第六章、假设检验

§6.1 问题的提法

- 例1.1. 某厂有一批产品，共200 件，次品率p 不超过1% 方可出厂. 抽取5件，发现含有次品. 问：这批产品能否出厂？
- 问题化为：根据抽样的结果判断$p \leqslant 0.01$是否成立.
- 例1.2. 温度真值：1277. 某仪器测量5 次，数据：1250, 1265, 1245, 1260, 1275. 问：该仪器测量结果有无系统偏差？
- 问题化为：判断等式$EX = 1277$成立与否.

- 例1.3. 某工厂近5 年发生了63 次事故, 在工作日的分布如下:

星期	一	二	三	四	五	六
次数	9	10	11	8	13	12

问: 事故的发生是否与星期几有关?

- 问题化为: 判断$P(X=i)\equiv\frac{1}{6}$是否成立, X: 事故发生日.
- 例1.4. 漂布工艺中温度对强力的影响.

70　20.5, 18.8, 19.8, 20.9, 21.5, 19.5, 21.0, 21.2

80　17.7, 20.3, 20.0, 18.8, 19.0, 20.1, 20.2, 19.1

- 问题变成: 判断$EX=EY$, $D(X)=D(Y)$成立与否.

- 例1.5. 新款吹风机失效率p_1; 旧款吹风机失效率p_2. 各取250件进行可靠性试验, 新款有11 个失效, 旧款有20 个失效. 问: 新款的可靠性是否不比旧款的差?
- 问题化为: 判断$p_1 \leqslant p_2$是否成立.
- 例1.6. 怎样根据样本值判断$X \sim N(\mu, \sigma^2)$?

  根据样本的特性判断$F_X(x) = F(x), \forall x$ 是否成立.

- 假设检验: 从样本值出发去判断一个看法是否成立.
- 例1.1 $\sim$ 1.6的看法:

  例1.1: $p \leqslant 0.01$;

  例1.2: $EX = 1277$;

  例1.3: $P(X=i) \equiv \frac{1}{6}$, $i = 1, 2, \cdots, 6$;

  例1.4: $EX = EY$, 进一步$D(X) = D(Y)$;

  例1.5: $p_1 \leqslant p_2$;

  例1.6: $F_X = F$.
- 看法又叫假设. 这些例子就是**假设检验**问题.

  判断所关心的假设是否成立.
- 一个总体的**参数**检验问题: 例1.1, 1.2, 1.3;

  二总体的检验问题: 例1.4, 1.5;

  总体分布检验: 例1.3, 1.6.

- 例1.1. 目标: 检验$p \leqslant 0.01$. 抽检: 200 件中抽查5 件,
  结果: $A =$ “含次品”发生.
- 假设总体中有$M \leqslant 2$ 件次品. 计算$P_M(A)$.
- $P_M(A) \leqslant 0.05$: 古典概型, $M \leqslant 2$, $200 - M = 200, 199, 198$,

$$P_M(A^c) = \frac{C_{200-M}^5}{C_{200}^5} \geqslant \frac{C_{198}^5}{C_{200}^5} \approx 0.9505.$$

- 单次试验中, 小概率事件发生, 不合理/矛盾!
- 不接受(拒绝/否定)假设$p \leqslant 0.01$, 这批产品不能出厂.

一般地,

- 目标: 检验H_0. 抽检: 得到数据$\vec{x}=(x_1,\cdots,x_n)$.
- 先假设H_0 成立.
  若实际情况在假设H_0 下是不合理的,
  (发现A 发生, 但$P_{H_0}(A)\leqslant\alpha$, 一般取$\alpha=0.05$,)
  则认为原来的假设是不正确的, **拒绝**原假设H_0;
- 否则, 认为样本没有不合理现象, **不拒绝(接受)** 原假设H_0.
- 注: 这不是纯粹的反证法. 不合理不是形式逻辑中绝对的矛盾, 而是认为小概率事件在一次观察中基本不可能发生.
- 但原假设成立的情况下, A 发生则错误地拒绝了H_0, 犯错的可能性$\leqslant\alpha$.
- 称α 为**检验水平/检验标准**.

§6.2 一个正态总体的假设检验

假设$X \sim N(\mu, \sigma^2)$, 四种假设检验问题:

- σ^2 已知, $H_0: \mu = \mu_0$ 或$H_0: \mu \leqslant \mu_0$ 或$H_0: \mu \geqslant \mu_0$;
- σ^2 未知, $\cdots\cdots$;
- μ 已知, $H_0: \sigma^2 = \sigma_0^2$ 或$H_0: \sigma^2 \leqslant \sigma_0^2$ 或$H_0: \sigma^2 \geqslant \sigma_0^2$;
- μ 未知, $\cdots\cdots$.

σ^2 已知，检验μ.

例2.1. 铜丝折断力$X \sim N(\mu, \sigma^2)$(单位:千克力), $\mu_0 = 570$, $\sigma = 8$. 换原料后，认为σ 不变. $n = 10$, 数据: 578, 572, 570, 568, 572, 570, 570, 572, 596, 584. 问: 折断力是否还一样?

- 已知方差$\sigma^2 = 8^2 = 64$. 检验假设$H_0 : \mu = \mu_0 = 570$.
- 在H_0 下，检验统计量 vs 枢轴量:

$$Z(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/10}} \overset{H_0}{=} \frac{\bar{X} - \mu}{\sqrt{\sigma^2/10}} = Z(\vec{X}, \mu) \sim N(0, 1).$$

- 查表$P(|Z| > 1.96) = 0.05$. (水平: $\alpha = 0.05$.)
- 代数据, $\bar{x} = 575.2$, $Z(\vec{x}) = 2.05 > 1.96$, 小概率事件发生了.
- 拒绝H_0: 折断力和原来的有显著差异.
- 调整水平α: $P(|Z| > 2.58) = 0.01$. A 不发生，接受H_0.

例2.1(续).

- 已知方差$\sigma^2 = 64$. 检验假设$H_0: \mu \leqslant \mu_0 = 570$.
- 在H_0 下, 比较检验统计量与枢轴量,

$$Z(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/10}} \overset{H_0}{\leqslant} \frac{\bar{X} - \mu}{\sqrt{\sigma^2/10}} = Z(\vec{X}, \mu) \sim N(0,1).$$

- 查(枢轴量的)表: $\lambda = z_{1-\alpha}$,

$$P_{H_0}(Z(\vec{X}) > \lambda) \leqslant P(Z(\vec{X}, \mu) > \lambda) \leqslant \alpha.$$

- 代数据: $Z(\vec{x}) = 2.05 > z_{0.975} = 1.96$, 小概率事件发生.
- 拒绝H_0: 折断力的大小比原来的有显著提高.
- 单边检验: H_0 选为希望得到的回答的反面.

  指望否定H_0, 于是得到希望得到的回答.

例2.2. 砖的抗断强度$X \sim N(\mu, 1.21)$. $n = 6$, 数据: 32.56, 29.66, 31.64, 30.00, 31.87, 31.03. 问: 能不能认为$H_0: \mu = \mu_0 = 32.80$?

