典型例题分析

例 1． 分别从方差为 20 和 35 的正态总抽取容量为 8 和 10 的两个样本，求第一个样本方差是第二个样本方差两倍的概率的范围。

解 以S_1^2和S_2^2分别表示两个〔修正〕样本方差。由$F=\dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}$知统计量

$$F=\frac{S_1^2/20}{S_2^2/35}=1.75\frac{S_1^2}{S_2^2}$$

服从F分布，自由度为〔7，9〕。

1）事件$\{S_1^2=2S_2^2\}$的概率

$$P\{S_1^2=2S_2^2\}=P\left\{\frac{S_1^2}{S_2^2}=2\right\}=P\left\{\frac{S_1^2/20}{S_2^2/35}=2\times\frac{35}{20}\right\}=P\{F=3.5\}=0$$

因为F是连续型随机变量，而任何连续型随机变量取任一给定值的概率都等于 0。

2）现在我们求事件$A=\{\text{第一样本方差不小于第二样本方差两倍}\}$的概率：

$$p=P\{S_1^2\ge 2S_2^2\}=P\{F\ge 3.5\}。$$

由附表可见，自由度$f_1=7,f_2=9$的F分布水平α上侧分位数$F_\alpha(f_1,f_2)$有如下数值：

$$F_{0.05}(7,9)=3.29<3.5<4.20=F_{0.025}(7,9)。$$

由此可见，事件A的概率p介于 0. 025 与 0. 05 之间；$0.025<p<0.05$。

例 2． 设X_1，X_2,，X_n是取自正态总体$N(\mu,\sigma^2)$的一个样本，s^2为样本方差，求满足不等式

$$P\left\{\frac{S^2}{\sigma_2}\le 1.5\right\}\ge 0.95$$

的最小n值。

解 由随机变量χ^2分布知，随机变量（$n-1$）S^2/σ服从χ^2分布，自由度$v=n-1$，于是，有

$$P=\left\{\frac{(n-1)S^2}{\sigma^2}\le 1.5(n-1)\right\}=P\{\chi_v^2\le 1.5(n-1)\}\ge P\{\chi_v^2\le\chi_{0.05,v}^2\}=0.95$$

其中χ_v^2表示自由度$v=n-1$的χ^2分布随机变量，$\chi_{0.05,v}^2$是自由度为$v=n-1$的水平$\alpha=0.05$的χ^2分布上侧分位数〔见附表〕。我们欲求满足

$$1.5(n-1)\ge\chi_{0.05,v}^2$$

的最小$n=v+1$值，由附表可见

$$1.5(27-1)=39>38.885=\chi^2_{0.05,26},$$

$$1.5(26-1)=40.5<37.652=\chi^2_{0.05,25}。$$

于是，所求$n=27$。

例 3. 假设随机变量X在区间$[\theta,\theta+1]$上有均匀分布，其中θ未知：$(X_1,\cdots,X_n)$是来自X的简单随机样本，$\overline{X}$是样本的均值，$X_{(1)}=\min\{X_1,\cdots,X_n\}$是最小观察值。证明

$$\hat{\theta}_1=\overline{X}-\frac{1}{2} \quad 和 \quad \hat{\theta}_2=X_{(1)}-\frac{1}{n+1}$$

都是θ的无偏估计量。

解 由X在$[\theta,\theta+1]$上均匀分布，知$EX_i=EX=(2\theta+1)/2$。

1）由

$$E\hat{\theta}_1=\frac{1}{n}\sum_{i=1}^{n}EX_i-\frac{1}{2}=\frac{1}{n}\sum_{i=1}^{n}\frac{2\theta+2}{2}-\frac{1}{2}=\theta+\frac{1}{2}-\frac{1}{2}=\theta,$$

可见$\hat{\theta}_1$是θ的无偏估计量。

2）为证明$\hat{\theta}_2$是θ的无偏估计。我们先求统计量$X_{(1)}$的概率分布。

$$F(x)=P\{X\le x\}=\begin{cases}0, & 若x<\theta;\\ x-\theta, & 若\theta\le x\le\theta+1;\\ 1, & 若\theta>1。\end{cases}$$

其密度为

$$f(x)=\begin{cases}1, & 若\theta\le x\le\theta+1,\\ 0, & 其他。\end{cases}$$

由于$X_1,\cdots,X_n$独立且与X同分布，知$X_{(1)}$的分布函数为

$$F_{(1)}(x)=P\{X_{(1)}\le x\}=1-P\{X_{(1)}>x\}=1-P\{X_1>x,\cdots,X_n>x\}$$

$$=1-P\{X_1>x\}\cdots P\{X_n>x\}$$

$$=1-[1-F(x)]^n;$$

$$f_{(1)}(x)=F'_{(1)}(x)=n[1-F(x)]^{n-1}f(x)=n(1+\theta-x)^{n-1} \qquad (\theta\le x\le\theta+1)$$

