，双侧假设的拒绝域在两侧，单侧假设的拒绝域在左侧或右侧，具体方向根据在备择假设 H_1 为真的情况下，检验统计量观测值最可能值所在方向而定。如假设 $H_0: \mu \ge \mu_0$ vs $H_1: \mu < \mu_0$，选取 U 统计量，在原假设 H_0 成立的条件下，$U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$，取 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$，但如果备择假设 H_1 为真，最可能出现 $\bar{x} < \mu_0$，从而检验统计量观测值 $u = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ 最可能为负，因此拒绝域在左侧，$W = \{u \le -u_{1-\alpha}\}$。

4． 作出决策

根据样本观测值计算检验统计量观测值，若检验统计量观测值不在拒绝域内，则接受原假设 H_0，拒绝备择假设 H_1；若检验统计量观测值落在拒绝域内，则拒绝原假设 H_0，接受备择假设 H_1。

双侧假设 $H_0: \mu = \mu_0$ vs $H_1: \mu \ne \mu_0$，选取 U 统计量。在 H_0 成立的条件下，$U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，取 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ 作为检验统计量。

单侧假设，如 $H_0: \mu \ge \mu_0$ vs $H_1: \mu < \mu_0$，选取 U 统计量，拒绝域 $W = \{u \le -u_{1-\alpha}\}$。在 H_0 成立的条件下，$U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$，且 $U \le U_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$，当 U_0 的观测值 $u_0 \le -u_{1-\alpha}$ 时，必有 U 的观测值 $u \le -u_{1-\alpha}$；

又如$H_0:\mu\le\mu_0$ vs $H_1:\mu>\mu_0$，拒绝域$W=\{u\ge u_{1-\alpha}\}$，有$U\ge U_0=\dfrac{\bar X-\mu_0}{\sigma/\sqrt n}$，当$U_0$的观测值$u_0\ge u_{1-\alpha}$时，必有$U$的观测值$u\ge u_{1-\alpha}$。可见对于单侧假设，按$\mu=\mu_0$计算统计量观测值$u$。当$U_0=\dfrac{\bar X-\mu_0}{\sigma/\sqrt n}$的观测值落入拒绝域$W$时，对在$H_0$成立条件下按参数$\mu$的任何值计算的统计量观测值$u$都将落入拒绝域$W$，从而都会作出拒绝$H_0$的决策；当$U_0=\dfrac{\bar X-\mu_0}{\sigma/\sqrt n}$的观测值没有落入拒绝域$W$时，在$H_0$成立条件下按参数$\mu$的其他值计算的统计量观测值$u$有些落入拒绝域$W$，有些没有落入，按不轻易作出拒绝$H_0$决策的原则，而作出接受$H_0$的决策。因此对于单侧假设$H_0:\mu\ge\mu_0$ vs $H_1:\mu<\mu_0$或$H_0:\mu\le\mu_0$ vs $H_1:\mu>\mu_0$，与双侧假设一样，都是取$U=\dfrac{\bar X-\mu_0}{\sigma/\sqrt n}$作为检验统计量（此时检验统计量并不一定服从相应的分布）。

根据假设检验所作出的决策并不能保证完全正确，可能发生两类错误。

（1）如果原假设H_0是正确的，但由于样本随机性，检验统计量观测值落入拒绝域，而作出拒绝H_0的决策，则称犯了第一类错误或拒真错误，犯拒真错误的概率就是显著水平$\alpha=P\{U\in W\mid H_0\}$。

（2）如果原假设H_0是错误的，但由于样本随机性，检验统计量观测值没有落入拒绝域，而作出接受H_0的决策，则称犯了第二类错误或受伪错误，犯受伪错误的概率记为$\beta=P\{U\notin W\mid H_1\}$。

即H_0为真，但拒绝H_0，犯拒真错误；H_0为假，但接受H_0，犯受伪错误。一般地，显著水平α越小，犯拒真错误概率越小，犯受伪错误概率越大。

例 设$X_1,X_2,\cdots,X_{16}$是来自正态总体$N(\mu,1)$的样本，检验假设$H_0:\mu\le 0$ vs $H_1:\mu>0$，显著水平$\alpha=0.05$，求拒绝域W，并问当$\mu=1$时，犯第二类错误的概率是多少？

解：假设$H_0:\mu\le 0$ vs $H_1:\mu>0$，选取统计量$U=\dfrac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$，检验统计量$U=\dfrac{\bar X-0}{\sigma/\sqrt n}$，显著水平$\alpha=0.05$，$u_{1-\alpha}=u_{0.95}=1.64$，拒绝域$W=\{u\ge 1.64\}$。

因$\sigma=1$，$n=16$，检验统计量

$$U=\frac{\bar X-0}{\sigma/\sqrt n}=4\bar X\ 。$$

但因$\mu=1$，实际上应为

$$U_1=\frac{\bar X-\mu}{\sigma/\sqrt n}=4(\bar X-1)\sim N(0,1)\ ，$$

故当 $\mu=1$时，犯第二类错误的概率

$$\beta=P\{U\notin W\mid\mu=1\}=P\{4\bar{X}<1.64\mid\mu=1\}=P\{4(\bar{X}-1)<-2.36\mid\mu=1\}$$

$$=\Phi(-2.36)=1-\Phi(2.36)=1-0.9909=0.0091。$$

一般地，检验统计量T 的观测值落入拒绝域W 的概率与参数θ的取值有关，是参数θ的函数，称为势函数，记为$g(\theta)$ 。对于假设$H_0:\theta\in\Theta_0$ vs $H_1:\theta\in\Theta_1$，有

$$g(\theta)=P\{T\in W\mid\theta\in\Theta\}=\begin{cases}\alpha(\theta), & \theta\in\Theta_0;\\ 1-\beta(\theta), & \theta\in\Theta_1.\end{cases}$$

7.1.3　　检验的 p 值

在假设检验问题中，根据不同的显著水平α，可得不同的拒绝域，对于同样的样本，由不同的显著水平就可能作出不同的决策。

如在U 检验法中，当显著水平$\alpha=0.05$时，双侧拒绝域$W_{0.05}=\{|u|\ge u_{0.975}\}=\{|u|\ge 1.96\}$；而当显著水平$\alpha=0.1$时，双侧拒绝域$W_{0.1}=\{|u|\ge u_{0.95}\}=\{|u|\ge 1.64\}$。若检验统计量观测值$u=1.8$，则当$\alpha=0.05$时，$u=1.8\notin W_{0.05}$，应接受原假设$H_0$；当$\alpha=0.1$时，$u=1.8\in W_{0.1}$，应拒绝原假设$H_0$。

事实上，此时如果取显著水平$\alpha=0.0718$，双侧拒绝域恰好是$W_{0.0718}=\{|u|\ge u_{0.9641}\}=\{|u|\ge 1.8\}$。因若显著水平$\alpha$越小，分位数$u_{1-\alpha/2}$越大，拒绝域$W=\{|u|\ge u_{1-\alpha/2}\}$的范围越小，故当$\alpha<0.0718$时，$W_\alpha$是$W_{0.0718}=\{|u|\ge 1.8\}$的真子集，有$u=1.8\notin W_\alpha$，应接受原假设$H_0$；反过来，若显著水平$\alpha$越大，分位数$u_{1-\alpha/2}$越小，拒绝域$W=\{|u|\ge u_{1-\alpha/2}\}$的范围越大，当$\alpha\ge 0.0718$时，都有$u=1.8\in W_{0.0718}\subset W_\alpha$，应拒绝原假设$H_0$。可见$\alpha=0.0718$是在检验统计量观测值$u=1.8$时作出拒绝原假设$H_0$决策的最小显著水平。

定义　假设检验问题中，根据检验统计量观测值作出拒绝原假设H_0决策的最小显著水平称为检验的 p 值。

如在U 检验法中，当检验统计量观测值$u=1.8$时，双侧检验的 p 值就是 0.0718。

根据统计量T 的分布与其观测值t计算检验的 p 值。对于左侧拒绝域，p 值就是 $p=P\{T\le t\}$；对于右侧拒绝域，p 值就是 $p=P\{T\ge t\}$；对于双侧拒绝域，p 值就是 $p=\min\{2P\{T\le t\},2P\{T\ge t\}\}$。

如在U 检验法中，统计量$U\sim N(0,1)$。若检验统计量观测值$u=1.8$，双侧拒绝域，则

$$p=\min\{2P\{U\le 1.8\},2P\{U\ge 1.8\}\}=2P\{U\ge 1.8\}=2[1-\Phi(1.8)]=2[1-0.9641]=0.0718。$$

若显著水平$\alpha=0.05$，有$\alpha<p$，则接受原假设H_0；若显著水平$\alpha=0.1$，有$\alpha>p$，则拒绝原假设H_0。

进行假设检验时，也可不必写出拒绝域，直接根据检验的 p 值与显著水平α进行比较而作出决策。当显著水平$\alpha<p$时，则接受原假设H_0；当显著水平$\alpha\ge p$时，则拒绝原假设H_0。特别是对于离散型总体，根据检验的 p 值作出决策，回避了检验统计量观测值落入拒绝域内的概率不能恰好等于显著水平α的问题，因此有时比根据拒绝域作出决策更为方便。

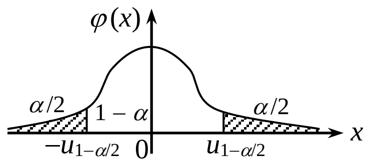
§7.2 正态总体参数的假设检验

7.2.1 单个正态总体参数的检验

正态总体 $N(\mu,\sigma^2)$， $X_1,X_2,\cdots,X_n$ 为样本，检验均值 μ 或方差 σ^2。

一. 已知方差 σ^2，检验均值 μ

步骤：



（1）假设 $H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$（或 $\mu<\mu_0$， $\mu>\mu_0$）。

（2）统计量 $U=\dfrac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$，检验统计量 $U=\dfrac{\bar X-\mu_0}{\sigma/\sqrt n}$。

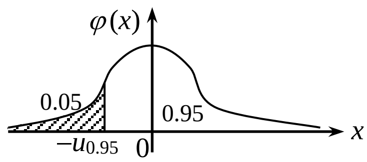
（3）拒绝域：$W=\{|u|\geq u_{1-\alpha/2}\}$（或 $\{u\leq -u_{1-\alpha}\}$，$\{u\geq u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这称之为 U 检验法。

例 某袋装食品正常情况下每袋重量（克）服从 $N(500,100)$，现抽取 25 袋测得平均重量为 495.3 克。问这种袋装食品是否重量不够？（$\alpha=0.05$）

解：单个正态总体，已知 σ^2，检验 μ，用 U 检验法。假设 $H_0:\mu\geq 500$ vs $H_1:\mu<500$，检验统计量 $U=\dfrac{\bar X-500}{\sigma/\sqrt n}$，显著水平 $\alpha=0.05$，$u_{1-\alpha}=u_{0.95}=1.64$，左侧拒绝域 $W=\{u\leq -1.64\}$。



因 $\bar x=495.3$，$\sigma=10$，$n=25$，则检验统计量观测值

$$u=\frac{495.3-500}{10/\sqrt{25}}=-2.35\in W,$$

并且左侧检验的 p 值

$$p=P\{U\leq -2.35\}=\Phi(-2.35)=1-0.9906=0.0094<\alpha=0.05,$$

故拒绝 H_0，接受 H_1。可以认为这种袋装食品重量不够。

例 从正态总体 $N(\mu,9)$ 中抽取容量为 n 的样本。问 n 不能超过多少才能在 $\bar x=21$ 的条件下接受假设 $H_0:\mu=21.5$，拒绝 $H_1:\mu\neq 21.5$。（$\alpha=0.05$）

解：单个正态总体，已知 σ^2，检验 μ，用 U 检验法。假设 $H_0:\mu=21.5$ vs $H_1:\mu\neq 21.5$，检验统计量 $U=\dfrac{\bar X-21.5}{\sigma/\sqrt n}$，显著水平 $\alpha=0.05$，$u_{1-\alpha/2}=u_{0.975}=1.96$，双侧拒绝域 $W=\{|u|\geq 1.96\}$。

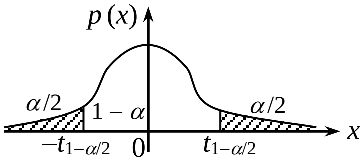
因 $\bar x=21$，$\sigma=3$，则检验统计量观测值

$$u=\frac{21-21.5}{3/\sqrt n}=-\frac{\sqrt n}{6}\notin W,$$

要接受 H_0，必须有 $|u|=\dfrac{\sqrt n}{6}<1.96$，即 $n<138.30$，故 n 不能超过 138。

二．未知方差σ^2，检验均值μ

步骤：



（1）假设$H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$（或$\mu<\mu_0$， $\mu>\mu_0$）。

（2）统计量$T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，检验统计量$T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$。

（3）拒绝域：$W=\{|t|\geq t_{1-\alpha/2}(n-1)\}$（或$\{t\leq -t_{1-\alpha}(n-1)\}$，$\{t\geq t_{1-\alpha}(n-1)\}$）。

（4）计算检验统计量观测值t与检验的p值，并作出决策。

这称之为T检验法。

例　已知某电子元件使用寿命（小时）服从$N(\mu,\sigma^2)$，随机抽取10只元件，测得使用寿命为：985, 1012, 1008, 993, 982, 990, 1005, 994, 982, 1002（小时），检验μ是否为1000小时。（$\alpha=0.1$）

