Chpt.8 Statistical Inference: Hypothesis Testing

第八章 假设检验

上节回顾

- 有关于正态总体方差的假设检验
- 非正态总体的假设检验

$$H_0\text{：}\ \theta=\theta_0 \qquad H_1\text{：}\ \theta\neq\theta_0$$

$$Z=\frac{\bar{X}-\mu(\theta_0)}{\sigma(\theta_0)/\sqrt{n}}\overset{\text{近似}}{\sim}N(0,1) \quad \text{拒绝域为 } |Z|\geq z_{\alpha/2}$$

- 分布拟合检验

例6 测量100个某种机械零件的质量(单位：g)，统计如下表

零件质量区间	频数	零件质量区间	频数
236.5~239.5	1	251.5~254.5	22
239.5~242.5	5	254.5~257.5	11
242.5~245.5	9	257.5~260.5	6
245.5~248.5	19	260.5~263.5	1
248.5~251.5	24	263.5~266.5	2

问： 这种机械零件的质量是否服从正态分布 (取显著水平$\alpha=0.05$)?

解： 依题意，检验假设

H_0：$X\sim N(\mu, \sigma^2)$，H_1：X不服从正态分布$N(\mu, \sigma^2)$.

已知 n=100，把各个区间的中点值取作x_i，分别计算参数μ，σ^2的极大似然估计值：

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 250.6 \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 26.82$$

$\hat{\sigma} = 5.18$，计算 $X \sim N(250.6,\ 5.18^2)$ 时，X 落在各个区间的概率

$$p_i = P\{\alpha_{i-1} \le X < \alpha_i\} = \Phi\left(\frac{\alpha_i - 250.6}{5.18}\right) - \Phi\left(\frac{\alpha_{i-1} - 250.6}{5.18}\right)$$

$$i = 1,2,\cdots,10$$

其中，Φ(X) 为标准正态分布的分布函数。

随机变量的取值为区间(-∞，+∞)， 应把第一个区间扩大为(-∞，239.5)，把最后一个区间扩大为(263.5，+∞) 。把检验所需数据整理并列表如下：

<table>
<tr><th>零件质量区间</th><th>m_i</th><th>p_i</th><th>np_i</th><th>$(m_i - np_i)^2 / np_i$</th></tr>
<tr><td>(-∞，239.5)</td><td rowspan="2">$\left.\begin{array}{c}1\\5\end{array}\right\}6$</td><td rowspan="2">0.0594</td><td rowspan="2">5.94</td><td rowspan="2">0.001</td></tr>
<tr><td>239.5~242.5</td></tr>
<tr><td>242.5~245.5</td><td>9</td><td>0.1041</td><td>10.41</td><td>0.191</td></tr>
<tr><td>245.5~248.5</td><td>19</td><td>0.1774</td><td>17.74</td><td>0.089</td></tr>
<tr><td>248.5~251.5</td><td>24</td><td>0.2266</td><td>22.66</td><td>0.079</td></tr>
<tr><td>251.5~254.5</td><td>22</td><td>0.2059</td><td>20.59</td><td>0.097</td></tr>
<tr><td>254.5~257.5</td><td>11</td><td>0.1348</td><td>13.48</td><td>0.456</td></tr>
<tr><td>257.5~260.5</td><td rowspan="3">$\left.\begin{array}{c}6\\1\\2\end{array}\right\}9$</td><td rowspan="3">0.0918</td><td rowspan="3">9.18</td><td rowspan="3">0.004</td></tr>
<tr><td>260.5~263.5</td></tr>
<tr><td>(263.5，+∞)</td></tr>
<tr><td>总计</td><td>100</td><td>1</td><td>100</td><td>0.917</td></tr>
</table>

由此得差异度$\chi^2=0.917$

因为合并后区间数为$k=7$，需要估计的参数个数 $r=2$，即χ^2

自由度 $l=k-r-1=4$

查表得 $\chi^2_{0.05}(4)=9.49$

因为 $\chi^2=0.917<\chi^2_{0.05}(2)=9.49$ ，所以接受原假设H_0

即在显著水平a=0.05下，可以认为这种机械零件的质量服从正

态分布.

习　题

从1500到1931年的432年间，每年爆发战争的次数可以看作一个随机变量，据统计，这432年间共爆发了299次战争，具体数据如下:

战争次数X	0	1	2	3	4
发生 X次战争的年数	223	142	48	15	4

通常假设每年爆发战争的次数服从泊松分布。那么上面的数据是否有充分的理由推翻每年爆发战争的次数服从泊松分布假设?