于是，有

$$EX_{(1)}=\int_{\theta}^{\theta+1}xf_{(1)}(x)dx=n\int_{\theta}^{\theta+1}x(1+\theta-x)^{n-1}dx$$

$$=n\int_{\theta}^{\theta+1}(1+\theta-x)^n d(1+\theta-x)+n\int_{\theta}^{\theta+1}(1+\theta)(1+\theta-x)^{n-1}dx$$

$$=n\left(-\frac{1}{n+1}+\frac{1+\theta}{n}\right)=\frac{1}{n+1}+\theta。$$

$$E\hat{\theta}_2=EX_{(1)}-\frac{1}{n+1}=\theta,$$

从而$\hat{\theta}_2$是θ的无偏估计。

在证$\hat{\theta}_2$的无偏估计时，先求估计量分布再求其数学期望。此外，下面将看到，$\hat{\theta}_1$是矩估计量，$X_{(1)}$是最大似然估计量。

3）有效性的验证，即验证两个无偏估计量哪一个更有效〔方差较小〕，只需计算它们的方差并加以比拟，验证估计量的最小方差超出了本课程的要求。读者只需了解一些常用的最小方差估计量。例如，对于正态分布总体$N(\mu,\sigma^2)$，样本均值$\overline{X}$和修正样本方差S^2相应为μ和σ^2的最小方差无偏估计量；事件频率$\hat{p}_n$是它的概率p的最小方差无偏估计量。

如果要求有效率，那么用公式$\dfrac{D_0(\theta)}{D(\hat{\theta})}$计算，其中$D(\hat{\theta})\ge D_0(\theta)=\dfrac{1}{nE\left[\dfrac{\partial}{\partial\theta}(\ln f(x,\theta))\right]^2}$——称为罗.克拉美不等式。

例4. 设总X服从正态分布N（μ，σ_0^2），其中方差σ_0^2为常数；关于未知数学期望μ有两个二者必居其一的假设：

$$H_0：\mu=\mu_0，\quad H_1：\mu=\mu_1，$$

其中μ_0和μ_1都有常数，并且$\mu_0<\mu_1$。根据来自总体X的简单随机样本X_1，X_2,…，X_n，确定假设H_0的α水平否认域〔即拒绝域〕，并计算第二类错误概率。

解 取统计量

$$U=\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}$$

做检验的统计量。在假设H_0：$\mu=\mu_0$成立的条件下，$U\sim N$（0,1）。

由于

$$P\{|U|\ge u_\alpha\}=P\{U\ge u_{2\alpha}\}=P\{U\le -u_{2\alpha}\}=p\{|U|\le u_{1-\alpha}\}=\alpha。$$

所以以下四种都是假设H_0的水平α的否认域：

$$V_1=\{|U|\ge u_\alpha\};\quad V_2=\{U\ge u_{2\alpha}\};$$

$$V_3 = \{U \le -u_{2\alpha}\}; \quad V_4 = \{|U| \le u_{1-\alpha}\},$$

其中u_α是标准正态分布α水平双侧分位数〔见附表〕。

在假设$H_1: \mu = \mu_1$成立的条件下，统计量$U \sim N(\Delta,1)$，其中$\Delta = \sqrt{n}(\mu_1 - \mu_0)/\sigma_0$。因此，以$V_i(i=1,2,3,4)$为假设否认域的检验的第二类错误概率为：

$$\beta_i = P\{\overline{V}|H_1\} = P\{\overline{V_i}|\mu = \mu_1\} = \frac{1}{\sqrt{2\pi}}\int_{\overline{V_i}} e^{-\frac{(x-\Delta)^2}{2}}dx\text{。}$$