- 检验统计量、枢轴量:

$$Z(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \overset{H_0}{=} \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} = Z(\vec{X}, \mu) \sim N(0, 1).$$

- 查(枢轴量的)表: $P(|Z| > 1.96) = 0.05$.
- 代数据: $Z(\vec{x}) = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{31.13 - 32.50}{\sqrt{1.21/6}} = -3.05$.
- 结论: $\underline{|Z(\vec{x})| = 3.05 > 1.96}$, 小概率事件<u>发生</u>.

  拒绝H_0, 不能认为$\mu = 32.50$.

- 注: 这批砖的平均抗断强度明显低于32.50.

检验水平与两类错误.

- 假设$\alpha < 0.5$. 取**临界值**$\lambda = \lambda(\alpha)$ 使得$P(|Z| > \lambda) = \alpha$.
- 检验: 计算检验统计量$Z = Z(\vec{x})$ 的值.
  若$|Z| > \lambda$, 则拒绝H_0; 否则, 不拒绝/接受H_0.
  注: 结果与α 的选择有关.
- **第一类错误**: H_0 成立, 拒绝H_0. α: 犯错概率的上界.
- **第二类错误**: H_0 不成立, 接受H_0. β: 犯错概率(的上界).
- 两类错误的犯错概率越小越好, 但两者互相矛盾.
- 经典的统计假设检验做法: 固定α, 然后尽可能设法减小β.
  如, 增大样本量, 设计更好的检验方案.
- 注: 拒绝H_0 vs 接受H_0. 实践中, **指望否定**H_0.
  报告偏差: 拒绝H_0 就报告, 接受H_0 则不报告.

例1.2. 测量值$X \sim N(\mu, \sigma^2)$. $n = 5$, 数据: 1250, 1265, 1245, 1260, 1275. σ^2 未知. 检验$H_0 : \mu = \mu_0 = 1277$.

- σ^2 未知: 用σ^2 的无偏估计S^2(样本方差) 代替σ^2.
  检验统计量、枢轴量:
  $$T(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \overset{H_0}{=} \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = T(\vec{X}, \mu) \sim t(n-1).$$
- 查$t(n-1)$ 分布表: $P(|T_4| > 2.776) = 0.05$.
- 代数据: $\bar{x} = 1259$, $s^2 = S^2(\vec{x}) = 142.5$, $T(\vec{x}) = -3.37$.
- 下结论: $|T(\vec{x})| > 2.776$, 小概率事件发生.

  拒绝H_0, (该仪器测温明显偏低).

例2.3. 钢筋强度$X \sim N(\mu, \sigma^2)$. $n = 6$. 数据: 48.5, 49.0, 53.5, 49.5, 56.0, 52.5. 检验$H_0 : \mu = 52.0$.

- 计算检验统计量$T(\vec{x})$: $\bar{x} = 51.5$, $s^2 = 8.9$,

$$T(\vec{x}) = \frac{51.5 - 52.0}{\sqrt{8.9/6}} = -0.411.$$

- 查表: $n = 6$, $P(|T_5| > 2.571) = 0.05$.
- 结论: $|T(\vec{x})| = 0.411 < 2.571$,

  H_0 是相容的, 不能否定$\mu = 52.0$.

例2.5. 罐头番茄汁中维C含量$X \sim N(\mu, \sigma^2)$. $\mu \geqslant 21$ 则合格.
$n = 17$, $\bar{x} = 23$, $S^2 = 3.98^2$. 问: 是否合格? ($\alpha = 0.05$).

- 希望“证明” $\mu \geqslant 21$, 从而选择H_0: $\mu \leqslant 21$.
- 统计量、枢轴量:

$$T(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \overset{H_0}{\leqslant} \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = T(\vec{X}, \mu) \sim t(n-1).$$

- 查$t(n-1)$ 的表: $\lambda = t_{1-\alpha}(n-1)$,

$$P_{H_0}(T(\vec{X}) > \lambda) \leqslant P_{H_0}(T(\vec{X}, \mu) > \lambda) = \alpha.$$

- 代数据: $n - 1 = 16$, $\lambda = 1.746$. $\underline{T(\vec{x}) = 2.07 > 1.746}$.
- 结论: 小概率事件发生, 拒绝H_0, 认为该批罐头合格.

例4.3. (成对数据比较). 9 批原料分别用两种工艺生产, 得到某指标的9 对数据(X,Y) 如下:

0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0.10	0.21	0.52	0.32	0.78	0.59	0.68	0.77	0.89

问: 两种工艺的指标有无显著差异?

- 假设$V = X - Y \sim N(\mu,\sigma^2)$, 检验$H_0: \mu = 0$.
- 计算9 对数据的差, 即V 的数据:

0.10	0.09	−0.12	0.18	−0.18	0.11	0.12	0.13	0.11

- 计算: $T(\vec{v}) = \dfrac{\bar{v}}{\sqrt{s^2/9}} = 1.467$.
- 查表得λ: $P(|T_8| > 2.306) = 0.05$.
- 下结论: $|T(\vec{v})| < \lambda$, H_0 相容, 两种工艺的指标无显著差异.

假设检验与置信区间.

- $H_0: \mu = \mu_0$ 的检验与μ 的置信区间有密切联系.
- 关键点: 枢轴量

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} = T(\vec{X},\mu) \sim t(n-1).$$

临界值: $P(|T_{n-1}| > \lambda) = \alpha$.

- μ 的置信水平为$1-\alpha$ 的置信区间:

$$|T(\vec{X},\mu)| \leqslant \lambda, \quad 即\ \left[\bar{x}-\lambda\sqrt{S^2/n}, \bar{x}+\lambda\sqrt{S^2/n}\right].$$

- 检验问题$H_0: \mu = \mu_0$. 将$\{|T_{n-1}| > \lambda\}$ 视为小概率事件, 故

$$当且仅当\ |T(\vec{X},\mu)| \leqslant \lambda\ 时,\ 接受H_0.$$

- 等价地: $\mu_0 \in [\star,\star]$ 时, 接受H_0; 否则, 拒绝H_0.
- 注: 可用置信区间来进行检验, 也可由检验法构造置信区间.

例2.4. 铜丝折断力$X \sim N(\mu, \sigma^2)$ (单位: 千克力). $n = 10$, 数据:
578, 572, 570, 568, 572, 570, 570, 570, 572, 596, 584.
检验$H_0 : \sigma^2 = \sigma_0^2 = 64$.

- 统计量、枢轴量:

$$K(\vec{X}) = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{=} \frac{(n-1)S^2}{\sigma^2} = K(\vec{X}, \sigma^2) \sim \chi^2(n-1).$$

- 查$\chi^2(n-1)$ 的表: $\lambda_1 = \chi^2_{\alpha/2}(n-1)$, $\lambda_2 = \chi^2_{1-\alpha/2}(n-1)$,

$$P(K_{n-1} < \lambda_1) = 0.025, \quad P(K_{n-1} > \lambda_2) = 0.025.$$

- 代数据: $\bar{x} = 575.2$, $s^2 = 75.73$, $K(\vec{x}) = \frac{9\times 75.73}{64} = 10.65$;
  $n - 1 = 9$, $\lambda_1 = 2.70$, $\lambda_2 = 19.0$.
- 下结论: $\lambda_1 < K(\vec{x}) < \lambda_2$, H_0 是相容的, 可以相信$\sigma^2 = 64$.