解：$H_0: X \sim \pi(\lambda)$，λ未知，$\hat{\lambda} = \overline{X} = 299/432 = 0.69.$

$$\hat{p}_i = \frac{\hat{\lambda}^i e^{-\hat{\lambda}}}{i!}, i = 0,1,2,3, \quad \hat{p}_4 = \sum_{j=4}^{\infty} \frac{\hat{\lambda}^j e^{-\hat{\lambda}}}{j!}.$$

战争次数x	0	1	2	3	$\geq$ 4
实测频数 m_i	223	142	48	15	4
概率估计 $\hat{p}_i$	0.502	0.346	0.119	0.027	0.006
理论频数 $n\hat{p}_i$	217	149	51	12	3

15

检验统计量的观察值为

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} = 1.74$$

即在显著性水平 $\alpha = 0.05$ 下临界值

$$\chi_\alpha^2(k-r-1) = \chi_{0.05}^2(4-1-1) = 5.991$$

于是， $1.74 < 5.991$,不能拒绝原假设。

假设强化训练

H0的选择非常重要。

- A：某种药为假药 B：某种药为真药

如果拒绝A带来的后果比拒绝B带来的后果更严重，则选择A作为原假设。
将A作为原假设，我们在检验时会控制第I类错误的概率，即当该药为假药错判为真药的概率不超过显著水平α，是个小概率事件。

- A：新技术为提高收益 B：新技术提高收益

在没有哪个假设带来的后果更严重需要规避时，常常选取H0为维持现状，即"无效益"，"无改进"，"无价值"等。这样，一旦拒绝H0，就表示有较强的理由去采取新技术。

- A: 红光的反应时间比绿光短 B: 红光的反应时间不小于绿光的反应时间

通常把我们想要支持的观点设为H1，这样一旦拒绝H0，就表示有较强的理由去支持我们想支持的观点。

假设强化训练

2. 如果一个矩形的宽度 w 与长度 l 的比 $w/l=\frac{1}{2}(\sqrt{5}-1)\approx 0.618$，这样的矩形称为黄金矩形. 这种尺寸的矩形使人们看上去有良好的感觉. 现代的建筑构件(如窗架)、工艺品(如图片镜框)，甚至司机的执照、商业的信用卡等常常都是采用黄金矩形. 下面列出某工艺品工厂随机取的 20 个矩形的宽度与长度的比值：

0.693	0.749	0.654	0.670	0.662	0.672	0.615	0.606	0.690	0.628
0.668	0.611	0.606	0.609	0.601	0.553	0.570	0.844	0.576	0.933

设这一工厂生产的矩形的宽度与长度的比值总体服从正态分布，其均值为 μ，方差为 σ^2，μ,σ^2 均未知. 试检验假设(取 $\alpha=0.05$)

$$H_0:\mu=0.618,\quad H_1:\mu\neq 0.618.$$

假设强化训练

3. 要求一种元件平均使用寿命不得低于 1 000 h，生产者从一批这种元件中随机抽取 25 件，测得其寿命的平均值为 950 h. 已知该种元件寿命服从标准差为 $\sigma=100$ h 的正态分布. 试在显著性水平 $\alpha=0.05$ 下判断这批元件是否合格？设总体均值为 μ，μ 未知. 即需检验假设

$$H_0:\mu\geqslant 1\,000,\ H_1:\mu<1\,000.$$

假设强化训练

4. 下面列出的是某工厂随机选取的 20 只部件的装配时间(min)：

9.8	10.4	10.6	9.6	9.7	9.9	10.9	11.1	9.6	10.2
10.3	9.6	9.9	11.2	10.6	9.8	10.5	10.1	10.5	9.7

设装配时间的总体服从正态分布 $N(\mu,\sigma^2)$，μ,σ^2 均未知. 是否可以认为装配时间的均值显著大于 10 (取 $\alpha=0.05$)?

$$H_0:\ \mu \le 10,\ H_1:\ \mu > 10$$

12. 某种导线，要求其电阻的标准差不得超过 0.005 Ω，今在生产的一批导线中取样品 9 根，测得 $s=0.007$ Ω，设总体为正态分布，参数均未知. 问在显著性水平 $\alpha=0.05$ 下能否认为这批导线的标准差显著地偏大?

$$H_0:\ \sigma < 0.005,\ H_1:\ \sigma \ge 0.005$$

假设强化训练

10. 为了试验两种不同的某谷物的种子的优劣，选取了10块土质不同的土地，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. 设在每块土地的两部分人工管理等条件完全一样. 下面给出各块土地上的单位面积产量：

土地编号 i	1	2	3	4	5	6	7	8	9	10
种子 $A(x_i)$	23	35	29	42	39	29	37	34	35	28
种子 $B(y_i)$	26	39	35	40	38	24	36	27	41	27