解：单个正态总体，未知σ^2，检验μ，用T检验法。假设$H_0:\mu=1000$ vs $H_1:\mu\neq1000$，检验统计量$T=\dfrac{\bar{X}-1000}{S/\sqrt{n}}$，显著水平$\alpha=0.1$，$n=10$，$t_{1-\alpha/2}(n-1)=t_{0.95}(9)=1.833$，双侧拒绝域$W=\{|t|\geq1.833\}$。

因$\bar{x}=995.3$，$s^2=10.92^2$，$n=10$，则检验统计量观测值

$$t=\frac{995.3-1000}{10.92/\sqrt{10}}=-1.36\notin W,$$

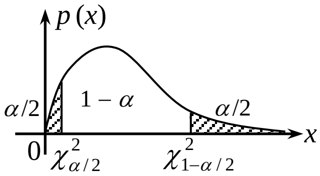
并且双侧检验的p值

$$p=\min\{2P\{T\leq-1.36\},2P\{T\geq-1.36\}\}=2P\{T\leq-1.36\}=0.2069>\alpha=0.1,$$

故接受H_0，拒绝H_1。可以认为μ是1000小时。

三．检验方差σ^2

步骤：



（1）假设$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2\neq\sigma_0^2$（或$\sigma^2<\sigma_0^2$，$\sigma^2>\sigma_0^2$）。

（2）统计量$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，检验统计量$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$。

（3）拒绝域：$W=\{\chi^2\leq\chi^2_{\alpha/2}(n-1)$ 或 $\chi^2\geq\chi^2_{1-\alpha/2}(n-1)\}$（或$\{\chi^2\leq\chi^2_{\alpha}(n-1)\}$，$\{\chi^2\geq\chi^2_{1-\alpha}(n-1)\}$）。

（4）计算检验统计量观测值χ^2与检验的p值，并作出决策。

这称之为χ^2检验法。

例　已知某电子元件使用寿命（小时）服从$N(\mu,\sigma^2)$，随机抽取10只元件，测得使用寿命为：985, 1012, 1008, 993, 982, 990, 1005, 994, 982, 1002（小时），检验σ^2是否为100。（$\alpha=0.1$）

解：单个正态总体，检验σ^2，用χ^2检验法。假设$H_0:\sigma^2=100$ vs $H_1:\sigma^2\neq100$，检验统计量

$\chi^2=\dfrac{(n-1)S^2}{100}$，显著水平 $\alpha=0.1$，$n=10$，$\chi^2_{\alpha/2}(n-1)=\chi^2_{0.05}(9)=3.3251$，$\chi^2_{1-\alpha/2}(n-1)=\chi^2_{0.95}(9)=16.919$，双侧拒绝域 $W=\{\chi^2\le 3.3251$ 或 $\chi^2\ge 16.919\}$。

因 $s^2=10.92^2$，$n=10$，则检验统计量观测值

$$\chi^2=\frac{9\times 10.92^2}{100}=10.7322\notin W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{\chi^2\le 10.7322\},2P\{\chi^2\ge 10.7322\}\}=2P\{\chi^2\ge 10.7322\}=0.5890>\alpha=0.1,$$

故接受 H_0，拒绝 H_1。可以认为 σ^2 是 100。

7.2.2　两个独立正态总体参数的检验

两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且相互独立。$X_1,X_2,\cdots,X_{n_1}$ 为来自总体 X 的样本，$Y_1,Y_2,\cdots,Y_{n_2}$ 为来自总体 Y 的样本，检验均值 $\mu_1=\mu_2$ 或方差 $\sigma_1^2=\sigma_2^2$。

一．已知方差 σ_1^2,σ_2^2，检验均值 $\mu_1=\mu_2$

步骤：

（1）假设 $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\ne\mu_2$（或 $\mu_1<\mu_2$，$\mu_1>\mu_2$）。

（2）统计量 $U=\dfrac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$，检验统计量 $U=\dfrac{\bar X-\bar Y}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$。

（3）拒绝域：$W=\{|u|\ge u_{1-\alpha/2}\}$（或 $\{u\le -u_{1-\alpha}\}$，$\{u\ge u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这是 U 检验法。

注：也可对 μ_1 与 μ_2 相差一个确定的常数进行假设检验，即检验假设 $H_0:\mu_1=\mu_2+C$，检验统计量为 $U=\dfrac{\bar X-\bar Y-C}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$。

例　现有新、旧两种电子元件，使用寿命（小时）都服从正态分布，且标准差分别为 100 小时，75 小时。现从新元件中抽取 15 个，测得平均寿命为 1352 小时；从旧元件中抽取 10 个，测得平均寿命为 1270 小时。问新元件寿命是否比旧元件长？（$\alpha=0.05$）

解：两个独立正态总体，已知 σ_1^2,σ_2^2，检验 $\mu_1=\mu_2$，用 U 检验法。假设 $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1>\mu_2$，检验统计量 $U=\dfrac{\bar X-\bar Y}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$，显著水平 $\alpha=0.05$，$u_{1-\alpha}=u_{0.95}=1.64$，右侧拒绝域 $W=\{u\ge 1.64\}$。

因 $\bar{x}=1352$， $\bar{y}=1270$， $\sigma_1^2=100^2$， $\sigma_2^2=75^2$， $n_1=15$， $n_2=10$，则检验统计量观测值

$$u=\frac{1352-1270}{\sqrt{\frac{100^2}{15}+\frac{75^2}{10}}}=2.34\in W,$$

并且右侧检验的 p 值

$$p=P\{U\geq 2.34\}=1-\Phi(2.34)=1-0.9904=0.0096<\alpha=0.05,$$

故拒绝 H_0，接受 H_1。可以认为新元件寿命比旧元件长。

二．未知方差 σ_1^2,σ_2^2，检验均值 $\mu_1=\mu_2$

1．方差 $\sigma_1^2=\sigma_2^2=\sigma^2$

步骤：

（1）假设 $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$（或 $\mu_1<\mu_2$， $\mu_1>\mu_2$）。

（2）统计量 $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$，其中 $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$，检验统计量 $T=\dfrac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$。

（3）拒绝域：$W=\{|t|\geq t_{1-\alpha/2}(n_1+n_2-2)\}$（或 $\{t\leq -t_{1-\alpha}(n_1+n_2-2)\}$，$\{t\geq t_{1-\alpha}(n_1+n_2-2)\}$）。

（4）计算检验统计量观测值 t 与检验的 p 值，并作出决策。

这是 T 检验法。

2．方差 $\dfrac{\sigma_2^2}{\sigma_1^2}=c$ 为已知常数

步骤：

（1）假设 $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$（或 $\mu_1<\mu_2$， $\mu_1>\mu_2$）。

（2）统计量 $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{c}{n_2}}}\sim t(n_1+n_2-2)$，其中 $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2/c}{n_1+n_2-2}}$，检验统计量 $T=\dfrac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{c}{n_2}}}$。

（3）拒绝域：$W=\{|t|\geq t_{1-\alpha/2}(n_1+n_2-2)\}$（或 $\{t\leq -t_{1-\alpha}(n_1+n_2-2)\}$，$\{t\geq t_{1-\alpha}(n_1+n_2-2)\}$）。

（4）计算检验统计量观测值t与检验的p值，并作出决策。

这是T检验法。

3． 未知二者关系，大样本场合

步骤：

（1）假设$H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$（或$\mu_1<\mu_2$，$\mu_1>\mu_2$）。

（2）统计量$U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}\sim N(0,1)$，检验统计量$U=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}$。

（3）拒绝域：$W=\{|u|\geq u_{1-\alpha/2}\}$（或$\{u\leq -u_{1-\alpha}\}$，$\{u\geq u_{1-\alpha}\}$）。

（4）计算检验统计量观测值u与检验的p值，并作出决策。

这是U检验法。

4． 未知二者关系，小样本场合

步骤：

（1）假设$H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$（或$\mu_1<\mu_2$，$\mu_1>\mu_2$）。

（2）统计量$T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}\sim t(l)$，$l\approx\dfrac{\left(\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}\right)^2}{\dfrac{S_x^4}{n_1^2(n_1-1)}+\dfrac{S_y^4}{n_2^2(n_2-1)}}$，检验统计量$T=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}$。

（3）拒绝域：$W=\{|t|\geq t_{1-\alpha/2}(l)\}$（或$\{t\leq -t_{1-\alpha}(l)\}$，$\{t\geq t_{1-\alpha}(l)\}$）。

（4）计算检验统计量观测值t与检验的p值，并作出决策。

这是T检验法。

三．检验方差$\sigma_1^2=\sigma_2^2$

步骤：

（1）假设$H_0:\sigma_1^2=\sigma_2^2$ vs $H_1:\sigma_1^2\neq\sigma_2^2$（或$\sigma_1^2<\sigma_2^2$，$\sigma_1^2>\sigma_2^2$）。

（2）统计量$F=\dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$，检验统计量$F=\dfrac{S_x^2}{S_y^2}$。

（3）拒绝域：$W=\{f\leq f_{\alpha/2}(n_1-1,n_2-1)$ 或 $f\geq f_{1-\alpha/2}(n_1-1,n_2-1)\}$（或$\{f\leq f_{\alpha}(n_1-1,n_2-1)\}$，$\{f\geq f_{1-\alpha}(n_1-1,n_2-1)\}$）。

（4）计算检验统计量观测值f与检验的p值，并作出决策。

这是F检验法。

注：也可对σ_1^2与σ_2^2相差一个确定的常数倍进行假设检验，即检验假设$H_0:c\sigma_1^2=\sigma_2^2$，检验统计量为$F=\dfrac{cS_x^2}{S_y^2}$。

例 两批钢丝抗断强度都服从正态分布，且方差相等。从第一批中取 16 根，测得样本均值 $\bar{x}=289.7$，样本方差 $s_x^2=10.2^2$；从第二批中取 12 根，测得样本均值 $\bar{y}=281.2$，样本方差 $s_y^2=9.5^2$。试比较两批钢丝抗断力的期望与方差是否有显著差异？（$\alpha=0.05$）。

解：先检验方差，两个独立正态总体，检验 $\sigma_1^2=\sigma_2^2$，用 F 检验法。假设 $H_0:\sigma_1^2=\sigma_2^2$ vs $H_1:\sigma_1^2\neq\sigma_2^2$，

检验统计量 $F=\dfrac{S_x^2}{S_y^2}$，显著水平 $\alpha=0.05$， $n_1=16$， $n_2=12$， $f_{1-\alpha/2}(n_1-1,n_2-1)=f_{0.975}(15,11)=3.3299$，

$f_{\alpha/2}(n_1-1,n_2-1)=f_{0.025}(15,11)=\dfrac{1}{f_{0.975}(11,15)}=0.3325$，双侧拒绝域 $W=\{f\le 0.3325$ 或 $f\ge 3.3299\}$。

因 $s_x^2=10.2^2$， $s_y^2=9.5^2$，则检验统计量观测值

$$f=\frac{10.2^2}{9.5^2}\approx 1.1528\notin W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{F\le 1.1528\},2P\{F\ge 1.1528\}\}=2P\{F\ge 1.1528\}=0.8272>\alpha=0.05,$$

故接受 H_0，拒绝 H_1。可以认为两批钢丝抗断力的方差没有显著差异。

再检验期望，未知 σ_1^2,σ_2^2，但 $\sigma_1^2=\sigma_2^2$，检验 $\mu_1=\mu_2$，用 T 检验法。假设 $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$，

检验统计量 $T=\dfrac{\bar{X}-\bar{Y}}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$，其中 $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$，显著水平 $\alpha=0.05$， $n_1=16$， $n_2=12$，

$t_{1-\alpha/2}(n_1+n_2-2)=t_{0.975}(26)=2.0555$，双侧拒绝域 $W=\{|t|\ge 2.0555\}$。

因 $\bar{x}=289.7$， $\bar{y}=281.2$， $s_x^2=10.2^2$， $s_y^2=9.5^2$， $n_1=16$， $n_2=12$，则检验统计量观测值

$$t=\frac{289.7-281.2}{\sqrt{\dfrac{15\times 10.2^2+11\times 9.5^2}{26}}\cdot\sqrt{\dfrac{1}{16}+\dfrac{1}{12}}}\approx 2.2461\in W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{T\le 2.2461\},2P\{T\ge 2.2461\}\}=2P\{T\ge 2.2461\}=0.0334<\alpha=0.05,$$

故拒绝 H_0，接受 H_1。可以认为两批钢丝抗断力的期望有显著差异。

7.2.3 成对数据检验

对两个正态总体 X 与 Y 的均值进行比较时，有时数据是成对出现的，此时两个正态总体 X 与 Y 的独立性不成立，前面的方法不再适用。

两个正态总体， $X\sim N(\mu_1,\sigma_1^2)$， $Y\sim N(\mu_2,\sigma_2^2)$， $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 为来自总体 X 与 Y 的成对数据样本，检验均值 $\mu_1=\mu_2$。

令 $D=X-Y$，设 D 服从正态分布 $N(\mu_d,\sigma_d^2)$，其中 $\mu_d=\mu_1-\mu_2$，σ_d^2 未知。再按单个正态总体未知方差，检验均值的 T 检验法，检验 $\mu_d=\mu_1-\mu_2=0$。

步骤：

（1）假设 $H_0:\mu_d=0$ vs $H_1:\mu_d\neq 0$（或 $\mu_d<0$，$\mu_d>0$）。

（2）统计量 $T=\dfrac{\bar{D}-\mu_d}{S_d/\sqrt{n}}\sim t(n-1)$，检验统计量 $T=\dfrac{\bar{D}}{S_d/\sqrt{n}}$。