特别〔设$\Phi(x)$是标准正态分布函数〕

$$\beta_1 = \frac{1}{\sqrt{2\pi}}\int_{-u_\alpha}^{u_\alpha} e^{-\frac{(x-\Delta)^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-u_\mu-\Delta}^{u_\alpha-\Delta} e^{\frac{u^2}{2}}du = \Phi(u_\alpha + \Delta) + \Phi(\mu_\alpha - \Delta) - 1;$$

$$\beta_2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u_{2\alpha}} e^{-\frac{(x-\Delta)^2}{2}}dx = \Phi(u_{2\alpha} - \Delta);$$

$$\beta_3 = \frac{1}{\sqrt{2\pi}}\int_{-u_{2\alpha}}^{+\infty} e^{-\frac{(x-\Delta)^2}{2}}dx = \Phi(u_{2\alpha} + \Delta);$$

$$\beta_4 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-u_{1-\alpha}} e^{-\frac{(x-\Delta)^2}{2}}dx + \frac{1}{\sqrt{2\pi}}\int_{u_{1-\alpha}}^{+\infty} e^{\frac{(x-\Delta)^2}{2}}dx = 2 - \Phi(u_{1-\alpha} + \Delta) - \Phi(\mu_{1-\alpha} - \Delta)\text{。}$$

为了便于比拟，设$\alpha = 0.1$，$\mu_0 = 0$，$\mu_1 = 1$，$\sigma_0 = 1$；$n = 9$，那么

$\Delta = 3, u_{0.1} = 1.65, u_{0.2} = 1.28, u_{0.9} = 0.13$。查附表并经计算，容易得到

$$\beta_1 = 0.0855,\quad \beta_2 = 0.0427,\quad \beta_3 = 0.9999,\quad \beta_4 = 0.9988\text{。}$$

计算结果说明，尽管四个检验的一类错误的概率都等于$\alpha = 0.1$，但它们的第二类错误的概率却不相同。以V_2为否认域的检验的第二类错误的概率最小，为我们所选用。

例 5．对二项分布$B(n,p)$作统计假设

$$H_0: p = 0.6,\quad H_1: p = 0.3\text{。}$$

假设H_0的否认域取为

$$V = \{\mu_n \le c_1\} \cup \{\mu_n \ge c_2\},$$

其中μ_n表示n次试验中成功的次数。对〔1〕$n = 10, c_1 = 1, c_2 = 9, \mu_n = 3;$

〔2〕$n = 20, c_1 = 7, c_2 = 17, \mu_n = 6$，求显著性水平$\alpha$和第二类错误的概率$\beta$。

解 〔1〕显著性水平α是第一类错误的概率，于是

$$\alpha = P\{\mu_n \in V | H_0\} = P\{\mu \in V | p = 0.6\}$$

$$= \sum_{i=0}^{1} C_{10}^{i} 0.6^{i} 0.4^{10-i} + \sum_{i=9}^{10} C_{10}^{i} 0.6^{i} 0.4^{10-i} \approx 0.0479。$$

$$\beta = P\{\mu_n \in \overline{V} | H_1\} = 1 - P\{\mu_n \in V | H_1\}$$

$$= 1 - P\{\mu_n \in V | p = 0.3\}$$

$$= 1 - \sum_{i=0}^{1} C_{10}^{i} 0.3^{i} 0.7^{10-i} - \sum_{i=9}^{10} C_{10}^{i} 0.3^{i} 0.7^{10-i} \approx 0.8506。$$

〔2〕

$$\alpha = P\{\mu_n \in V | H_0\} = P\{\mu_n \in V | p = 0.6\}$$

$$= \sum_{i=0}^{7} C_{20}^{i} 0.6^{i} 0.4^{20-i} + \sum_{i=17}^{20} C_{20}^{i} 0.6^{i} 0.4^{i} \approx 0.0370。$$

$$\beta = P\{\mu_n \in \overline{V} | H_1\} = 1 - P\{\mu_n \in V | p = 0.3\}$$

$$= 1 - \sum_{i=0}^{7} C_{10}^{i} 0.3^{i} 0.7^{20-i} - \sum_{i=17}^{20} C_{20}^{i} 0.3^{i} 0.7^{20-i} \approx 0.2277。$$

例 6. 谋装置的平均工作温度据制造厂家称不高于 190℃。今从一个由 16 台装置构成的随机样本册的工作温度的平均值和标准差分别为 195℃和 8℃。根据这些数据能否说明平均工作温度比制造厂所说的要高？设$\alpha = 0.05$，并假定工作温度服从正态分布。