例. $X \sim N(\mu, \sigma^2)$, σ^2: 工艺精度. $H_0: \sigma^2 \geqslant \sigma_0^2$, 或$H_0: \sigma^2 \leqslant \sigma_0^2$.

- 指望否定H_0, “证明”精度有改进,
  指望接受H_0, 以“防止”精度太差.
- 统计量、枢轴量:

$$K(\vec{X}) = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{\geqslant} \frac{(n-1)S^2}{\sigma^2} = K(\vec{X}, \sigma^2) \sim \chi^2(n-1).$$

- 查$\chi^2(n-1)$ 的表: $\lambda = \chi^2_\alpha(n-1)$,

$$P_{H_0}(\underset{\sim\sim\sim}{K(\vec{X}) < \lambda}) \leqslant P(K(\vec{X}, \sigma^2) < \lambda) = \alpha.$$

- 结论: 如果$\underline{K(\vec{x}) < \lambda}$, 则拒绝$H_0$, “证明” $\sigma^2 < \sigma_0^2$;
  否则, 不拒绝H_0, H_0 与数据相容.

检验$H_0: \mu = \mu_0$ (或 $\mu \leqslant \mu_0$, 或 $\mu \geqslant \mu_0$):

(1) 统计量、枢轴量: $U \overset{H_0}{=}, \overset{H_0}{\leqslant}, \overset{H_0}{\geqslant} V$,

$$U = U(\vec{X}) = \frac{\bar{X} - \mu_0}{\sqrt{\delta/n}}, \quad V = V(\vec{X}, \mu) = \frac{\bar{X} - \mu}{\sqrt{\delta/n}}.$$

(2) σ^2 已知: $\delta = \sigma^2$, $V = Z \sim N(0,1)$,

σ^2 未知: $\delta = S^2$, $V = T_{n-1} \sim t(n-1)$.

(3) 查表得临界值: $\lambda = z_{1-\alpha/2}$ 或, 或 $z_{1-\alpha}$;

$$\lambda = t_{1-\alpha/2}(n-1) \text{ 或, 或 } t_{1-\alpha}(n-1).$$

(4) 当$|U(\vec{x})| > \lambda$ (或 $U(\vec{x}) > \lambda$, 或 $U(\vec{x}) < -\lambda$) 时, 拒绝H_0. 否则, 接受H_0.

检验$H_0: \sigma^2 = \sigma_0^2$ (或 $H_0: \sigma^2 \geqslant \sigma_0^2$, 或 $H_0: \sigma^2 \leqslant \sigma_0^2$):

(1) 统计量、枢轴量: K =, $\geqslant$, $\leqslant$ K_{n-1},

$$K(\vec{X}) = \frac{(n-1)S^2}{\sigma_0^2}, \quad \frac{(n-1)S^2}{\sigma^2} = K(\vec{X}, \sigma^2) \sim \chi^2(n-1).$$

(2) 查表得临界值: $\lambda_1 = \chi^2_{\alpha/2}(n-1)$ (或 $\chi^2_{\alpha}(n-1)$), $\lambda_2 = \chi^2_{1-\alpha/2}(n-1)$ (或 $\chi^2_{1-\alpha}(n-1)$).

(3) 当$K(\vec{x}) < \lambda_1$ 或$K(\vec{x}) > \lambda_2$ 时(或 $K(\vec{x}) < \lambda_1$, 或 $K(\vec{x}) > \lambda_2$), 拒绝H_0. 否则, 接受H_0.

- 注: 若μ 已知, 则$(n-1)S^2$ 改为$\sum_{i=1}^{n}(X_i-\mu)^2$, $\chi^2(n-1)$ 改为$\chi^2(n)$.

非正态总体的均值假设检验

- 对于非正态的总体，当$n \to \infty$ 时，根据中心极限定理以及相关的概率极限理论可以证明：在$H_0 : \mu = \mu_0$ 下，统计量$Z(\vec{X})$ 近似服从N(0, 1).
- 当样本量n 不小于30，最好是50以上，甚至100 以上时，可近似使用正态总体的结论.

§6.3 假设检验的某些概念和数学描述

- H_0 是关于X 的分布的 “看法”.
- $F_X(x) = F(x,\theta)$, $\theta \in \Theta$.
- 零假设/原假设, $H_0 : \theta \in \Theta_0$,
  对立假设/备择假设, $H_a, H_1 : \theta \in \Theta_1 = \Theta \backslash \Theta_0$.
- 例2.1. 铜丝折断力$X \sim N(\mu, \sigma^2)$, $\sigma^2 = 8^2$ (已知),

$$H_0 : \mu = \mu_0 \longleftrightarrow H_1 : \mu \neq \mu_0.$$

检验法

- 检验法就是给出一个**规则**, 对给定的样本值$\vec{x}=(x_1,\ldots,x_n)$, 进行**表态**: 拒绝假设H_0还是接受假设H_0.
- 将$\mathbb{R}^n$**划分**为两个区域: $\mathcal{W}$, $A=\mathbb{R}^n\backslash\mathcal{W}$.
  规则: 若$\vec{x}\in\mathcal{W}$, 则拒绝H_0; 若$\vec{x}\in A$ 则接受H_0.
- 称$\mathcal{W}$ 为否定域或拒绝域; 称A 为接受域.
- 检验法= 拒绝域/否定域: $\mathcal{W}\subseteq\mathbb{R}^n$.
- $\mathcal{W}$ 通常由检验统计量给出.
  例: $\mathcal{W}=\{\vec{x}:|Z(\vec{x})|>\lambda\}$, $\quad Z(\vec{x})=\frac{\bar{x}-\mu_0}{\sqrt{\sigma^2/n}}$.
- 检验结果与检验方法$\mathcal{W}$ 的选择有关; 先定方法, 再采集数据.

- 检验法, 拒绝域: $\mathcal{W} \subseteq \mathbb{R}^n$.

  结果: $\vec{x} \in \mathcal{W}$ 则拒绝H_0; $\vec{x} \notin \mathcal{W}$ 则不拒绝H_0.

- 两类错误:

真情\ 结果	H_0	H_1
H_0	✓	× (第I类)
H_1	× (第II类)	✓

- 犯错概率$\alpha_{\mathcal{W}}(\theta)$ 与$\beta_{\mathcal{W}}(\theta)$:

$$\alpha_{\mathcal{W}}(\theta) = P_\theta(\vec{X} \in \mathcal{W}) =: M_{\mathcal{W}}(\theta),\ \theta \in \Theta_0;$$

$$\beta_{\mathcal{W}}(\theta) = P_\theta(\vec{X} \notin \mathcal{W}) = 1 - M_{\mathcal{W}}(\theta),\ \theta \in \Theta_1$$

- 称$M_{\mathcal{W}}(\theta)$ 为否定域$\mathcal{W}$ (或对应的检验法的)功效函数.