（3）拒绝域：$W=\{|t|\geq t_{1-\alpha/2}(n-1)\}$（或 $\{t\leq -t_{1-\alpha}(n-1)\}$，$\{t\geq t_{1-\alpha}(n-1)\}$）。

（4）计算检验统计量观测值 t 与检验的 p 值，并作出决策。

这是 T 检验法。

例　为了比较两种谷物种子的优劣，特选取 10 块土质不全相同的土地，并将每块土地分为面积相同的两部分，分别种植这两种种子，施肥与田间管理在 20 小块土地上都是一样，下面是各小块上的单位产量：

土地	1	2	3	4	5	6	7	8	9	10
种子一	23	35	29	42	39	29	37	34	35	28
种子二	30	39	35	40	38	34	36	33	41	31
差	−7	−4	−6	2	1	−5	1	1	−6	−3

假定单位产量服从正态分布，试问：两种种子的平均产量在显著水平 $\alpha=0.05$ 下有无显著差异？

解： 两个正态总体成对数据检验 $\mu_1=\mu_2$。令 $D=X-Y$，用 T 检验法。假设 $H_0:\mu_d=0$ vs $H_1:\mu_d\neq 0$，检验统计量 $T=\dfrac{\bar{D}}{S_d/\sqrt{n}}$，显著水平 $\alpha=0.05$，$n=10$，$t_{1-\alpha/2}(n-1)=t_{0.975}(9)=2.2622$，双侧拒绝域 $W=\{|t|\geq 2.2622\}$。

因 $\bar{d}=-2.6$，$s_d^2=3.5024^2$，$n=10$，则检验统计量观测值

$$t=\frac{-2.6}{3.5024/\sqrt{10}}=-2.3475\in W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{T\leq -2.3475\},2P\{T\geq -2.3475\}\}=2P\{T\leq -2.3475\}=0.0435<\alpha=0.05,$$

故拒绝 H_0，接受 H_1。可以认为两种种子的平均产量有显著差异。

§7.3　其它分布参数的假设检验及似然比检验

7.3.1　　指数分布参数的假设检验

总体 X 服从指数分布 $Exp\left(\frac{1}{\theta}\right)$， $X_1, X_2, \cdots, X_n$ 为样本，检验参数 θ 。参数 θ 的点估计为 $\bar{X}$ 。

因自由度为 n 的 χ^2 分布密度函数为

$$p(y)=\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} I_{y>0},$$

当 $n=2$ 时， χ^2 分布 $\chi^2(2)$ 的密度函数 $p(y)=\frac{1}{2} e^{-\frac{y}{2}} I_{y>0}$ ，这是指数分布 $Exp\left(\frac{1}{2}\right)$ 的密度函数。

因 $X_i \sim Exp\left(\frac{1}{\theta}\right)$，密度函数为 $p_X(x_i)=\frac{1}{\theta} e^{-\frac{x_i}{\theta}} I_{x_i>0}$ ，可知 $Y_i=\frac{2X_i}{\theta}$ 的密度函数为

$$p_Y(y_i)=\frac{1}{\theta} e^{-\frac{1}{\theta}\cdot\frac{\theta y_i}{2}} I_{\frac{\theta y_i}{2}>0} \cdot \frac{\theta}{2}=\frac{1}{2} e^{-\frac{y_i}{2}} I_{y_i>0},$$

则 $Y_i=\frac{2X_i}{\theta} \sim Exp\left(\frac{1}{2}\right)$，即 $\frac{2X_i}{\theta} \sim \chi^2(2),\quad i=1, 2, \cdots, n$ ，并且相互独立，可得

$$\sum_{i=1}^{n} \frac{2X_i}{\theta}=\frac{2n\bar{X}}{\theta} \sim \chi^2(2n) 。$$

步骤：

（1）假设 $H_0: \theta=\theta_0$ vs $H_1: \theta \neq \theta_0$（或 $\theta<\theta_0$， $\theta>\theta_0$）。

（2）统计量 $\chi^2=\frac{2n\bar{X}}{\theta} \sim \chi^2(2n)$ ，检验统计量 $\chi^2=\frac{2n\bar{X}}{\theta_0}$ 。

（3）拒绝域：$W=\{\chi^2 \leq \chi^2_{\alpha/2}(2n)$ 或 $\chi^2 \geq \chi^2_{1-\alpha/2}(2n)\}$（或 $\{\chi^2 \leq \chi^2_{\alpha}(2n)\}$， $\{\chi^2 \geq \chi^2_{1-\alpha}(2n)\}$）。

（4）计算检验统计量观测值 χ^2 与检验的 p 值，并作出决策。

这是 χ^2 检验法。

注：如果总体 X 服从指数分布 $Exp(\lambda)$， $X_1, X_2, \cdots, X_n$ 为样本，检验参数 λ ，参数 λ 的点估计为 $\frac{1}{\bar{X}}$ 。此时统计量 $\chi^2=2n\lambda\bar{X} \sim \chi^2(2n)$ 。如对于单侧假设 $H_0: \lambda \geq \lambda_0$ vs $H_1: \lambda<\lambda_0$ ，检验统计量 $\chi^2=2n\lambda_0\bar{X}$ ，但如果备择假设 H_1 为真，最可能出现 $\bar{x}>1/\lambda_0$ ，从而检验统计量观测值 $\chi^2=2n\lambda_0\bar{x}$ 最可能大于 $2n$ ，因此拒绝域在右侧，$W=\{\chi^2 \geq \chi^2_{1-\alpha}(2n)\}$ 。

可见，拒绝域的方向并不一定与H_1中的不等式方向一致，应该根据具体情况而定。

例 设某种电子元件的平均寿命服从指数分布。随机抽取 5 个元件，测得失效时间（小时）为 395，4094，119，11572，6133。检验这种元件的平均寿命是否不小于 6000 小时。（$\alpha=0.05$）

解：总体服从指数分布，用χ^2检验法。假设$H_0:\theta\geq 6000$ vs $H_1:\theta<6000$，检验统计量$\chi^2=\dfrac{2n\bar{X}}{6000}$，显著水平$\alpha=0.05$，$n=5$，$\chi^2_\alpha(2n)=\chi^2_{0.05}(10)=3.9403$，左侧拒绝域$W=\{\chi^2\leq 3.9403\}$。

因$\bar{x}=4462.6$，$n=5$，则检验统计量观测值

$$\chi^2=\frac{2\times 5\times 4462.6}{6000}=7.4377\notin W，$$

并且左侧检验的p值

$$p=P\{\chi^2\leq 7.4377\}=0.3164>\alpha=0.05，$$

故接受H_0，拒绝H_1。可以认为这种元件的平均寿命不小于 6000 小时。

7.3.2 比例 p 的假设检验

检验事件A在一次试验中发生的概率p，可以看作总体X服从二项分布$b(1,p)$，样本是n重伯努利试验每一次试验中事件A的发生次数，检验概率p。在n重伯努利试验中事件A的发生次数为$\sum_{i=1}^{n}X_i=n\bar{X}$，且$n\bar{X}\sim b(n,p)$，概率$p$的点估计为频率$\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$。但由于二项分布是离散型分布，确定拒绝域时，往往不能使得在H_0成立时检验统计量观测值落入拒绝域内的概率恰好等于显著水平α，因此选取的拒绝域是使得检验统计量观测值落入其中的概率不超过α。

步骤：

（1）假设$H_0:p=p_0$ vs $H_1:p\neq p_0$（或$p<p_0$，$p>p_0$）。

（2）统计量$n\bar{X}\sim b(n,p)$，检验统计量$n\bar{X}$。

（3）拒绝域：$W=\{n\bar{x}\leq c_{\alpha/2}$ 或 $n\bar{x}\geq c_{1-\alpha/2}\}$（或$\{n\bar{x}\leq c_\alpha\}$，$\{n\bar{x}\geq c_{1-\alpha}\}$）。其中

$$c_{\alpha/2}=\max_c\left\{c:\sum_{i=0}^{c}C_n^i p_0^i(1-p_0)^{n-i}\leq\frac{\alpha}{2}\right\},$$

$$c_{1-\alpha/2}=\min_c\left\{c:\sum_{i=c}^{n}C_n^i p_0^i(1-p_0)^{n-i}\leq\frac{\alpha}{2}\right\}=\min_c\left\{c:\sum_{i=0}^{c-1}C_n^i p_0^i(1-p_0)^{n-i}\geq 1-\frac{\alpha}{2}\right\},$$

并且类似定义c_α与$c_{1-\alpha}$。

（4）计算检验统计量观测值$n\bar{x}$与检验的p值，并作出决策。

例 某厂生产的产品优质品率一直保持在 40%，近期对该厂生产的该类产品抽检 20 件，其中优质品 7 件，在$\alpha=0.05$下能否认为优质品率仍保持在 40%？

解：比例p假设检验。假设$H_0:p=0.4$ vs $H_1:p\neq 0.4$，检验统计量$n\bar{X}$，显著水平$\alpha=0.05$，$n=20$，

$p=0.4$，有

$$P\{n\bar{x}\le 3\}=0.016<\frac{\alpha}{2}=0.025<P\{n\bar{x}\le 4\}=0.051,$$

$$P\{n\bar{x}\ge 12\}=0.0565>\frac{\alpha}{2}=0.025>P\{n\bar{x}\ge 13\}=0.021,$$

即 $c_{\alpha/2}=c_{0.025}=3$， $c_{1-\alpha/2}=c_{0.975}=13$，双侧拒绝域 $W=\{n\bar{x}\le 3$ 或 $n\bar{x}\ge 13\}$。检验统计量观测值

$$n\bar{x}=7\notin W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{n\bar{x}\le 7\},2P\{n\bar{x}\ge 7\}\}=2P\{n\bar{x}\le 7\}=0.8318>\alpha=0.05,$$

故接受 H_0，拒绝 H_1。可以认为优质品率仍保持在 40%。

补充：泊松分布参数的假设检验

总体 X 服从泊松分布 $P(\lambda)$， $X_1,X_2,\cdots,X_n$ 为样本，检验参数 λ，参数 λ 的点估计为 $\bar{X}$。

因 $X_i\sim P(\lambda),\quad i=1,2,\cdots,n$，并且相互独立，则

$$\sum_{i=1}^{n}X_i=n\bar{X}\sim P(n\lambda)。$$

步骤：

（1）假设 $H_0:\lambda=\lambda_0$ vs $H_1:\lambda\ne\lambda_0$（或 $\lambda<\lambda_0$， $\lambda>\lambda_0$）。

（2）统计量 $n\bar{X}\sim P(n\lambda)$，检验统计量 $n\bar{X}$。

（3）拒绝域：$W=\{n\bar{x}\le c_{\alpha/2}$ 或 $n\bar{x}\ge c_{1-\alpha/2}\}$（或 $\{n\bar{x}\le c_{\alpha}\}$，$\{n\bar{x}\ge c_{1-\alpha}\}$）。其中

$$c_{\alpha/2}=\max_c\left\{c:\sum_{i=0}^{c}\frac{(n\lambda)^i}{i!}e^{-n\lambda}\le\frac{\alpha}{2}\right\},$$

$$c_{1-\alpha/2}=\min_c\left\{c:\sum_{i=c}^{n}\frac{(n\lambda)^i}{i!}e^{-n\lambda}\le\frac{\alpha}{2}\right\}=\min_c\left\{c:\sum_{i=0}^{c-1}\frac{(n\lambda)^i}{i!}e^{-n\lambda}\ge 1-\frac{\alpha}{2}\right\},$$

并且类似定义 c_α 与 $c_{1-\alpha}$。

（4）计算检验统计量观测值 $n\bar{x}$ 与检验的 p 值，并作出决策。

例 设一场足球比赛的进球数服从泊松分布 $P(\lambda)$。2018 年世界杯足球比赛共进行了 64 场比赛，进球数的情况如下表：

进球数	0	1	2	3	4	5	6	7
场次	1	15	17	19	5	2	2	3

据此检验能否认为每场足球比赛的平均进球数是 2.5 个？（$\alpha=0.05$）

解：总体服从泊松分布。假设 $H_0:\lambda=2.5$ vs $H_1:\lambda\ne 2.5$，检验统计量 $n\bar{X}$，显著水平 $\alpha=0.05$，$n=64$，$\lambda=2.5$，有

$$P\{n\bar{X}\le 135\}=0.0241<\frac{\alpha}{2}=0.025<P\{n\bar{X}\le 136\}=0.0292,$$

$$P\{n\bar{X}\ge 185\}=0.0285>\frac{\alpha}{2}=0.025>P\{n\bar{X}\ge 186\}=0.0239,$$

即 $c_{\alpha/2}=c_{0.025}=135$，$c_{1-\alpha/2}=c_{0.975}=186$，双侧拒绝域 $W=\{n\bar{x}\le 135$ 或 $n\bar{x}\ge 186\}$。检验统计量观测值

$$n\bar{x}=169\notin W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{n\bar{x}\le 169\},2P\{n\bar{x}\ge 169\}\}=2P\{n\bar{x}\ge 169\}=0.4970>\alpha=0.05,$$