解 设工作温度为X，根据题设$X \sim N(\mu, \sigma^2)$。考虑假设

$$H_0 : \mu \le 190, \quad H_1 > 190$$

由于总体方差σ^2未知，故用t检验。

这里$n = 16, v = n - 1 = 15,$对给定的$\alpha = 0.05$，查表得$t_{20,v} = t_{0.1,1.5} = 1.75$。于是由表情形知假设$H_0$的否认域为

$$V = \{t \ge 1.75\}。$$

由条件和H_0知$\mu_0 = 190, \overline{X} = 195, S = 8$，因此

$$t = \frac{195 - 190}{8 / \sqrt{16}} = 2.5。$$

由于$t = 2.5 > 1.75$，所以否认域假设H_0，说明平均工作温度比制造厂说的要高。

例7 某 交换台在一小时〔60 分钟〕内每分钟接到 用户的呼唤次数有如下纪录：

呼唤次数 k	0	1	2	3	4	5	6	≥ 7
实际频数 v_k	8	16	17	10	6	2	1	0

问统计资料是否可以说明，每分钟 呼唤次数服从泊松分布？〔 $\alpha = 0.05$）

解 设 X 表示每分钟 呼唤次数，需要检验的假设

H_0：X 服从泊松分布。

泊松分布中未知参数 λ 的最大似然估计为

$$\hat{\lambda} = \frac{1}{60}\sum_{k=0}^{6} k v_k = 2。$$

我们用

$$\hat{p}_k = \frac{2^k}{k!}e^{-k} \qquad (k = 0,1,\cdots,6)$$

估计概率 $p_k = P\{X = k\}(k = 0,1,\cdots,6)$；用 $E_k = n\hat{p}_k (k = 0,1,2,3,4)$ 估计 $\{X = k\}$ 的期望频数。为防止期望频数太小，将呼唤次数为 5 和 6 的情况，合并为 $X \geq 5$ 的情况，为第 6 组: 其实际频数为 2+1=3，期望频数为

$$E_5 = n(p_5 + p_6) = 3.16。$$

计算结果列入下表：

分组 k	0	1	2	3	4	5
实际频数 v_k	8	16	17	10	6	3
期望频数 E_k						
$(v_k - E_k)^2$						
$\frac{(v_k - E_k)^2}{E_k}$						

所以统计量

$$\chi^2 = \sum_{k=0}^{5}\frac{(v_k - E_k)^2}{E_k} = 0.1762。$$

统计量 χ^2 的自由度 $v = 6 - m - 1$，其中 $m = 1$ 是用到参数估计值的个数，故 $v = 4$。对于，

$\alpha = 0.05$，查表得 $\chi^2_{0.05,4} = 9.488$；假设 H_0 的否认域为

$$V = \{\chi^2 \geq 9.488\}。$$

由于 χ^2=0. 1762<9. 488, 所以不否认假设 H_0，即可以认为　呼唤次数服从泊松分布。

例7 对 200 个电池左寿命试验，得如下统计分布：

使用寿命小时	0-5	5-10	10-15	15-20	20-25	25-30
电池个数	133	45	15	4	2	1

试求所得统计分布与指数分布的拟合优度。

解　设 X 表示电池的寿命，需要检验假设 $H_0: X$ 服从指数分布。指数分布中未知参数 λ 需要用其最大似然估计 $\lambda = 1/\overline{X}$ 来估计。在这里

$$\overline{X} = \frac{1}{200}(2.5\times133 + 7.5\times45 + 12.5\times15 + 17.5\times4 + 22.5\times2 + 27.5\times1) = 5。$$

所以 $\hat{\lambda} = 1/5$。在 H_0:“X服从指数分布，参数为$1/5$”成立前提下，观察值落入各组的概率

$$\hat{p}_i = P\{u_{i-1} < X \leq u_i\} = \int_{u_{i-1}}^{u_i} \frac{1}{5} e^{-\frac{x}{5}} dx = e^{-\frac{u_{i-1}}{5}} - e^{-\frac{u_i}{5}} \qquad (i = 1,2,\cdots,6)。$$

计算结果列入下表：

分组 $u_{i-1} - u_i$	0-5	5-10	10-15	15-20	20-25	25-∞
实际频率 v_i	133	45	15	4	2	1
概　率 $\hat{p}_i$						
期望概率 E_i						
$\dfrac{(v_i - E_i)^2}{E_i}$						

所以统计量

$$\chi^2 = \sum \frac{(v_i - E_i)^2}{E_i} = 1.6279。$$

统计量 χ^2 的自由度 $v=6-1-1=4$，查表得 $\chi^2_{0.94,4}$=1.064， $\chi^2_{0.7,4}=2.195$。由于 1.064<1.6297<2.195,的可得统计分布与指数分布的拟合优度不小于0.70。

例9 设随机变量X和Y相互独立，$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$。$X_1,X_2,\cdots,X_{16}$ 是X的一个样本，$Y_1,Y_2,\cdots,Y_{10}$ 是 Y 的一个样本，测得数据

$$\sum_{i=1}^{16}x_i=84,\sum_{i=1}^{16}x_i^2=563,\sum_{i=1}^{10}y_i=18,\sum_{i=1}^{10}y_i^2=72$$