- $\mathcal{W}$ 是检验/显著性水平为α **的否定域**: $\alpha_{\mathcal{W}}(\theta) \leqslant \alpha,\ \forall \theta \in \Theta_0$.

  α 为W **的精确检验水平**: $\alpha = \sup_{\theta \in \Theta_0} \alpha_{\mathcal{W}}(\theta)$. (定义3.1)

- 经典做法：固定α；尽量降低$\beta_{\mathcal{W}}(\theta)$, $\theta \in \Theta_1$. (功效尽量大).
- 小概率事件: $P_\theta(\vec{X} \in \mathcal{W}) \leqslant \alpha$, $\theta \in \Theta_0$.
- 结果: 拒绝H_0. 犯错概率$\leqslant \alpha$. 强拒绝.
- 结果: 接受H_0, 犯错概率$\beta_{\mathcal{W}}(\theta)$, $\theta \in \Theta_1$.

  “没推出矛盾”不代表“假设成立”. 弱接受.
- 否定域$\mathcal{W}$, 检验统计量 $\varphi(\vec{X})$.
- 单边否定域: 临界值$\lambda = \lambda(\alpha)$,

$$\mathcal{W} = \{\vec{x} : \varphi(\vec{x}) > \lambda\} \text{ 或 } \mathcal{W} = \{\vec{x} : \varphi(\vec{x}) < \lambda\}.$$

- 双边否定域: 临界值$\lambda_1 = \lambda_1(\alpha)$, $\lambda_2 = \lambda_2(\alpha)$,

$$\mathcal{W} = \{\vec{x} : \varphi(\vec{x}) < \lambda_1\} \bigcup \{\vec{x} : \varphi(\vec{x}) > \lambda_2\}.$$

- 单边. $\mathcal{W}=\{\vec{x}:\varphi(\vec{x})>\lambda\}$. 找$\lambda$ 满足
$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})>\lambda)=\alpha. \tag{3.3}$$
- 若是连续型, 则这样的λ 往往存在.
  可根据水平α, 检验统计量 & 枢轴量来确定.
- 例: 正态总体, σ^2 已知. $H_0:\mu\leqslant\mu_0$.
  - 检验统计量$Z(\vec{x})=\frac{\bar{x}-\mu_0}{\sqrt{\sigma^2/n}}$, 否定域: $\mathcal{W}=\{\vec{x}:Z(\vec{x})>\lambda\}$.
  - 枢轴量: $Z(\vec{x},\mu):=\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$.
  - $Z(\vec{x})\leqslant Z(\vec{x},\mu)$, $\forall\mu\leqslant\mu_0$.
    故 $\sup_{\mu\leqslant\mu_0} P_\mu(Z(\vec{X})>\lambda)=P_{\mu_0}(Z(\vec{X})>\lambda)=P(Z(\vec{X},\mu)>\lambda)$.
- 若是离散型, 则满足(3.3) 的λ 不一定存在. 此时, 找λ 使得
$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})>\lambda)\leqslant\alpha<\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})\geqslant\lambda). \tag{3.4}$$

- **双边.** $\mathcal{W}=\{\vec{x}:\varphi(\vec{x})<\lambda_1$ 或 $\varphi(\vec{x})>\lambda_2\}$.
  找λ_1和λ_2 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})<\lambda_1)=\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})>\lambda_2)=\frac{\alpha}{2}. \quad (3.5, 3.6)$$

- 若是连续型, 这样的λ_1 和λ_2 往往存在.
- 若是离散型, 则不一定存在, 这时找λ_1 和λ_2 使得

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})<\lambda_1)\leqslant\frac{\alpha}{2}<\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})\leqslant\lambda_1), \quad (3.7)$$

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})>\lambda_2)\leqslant\frac{\alpha}{2}<\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{X})\geqslant\lambda_2). \quad (3.8)$$

- 根据检验水平确定临界值从而获得否定域的方法, 称为**临界值方法**.

例3.1. 设$X \sim N(\mu,\sigma^2)$, σ^2 未知. $H_0:\mu\leqslant\mu_0$.

- $\theta=(\mu,\sigma^2)$, $\Theta=\mathbb{R}\times\mathbb{R}_+$, $\Theta_0=(-\infty,\mu_0]\times\mathbb{R}_+$.
- $H_0:\theta\in\Theta_0$, 即$H_0:\mu\leqslant\mu_0 \longleftrightarrow H_a:\mu>\mu_0$.
- 取检验统计量$\varphi(\vec{X})=T(\vec{X}):=\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}}$.
- 直观: μ 越小, $\bar{X}$ 越小, 从而$T(\vec{X})$ 越小.
  因此, $T(\vec{x})$ 越大, 数据越倾向于否定H_0.
- 从而, 取单边否定域:
  $$\mathcal{W}=\{\vec{x}:\varphi(\vec{x})>\lambda\}=\{\vec{x}:T(\vec{x})>\lambda\}.$$
- 临界值$\lambda=t_{1-\alpha}(n-1)$: $\sup_{\theta\in\Theta_0}P_\theta(\varphi(\vec{X})>\lambda)=\alpha$.
  $T(\vec{X})\overset{H_0}{\leqslant}T(\vec{X},\mu)=\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$, (等号在$\mu=\mu_0$ 处达到),
  $$P_\theta\left(T(\vec{X})>\lambda\right)\overset{H_0}{\leqslant}P_\theta\left(T(\vec{X})>\lambda\right)=P(T_{n-1}>\lambda).$$

单边.

- 例3.1. $X \sim \mathrm{N}(\mu, \sigma^2)$, σ^2 已知. 希望证明μ 较大.
  取$H_0 : \mu \leqslant \mu_0$. 若$H_0$ 为真, 则$\varphi(\vec{X}) = \bar{X}$ 比较小.
- 否定域形如$\mathcal{W} = \mathcal{W}_\lambda = \{\vec{x} : \varphi(\vec{x}) > \lambda\}$. 指望数据否定$H_0$.
- 根据α 找临界值. 设$\exists!\lambda = \lambda(\alpha)$ 使得
$$\alpha = \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{X}) > \lambda \right).$$
- 定义3.2. 称如下定义的$p(\vec{x})$ 为样本值$\vec{x}$ 的p **值**.
$$p(\vec{x}) := \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{X}) \geqslant \varphi(\vec{x}) \right).$$
- 注: $p(\vec{x}) \in [0, 1]$; $p(\vec{X})$ 是统计量.
- **直观**: $p(\vec{x})$ 是能拒绝H_0 的最小α, 否定H_0 的强烈程度.
  - 若取$\alpha < p(\vec{x})$, 对则应的$\lambda > \varphi(\vec{x})$, 接受$H_0$.
  - 若取$\alpha > p(\vec{x})$, 对则应的$\lambda < \varphi(\vec{x})$, 否定$H_0$.