设 $D_i=X_i-Y_i$ $(i=1,2,\cdots,10)$是来自正态总体 $N(\mu_D,\sigma_D^2)$的样本，μ_D,σ_D^2 均未知. 问以这两种种子种植的谷物的产量是否有显著的差异(取 $\alpha=0.05$)？

$$H_0: \mu_D = 0,\ H_1: \mu_D \neq 0$$

假设强化训练

9. 为了比较用来做鞋子后跟的两种材料的质量，选取了 15 名男子(他们的生活条件各不相同)，每人穿一双新鞋，其中一只是以材料 A 做后跟，另一只以材料 B 做后跟，其厚度均为 10 mm. 过了一个月再测量厚度，得到数据如下：

男子	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
材料 $A(x_i)$	6.6	7.0	8.3	8.2	5.2	9.3	7.9	8.5	7.8	7.5	6.1	8.9	6.1	9.4	9.1
材料 $B(y_i)$	7.4	5.4	8.8	8.0	6.8	9.1	6.3	7.5	7.0	6.5	4.4	7.7	4.2	9.4	9.1

设 $D_i=X_i-Y_i\quad(i=1,2,\cdots,15)$是来自正态总体 $N(\mu_D,\sigma_D^2)$的样本，μ_D,σ_D^2 均未知. 问是否可以认为以材料 A 制成的后跟比材料 B 的耐穿(取 $\alpha=0.05$)?

$$H_0\colon \mu_D \le 0,\ H_1\colon\ \mu_D > 0$$

检验假设的设计

对于某研究问题，假设某样本预期服从$N(10,3.5^2)$正态分布，其采样均值$\bar{x}=12$，同时已知$\mathrm{n}=9$

请做如下判定：在显著性0.05情况下，1. 本次采样的结果（均值）是否发生变化； 2. 本次样本的均值是否变大

第一问： $H_0: u=u_0; H_1: u \neq u_0$

第二问： $H_0: u \leq u_0; H_1: u > u_0$

“工具”随机变量 $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 带入数据： $\frac{12-10}{3.5/\sqrt{9}}=\frac{2*3}{3.5}=1.7$

对于第一问： 双尾检验， $Z_{\alpha/2}=1.96$

所以有 $-1.96<z=1.7<1.96$ 结论:

$u=u_0$

对于第二问： 单尾检验， $Z_{\alpha}=1.64$

所以有 $z=1.7>1.64$ 结论:

$u>u_0$

检验结果是高度依赖检验设计的。当有可能时，应该选择更为严格（或者说更明确）的H1.

置信区间与假设检验的关系

我们观测到，在进行置信区间估计和假设检验时，使用了同样的统计量选用方法，那么二者是否存在联系？

注意： 首先分辨两者的区别！

8.5* 置信区间与假设检验的关系

A：从置信区间出发

设$X\sim f(\theta)$，对参数θ应用某统计量进行区间估计，得到$1-\alpha$置信区间$(\underline{\theta},\overline{\theta})$，按照置信区间定义，有

$$P\{\underline{\theta}<\theta<\overline{\theta}\}\geq 1-\alpha$$

此时对于原假设$H_0:\theta=\theta_0$，若其满足，则应该有$P\{\underline{\theta}<\theta_0<\overline{\theta}\}\geq 1-\alpha$

对上述随机变量运算进行处理

$$P\{(\underline{\theta}<\theta_0)\cap(\theta_0<\overline{\theta})\}\geq 1-\alpha$$

或者有

$$P\{(\theta_0<\overline{\theta})\cap(\theta_0>\underline{\theta})\}\geq 1-\alpha$$

若满足任意性，则上述三个命题等价

8.5* 置信区间与假设检验的关系

$P\{\underline{\theta} < \theta < \overline{\theta}\} \geq 1-\alpha$ 讨论的是θ的$1-\alpha$的置信区间；而

$P\{(\theta_0 < \overline{\theta}) \cap (\theta_0 > \underline{\theta})\} \geq 1-\alpha$讨论的是 θ_0落在两个（关于X的统计量）的区间之内，

则H_0则为大概率事件，接受H0。

对此我们得到另一种假设检验的方法：

计算θ的置信区间，然后观察θ_0是否落入该置信区间内，如果落入则接受$H_0: \theta = \theta_0$

对比原来方法求拒绝域：

我们首先定义一个包含θ_0和X的，服从基本分布的统计量，然后计算该统计量的分位值及相应拒绝域。观察样本值造成的统计量的数值，是否落入该拒绝域。

进而拓展讨论：如何处理讨论单尾检验vs. 单侧置信区间

8.5* 置信区间与假设检验的关系

B：从假设检验出发（H_0：$\theta_0=\theta$）

对于θ_0的接受域有，$P\{(\theta_0<\overline{\theta})\cap(\theta_0>\underline{\theta})\}\geq 1-\alpha$，