故接受 H_0，拒绝 H_1。可以认为每场足球比赛的平均进球数是 2.5 个。

7.3.3 大样本检验

假设检验问题中对总体参数 θ 进行检验，一般根据 θ 的最大似然估计 $\hat{\theta}$ 构造检验统计量。在大样本场合下，可用 $\hat{\theta}$ 的极限分布处理。特别是当 $\hat{\theta}$ 为独立随机变量和的形式时，由中心极限定理知可用正态分布近似处理。设 $\hat{\theta}$ 的数学期望为 θ，方差为 $\sigma^2(\theta)$。当 n 很大时，有

$$U=\frac{\hat{\theta}-\theta}{\sigma(\theta)}\sim N(0,1),$$

对于假设 $H_0:\theta=\theta_0$ vs $H_1:\theta\ne\theta_0$（或 $\theta<\theta_0$，$\theta>\theta_0$），根据 U 检验法进行检验。一般检验统计量取为

$$U=\frac{\hat{\theta}-\theta_0}{\sigma(\theta_0)} \text{ 或 } U=\frac{\hat{\theta}-\theta_0}{\sigma(\hat{\theta})},$$

一．指数分布参数的大样本检验

总体 X 服从指数分布 $Exp\left(\frac{1}{\theta}\right)$，$X_1,X_2,\cdots,X_n$ 为样本，检验参数 θ。参数 θ 的点估计为 $\bar{X}$，有

$$E(\bar{X})=E(X)=\theta,\quad \mathrm{Var}(\bar{X})=\frac{\mathrm{Var}(X)}{n}=\frac{\theta^2}{n},$$

当 n 很大时，有

$$U=\frac{\bar{X}-\theta}{\theta/\sqrt{n}}\sim N(0,1).$$

步骤：

（1）假设 $H_0:\theta=\theta_0$ vs $H_1:\theta\ne\theta_0$（或 $\theta<\theta_0$，$\theta>\theta_0$）。

（2）统计量 $U=\dfrac{\bar{X}-\theta}{\theta/\sqrt{n}}\sim N(0,1)$，检验统计量 $U=\dfrac{\bar{X}-\theta_0}{\theta_0/\sqrt{n}}$ 或 $U=\dfrac{\bar{X}-\theta_0}{\bar{X}/\sqrt{n}}$。

（3）拒绝域：$W=\{|u|\ge u_{1-\alpha/2}\}$（或 $\{u\le -u_{1-\alpha}\}$，$\{u\ge u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这是 U 检验法。

二．比例 p 的大样本检验

检验事件 A 在一次试验中发生的概率 p，可以看作总体 X 服从二项分布 $b(1,p)$，样本是 n 重伯努利试

验每一次试验中事件 A 的发生次数，检验概率 p 。设 n 重伯努利试验中事件 A 的发生次数为 $\sum_{i=1}^{n} X_i = n\bar{X}$ ，且 $n\bar{X} \sim b(n,\ p)$ ，概率 p 的点估计为频率 $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ ，有

$$E(\bar{X}) = E(X) = p，\quad \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{p(1-p)}{n}，$$

当 n 很大时，有

$$U = \frac{\bar{X} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \stackrel{\cdot}{\sim} N(0,1)。$$

步骤：

（1）假设 $H_0: p = p_0$ vs $H_1: p \ne p_0$（或 $p < p_0$， $p > p_0$）。

（2）统计量 $U = \dfrac{\bar{X} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \stackrel{\cdot}{\sim} N(0,1)$，检验统计量 $U = \dfrac{\bar{X} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ 或 $U = \dfrac{\bar{X} - p_0}{\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}}$ 。

（3）拒绝域：$W = \{|u| \ge u_{1-\alpha/2}\}$（或 $\{u \le -u_{1-\alpha}\}$，$\{u \ge u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这是 U 检验法。

例 掷一枚硬币 100 次，结果正面出现了 65 次，能否认为这枚硬币均匀？（$\alpha = 0.05$）

解： 比例 p 的假设检验，大样本场合。假设 $H_0: p = 0.5$ vs $H_1: p \ne 0.5$，检验统计量 $U = \dfrac{\bar{X} - 0.5}{\sqrt{\dfrac{0.25}{n}}}$，显著水平 $\alpha = 0.05$， $u_{1-\alpha/2} = u_{0.975} = 1.96$，双侧拒绝域 $W = \{|u| \ge 1.96\}$ 。

因 $\bar{x} = \dfrac{65}{100} = 0.65$， $n = 100$，则检验统计量观测值

$$u = \frac{0.65 - 0.5}{\sqrt{\dfrac{0.25}{100}}} = 3 \in W，$$

并且双侧检验的 p 值

$$p = \min\{2P\{U \le 3\}, 2P\{U \ge 3\}\} = 2P\{U \ge 3\} = 2 - 2\Phi(3) = 0.0026 < \alpha = 0.05，$$

故拒绝 H_0，接受 H_1。可以认为这枚硬币不均匀。

三．两个比例 p_1 与 p_2 比较的大样本检验

检验事件 A 与 B 在一次试验中发生的概率 p_1 与 p_2 是否相等，可以看作总体 X 与 Y 分别服从二项分布 $b(1,\ p_1)$ 与 $b(1,\ p_2)$，且相互独立，样本分别是 n_1 与 n_2 重伯努利试验各自的每一次试验中事件 A 与 B 的发生次数，检验概率 $H_0: p_1 = p_2$ 。事件 A 与 B 的发生次数分别为 $\sum_{i=1}^{n_1} X_i = n_1\bar{X}$ 与 $\sum_{j=1}^{n_2} Y_j = n_2\bar{Y}$ ，且 $n_1\bar{X} \sim b(n_1,\ p_1)$，

$n_2\bar{Y} \sim b(n_2, p_2)$，概率 p_1 与 p_2 的点估计分别为频率 $\frac{1}{n_1}\sum_{i=1}^{n_1} X_i = \bar{X}$ 与 $\frac{1}{n_2}\sum_{j=1}^{n_2} Y_j = \bar{Y}$，并且有

$$E(\bar{X} - \bar{Y}) = E(X) - E(Y) = p_1 - p_2,$$

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{\mathrm{Var}(X)}{n_1} + \frac{\mathrm{Var}(Y)}{n_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2},$$

当 n_1 与 n_2 很大时，有

$$U = \frac{(\bar{X} - \bar{Y}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \dot{\sim} N(0,1)\text{。}$$

在 $H_0: p_1 = p_2$ 成立的条件下，记 $p_1 = p_2 = p$，有

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{p(1-p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \dot{\sim} N(0,1),$$

概率 p 未知，用总频率 $\hat{p} = \frac{n_1\bar{X} + n_2\bar{Y}}{n_1 + n_2}$ 替换，则

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \dot{\sim} N(0,1)\text{。}$$

步骤：

（1）假设 $H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$。

（2）检验统计量 $U = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$，其中 $\hat{p} = \frac{n_1\bar{X} + n_2\bar{Y}}{n_1 + n_2}$。

（3）拒绝域：$W = \{|u| \geq u_{1-\alpha/2}\}$（或 $\{u \leq -u_{1-\alpha}\}$，$\{u \geq u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这是 U 检验法。

四．泊松分布参数的大样本检验

总体 X 服从泊松分布 $P(\lambda)$，$X_1, X_2, \cdots, X_n$ 为样本，检验参数 λ，参数 λ 的点估计为 $\bar{X}$，有

$$E(\bar{X}) = E(X) = \lambda, \quad \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{\lambda}{n},$$

当 n 很大时，有

$$U = \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \dot{\sim} N(0,1)\text{。}$$

步骤：

（1）假设 $H_0:\lambda=\lambda_0$ vs $H_1:\lambda\neq\lambda_0$（或 $\lambda<\lambda_0$，$\lambda>\lambda_0$）。

（2）统计量 $U=\dfrac{\bar{X}-\lambda}{\sqrt{\lambda/n}}\sim N(0,1)$，检验统计量 $U=\dfrac{\bar{X}-\lambda_0}{\sqrt{\lambda_0/n}}$ 或 $U=\dfrac{\bar{X}-\lambda_0}{\sqrt{\bar{X}/n}}$。

（3）拒绝域：$W=\{|u|\geq u_{1-\alpha/2}\}$（或 $\{u\leq -u_{1-\alpha}\}$，$\{u\geq u_{1-\alpha}\}$）。

（4）计算检验统计量观测值 u 与检验的 p 值，并作出决策。

这是 U 检验法。

例 设一场足球比赛的进球数服从泊松分布 $P(\lambda)$。以 2018 年世界杯足球赛进球数的情况，检验每场足球比赛的平均进球数是否为 2.5 个。（$\alpha=0.05$）

解：总体服从泊松分布，大样本场合。假设 $H_0:\lambda=2.5$ vs $H_1:\lambda\neq 2.5$，检验统计量 $U=\dfrac{\bar{X}-2.5}{\sqrt{2.5/n}}$，

显著水平 $\alpha=0.05$，$u_{1-\alpha/2}=u_{0.975}=1.96$，双侧拒绝域 $W=\{|u|\geq 1.96\}$。

因 $\bar{x}=\dfrac{169}{64}=2.6406$，$n=64$，则检验统计量观测值

$$u=\frac{2.6406-2.5}{\sqrt{\dfrac{2.5}{64}}}=0.7115\notin W,$$

并且双侧检验的 p 值

$$p=\min\{2P\{U\leq 0.7115\},2P\{U\geq 0.7115\}\}=2P\{U\geq 0.7115\}=0.4768>\alpha=0.05,$$

故接受 H_0，拒绝 H_1。可以认为足球比赛的平均进球数为 2.5 个。

7.3.4 似然比检验

进行假设检验时，关键是如何构造检验统计量，一般是利用未知参数的最大似然估计，根据其分布作适当变化而得到，但究竟怎样变化，前面的内容中都是根据具体的参数而定，没有一个统一的方法。这里给出似然比检验统计量，则是按照一个统一的方法进行构造。

基本思想是利用似然函数的上确界构造，似然函数在整个参数空间的上确界与其在原假设 H_0 成立条件下的上确界之比构成似然比检验统计量。

定义 设总体 X 中未知参数为 θ，似然函数为 $L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)$。检验假设 $H_0:\theta\in\Theta_0$，令

$$\Lambda(X_1,X_2,\cdots,X_n)=\frac{\sup\limits_{\theta\in\Theta}L(\theta)}{\sup\limits_{\theta\in\Theta_0}L(\theta)},$$

称为似然比（Likelihood Ratio）。

以似然比 Λ 作为检验统计量，显然 $\Lambda\geq 1$。如果原假设 H_0 为真，则 $\sup\limits_{\theta\in\Theta}L(\theta)$ 与 $\sup\limits_{\theta\in\Theta_0}L(\theta)$ 应相差不大，即 Λ 的观测值接近于 1 才"合理"。可见当 Λ 的观测值接近于 1 时，应接受 H_0；当 Λ 的观测值远大于 1 时，应拒绝 H_0。取右侧拒绝域，形式为 $W=\{\Lambda\geq\Lambda_{1-\alpha}\}$，其中 $\Lambda_{1-\alpha}$ 是似然比 Λ 的 $1-\alpha$ 分位数。

NP 引理 设总体 X 的密度函数或质量函数为 $p(x;\theta)$，θ 为未知参数，似然函数为 $L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)$。给定 $\alpha\in(0,1)$，设 $W_0=\left\{(X_1,X_2,\cdots,X_n):\frac{L(\theta_1)}{L(\theta_0)}\ge\lambda_0\right\}$满足

$$P\{(X_1,X_2,\cdots,X_n)\in W_0\mid\theta=\theta_0\}=\int\cdots\int_{W_0}L(\theta_0)dx_1\cdots dx_n=\alpha,$$

则对任何拒绝域 $W\subset R^n$，只要 $P\{(X_1,X_2,\cdots,X_n)\in W\mid\theta=\theta_0\}\le\alpha$，则必有

$$P\{(X_1,X_2,\cdots,X_n)\notin W_0\mid\theta=\theta_1\}\le P\{(X_1,X_2,\cdots,X_n)\notin W\mid\theta=\theta_1\}。$$

即 W_0 是所有显著水平为 α 的拒绝域中犯第二类错误的概率 β 最小的一个。

判断一个检验统计量的优劣，理论上是根据在相同的显著水平 α 下，犯第二类错误概率 β 的大小而定，β 越小，检验统计量越好。根据 NP 引理，似然比 Λ 作为检验统计量，在相同的显著水平 α 下，犯第二类错误概率 β 最小，可见似然比 Λ 是最好的检验统计量。

但似然比统计量 Λ 的精确分布一般比较复杂，难以得到分位数 $\Lambda_{1-\alpha}$ 的精确值。设函数 f 可逆，统计量 $f(\Lambda)$ 与 Λ 作为检验统计量得到的拒绝域是等价的。因此通常是找到一个与似然比检验统计量 Λ 具有函数关系并且精确分布比较简单的统计量 $f(\Lambda)$ 作为检验统计量。从而 $f(\Lambda)$ 在理论上和实际应用上都是最好的检验统计量。

对于简单原假设 $H_0:\theta=\theta_0$，若只有 θ 一个未知参数，且 θ 的最大似然估计为 $\hat\theta$，则 $\sup_{\theta\in\Theta}L(\theta)=L(\hat\theta)$，$\sup_{\theta\in\Theta_0}L(\theta)=L(\theta_0)$，即似然比

$$\Lambda(X_1,X_2,\cdots,X_n)=\frac{L(\hat\theta)}{L(\theta_0)};$$

若不只 θ 一个未知参数，不妨设有 θ 和 φ 两个参数，θ 与 φ 的最大似然估计分别为 $\hat\theta$ 与 $\hat\varphi$，且当 $\theta=\theta_0$ 时 φ 的最大似然估计为 $\hat\varphi_0$，则 $\sup_{\theta\in\Theta}L(\theta)=L(\hat\theta,\hat\varphi)$，$\sup_{\theta\in\Theta_0}L(\theta)=L(\theta_0,\hat\varphi_0)$，即似然比

$$\Lambda(X_1,X_2,\cdots,X_n)=\frac{L(\hat\theta,\hat\varphi)}{L(\theta_0,\hat\varphi_0)}。$$