〔1〕分别求 μ_1,μ_2 的矩估计量;〔2〕分别求 σ_1^2，σ_2^2 的极大似然估计值;

〔3〕在显著水平 $\alpha=0.05$ 下检验假设 H_0：$\sigma_1^2\le\sigma_2^2$，H_1：$\sigma_1^2>\sigma_2^2$。

解 〔1〕用样本一阶原点矩估计总体一阶矩，即得 μ_1 和 μ_2 的矩估计值：

$$\hat{\mu}_1=\bar{x}=\frac{1}{16}\sum_{i=1}^{16}x_i=5.25,\quad \hat{\mu}_2=\frac{1}{10}\sum_{i=1}^{10}y_i=1.8。$$

〔2〕正态总体 $X\sim N(\mu,\sigma^2)$ 的参数 σ^2 的极大似然估计量为

$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}=(X_i-\bar{X})^2$。因此 σ_1^2 和 σ_2^2 的极大似然估计值为

$$\hat{\sigma}_1^2=\frac{1}{16}\sum_{i=1}^{16}=(x_i-\bar{x})^2=\frac{1}{16}\left(\sum_{i=1}^{n}x_i^2-16\bar{x}^2\right)=7.625$$

$$\hat{\sigma}_2^2=\frac{1}{10}\sum_{i=1}^{10}=(y_i-\bar{y})^2=\frac{1}{10}\left(\sum_{i=1}^{n}y_i^2-16\bar{y}^2\right)=3.96$$

〔3〕是 μ_1,μ_2 未知，双总体方差的假设检验。待检假设 H_0：$\sigma_1^2\le\sigma_2^2$；H_1：$\sigma_1^2>\sigma_2^2$，是在 $\alpha=0.05$ 下的单侧检验。

因为 $S_1^2=\frac{1}{15}\sum_{i=1}^{n}(x_i-\bar{x}^2)=8.15$，$S_2^2=\frac{1}{9}\sum_{i=1}^{n}(y_1-\bar{y})^2=4.4$。所以F同机量得值

$$F=\frac{S_1^2}{S_2^2}=\frac{8.15}{4.4}=1.847$$

查F分布表，得 $F_{0.05}(15,9)=3.01$. 经比拟知，$F=1.847<F_{0.05}(15,9)=3.01$，故接受 H_0，认为 σ_1^2 不比 σ_2^2 大。

例10 有三台机器，生产同一种规格的铝合金薄板，测量三台机器所生产的

薄板厚度〔单位：厘米〕，得结果如表所示。

机器 1	机器 2	机器 3

试考察机器对薄板厚度有无显著的影响（$\alpha = 0.05$）。

解　检验假设 H_0：$\mu_1 = \mu_2 = \mu_3$。μ_i 是各台机器生产的薄板总体的均值。

经计算 $s = 3, n_1 = n_2 = n_3 = 5, n = 15$，

$$\sum_{j=1}^{3}\sum_{i=1}^{5} x_{ij}^2 = 0.963912, T = 3.8, \sum_{j=1}^{3} T_{\cdot j}^2 = 4.8102 。$$

$$S_T = \sum_{j=1}^{3}\sum_{i=1}^{5} x_{ij}^2 - \frac{1}{15}T^2 = 0.0012453,$$

$$S_A = \frac{1}{5}\sum_{j=1}^{3} T_{\cdot j}^2 - \frac{1}{15}T^2 = 0.0010533,$$

$$S_E = S_T - S_E = 0.000192.$$

列出方差分析表如下

方差来源	平均和	自由度	均方	F 比	结论
因素	0.0010533	2			显著
误差		12			
总	0.0012453	14			

因为 $F_{0.05}(2,12) = 3.89 < F_{比} = 32.92$，故拒绝 H_0，认为各台机器生产的薄板厚度有显著差异。

在进行方差分析时，还常要对未知参数进行估计。下面写出常用的几个估计：①

$\hat{\sigma}^2 = \dfrac{S_E}{n-s}$ 是的无偏估计。

② $\hat{\mu} = \bar{x}, \hat{\mu}_j = \bar{x}_{\cdot j}$ 分别是 μ, μ_j 的无偏估计。

③ $\hat{\sigma}_j = \bar{x}_{\cdot j} - \bar{x}$ 是 δ_j 的无偏估计，且 $\sum n_j \delta_j = 0$。

④两总体 $N(\mu_j, \sigma^2)$ 与 $N(\mu_K, \sigma^2)$ 的均差值 $\mu_j - \mu_k$ 的置信度为 $1-\alpha$ 的置信区间为