引理3.1. 设$\forall\alpha\in(0,1)$, $\exists!\lambda=\lambda(\alpha)$ 满足:

$$\alpha=\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})>\lambda\right).$$

则$\varphi(\vec{x})>\lambda(\alpha)$ 当且仅当$p(\vec{x}):=\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})\geqslant\varphi(\vec{x})\right)<\alpha$.

- 注: p 值方法即为否定域取为$\mathcal{W}=\{\vec{x}:p(\vec{x})<\alpha\}$.
- 注: α 为精确检验水平: $\mathcal{W}=\{\vec{x}:p(\vec{x})<\alpha\}$,
$$\sup_{\theta\in\Theta_0}P_\theta(\vec{X}\in\mathcal{W})=\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})>\lambda(\alpha)\right)=\alpha.$$
- 证: 设$p(\vec{x})<\alpha$. 则$\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})>\varphi(\vec{x})\right)\leqslant\star\star<\alpha$,

  故$\varphi(\vec{x})>\lambda$.
- 反过来, 设$\varphi(\vec{x})>\lambda=\lambda(\alpha)$. 则$\exists\varepsilon>0$ 使$\varphi(\vec{x})-\varepsilon>\lambda$. 于是
$$\star\star\leqslant\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})>\varphi(\vec{x})-\varepsilon\right)\leqslant\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})>\lambda\right)=\alpha.$$
- $\lambda=\lambda(\alpha)$ 是唯一的, 因此 “$=$” 不成立. 即$p(\vec{x})<\alpha$.

引理3.2. 设$\forall\alpha\in(0,1)$, 存在λ (未必唯一)满足:

$$\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X})>\lambda\right)\leqslant\alpha<\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X})\geqslant\lambda\right).$$

则 $\varphi(\vec{x})>\lambda$ 当且仅当$p(\vec{x}):=\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X})\geqslant\varphi(\vec{x})\right)\leqslant\alpha$.

- 证: 一方面, 若$\varphi(\vec{x})>\lambda$. 则
$$p(\vec{x})\leqslant\sup_{\theta\in\Theta_0} P_\theta(\varphi(\vec{x})>\lambda)\leqslant\alpha.$$
- 另一方面, 若$\varphi(\vec{x})\leqslant\lambda$, 则
$$p(\vec{x})\geqslant\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X})\geqslant\lambda\right)>\alpha.$$
- 注: 否定域$\mathcal{W}=\{\vec{x}:\star\}=\{\vec{x}:p(\vec{x})\leqslant\alpha\}$. 精确检验水平$\leqslant\alpha$.

例3.2. 设$X \sim N(\mu,\sigma^2)$, σ^2 未知. 检验$H_0: \mu \leqslant \mu_0$.

- 在H_0 下,

$$T(\vec{X}) = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \overset{H_0}{\leqslant} \frac{\bar{X}-\mu}{\sqrt{S^2/n}} = T(\vec{X},\mu) \sim t(n-1).$$

- 否定域: 取$\lambda = t_{1-\alpha}(n-1)$,

$$\mathcal{W} = \{\vec{x} : T(\vec{x}) > \lambda\} = \{\vec{x} : p(\vec{x}) < \alpha\}.$$

- 计算p 值:

$$p(\vec{x}) = \sup_{\theta\in\Theta_0} P_\theta\Big(T(\vec{X}) \geqslant T(\vec{x})\Big) = P\big(T_{n-1} \geqslant T(\vec{x})\big).$$

- 例, $\mu_0 = 25$, $n = 64$, $\bar{x} = 25.9$, $s^2 = 17.3$.

$$\varphi(\vec{x}) = 1.731, \quad p(\vec{x}) = P(T_{63} \geqslant 1.731) = 0.044.$$

- 若水平$\alpha = 0.05$, 则拒绝H_0. 若水平$\alpha = 0.03$, 则接受H_0.

双边.

- 否定域形如$\mathcal{W} = \{\vec{x}: \ \varphi(\vec{x}) < \lambda_1 \text{ 或 } \varphi(\vec{x}) > \lambda_2\}$.
- (任)取$\lambda_0 \in [\lambda_1, \lambda_2)$. 定义3.3. 称如下定义的$p(\vec{x})$ 为$\vec{x}$ 的p 值.

$$p(\vec{x}) = \begin{cases} \min\left\{2 \sup\limits_{\theta\in\Theta_0} P_\theta\Big(\varphi(\vec{X}) \leqslant \varphi(\vec{x})\Big), \ 1\right\}, & \varphi(\vec{x}) \leqslant \lambda_0; \\ \min\left\{2 \sup\limits_{\theta\in\Theta_0} P_\theta\Big(\varphi(\vec{X}) \geqslant \varphi(\vec{x})\Big), \ 1\right\}, & \varphi(\vec{x}) > \lambda_0. \end{cases}$$

引理3.3. 设$\exists \alpha \in (0,1)$, $\exists ! \lambda_1, \lambda_2$ 满足下式, 则

$$\sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{x}) < \lambda_1 \right) = \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{x}) > \lambda_2 \right) = \frac{\alpha}{2}.$$

则 “$\varphi(\vec{x}) < \lambda_1$ 或$\varphi(\vec{x}) > \lambda_2$” 当且仅当$p(\vec{x}) < \alpha$.

- 证: 若$\varphi(\vec{x}) \leqslant \lambda_0$, 则$\varphi(\vec{x}) < \lambda_1$ 当且仅当$p(\vec{x}) < \alpha$: λ_1 唯一,

$$p(\vec{x}) \leqslant 2 \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{X}) \leqslant \varphi(\vec{x}) \right) \leqslant 2 \sup_{\theta \in \Theta_0} P_\theta (\varphi(\vec{x}) < \lambda_1 - \varepsilon) < \alpha.$$

- 若$\varphi(\vec{x}) > \lambda_0$, 则$\varphi(\vec{x}) > \lambda_2$ 当且仅当$p(\vec{x}) < \alpha$: λ_2 唯一,

$$p(\vec{x}) \leqslant 2 \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{X}) \geqslant \varphi(\vec{x}) \right) \leqslant 2 \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(\vec{x}) > \lambda_2 + \varepsilon \right) < \alpha.$$

引理3.4. 设λ_1 和λ_2 (未必唯一)满足:

$$\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X}) < \lambda_1\right) \leqslant \frac{\alpha}{2} < \sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X}) \leqslant \lambda_1\right),$$

$$\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X}) > \lambda_2\right) \leqslant \frac{\alpha}{2} < \sup_{\theta\in\Theta_0} P_\theta\left(\varphi(\vec{X}) \geqslant \lambda_2\right).$$

则 "$\varphi(x) < \lambda_1$ 或$\varphi(x) > \lambda_2$" 当且仅当$p(\vec{x}) \leqslant \alpha$.

- 证明略.

例3.3. 设$X \sim N(\mu,\sigma^2)$. 检验$H_0:\sigma^2=\sigma_0^2$.