例 总体 X 服从正态分布 $N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 为样本。已知 σ^2，求检验 $H_0:\mu=\mu_0$ 的似然比检验统计量。

解：已知 σ^2，似然函数

$$L(\mu)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\frac{1}{(\sqrt{2\pi})^n\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

只有 μ 一个未知参数，μ 的最大似然估计为 $\bar X$。对于简单原假设 $H_0:\mu=\mu_0$，似然比检验统计量

$$\Lambda(X_1, X_2, \cdots, X_n) = \frac{L(\bar{X})}{L(\mu_0)} = \frac{\dfrac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2}}{\dfrac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu_0)^2}} = e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (X_i - \bar{X})^2 - \sum_{i=1}^n (X_i - \mu_0)^2\right]}。$$

因

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n [(X_i - \mu_0) - (\bar{X} - \mu_0)]^2 = \sum_{i=1}^n (X_i - \mu_0)^2 - n(\bar{X} - \mu_0)^2,$$

故

$$\Lambda(X_1, X_2, \cdots, X_n) = e^{\frac{n(\bar{X} - \mu_0)^2}{2\sigma^2}} = e^{\frac{1}{2}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right)^2} = e^{\frac{1}{2}U^2}。$$

这与已知σ^2，检验μ的U统计量具有函数关系，因此$U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$是已知$\sigma^2$，检验$\mu$最好的检验统计量。

例　总体X服从正态分布$N(\mu, \sigma^2)$，$X_1, X_2, \cdots, X_n$为样本。未知σ^2，求检验$H_0: \mu = \mu_0$的似然比检验统计量。

解：未知σ^2，似然函数

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2},$$

有μ和σ^2两个未知参数，μ与σ^2的最大似然估计分别为$\bar{X}$与$S^{*2} = \dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$，当$\mu = \mu_0$时，$\sigma^2$的最大似然估计为$\hat{\sigma}_0^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$。对于简单原假设$H_0: \mu = \mu_0$，似然比检验统计量

$$\Lambda(X_1, X_2, \cdots, X_n) = \frac{L(\bar{X}, S^{*2})}{L(\mu_0, \hat{\sigma}_0^2)} = \frac{\dfrac{1}{(\sqrt{2\pi})^n (S^{*2})^{\frac{n}{2}}} e^{-\frac{1}{2S^{*2}}\sum_{i=1}^n (X_i - \bar{X})^2}}{\dfrac{1}{(\sqrt{2\pi})^n (\hat{\sigma}_0^2)^{\frac{n}{2}}} e^{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (X_i - \mu_0)^2}} = \left[\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right]^{\frac{n}{2}}。$$

因

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n [(X_i - \mu_0) - (\bar{X} - \mu_0)]^2 = \sum_{i=1}^n (X_i - \mu_0)^2 - n(\bar{X} - \mu_0)^2,$$

故

$$\Lambda(X_1, X_2, \cdots, X_n) = \left[\frac{(n-1)S^2 + n(\bar{X} - \mu_0)^2}{(n-1)S^2}\right]^{\frac{n}{2}} = \left[1 + \frac{1}{n-1}\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}}\right)^2\right]^{\frac{n}{2}} = \left(1 + \frac{T^2}{n-1}\right)^{\frac{n}{2}}。$$

这与未知σ^2，检验μ的T统计量具有函数关系，因此$T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$是未知$\sigma^2$，检验$\mu$最好的检验统计量。

§7.4 分布拟合检验

前几节是检验总体参数；这一节是检验总体分布，设$F(x)$为总体分布函数，$F_0(x)$为某指定的分布函数，检验假设$H_0: F(x)=F_0(x)$。（当然也可以对质量函数或密度函数进行检验）

直观上，这是非参数检验，但可以通过分类数据的拟合优度检验转化为多参数检验。

7.4.1 分类数据的 χ^2 拟合优度检验

孟德尔一次豌豆试验中收获了 556 颗豌豆，其中黄圆、绿圆、黄皱、绿皱分别有 315、108、101、32 颗，能否认为这四类的比例符合根据遗传学原理判断的 9:3:3:1？

当检验数据分成r类，设取得第i类数据的概率为

$$p_i,\quad i=1,2,\cdots,r\text{，且 }0<p_i<1\text{，}\quad p_1+p_2+\cdots+p_r=1\text{。}$$

对容量为n的数据样本，设其中属于第i类的样品个数为

$$n_i,\quad i=1,2,\cdots,r\text{，且 }n_i\ge 0\text{，}\quad n_1+n_2+\cdots+n_r=n\text{，}$$

显然$n_i\sim b(n,\,p_i)$，有$E(n_i)=np_i$，$\mathrm{Var}(n_i)=np_i(1-p_i)$，由中心极限定理可得，当$n$很大时，有

$$\frac{n_i-np_i}{\sqrt{np_i(1-p_i)}}\sim N(0,1)\text{，}$$

但由于$n_i,\quad i=1,2,\cdots,r$不是相互独立，不能推出它们的平方和服从χ^2分布，但可以证明

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\sim\chi^2(r-1)\text{，}$$

显然χ^2越小，说明总体分布与指定的分布吻合得越好，假设检验时取右侧拒绝域$W=\{\chi^2\ge\chi^2_{1-\alpha}(r-1)\}$。

检验取得第i类数据的概率p_i是否为给定的概率$\hat{p}_i,\quad i=1,2,\cdots,r$，步骤：

（1）假设$H_0: p_i=\hat{p}_i,\quad i=1,2,\cdots,r$。

（2）检验统计量$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\sim\chi^2(r-1)$。

（3）拒绝域：$\{\chi^2\ge\chi^2_{1-\alpha}(r-1)\}$。

（4）计算检验统计量观测值χ^2与检验的p值，并作出决策。

这是χ^2检验法。

注：（1）若给定概率$\hat{p}_i,\quad i=1,2,\cdots,r$的计算依赖于$k$个未知参数，此时$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}$不是统计

量。先对这 k 个未知参数作出最大似然估计，再由此计算出各个未知参数 p_i 的估计值 $\hat{p}_i$， $i=1,2,\cdots,r$，用 $\hat{p}_i$ 替换 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}$ 中的 p_i，可以证明 $\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$ 仍然近似服从 χ^2 分布，但自由度将减少 k 个，即

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\sim\chi^2(r-k-1)。$$

（2）这里的 χ^2 统计量都是采用渐近分布，因此样本容量不能太小，通常要求每一类别样本数据都不小于 5。

7.4.2　分布的 χ^2 拟合优度检验

为了检验总体 X 的分布是否为某指定的分布，总是可以将其取值分成有限类，进行分类数据的 χ^2 拟合优度检验。对于离散分布，可将每一个取值或合并相邻的某几个取值分别作为一类；对于连续分布，可将实数轴化分成若干个区间，将每一个区间的取值分别作为一类。

一．总体分布为离散分布

总体 X 的分布为离散分布，将每一个取值或合并相邻的某几个取值分别作为一类，使得每一类别样本数据都不小于 5。设共分成 r 个类别，这样将原假设 $H_0:F(x)=F_0(x)$ 改为 $H_0:p_i=\hat{p}_i$， $i=1,2,\cdots,r$。并设概率 $\hat{p}_i$， $i=1,2,\cdots,r$ 的计算依赖于 k 个未知参数（$0\le k<r-1$）。

步骤：

（1）假设 $H_0:p_i=\hat{p}_i$， $i=1,2,\cdots,r$。

（2）检验统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\sim\chi^2(r-k-1)$。

（3）拒绝域：$\{\chi^2\ge\chi^2_{1-\alpha}(r-k-1)\}$。

（4）计算检验统计量观测值 χ^2 与检验的 p 值，并作出决策。

这是 χ^2 检验法。

例　掷一枚骰子 100 次，结果如下：

点数	1	2	3	4	5	6
次数	14	25	17	10	23	11

可否认为这枚骰子不均匀？（$\alpha=0.05$）

解：检验骰子是否均匀，即检验每个点数出现的概率是否都是 $\frac{1}{6}$，用 χ^2 检验法。分成 $r=6$ 个类别，概率 $p_{i0}=\frac{1}{6}$ 不依赖于未知参数（$k=0$）。假设 $H_0:p_i=\frac{1}{6}$， $i=1,2,\cdots,6$，检验统计量 $\chi^2=\sum_{i=1}^{6}\frac{(n_i-n/6)^2}{n/6}$，

显著水平 $\alpha=0.05$， $r=6$， $\chi^2_{1-\alpha}(r-1)=\chi^2_{0.95}(5)=11.0705$，右侧拒绝域 $W=\{\chi^2\ge 11.0705\}$。列表计算

点数	1	2	3	4	5	6	Σ
次数 n_i	14	25	17	10	23	11	100
p_i	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	1
np_i	16.6667	16.6667	16.6667	16.6667	16.6667	16.6667	100
$(n_i-np_i)^2/(np_i)$	0.4267	4.1667	0.0067	2.6667	2.4067	1.9267	11.6

检验统计量观测值

$$\chi^2=\sum_{i=1}^{6}\frac{(n_i-n/6)^2}{n/6}=11.6\in W,$$

并且右侧检验的 p 值

$$p=P\{\chi^2\ge 11.6\}=0.0407<\alpha=0.05,$$

故拒绝 H_0，接受 H_1。可以认为这枚骰子不均匀。

例 以 2018 年世界杯足球比赛进球数的情况检验一场足球比赛进球数是否服从泊松分布。（$\alpha=0.05$）

进球数	0	1	2	3	4	5	6	7
场次	1	15	17	19	5	2	2	3

解： 离散分布，用 χ^2 检验法。按进球数合并后分成 $r=5$ 类：$(\le 1), 2, 3, 4, (\ge 5)$，如下表：

进球数	≤ 1	2	3	4	≥ 5
场次	16	17	19	5	7

计算概率 $\hat{p}_i$ 需要估计 $k=1$ 个参数 λ，其最大似然估计为

$$\hat{\lambda}=\overline{x}=\frac{169}{64},$$

可得

$$\hat{p}_1=e^{-\hat{\lambda}}+\hat{\lambda}e^{-\hat{\lambda}}=0.2596,\quad \hat{p}_2=\frac{\hat{\lambda}^2}{2}e^{-\hat{\lambda}}=0.2486,\quad \hat{p}_3=\frac{\hat{\lambda}^3}{3!}e^{-\hat{\lambda}}=0.2189,$$

$$\hat{p}_4=\frac{\hat{\lambda}^4}{4!}e^{-\hat{\lambda}}=0.1445,\quad \hat{p}_5=\sum_{j=5}^{+\infty}\frac{\hat{\lambda}^j}{j!}e^{-\hat{\lambda}}=0.1284,$$

假设 $H_0: p_1=0.2596,\ p_2=0.2486,\ p_3=0.2189,\ p_4=0.1445,\ p_5=0.1284$，检验统计量 $\chi^2=\sum_{i=1}^{5}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$，

显著水平 $\alpha=0.05$， $r=5$， $k=1$， $\chi^2_{1-\alpha}(r-k-1)=\chi^2_{0.95}(3)=7.8147$，右侧拒绝域 $W=\{\chi^2\ge 7.8147\}$。列表计算

进球数	≤ 1	2	3	4	≥ 5	Σ
场次 n_i	16	17	19	5	7	64
$\hat{p}_i$	0.2596	0.2486	0.2189	0.1445	0.1284	1
$n\hat{p}_i$	16.6168	15.9131	14.0068	9.2467	8.2166	64
$(n_i-n\hat{p}_i)^2/(n\hat{p}_i)$	0.0229	0.0742	1.7800	1.9504	0.1801	4.0076

检验统计量观测值

$$\chi^2=\sum_{i=1}^{5}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}=4.0076\notin W\text{，}$$

并且右侧检验的 p 值

$$p=P\{\chi^2\ge 4.0076\}=0.2606>\alpha=0.05\text{，}$$

故接受 H_0，拒绝 H_1。可以认为一场足球比赛的进球数服从泊松分布。

二．总体分布为连续分布

对样本进行分组。先确定组数，通常在 5 到 20 个之间。对容量较小的样本分为 5 到 6 组，容量为 100 左右的样本分为 7 到 10 组，容量为 200 左右的样本分为 9 到 13 组，容量为 300 及以上的样本分为 12 到 20 组。再确定组距，通常选取相同长度的区间进行分组，组距约为样本极差与组数之商。最后确定组限，即具体的分组区间，要求每一分组区间的样本数据都不小于 5。

每一分组区间作为一类，设分成 r 个类别，将原假设 $H_0:F(x)=F_0(x)$ 改为 $H_0:p_i=\hat{p}_i,\quad i=1,2,\cdots,r$。

并设概率 $\hat{p}_i,\quad i=1,2,\cdots,r$ 的计算依赖于 k 个未知参数（$0\le k<r-1$）。

步骤：

（1）假设 $H_0:p_i=\hat{p}_i,\quad i=1,2,\cdots,r$。

（2）检验统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\sim\chi^2(r-k-1)$。

（3）拒绝域：$\{\chi^2\ge\chi^2_{1-\alpha}(r-k-1)\}$。

（4）计算检验统计量观测值 χ^2 与检验的 p 值，并作出决策。

这是 χ^2 检验法。

例　某气象站收集了 44 个独立的年降雨量数据，资料如下：

520	556	561	616	635	669	686	692	704	707	711
713	714	719	727	735	740	744	745	750	776	777
786	786	791	794	821	822	826	834	837	851	862
873	879	889	900	904	922	926	952	963	1056	1074