由于对于任意θ_0，θ都成立，那么对于$\theta_0=\theta$的情况，只要θ落在上述区间$(\underline{\theta},\overline{\theta})$中（注意，此时区间$(\underline{\theta},\overline{\theta})$是$\theta_0$的接受域）

仿照前文讨论，进行部分替换后有下面公式

$$P\{\underline{\theta}<\theta<\overline{\theta}\}\geq 1-\alpha$$

此时接受域$(\underline{\theta},\overline{\theta})$恰好符合$\theta$置信空间的定义，所以$(\underline{\theta},\overline{\theta})$可以作为$\theta$的$1-\alpha$置信空间

所以，若我们将假设检验研究中的接受域$Z_{1-\alpha/2}<$统计量$(X,\theta_0)<Z_{\alpha/2}$，转化为

$$\underline{\theta}=\text{统计量}(X,Z_{1-\alpha/2})<\theta_0<\overline{\theta}=\text{统计量}(X,Z_{\alpha/2})$$

则$(\underline{\theta},\overline{\theta})$也可以作为$\theta$的置信区间（当$H_0$满足时）

置信区间与假设检验的关系

例 1 设 $X\sim N(\mu,1)$，μ 未知，$\alpha=0.05$，$n=16$，且由一样本算得 $\bar{x}=5.20$，于是得到参数 μ 的一个置信水平为 0.95 的置信区间

$$\left(\bar{x}-\frac{1}{\sqrt{16}}z_{0.025},\bar{x}+\frac{1}{\sqrt{16}}z_{0.025}\right)=(5.20-0.49,5.20+0.49)$$

$$=(4.71,5.69).$$

现在考虑检验问题 $H_0:\mu=5.5$，$H_1:\mu\neq 5.5$. 由于 $5.5\in(4.71,5.69)$，故接受 H_0. □

例 2 数据如上例. 试求右边检验问题 $H_0:\mu\leqslant\mu_0$，$H_1:\mu>\mu_0$ 的接受域，并求 μ 的单侧置信下限($\alpha=0.05$).

解 检验问题的拒绝域为 $z=\dfrac{\bar{x}-\mu_0}{1/\sqrt{16}}\geqslant z_{0.05}$，或即 $\mu_0\leqslant 4.79$. 于是检验问题的接受域为 $\mu_0>4.79$. 这样就得到 μ 的单侧置信区间 $(4.79,\infty)$，单侧置信下限 $\underline{\mu}=4.79$. □

p值法

例 1 设总体 $X \sim N(\mu,\sigma^2)$，μ 未知，$\sigma^2=100$，现有样本 $x_1,x_2,\cdots,x_{52}$，算得 $\bar{x}=62.75$. 现在来检验假设

$$H_0:\mu=\mu_0=60,\qquad H_1:\mu>60.$$

采用 Z 检验法，检验统计量为

$$Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}.$$

以数据代入，得 Z 的观察值为

$$z_0=\frac{62.75-60}{10/\sqrt{52}}=1.983.$$

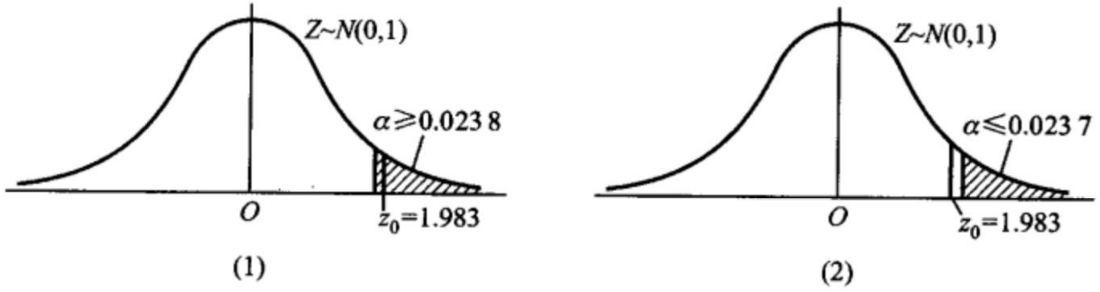
概率

$$P\{Z\geqslant z_0\}=P\{Z\geqslant 1.983\}=1-\Phi(1.983)=0.0238.$$

此即为图 8—7 中标准正态曲线下位于 z_0 右边的尾部面积.

此概率称为 Z 检验法的右边检验的 p 值. 记为

$$P\{Z\geqslant z_0\}=p\text{ 值}(=0.0237).$$



*样本容量的选取
*秩和检验