- 检验统计量、枢轴量$K_{n-1}\sim\chi^2(n-1)$:

$$\varphi(\vec{X})=K(\vec{X})=\frac{1}{\sigma_0^2}\sum_{i=1}^n(X_i-\bar{X})^2\overset{H_0}{=}\frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\bar{X})^2=K(\vec{X},\theta).$$

- 双边否定域: $\mathcal{W}=\{\vec{x}:\varphi(\vec{x})<\lambda_1\ 或\varphi(\vec{x})>\lambda_2\}$.
- 直观: $K_{n-1}\approx n-1$ (此即λ_0). $\lambda_1<n-1<\lambda_2$.
- $\varphi(\vec{x})\leqslant n-1$ 当且仅当$s^2\leqslant\sigma_0^2$:

$$\sup_{\theta\in\Theta_0}P_\theta\left(\varphi(\vec{X})\leqslant\varphi(\vec{x})\right)=P\left(K_{n-1}\leqslant\varphi(\vec{x})\right).$$

- p 值$p(\vec{x}):=\begin{cases}\min\{2P\left(K_{n-1}\leqslant\varphi(\vec{x})\right),1\}, & s^2\leqslant\sigma_0^2;\\ \min\{2P\left(K_{n-1}\geqslant\varphi(\vec{x})\right),1\}, & s^2>\sigma_0^2.\end{cases}$
- 例2.4. 铜丝折断力$X\sim N(\mu,\sigma^2)$. $n=10$, 数据: $\cdots\cdots$.

  检验$H_0:\sigma^2=\sigma_0^2=64$. $\varphi(\vec{x})=\frac{1}{\sigma_0^2}\sum_{i=1}^{10}(x_1-\bar{x})^2=10.65$.
- p 值为0.60: $s^2=75.7>64$, $P\left(K_9\geqslant\varphi(\vec{x})\right)=0.30$.
- 只要$\alpha<0.60$ (一般地, $\alpha<0.5$), H_0 都与$\vec{x}$ 相容.

p **值方法的总结.**

- 单边否定域:
  $\mathcal{W} = \{\vec{x} : p(\vec{x}) < \alpha\}$ (引理3.1),
  $\mathcal{W} = \{\vec{x} : p(\vec{x}) \leqslant \alpha\}$ (引理3.2).
- 双边否定域:
  $\mathcal{W} = \{\vec{x} : p(\vec{x}) < \alpha\}$ (引理3.3),
  $\mathcal{W} = \{\vec{x} : p(\vec{x}) \leqslant \alpha\}$ (引理3.4).
- (精确)检验水平均$\leqslant \alpha$.
- 优点: 计算$p(\vec{x})$.
  对任意**预先给定的** α 适用, 不用对每个α 都计算λ 再判断$\star$.
  $p(\vec{x})$ 给出拒绝H_0 的强烈程度.

假设检验与置信区间的联系

- 设总体的分布函数为$F(x,\theta)$, 其中θ 为未知参数, 范围为Θ.
  考虑$H_0: \theta = \theta_0 \longleftrightarrow H_a: \theta \neq \theta_0$.
- 记$\mathcal{A} = \mathcal{A}(\theta_0) = \mathcal{W}^c$,

$$P_{\theta_0}(\vec{X} \in \mathcal{W}) \leqslant \alpha, \quad P_{\theta_0}(\vec{X} \in \mathcal{A}(\theta_0)) \geqslant 1-\alpha.$$

- $\vec{x} \in \mathcal{A}(\theta_0)$ 当且仅当$\theta_0 \in S(\vec{x})$. 其中,

$$S(\vec{x}) := \{\theta : \vec{x} \in \mathcal{A}(\theta)\}.$$

- $P_{\theta_0}\left(\theta_0 \in S(\vec{X})\right) \geqslant 1-\alpha, \quad \forall \theta_0 \in \Theta.$
- $S(\vec{x})$ 是θ 的置信水平为$1-\alpha$ 的置信“区间”.
- 注: 反过来, 可用置信“区间”造$\mathcal{W}$.

例3.4. 设$X \sim N(\mu, 1)$. 检验$H_0 : \mu = \mu_0$.

- 参数: $\theta = \mu$, 范围: $\mathbb{R}$. $\theta_0 = \mu_0$.
- 接受域: 精确水平为α,

$$\mathcal{A}(\theta_0) = \left\{\vec{x} : |\bar{x} - \theta_0| \leqslant \lambda = z_{1-\frac{\alpha}{2}}/\sqrt{n}.\right\}$$

- $P_{\mu_0}(|\bar{X} - \mu_0| \leqslant \lambda) = 1 - \alpha$.
- $\forall \mu$, $P_{\mu}(|\bar{X} - \mu| \leqslant \lambda) = 1 - \alpha$.
- 置信区间: $[\bar{X} - \lambda,\ \bar{X} + \lambda]$.
- 否定域/接受域: 接受H_0 当且仅当$\mu_0 \in [\bar{X} - \lambda,\ \bar{X} + \lambda]$.

§6.4 两个正态总体的假设检验

设$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且$X, Y$ 相互独立.

(1) σ_1^2, σ_2^2 未知, 但知道$\sigma_1^2 = \sigma_2^2$. $H_0 : \mu_1 = \mu_2$, 或$H_0 : \mu_1 \leqslant \mu_2$.

(2) μ_1, μ_2 未知. $H_0 : \sigma_1^2 = \sigma_2^2$, 或$H_0 : \sigma_1^2 \leqslant \sigma_2^2$.

(3) σ_1^2, σ_2^2 未知, 但知道$\sigma_1^2 \neq \sigma_2^2$. $H_0 : \mu_1 = \mu_2$, 或$H_0 : \mu_1 \leqslant \mu_2$.

(1) σ_1^2, σ_2^2 未知, $\sigma_1^2=\sigma_2^2$. $H_0: \mu_1=\mu_2$ 或$H_0: \mu_1 \leqslant \mu_2$.

- 数据: $X_1, \cdots, X_{n_1}; Y_1, \cdots, Y_{n_2}$.
- $\bar{X}$, $\sum_{i=1}^{n_1}(X_i-\bar{X})^2$; $\bar{Y}$, $\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$ 相互独立,

$$\bar{X} \sim \mathrm{N}(\mu_1, \frac{1}{n_1}\sigma^2), \quad \bar{Y} \sim \mathrm{N}(\mu_2, \frac{1}{n_2}\sigma^2),$$
$$\frac{\star}{\sigma^2} \sim \chi^2(n_1-1), \quad \frac{\star}{\sigma^2} \sim \chi^2(n_2-1).$$

- 分子:

$$(\bar{X}-\bar{Y})-(\mu_1-\mu_2) \sim \mathrm{N}\left(0, \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2\right).$$

- 分母: $\frac{\star+\star}{\sigma^2} \sim \chi^2(n_1+n_2-2)$.
  σ^2 的无偏估计为$\hat{\sigma}^2=\frac{1}{n}(\star+\star)$.
- 重点: 分子与分母相互独立.