检验这批数据是否服从正态分布？（$\alpha=0.05$）

解：连续分布，用 χ^2 检验法。对这批数据分组，44 个数据分成 5 组，分组区间为

$$(-\infty,650],\quad(650,740],\quad(740,830],\quad(830,920],\quad(920,+\infty)\text{。}$$

各区间的频数依次为 $5,12,12,9,6$，并且计算概率 $\hat{p}_i$ 需要估计 $k=2$ 个参数 μ 与 σ^2，其最大似然估计为

$$\hat{\mu}=\bar{x}=785.1136\text{，}\quad\hat{\sigma}^2=s^{*2}=119.7414^2\text{，}$$

可得

$$\hat{p}_1=\Phi\left(\frac{650-\hat{\mu}}{\hat{\sigma}}\right)=0.1296\text{，}\quad\hat{p}_2=\Phi\left(\frac{740-\hat{\mu}}{\hat{\sigma}}\right)-\Phi\left(\frac{650-\hat{\mu}}{\hat{\sigma}}\right)=0.2236\text{，}$$

$$\hat{p}_3 = \Phi\left(\frac{830-\hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{740-\hat{\mu}}{\hat{\sigma}}\right) = 0.2929，\quad \hat{p}_4 = \Phi\left(\frac{920-\hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{830-\hat{\mu}}{\hat{\sigma}}\right) = 0.2239，$$

$$\hat{p}_5 = 1 - \Phi\left(\frac{920-\hat{\mu}}{\hat{\sigma}}\right) = 0.1300，$$

假设 $H_0: p_1 = 0.1296,\ p_2 = 0.2236,\ p_3 = 0.2929,\ p_4 = 0.2239,\ p_5 = 0.1300$，检验统计量 $\chi^2 = \sum_{i=1}^{5} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$，

显著水平 $\alpha = 0.05$，$r = 5$，$k = 2$，$\chi^2_{1-\alpha}(r-k-1) = \chi^2_{0.95}(2) = 5.9915$，右侧拒绝域 $W = \{\chi^2 \ge 5.9915\}$。列表计算

年降雨量	$(-\infty, 650]$	$(650, 740]$	$(740, 830]$	$(830, 920]$	$(920, +\infty)$	Σ
次数 n_i	5	12	12	9	6	44
$\hat{p}_i$	0.1296	0.2236	0.2929	0.2239	0.1300	1
$n\hat{p}_i$	5.7024	9.8384	12.8876	9.8516	5.7200	44
$(n_i - n\hat{p}_i)^2 / (n\hat{p}_i)$	0.0865	0.4749	0.0611	0.0736	0.0137	0.7099

检验统计量观测值

$$\chi^2 = \sum_{i=1}^{5} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 0.7099 \notin W，$$

并且右侧检验的 p 值

$$p = P\{\chi^2 \ge 0.7099\} = 0.7012 > \alpha = 0.05，$$

故接受 H_0，拒绝 H_1。可以认为这批数据服从正态分布。

7.4.3　列联表独立性检验

检验一个总体的两种属性 A 与 B 是否独立。设属性 A 有 r 个类别 $A_1, A_2, \cdots, A_r$，属性 B 有 c 个类别 $B_1, B_2, \cdots, B_c$，且总体中每一个个体必属于属性 A 中某一类，也必属于属性 B 中某一类。总体中一个个体既属于 A_i 类又属于 B_j 类的概率为 p_{ij}，$i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c$，满足 $\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1$，则属于 A_i 类的概率为 $p_{i\cdot} = \sum_{j=1}^{c} p_{ij}$，$i = 1, 2, \cdots, r$，有 $\sum_{i=1}^{r} p_{i\cdot} = 1$，属于 B_j 类的概率为 $p_{\cdot j} = \sum_{i=1}^{r} p_{ij}$，$j = 1, 2, \cdots, c$，有 $\sum_{j=1}^{c} p_{\cdot j} = 1$。检验 A 与 B 是否独立，即检验 $H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$，$i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c$。总体的取值按属性 A 与 B 的类别，分成 rc 个类别，采用分类数据的 χ^2 拟合优度检验。概率 $p_{i\cdot}$ 与 $p_{\cdot j}$ 未知，用频率作为其点估计。

随机抽取容量为 n 的样本，设既属于 A_i 类又属于 B_j 类的样品个数为 n_{ij}，$i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c$，满足 $\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} = n$，属于 A_i 类的样品个数为 $n_{i\cdot} = \sum_{j=1}^{c} n_{ij}$，$i = 1, 2, \cdots, r$，有 $\sum_{i=1}^{r} n_{i\cdot} = n$，属于 B_j 类的样品个数

为 $n_{\cdot j}=\sum_{i=1}^{r}n_{ij}$， $j=1,2,\cdots,c$，有 $\sum_{j=1}^{c}n_{\cdot j}=n$。概率 $p_{i\cdot}$ 与 $p_{\cdot j}$ 的点估计分别为频率

$$\hat{p}_{i\cdot}=\frac{n_{i\cdot}}{n}=\frac{1}{n}\sum_{j=1}^{c}n_{ij},\quad i=1,2,\cdots,r \text{ 与 } \hat{p}_{\cdot j}=\frac{n_{\cdot j}}{n}=\frac{1}{n}\sum_{i=1}^{r}n_{ij},\quad j=1,2,\cdots,c,$$

$$\hat{p}_{ij}=\hat{p}_{i\cdot}\cdot\hat{p}_{\cdot j},\quad i=1,2,\cdots,r;\ j=1,2,\cdots,c,$$

因 $\sum_{i=1}^{r}p_{i\cdot}=1$ 且 $\sum_{j=1}^{c}p_{\cdot j}=1$，故总共需要估计 $(r-1)+(c-1)=r+c-2$ 个参数。统计量 $\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$，近似服从 χ^2 分布，自由度为 $rc-(r+c-2)-1=(r-1)(c-1)$。

步骤：

（1）假设 $H_0: p_{ij}=p_{i\cdot}\cdot p_{\cdot j},\quad i=1,2,\cdots,r;\ j=1,2,\cdots,c$。

（2）检验统计量 $\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\sim\chi^2((r-1)(c-1))$。

（3）拒绝域：$W=\{\chi^2\geq\chi^2_{1-\alpha}((r-1)(c-1))\}$。

（4）计算检验统计量观测值 χ^2 与检验的 p 值，并作出决策。

这是 χ^2 检验法。

例 为了检验性别（男或女）与色觉（正常或色盲）有无关系，随机抽取 1000 人按性别及色觉两个属性分类，得二维列联表

性别＼色觉	正常	色盲
男	535	65
女	382	18

检验二者有无显著关系。（$\alpha=0.05$）

解：列联表独立性检验，用 χ^2 检验法。计算概率 $\hat{p}_{ij}$ 需要估计 $k=2$ 个参数，

$$\hat{p}_{1\cdot}=\frac{n_{1\cdot}}{n}=\frac{535+65}{1000}=0.6,\ \text{且 } \hat{p}_{2\cdot}=1-\hat{p}_{1\cdot}=0.4,$$

$$\hat{p}_{\cdot 1}=\frac{n_{\cdot 1}}{n}=\frac{535+382}{1000}=0.917,\ \text{且 } \hat{p}_{\cdot 2}=1-\hat{p}_{\cdot 1}=0.083$$

可得

$$\hat{p}_{11}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 1}=0.6\times0.917=0.5502,\quad \hat{p}_{12}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 2}=0.6\times0.083=0.0498,$$

$$\hat{p}_{21}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 1}=0.4\times0.917=0.3668,\quad \hat{p}_{22}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 2}=0.4\times0.083=0.0332,$$

假设 $H_0: p_{ij}=p_{i\cdot}\cdot p_{\cdot j},\quad i=1,2;\ j=1,2$，检验统计量 $\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\sim\chi^2(1)$，显著水平 $\alpha=0.05$，

$r=2$，$c=2$，$\chi^2_{1-\alpha}((r-1)(c-1))=\chi^2_{0.95}(1)=3.8145$，右侧拒绝域$W=\{\chi^2\ge 3.8145\}$。

列表计算

性别	男		女		Σ
色觉	正常	色盲	正常	色盲	
人数 n_{ij}	535	65	382	18	1000
$\hat{p}_{ij}$	0.5502	0.0498	0.3668	0.0332	1
$n\hat{p}_{ij}$	550.2	49.8	366.8	33.2	1000
$(n_{ij}-n\hat{p}_{ij})^2\big/(n\hat{p}_{ij})$	0.4199	4.6394	0.6299	6.9590	12.6482

检验统计量观测值

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}=12.6482\in W,$$

并且右侧检验的 p 值

$$p=P\{\chi^2\ge 12.6482\}=0.0004<\alpha=0.05,$$

故拒绝H_0，接受H_1。可以认为性别与色觉有显著关系。

例 为研究儿童智力发展与营养的关系，某研究机构调查了 1436 名儿童，得到如下数据，在显著水平$\alpha=0.05$下判断智力发展与营养有无显著关系。

营养 \ 智商	<80	$80\sim 89$	$90\sim 99$	≥ 100	合计
营养良好	367	342	266	329	1304
营养不良	56	40	20	16	132
合计	423	382	286	345	1436

解：列联表独立性检验，用χ^2检验法。计算概率$\hat{p}_{ij}$需要估计$k=4$个参数，

$$\hat{p}_{1\cdot}=\frac{n_{1\cdot}}{n}=\frac{1304}{1436}=0.9081，\text{且 }\hat{p}_{2\cdot}=1-\hat{p}_{1\cdot}=0.0919,$$

$$\hat{p}_{\cdot 1}=\frac{n_{\cdot 1}}{n}=\frac{423}{1436}=0.2946,\quad \hat{p}_{\cdot 2}=\frac{n_{\cdot 2}}{n}=\frac{382}{1436}=0.2660,\quad \hat{p}_{\cdot 3}=\frac{n_{\cdot 3}}{n}=\frac{286}{1436}=0.1992,$$

$$\text{且 }\hat{p}_{\cdot 4}=1-\hat{p}_{\cdot 1}-\hat{p}_{\cdot 2}-\hat{p}_{\cdot 3}=0.2402$$

可得

$$\hat{p}_{11}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 1}=0.9081\times 0.2946=0.2675,\quad \hat{p}_{12}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 2}=0.9081\times 0.2660=0.2416,$$

$$\hat{p}_{13}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 3}=0.9081\times 0.1992=0.1809,\quad \hat{p}_{14}=\hat{p}_{1\cdot}\cdot\hat{p}_{\cdot 4}=0.9081\times 0.2402=0.2181,$$

$$\hat{p}_{21}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 1}=0.0919\times 0.2946=0.0271,\quad \hat{p}_{22}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 2}=0.0919\times 0.2660=0.0244,$$

$$\hat{p}_{23}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 3}=0.0919\times 0.1992=0.0183,\quad \hat{p}_{24}=\hat{p}_{2\cdot}\cdot\hat{p}_{\cdot 4}=0.0919\times 0.2402=0.0221,$$

假设 $H_0: p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$， $i=1,2;\ j=1,2,3,4$， 检验统计量 $\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\sim\chi^2(3)$， 显著水平

$\alpha=0.05$， $r=2$， $c=4$， $\chi^2_{1-\alpha}((r-1)(c-1))=\chi^2_{0.95}(3)=7.8147$， 右侧拒绝域$W=\{\chi^2\ge 7.8147\}$。

列表计算

营养	营养良好				营养不良				Σ
智商	<80	80−89	90−99	≥ 100	<80	80−89	90−99	≥ 100	
人数 n_{ij}	367	342	266	329	56	40	20	16	1436
$\hat{p}_{ij}$	0.2675	0.2416	0.1809	0.2182	0.0271	0.0245	0.0183	0.0221	1
$n\hat{p}_{ij}$	384.12	346.89	259.71	313.29	38.883	35.114	26.290	31.713	1436
$(n_{ij}-n\hat{p}_{ij})^2\big/(n\hat{p}_{ij})$	0.7628	0.0688	0.1523	0.7881	7.5352	0.6798	1.5048	7.7855	19.2773

检验统计量观测值

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}=19.2773\in W，$$

并且右侧检验的 p 值

$$p=P\{\chi^2\ge 19.2773\}=0.0002<\alpha=0.05，$$

故拒绝 H_0，接受 H_1。可以认为智力发展与营养有显著关系。

§7.5 正态性检验

正态分布是最常用的分布，检验总体是否服从正态分布非常重要，正态性检验除了可用分组的 χ^2 检验法之外还有一些特殊的检验方法。

7.5.1 正态概率图

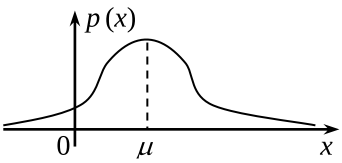
对于样本 $X_1, X_2, \cdots, X_n$，相应的顺序统计量为 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$。构造样本经验分布函数

