§8.5 关于比率的检验

$X \sim B(n,p)$, 参数: p, 范围: $[0,1]$. 检验问题:

$$H_0 : p = p_0 \quad 或 \quad H_0 : p \geqslant p_0 \quad 或 \quad H_0 : p = p_0.$$

- 大样本方法: 当$n \gg 1$ 且$np_0 \geqslant 5$ 时, 用中心极限定理.
- 小样本方法. 联系F 分布.

例5.1. 50 ~ 54 岁的美国妇女乳腺癌发病率为$p_0 = 2\%$. 调查10000 名50 ~ 54 岁的母亲患有乳腺癌的妇女, 发现其中有400 名患乳腺癌. 检验

$$H_0 : p = 2\% \leftrightarrow H_1 : p \neq 2\%.$$

- $X \sim B(n,p)$: 在H_0 下, 近似地有
$$\frac{X - np}{\sqrt{npq}} \sim N(0,1), \quad (q = 1 - p).$$
- 否定域:
$$\mathcal{W} = \left\{ x : \left| \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \right| > z_{1-\alpha/2} \right\}.$$
- $\frac{x-np_0}{\sqrt{np_0q_0}} = 14.28 > 1.96 = z_{0.975}$, 否定$H_0$.

例5.2. 学生刻苦(复习时间 : 上课时间 $> 1:1$)的概率为p. 在132份调查问卷中发现有127 人刻苦. 检验

$$H_0 : p \leqslant p_0 = 0.9 \leftrightarrow H_1 : p > p_0.$$

- $X \sim B(n,p)$. 小样本方法: 否定域为$\mathcal{W} = \{x : x \geqslant i\}$.
- i 满足$f(i,p_0) \leqslant \alpha < f(i-1,p_0)$: 由例3.6.7,
  $f(i,p) = P_p(X \geqslant i) = \int_0^p \frac{n!}{(i-1)!(n-i)!} u^{i-1}(1-u)^{n-i}du.$
- $\mathcal{W} = \{x : f(x,p_0) \leqslant \alpha\}$. 因为, $x \geqslant i \Leftrightarrow f(x,p_0) \leqslant f(i,p_0)$.
- $f(x,p_0) = P(Y \geqslant y)$, 其中$Y \sim F = F(2(n-x+1), 2x)$,
  $$\varphi(y) := (1 + \tfrac{n-x+1}{x} y)^{-1} = p_0.$$
- $\mathcal{W} = \{x : p(x) \geqslant p_0\}$, 其中, $p(x) := \varphi(F_{1-\alpha})$:
  $$f(x,p_0) \leqslant \alpha \text{ iff } y \geqslant F_{1-\alpha} \text{ iff } p_0 \leqslant \varphi(F_{1-\alpha}).$$
- 否定H_0:
  $$p(127) = \left(1 + \frac{132-127+1}{127} F_{0.95}\right)^{-1} = 0.9220 > p_0.$$

2. 两总体比较的假设检验问题.

总体: $X \sim B(n_1, p_1)$, $Y \sim B(n_2, p_2)$. 检验问题:

$$H_0: p_1 \leqslant p_2 \quad 或\ H_0: p_1 \geqslant p_2 \quad 或\ H_0: p_1 = p_2.$$

- CLT: $\frac{X - n_1 p_1}{\sqrt{n_1 p_1 q_1}} \overset{d}{\approx} Z \sim N(0, 1)$. SLLN: $\hat{p}_1 = \frac{X}{n} \approx p_1$.
- $\frac{X}{n_1} - p_1 \overset{d}{\approx} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}} Z_1$, $\frac{Y}{n_2} - p_2 \overset{d}{\approx} \sqrt{\frac{\hat{p}_2 \hat{q}_2}{n_2}} Z_2$.
- $(\frac{X}{n_1} - \frac{Y}{n_2}) - (p_1 - p_2) \overset{d}{\approx} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} Z$.
- $\xi = \left((\frac{X}{n_1} - \frac{Y}{n_2}) - (p_1 - p_2) \right) / \star\star$ 近似服从 $N(0, 1)$.
- $\eta = (\bar{X} - \bar{Y}) / \star\star \overset{H_0}{\leqslant} \xi$.
- 否定域为: $\mathcal{W} = \{x : \eta \geqslant z_{1-\alpha}\}$.

$$H_0 : p_1 = p_2 \leftrightarrow H_1 : p_1 \neq p_2.$$

- 在H_0 下, $p_1 = p_2 = p$. CLT:

$$\frac{X}{n_1} - p \overset{d}{\approx} \sqrt{\frac{\hat{p}\hat{q}}{n_1}} Z_1, \quad \frac{Y}{n_2} - p \overset{d}{\approx} \sqrt{\frac{\hat{p}\hat{q}}{n_2}} Z_2,$$

$$\frac{X}{n_1} - \frac{Y}{n_2} \overset{d}{\approx} \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\hat{p}\hat{q}} Z.$$

- SLLN: $\hat{p} = \frac{X+Y}{n_1+n_2} \approx p.$
- $\zeta = (\bar{X} - \bar{Y})/\star\star$. 否定域: $\mathcal{W} = \{x : |\zeta| \geqslant z_{1-\alpha/2}\}$.

例5.3. 研究口服避孕药对40 $\sim$ 44 岁年龄段妇女心脏的影响. 5000 位使用者三年内心梗死13 人; 10000 位不使用者三年内心梗死7 人. 检验$H_0: p_1 = p_2$. ($\alpha = 0.01$.)

- $\hat{p} = (13+7)/(5000+10000) = 20/15000$.
- $\zeta = (\bar{x}-\bar{y})/\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})\hat{p}\hat{q}} = 3.01 > 2.58 = z_{0.995}$, 否定$H_0$.
- 改为验证$H_0: p_1 \leqslant p_2$.
- $\hat{p}_1 = 13/5000$, $\hat{p}_2 = 7/10000$.
- $\eta = (\bar{x}-\bar{y})/\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} = 2.48 > 2.33 = z_{0.99}$, 否定$H_0$.