- $\frac{1}{\sigma}\left((\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right) \sim \mathrm{N}\left(0, \frac{1}{n_1}+\frac{1}{n_2}\right)$,
  $\frac{1}{\sigma^2}\left(\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2\right) \sim \chi^2(n_1+n_2-2)$,
  且以上两个随机变量相互独立.
- 检验统计量$T(\vec{X},\vec{Y})$ 与枢轴量$T(\vec{X},\vec{Y},\theta)$:

$$T(\vec{X},\vec{Y}) = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}},$$

$$T(\vec{X},\vec{Y};\theta) = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}} \sim t(n), \quad \hat{\sigma}^2 = \frac{1}{n}(\star+\star).$$

- 根号下:

$$\star\cdot\frac{1}{n}(\star+\star) = \frac{n_1+n_2}{n_1 n_2}\frac{1}{n_1+n_2-2}(\star+\star) = \frac{n_1+n_2}{n_1 n_2(n_1+n_2-2)}(\star+\star).$$

- $n_1 = n_2$ 时, $\frac{n_1+n_2}{n_1 n_2(n_1+n_2-2)} = \frac{2n}{n^2\cdot 2(n-1)} = \frac{1}{n(n-1)}$.

- 检验统计量$T(\vec{X},\vec{Y})$ 与枢轴量$T(\vec{X},\vec{Y},\theta)$: $n=n_1+n_2-2$,

$$T(\vec{X},\vec{Y})=\frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}},$$

$$T(\vec{X},\vec{Y};\theta)=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}}\sim t(n),\quad \hat{\sigma}^2=\frac{1}{n}(\star+\star).$$

- **双边.** 若$H_0:\mu_1=\mu_2$ 成立, 则$T(\vec{X},\vec{Y})=T(\vec{X},\vec{Y};\theta)$.
- 查$t(n_1+n_2-2)$ 的表得临界值λ: $\lambda=t_{1-\alpha/2}(n)$.
- 否定域:

$$\mathcal{W}=\{(\vec{x},\vec{y}):|T(\vec{x},\vec{y})|>\lambda\}.$$

- 该检验法称为**两样本t检验**, 也称为**平均数的显著性鉴定**.
- 拒绝$H_0:\mu_1=\mu_2$, 则一般称(在α 水平下)两个总体的平均数**有显著(性)差异**.

- 检验统计量$T(\vec{X},\vec{Y})$ 与枢轴量$T(\vec{X},\vec{Y},\theta)$: $n = n_1 + n_2 - 2$,

$$T(\vec{X},\vec{Y}) = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}},$$

$$T(\vec{X},\vec{Y};\theta) = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{\sigma}^2}} \sim t(n), \quad \hat{\sigma}^2 = \frac{1}{n}(\star+\star).$$

- **单边.** 若$H_0: \mu_1 \leqslant \mu_2$ 成立, 则$T(\vec{X},\vec{Y}) \leqslant T(\vec{X},\vec{Y};\theta)$.
- 否定域:

$$\mathcal{W} = \{(\vec{x},\vec{y}) : T(\vec{x},\vec{y}) > \lambda\}.$$

- $\sup_{\theta\in\Theta_0} P_\theta(\vec{X}\in\mathcal{W}) = P(T_n > \lambda)$, 故取$\lambda = t_{1-\alpha}(n)$.

例4.1(例1.4). 漂布工艺中温度对强力X, Y 的影响.

$$70, X:\ 20.5,\ 18.8,\ 19.8,\ 20.9,\ 21.5,\ 19.5,\ 21.0,\ 21.2$$

$$80, Y:\ 17.7,\ 20.3,\ 20.0,\ 18.8,\ 19.0,\ 20.1,\ 20.2,\ 19.1$$

$\sigma_1^2=\sigma_2^2=\sigma^2$. $H_0: \mu_1=\mu_2 \longleftrightarrow H_a: \mu_1\neq\mu_2$.

- 代数据: $n_1=n_2=8$, $\bar{x}=20.4$, $\bar{y}=19.4$,
  $\sum_{i=1}^n(x_i-\bar{x})^2=6.20$, $\sum_{i=1}^n(y_i-\bar{y})^2=5.80$;
  $\hat{\sigma}^2=\frac{1}{2\times 8-2}(6.20+5.80)=0.8571$,
  $T(\vec{x},\vec{y})=\frac{20.4-19.4}{\sqrt{2*0.8571/8}}=2.16$.
- 或者, 查t 分布表, 自由度是$n_1+n_2-2=14$, 取$\alpha=0.05$, 得$\lambda=2.145$. **下结论**: $|T(\vec{x},\vec{y})|=2.16>\lambda$, 拒绝$H_0$.
- 或者, 计算p 值: $p(\vec{x})=P(|K_{2n-2}|\geqslant|2.16|)=0.0486$, **下结论**: $p(\vec{x})<0.05=\alpha$, 拒绝H_0.
- 注: 可令$H_0: \mu_1\leqslant\mu_2$. **下结论**: 否定H_0.

例4.2. 研究口服避孕药对妇女血压影响, X, Y: 收缩压的值.

$n_1 = 8$ 人服药, $\bar{x} = 132.86$, $s_X = 15.35$;

$n_2 = 21$ 人未服药, $\bar{y} = 127.44$, $s_Y = 18.23$.

假定$\sigma_1^2 = \sigma_2^2 = \sigma^2$. 检验$H_0 : \mu_1 = \mu_2$.

- 计算:
  $\hat{\sigma}^2 = \frac{1}{8+21-2}((8-1) \times 15.35^2 + (21-1) \times 18.23^2) = 294.95.$
- $T(\vec{x}, \vec{y}) = \frac{132.86-127.44}{\sqrt{\left(\frac{1}{8}+\frac{1}{21}\right)294.95}} = 0.760.$
- 查t 分布表: 自由度$n = 27$, $\alpha = 0.05$, 得$\lambda = 2.052$.
- 下结论: $|T(\vec{x}, \vec{y})| < \lambda$, H_0 相容, 两平均值无显著差异.

(2) μ_1, μ_2 未知. $H_0: \sigma_1^2 = \sigma_2^2$, 或$H_0: \sigma_1^2 \leqslant \sigma_2^2$.