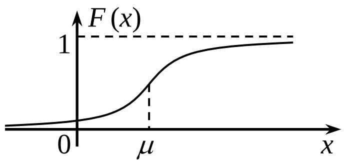
$$F_n(x)=\begin{cases}0, & x<X_{(1)};\\ \cdots & \cdots\\ \dfrac{k}{n}, & X_{(k)}\le x<X_{(k+1)};\\ \cdots & \cdots\\ 1, & x\ge X_{(n)}\end{cases}$$

可以证明，当 $n\to\infty$ 时，样本经验分布函数 $F_n(x)$ 几乎处处一致收敛于总体分布函数 $F(x)$。

$P\{\sup\limits_{-\infty<x<+\infty}\left|F_n(x)-F(x)\right|\to 0\}=1$，sup 表示上确界。

如果总体 X 服从正态分布 $N(\mu,\sigma^2)$，则当样本容量 n 很大时，其样本经验分布函数 $F_n(x)$ 近似为一个正态分布函数 $F(x)$。





要直观判断 $F_n(x)$ 是否近似为正态分布函数曲线 $F(x)$ 是很困难的，故需要通过坐标刻度的缩放将曲线转化成“直线”后再加以判断。正态分布函数曲线 $F(x)$ 两端缓慢增加，中间快速增加，因此将纵坐标在 0 和 1 附近的刻度拉长，在 0.5 附近的刻度压缩，制作正态概率图，使得正态分布函数曲线 $F(x)$ 在正态概率图上是一条直线。根据样本的经验分布函数 $F_n(x)$ 在正态概率图上是否近似为一条直线，判断总体是否服从正态分布。

正态概率图是根据标准正态分布函数反函数 $\Phi^{-1}(\cdot)$ 制作的，如果总体 X 服从正态分布 $N(\mu,\sigma^2)$，则当 $n\to\infty$ 时，样本经验分布函数 $F_n(x)$ 几乎处处一致收敛于正态分布函数 $F(x)$，于是 $\Phi^{-1}(F_n(x))$ 几乎处处一致收敛于

$$\Phi^{-1}(F(x))=\Phi^{-1}\left(\Phi\left(\frac{x-\mu}{\sigma}\right)\right)=\frac{x-\mu}{\sigma},$$

即 $y=\Phi^{-1}(F_n(x))$ 近似为一条直线 $y=\dfrac{x-\mu}{\sigma}$。对于分布函数 $y=F(x)$，以 $\Phi^{-1}(y)$ 作为纵坐标的单位作图，得到如下的正态概率图。

由于经验分布函数 $F_n(x)$ 在区间 $[x_{(k)}, x_{(k+1)})$ 上的取值都等于累计频率 $\dfrac{k}{n}$，即 $F_n(x)$ 的图形为阶梯型，通常是在正态概率图上描点代替 $F_n(x)$ 的直线段。但取点 $\left(x_{(k)}, \dfrac{k}{n}\right)$ 并不合适，因为对应的分布函数在区间 $[x_{(k)}, x_{(k+1)})$ 左端点 $x_{(k)}$ 处的值应该略小于 $\dfrac{k}{n}$ 才合理。通常需要对累计频率 $\dfrac{k}{n}$ 进行修正，使之略微缩小。常用的修正累计频率有以下三种

$$\hat{F}(x_{(k)}) = \frac{k}{n+1}，\quad \hat{F}(x_{(k)}) = \frac{k-1/2}{n}，\quad \hat{F}(x_{(k)}) = \frac{k-3/8}{n+1/4}。$$

目前较为普遍采用的是第三种修正累计频率。利用正态概率图作正态性检验时，以样本的顺序统计量观测值 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 作为横坐标，以相应的修正累计频率

$$\frac{k-3/8}{n+1/4}, \quad k = 1, 2, \cdots, n$$

作为纵坐标在正态概率图上描点

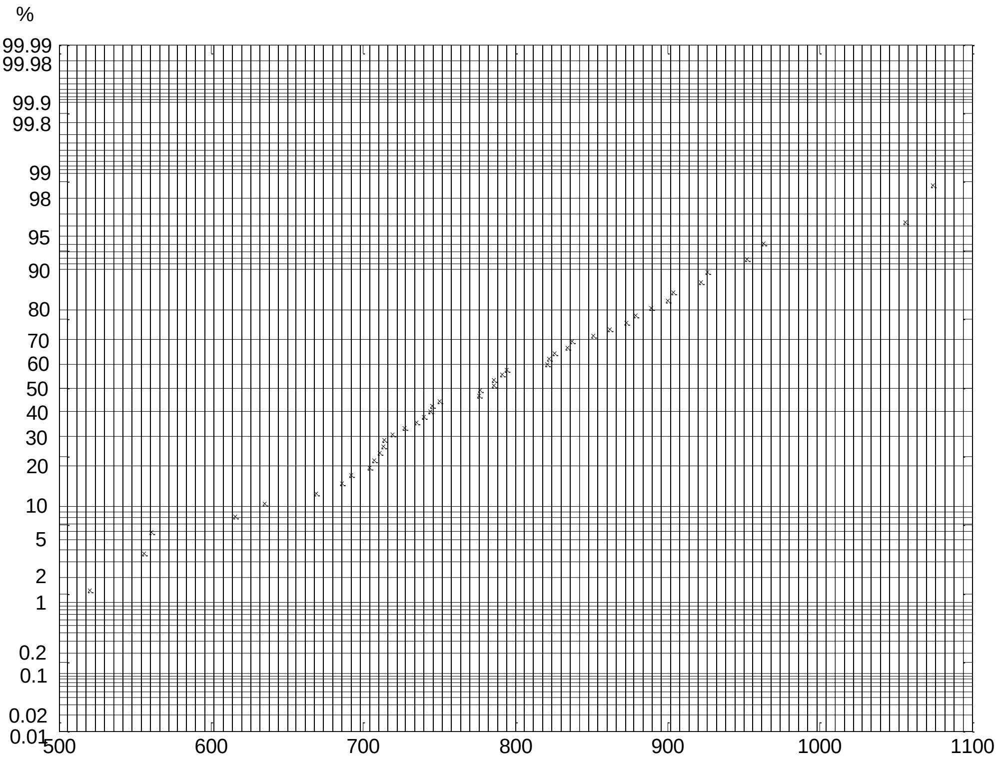
$$\left(x_{(k)}, \frac{k-3/8}{n+1/4}\right), \quad k = 1, 2, \cdots, n，$$

根据这些点是否近似在一条直线上来判断总体是否服从正态分布。

例 某气象站收集了 44 个独立的年降雨量数据，用正态概率图检验这批数据是否服从正态分布。

520	713	786	873	556	714	786	879	561	719	791
889	616	727	794	900	635	735	821	904	669	740
822	922	686	744	826	926	692	745	834	952	704
750	837	963	707	776	851	1056	711	777	862	1074

解： 在正态概率纸上描点 $\left(x_{(k)}, \dfrac{k-3/8}{44+1/4}\right), \ k = 1, 2, \cdots, 44$，



可以看出，这些点在正态概率图上近似在一条直线上，可以认为总体服从正态分布。

7.5.2　　W 检验

W 检验，又称夏皮洛-威尔克检验，它是基于一般正态分布可经过一个线性变换（标准化）转化为标准正态分布而提出的。

设总体 X 服从正态分布，$X_1, X_2, \cdots, X_n$ 为样本，将其标准化后，可看作取自标准正态总体的一个样本 $Y_1, Y_2, \cdots, Y_n$，其顺序统计量分别为 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 与 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$。设 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ 的均值向量和协方差矩阵为

$$m = \begin{pmatrix} E(Y_{(1)}) \\ E(Y_{(2)}) \\ \vdots \\ E(Y_{(n)}) \end{pmatrix}, \quad V = \begin{pmatrix} \mathrm{Var}(Y_{(1)}) & \mathrm{Cov}(Y_{(1)}, Y_{(2)}) & \cdots & \mathrm{Cov}(Y_{(1)}, Y_{(n)}) \\ \mathrm{Cov}(Y_{(2)}, Y_{(1)}) & \mathrm{Var}(Y_{(2)}) & \cdots & \mathrm{Cov}(Y_{(2)}, Y_{(n)}) \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}(Y_{(n)}, Y_{(1)}) & \mathrm{Cov}(Y_{(n)}, Y_{(2)}) & \cdots & \mathrm{Var}(Y_{(n)}) \end{pmatrix},$$

记 $\vec{Y} = (Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)})^T$，令 $\vec{Z} = (Z_{(1)}, Z_{(2)}, \cdots, Z_{(n)})^T = V^{-1}(Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)})^T = V^{-1}\vec{Y}$，可得

$$\mathrm{Cov}(\vec{Y}, \vec{Z}) = \mathrm{Cov}(\vec{Y}, V^{-1}\vec{Y}) = \mathrm{Cov}(\vec{Y}, \vec{Y})(V^{-1})^T = VV^{-1} = E,$$

$$\mathrm{Cov}(Y_{(i)}, Z_{(j)}) = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

由于 $Y_{(i)}$ 是 $X_{(i)}$ 的线性函数，可知 $X_{(i)}$ 与 $Z_{(i)}$ 线性相关，但当 $i \neq j$ 时，$X_{(i)}$ 与 $Z_{(j)}$ 不相关。根据这个原理进行正态性检验，首先对 $\vec{Z}$ 求数学期望，并单位化，令

$$(a_1, a_2, \cdots, a_n)^T = \frac{E(\vec{Z})}{\| E(\vec{Z}) \|} = \frac{V^{-1}E(\vec{Y})}{\| V^{-1}E(\vec{Y}) \|} = \frac{V^{-1}m}{\| V^{-1}m \|}$$ 。

从待检验的总体中抽取样本，设其顺序统计量的观测值为 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$。如果总体服从正态分布，样本的顺序统计量观测值 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 与 $a_1, a_2, \cdots, a_n$ 呈线性相关关系，检验统计量取为它们的样本相关系数的平方

$$W = \frac{\left[\sum_{k=1}^{n}(a_k - \bar{a})(x_{(k)} - \bar{x})\right]^2}{\sum_{k=1}^{n}(a_k - \bar{a})^2 \sum_{k=1}^{n}(x_{(k)} - \bar{x})^2},$$

显然 $W \in [0,1]$，且 W 越接近 1，表明 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 与 $a_1, a_2, \cdots, a_n$ 线性相关关系越强，总体 X 更可能服从正态分布，因此取左侧拒绝域 $\{W \le W_\alpha(n)\}$。检验系数 $a_1, a_2, \cdots, a_n$ 可查 W 统计量的系数数值表（附表 6），对于显著水平 α 相应的 α 分位数 $W_\alpha(n)$ 可查 W 统计量的分位数表（附表 7）。当 $W > W_\alpha(n)$ 时，则可以认为总体服从正态分布。

由于标准正态分布的对称性以及 $(a_1, a_2, \cdots, a_n)^T$ 为单位向量，可得数 $a_1, a_2, \cdots, a_n$ 满足

$$a_k = a_{n+1-k}, \quad k = 1, 2, \cdots, \left[\frac{n}{2}\right], \quad \sum_{k=1}^{n} a_k = 0, \quad \sum_{k=1}^{n} a_k^2 = 1,$$

可得 $\bar{a} = 0$，统计量 W 可简化为

$$W = \frac{\left[\sum_{k=1}^{n} a_k (x_{(k)} - \bar{x})\right]^2}{\sum_{k=1}^{n}(x_{(k)} - \bar{x})^2} = \frac{\left[\sum_{k=1}^{[n/2]} a_k (x_{(k)} - \bar{x}) + \sum_{k=1}^{[n/2]}(-a_k)(x_{(n+1-k)} - \bar{x})\right]^2}{\sum_{k=1}^{n}(x_{(k)} - \bar{x})^2} = \frac{\left[\sum_{k=1}^{[n/2]} a_k (x_{(n+1-k)} - x_{(k)})\right]^2}{\sum_{k=1}^{n}(x_{(k)} - \bar{x})^2}$$ 。

步骤：

（1）假设 H_0:总体服从正态分布。

（2）检验统计量 $W = \dfrac{\left[\sum_{k=1}^{[n/2]} a_k (x_{(n+1-k)} - x_{(k)})\right]^2}{\sum_{k=1}^{n}(x_{(k)} - \bar{x})^2}$，其中 a_k 可查 W 统计量的系数数值表（附表 6）。

（3）拒绝域：$\{W \le W_\alpha(n)\}$，其中 $W_\alpha(n)$ 可查 W 统计量的分位数表（附表 7）。

（4）计算检验统计量观测值 W，并作出决策，也可根据检验的 p 值作出决策。

W 检验法用于检验样本容量在 8—50 之间的情况。样本容量太少时，检验不太有效；样本容量太大时，系数 a_k 与分位数 $W_\alpha(n)$ 的计算很复杂。

例 某气象站收集了 44 个独立的年降雨量数据，用 W 检验法检验这批数据是否服从正态分布

（$\alpha = 0.05$）。

520	713	786	873	556	714	786	879	561	719	791
889	616	727	794	900	635	735	821	904	669	740
822	922	686	744	826	926	692	745	834	952	704
750	837	963	707	776	851	1056	711	777	862	1074

解： 假设 H_0 : 总体服从正态分布，统计量 $W = \dfrac{\left[\sum_{k=1}^{22} a_k (x_{(n+1-k)} - x_{(k)})\right]^2}{\sum_{k=1}^{44} (x_{(k)} - \bar{x})^2}$，显著水平 $\alpha = 0.05$，分位数 $W_\alpha(n) = W_{0.05}(44) = 0.944$，左侧拒绝域为 $\{W \le 0.944\}$。

因 $\bar{x} = 785.1136$，$\sum_{k=1}^{44} (x_{(k)} - \bar{x})^2 = 43s^2 = 630872.4318$，将样本观测值按由小到大顺序排列，并查 W 统计量的系数数值表得 $a_k = 1, 2, \cdots, 22$。

k	$x_{(k)}$	$x_{(n+1-k)}$	d_k	a_k	$a_k d_k$	k	$x_{(k)}$	$x_{(n+1-k)}$	d_k	a_k	$a_k d_k$
1	520	1074	554	0.3872	214.5088	12	713	862	149	0.0943	14.0507
2	556	1056	500	0.2667	133.3500	13	714	851	137	0.0842	11.5354
3	561	963	402	0.2323	93.3846	14	719	837	118	0.0745	8.7910
4	616	952	336	0.2072	69.6192	15	727	834	107	0.0651	6.9657
5	635	926	291	0.1868	54.3588	16	735	826	91	0.0560	5.0960
6	669	922	253	0.1695	42.8835	17	740	822	82	0.0471	3.8622
7	686	904	218	0.1542	33.6156	18	744	821	77	0.0383	2.9491
8	692	900	208	0.1405	29.2240	19	745	794	49	0.0296	1.4504
9	704	889	185	0.1278	23.6430	20	750	791	41	0.0211	0.8651
10	707	879	172	0.1160	19.9520	21	776	786	10	0.0126	0.1260
11	711	873	162	0.1049	16.9938	22	777	786	9	0.0042	0.0378
Σ											787.2627

计算可得

$$W = \frac{787.2627^2}{630872.4318} = 0.9824 > 0.944,$$

故接受 H_0，拒绝 H_1，可以认为这批数据服从正态分布。

7.5.3 EP 检验

EP 检验的基本思想是样本特征函数收敛于总体特征函数，根据它们之差模的平方定义检验统计量。检验总体是否服从正态分布时，检验统计量为

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\exp\left\{-\frac{(x_j - x_i)^2}{2s^{*2}}\right\} - \sqrt{2}\sum_{i=1}^{n}\exp\left\{-\frac{(x_i - \bar{x})^2}{4s^{*2}}\right\},$$

拒绝域为 $W = \{T_{EP} \ge T_{1-\alpha,\,EP}(n)\}$，$T_{1-\alpha,\,EP}(n)$ 是样本容量为 n 的 T_{EP} 统计量的 $1-\alpha$ 分位数，可查附表 11 得到。

§7.6　其它非参数检验

7.6.1　　游程检验

游程检验用于检验样本数据是否是随机抽样获得的。

对于一个样本 $X_1, X_2, \cdots, X_n$，样本中位数为 $m_{0.5}$，将样本观测值序列 $x_1, x_2, \cdots, x_n$ 中大于等于 $m_{0.5}$ 的数记为 1，小于 $m_{0.5}$ 的数记为 0，这样获得一个由 0 与 1 组成的序列，通常称一串连续的 0 或一串连续的 1 为一个游程。由 0 与 1 组成的序列可看作由若干个 0 游程和 1 游程组成。

设容量为 n 的序列中分别有 n_1 个 0 和 n_2 个 1（$n_1+n_2=n$，通常 n_1 和 n_2 都大于 0），R 为 0 游程和 1 游程的总游程数。如果样本数据是随机抽样获得的，则游程数 R 不应太小或太大。若 R 太小，说明样本数据中较小的数据与较大的数据分别太集中；若 R 太大，说明样本数据中较小的数据与较大的数据交替出现，有周期变化的趋势。可以认为这两种情况都不符合随机抽样的原则。为了检验样本数据是否符合随机抽样的原则，取游程数 R 为检验统计量，其分布可根据重复组合得到。

将 n_1 个 0 和 n_2 个 1 放在一起的方法总数为 $n=n_1+n_2$ 选 n_1 次的组合数，即 $C_n^{n_1}$。

当游程数 $R=2k$ 为偶数时，显然 0 游程和 1 游程分别各有 k 个。将 n_1 个 0 放入 k 个 0 游程且每个 0 游程至少有一个 0，先将每个 0 游程各放一个 0，再将其余 n_1-k 个 0 任意放入这 k 个 0 游程，即 k 取 n_1-k 次的重复组合，方法总数为 $C_{k+(n_1-k)-1}^{n_1-k}=C_{n_1-1}^{n_1-k}=C_{n_1-1}^{k-1}$；同样将 n_2 个 1 放入 k 个 1 游程方法总数应为 $C_{n_2-1}^{k-1}$；将 0 游程和 1 游程合在一起，有第一个游程是 0 游程与第一个游程是 1 游程两种方法。这样游程数 $R=2k$ 的方法总数为 $2C_{n_1-1}^{k-1}C_{n_2-1}^{k-1}$，概率为

$$P\{R=2k\}=\frac{2C_{n_1-1}^{k-1}C_{n_2-1}^{k-1}}{C_n^{n_1}},\quad k=1,2,\cdots,\left[\frac{n}{2}\right]。$$

当游程数 $R=2k+1$ 为奇数时，有两种情形：一是 k 个 0 游程和 $k+1$ 个 1 游程，二是 $k+1$ 个 0 游程和 k 个 1 游程。第一种情形方法总数为 $C_{n_1-1}^{k-1}C_{n_2-1}^{k}$，第二种情形方法总数为 $C_{n_1-1}^{k}C_{n_2-1}^{k-1}$，这样游程数 $R=2k+1$ 的方法总数为 $C_{n_1-1}^{k-1}C_{n_2-1}^{k}+C_{n_1-1}^{k}C_{n_2-1}^{k-1}$，概率为

$$P\{R=2k+1\}=\frac{C_{n_1-1}^{k-1}C_{n_2-1}^{k}+C_{n_1-1}^{k}C_{n_2-1}^{k-1}}{C_n^{n_1}},\quad k=1,2,\cdots,\left[\frac{n-1}{2}\right]。$$

步骤：

（1）假设 H_0: 样本数据符合随机抽样的原则。

（2）检验统计量游程数 R，$P\{R=m\}=\begin{cases}\dfrac{2C_{n_1-1}^{k-1}C_{n_2-1}^{k-1}}{C_n^{n_1}}, & m=2k,\quad k=1,2,\cdots,\left[\dfrac{n}{2}\right];\\[2ex] \dfrac{C_{n_1-1}^{k-1}C_{n_2-1}^{k}+C_{n_1-1}^{k}C_{n_2-1}^{k-1}}{C_n^{n_1}}, & m=2k+1,\quad k=1,2,\cdots,\left[\dfrac{n-1}{2}\right].\end{cases}$

（3）拒绝域：$W=\{R\le c_1$ 或 $R\ge c_2\}$，其中 c_1 与 c_2 可查游程总数检验临界值表（附表 12）。

（4）计算检验统计量观测值 R 与检验的 p 值，并作出决策。

7.6.2 符号检验

符号检验用于检验总体的 p 分位数 x_p 是否为给定的值 x_0。

对于样本 $X_1, X_2, \cdots, X_n$，构造符号统计量 $S^+=\sum_{i=1}^{n} I_{X_i>x_0}$，即 S^+ 表示样本数据中大于 x_0 的样品个数。如果总体的 p 分位数 x_p 等于 x_0，即总体分布函数值 $F(x_0)=p$，有 $X_i>x_0$ 的概率 $P\{X_i>x_0\}=1-p$，则 S^+ 服从二项分布 $b(n,1-p)$。若 S^+ 太小，说明 $X_i>x_0$ 的概率应小于 $1-p$，即 $X_i\le x_0$ 的概率应大于 p，可见 $x_p<x_0$ 才合理；反之，若 S^+ 太大，说明 $x_p>x_0$ 才合理。

步骤：

（1）假设 $H_0: x_p=x_0$ vs $H_1: x_p\ne x_0$（或 $x_p<x_0$， $x_p>x_0$）。

（2）检验统计量 $S^+=\sum_{i=1}^{n} I_{X_i>x_0}\sim b(n,1-p)$。

（3）拒绝域：$W=\{S^+\le c_{\alpha/2}$ 或 $S^+\ge c_{1-\alpha/2}\}$（或 $\{S^+\le c_\alpha\}$，$\{S^+\ge c_{1-\alpha}\}$）。其中

$$c_{\alpha/2}=\max_c\left\{c:\sum_{i=0}^{c} C_n^i(1-p)^i p^{n-i}\le\frac{\alpha}{2}\right\},$$

$$c_{1-\alpha/2}=\min_c\left\{c:\sum_{i=c}^{n} C_n^i(1-p)^i p^{n-i}\le\frac{\alpha}{2}\right\}=\min_c\left\{c:\sum_{i=0}^{c-1} C_n^i(1-p)^i p^{n-i}\ge 1-\frac{\alpha}{2}\right\},$$

并且类似定义 c_α 与 $c_{1-\alpha}$。

（4）计算检验统计量观测值 S^+ 与检验的 p 值，并作出决策。

7.6.3 秩和检验

秩和检验用于检验对称分布总体的对称点 θ 是否为给定的值 θ_0（即总体中位数是否为 θ_0），以及分布类型相同的两个总体是否存在位置差异。

检验对称分布总体的中位数可以采用符号检验，但符号检验只考虑到样本观测值与中位数之差的符号，而没有考虑到二者之间距离的大小。秩和检验则是对其改进而引入反映距离大小的量——秩。

定义 对于样本 $X_1, X_2, \cdots, X_n$，按由小到大顺序排列成 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$，则样品 X_i 在有序样本中的序号 r 称为 X_i 的秩，记为 $R_i=r$。由样本 $X_1, X_2, \cdots, X_n$ 的秩 $R_1, R_2, \cdots, R_n$ 导出的统计量称为秩统计量。

检验对称分布总体 X 的对称点 θ 是否为给定的值 θ_0（即中位数是否为 θ_0），抽取样本 $X_1, X_2, \cdots, X_n$，令 $Y_i=X_i-\theta_0$，转化为检验总体 Y 的中位数 $y_{0.5}$ 是否为 0，将 $|Y_1|, |Y_2|, \cdots, |Y_n|$ 按由小到大顺序排列，设其

秩为 $R_1, R_2, \cdots, R_n$，构造秩和统计量 $W^+ = \sum_{i=1}^{n} R_i I_{Y_i > 0}$，即 W^+ 表示 $Y_1, Y_2, \cdots, Y_n$ 中大于 0 的样品关于 $|Y_1|, |Y_2|, \cdots, |Y_n|$ 的秩之和。若 W^+ 太小，说明总体 Y 的中位数 $y_{0.5}$ 应小于，即 $\theta < \theta_0$ 才合理；反之，若 W^+ 太大，说明 $\theta > \theta_0$ 才合理。

步骤：

（1）假设 $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$（或 $\theta < \theta_0$，$\theta > \theta_0$）。

（2）检验统计量 $W^+ = \sum_{i=1}^{n} R_i I_{Y_i > 0}$。

（3）拒绝域：$W = \{W^+ \le W^+_{\alpha/2}(n)$ 或 $W^+ \ge W^+_{1-\alpha/2}(n)\}$（或 $\{W^+ \le W^+_{\alpha}(n)\}$，$\{W^+ \ge W^+_{1-\alpha}(n)\}$）。其中 $W^+_{\alpha}(n)$ 可查符号秩和检验统计量分位数表（附表 13），且 $W^+_{1-\alpha}(n) = \frac{1}{2}n(n+1) - W^+_{\alpha}(n)$。

（4）计算检验统计量观测值 W^+ 与检验的 p 值，并作出决策。

注： 秩 $R_1, R_2, \cdots, R_n$ 的总和为 $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$，由对称性有 $W^+_{1-\alpha}(n) = \frac{1}{2}n(n+1) - W^+_{\alpha}(n)$。

检验分布类型相同的两个总体 X 与 Y 是否存在位置差异，设总体 X 的分布函数为 $F(x-\theta_1)$，取得样本 $X_1, X_2, \cdots, X_m$，总体 Y 的分布函数为 $F(y-\theta_2)$，取得样本 $Y_1, Y_2, \cdots, Y_n$，检验 θ_1 与 θ_2 是否相等。将两个样本合并为 $X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n$，再按由小到大顺序排列，设其秩为 $Q_1, Q, \cdots, Q_m, R_1, R_2, \cdots, R_n$，构造秩和统计量 $W = \sum_{i=1}^{n} R_i$，即 $W = \sum_{i=1}^{n} R_i$ 表示 $Y_1, Y_2, \cdots, Y_n$ 在 $X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n$ 中的秩之和。若 W 太小，说明总体 Y 的均值应小于总体 X 的均值，即 $\theta_1 > \theta_2$ 才合理；反之，若 W 太大，说明 $\theta_1 < \theta_2$ 才合理。

步骤：

（1）假设 $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 \neq \theta_2$（或 $\theta_1 < \theta_2$，$\theta_1 > \theta_2$）。

（2）检验统计量 $W = \sum_{i=1}^{n} R_i$。

（3）拒绝域：$W = \{W \le W_{\alpha/2}(m,n)$ 或 $W \ge W_{1-\alpha/2}(m,n)\}$（或 $\{W \ge W_{1-\alpha}(m,n)\}$，$\{W \le W_{\alpha}(m,n)\}$）。其中 $W_{\alpha}(m,n)$ 可查秩和检验临界值表（附表 14），且 $W_{1-\alpha}(m,n) = n(m+n+1) - W_{\alpha}(m,n)$。

（4）计算检验统计量观测值 W 与检验的 p 值，并作出决策。