- σ_1^2 和σ_2^2 的无偏估计:

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$

- 检验统计量: $F(\vec{X}, \vec{Y}) = S_1^2/S_2^2$.
- 枢轴量: $F(\vec{X}, \vec{Y}; \theta) = \frac{S_1^2}{\sigma_1^2}/(\frac{S_2^2}{\sigma_2^2})$.
- 第一自由度为n, 第二自由度为m 的F 分布, 记为$F(n, m)$: 指$\frac{1}{n}K_n/(\frac{1}{m}K_m)$ 的分布, 其中$K_n \sim \chi^2(n)$, $K_m \sim \chi^2(m)$, 且相互独立. 其密度如下(定义4.1, 用§4.2的方法推导): $x > 0$,

$$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}.$$

- $F(\vec{X},\vec{Y})=S_1^2/S_2^2$,
  $F(\vec{X},\vec{Y};\theta)=\frac{S_1^2}{\sigma_1^2}/(\frac{S_2^2}{\sigma_2^2})=\frac{\sigma_2^2}{\sigma_1^2}F(\vec{X},\vec{Y})\sim F(n_1-1,n_2-1)$.
  $K_n\sim\chi^2(n)$, $K_m\sim\chi^2(m)$, 且独立. $F_{n,m}=\frac{1}{n}K_n/(\frac{1}{m}K_m)$.
- **双边.** 若$H_0:\sigma_1^2=\sigma_2^2$ 成立, 则$F(\vec{X},\vec{Y})=F(\vec{X},\vec{Y};\theta)$.
  否定域: $\mathcal{W}=\{(\vec{x},\vec{y}):F(\vec{x},\vec{y})<\lambda_1$ 或 $>\lambda_2\}$.
  其中, $\lambda_1=F_{\alpha/2}(n_1-1,n_2-1)$, $\lambda_2=F_{1-\alpha/2}(n_1-1,n_2-1)$.
- **单边.** 若$H_0:\sigma_1^2\leqslant\sigma_2^2$ 成立, 则$F(\vec{X},\vec{Y})\leqslant F(\vec{X},\vec{Y};\theta)$.
  否定域: $\mathcal{W}=\{(\vec{x},\vec{y}):F(\vec{x},\vec{y})>\lambda\}$,
  故取$\lambda=F_{1-\alpha}(n_1-1,n_2-1)$.
  (注: $\max\limits_{\theta\in\Theta_0}P_\theta\big(F(\vec{X},\vec{Y})>\lambda\big)=P(F_{n_1-1,n_2-1}>\lambda)$.)
- 附表只直接给出右临界值λ_2 和λ.
- $\lambda_1=1/F_{1-\alpha/2}(n_2-1,n_1-1)$:
  $$\frac{1}{F_{n,m}}\sim F(m,n) \text{ 且 } P(F_{n,m}<\lambda_1)=P\left(\frac{1}{F_{n,m}}>\frac{1}{\lambda_1}\right)\Rightarrow\frac{1}{\lambda}=\star.$$

例4.1(续). 考虑70 度与80 度下强力的方差的比较. $H_0 : \sigma_1^2 = \sigma_2^2$.

- 代数据: $n_1 = n_2 = 8$, $s_1^2 = 6.20/7$, $s_2^2 = 5.80/7$,

$$F(\vec{x}, \vec{y}) = \frac{s_1^2}{s_2^2} = \frac{6.20}{5.80} = 1.07.$$

- 查$F(7, 7)$ 表: $\alpha = 0.05$, 得$\lambda_2 = 4.99$.
- 查$F(7, 7)$ 表: $\alpha = 0.05$, 的$1/\lambda_1 = 4.99$, 故$\lambda_1 = 0.200$.
- 下结论: $\lambda_1 < F(\vec{x}, \vec{y}) < \lambda_2$, 数据与$H_0$ 相容.
- 解释: 没有明显的证明表明H_0 不成立, 故不能否认$\sigma_1^2 = \sigma_2^2$, 称σ_1^2 与σ_2^2 **无显著差异**.
- 进一步, 可假设$\sigma_1^2 = \sigma_2^2$, 并进行(1) 中的均值检验.

例4.4.(例4.2续). 服药/不服药$H_0 : \sigma_1^2 = \sigma_2^2$.

- 代数据: $n_1 = 8,\ n_2 = 21,\ s_1^2 = (15.35)^2,\ s_2^2 = (18.23)^2$,

$$F(\vec{x}, \vec{y}) = \frac{s_1^2}{s_2^2} = 0.709.$$

- 查$F(7, 20)$ 的表: 水平$= \frac{\alpha}{2} = 0.025$, 得$\lambda_2 = 3.01$,
- 查$F(20, 7)$ 的表: 水平$= 0.025$. 表中没有20.
- 解决方案:
  - 用最近的代替: 自由度$(24, 7)$, 令$\lambda_1 = 1/4.42 = 0.226$.
  - 线性插值: 自由度$(12, 7)$, $(24, 7)$ 最接近,

$$1/\lambda_1 = 4.67 + \frac{4.42-4.67}{24-12}(20-12) = 4.50.$$

  - 注: 若两个自由度都不在表格中, 则这种方法要两次插值, 精度也只是略有提高.

- 下结论: $\lambda_1 < F(\vec{x}, \vec{y}) < \lambda_2$, 数据与$H_0$ 相容.

例4.5. 滚珠直径X, Y. $n_1 = 8$, $n_2 = 9$, 数据:

$$X : 15.0, 14.5, 15.2, 15.5, 14.8, 15.1, 15.2, 14.8$$

$$Y : 15.2, 15.0, 14.8, 15.2, 15.0, 15.0, 14.8, 15.1, 14.8$$

问: 乙车床产品直径的方差是否比甲车床的小?

- 想得到结论$\sigma_2^2 < \sigma_1^2$, 把它作为H_a. $H_0 : \sigma_1^2 \leqslant \sigma_2^2$.
- 代数据: $n_1 = 8$, $n_2 = 9$, $s_1^2 = 0.09554$, $s_2^2 = 0.02611$,
  $F(\vec{x}, \vec{y}) = 3.66$.
- 查$F(7, 8)$ 的表: $\alpha = 0.05$, 得$\lambda = 3.50$.
- 下结论: $F(\vec{x}, \vec{y}) = 3.66 > 3.50$, 否定$H_0$,
  在0.05水平下, 认为乙车床产品直径的方差显著小于甲车床产品直径的方差.

例4.6. 赈灾捐赠的男女差异. 男: X, $n_1 = 25$, $\bar{x} = 12.40$, $s_X = 2.50$; 女: Y, $n = 25$, $\bar{y} = 8.90$, $s_Y = 1.34$. 问: 男士捐赠额的方差是否大于女士捐赠额的方差?

- 令$H_0 : \sigma_1^2 \leqslant \sigma_2^2$.
- 计算$F(\vec{x}, \vec{y}) = (2.50)^2/(1.34)^2 = 3.48$.
- 查表, $F(24, 24)$, $\alpha = 0.01$, 得$\lambda = 2.66$.
- 下结论: $F(\vec{x}, \vec{y}) = 3.48 > 2.66$, 拒绝$H_0$.
  在0.01水平下, 认为男士捐赠额的方差显著大于女士捐赠额的方差.

(3) σ_1^2, σ_2^2 未知, $\sigma_1^2 \neq \sigma_2^2$. $H_0: \mu_1 = \mu_2$, 或$H_0: \mu_1 \leqslant \mu_2$.

- 该是著名的Behrens–Fisher 问题.
- 数据: $X_1, \cdots, X_{n_1}; Y_1, \cdots, Y_{n_2}$.
- $\bar{X}$, $\sum_{i=1}^{n_1}(X_i - \bar{X})^2$; $\bar{Y}$, $\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ 相互独立,

$$\bar{X} \sim \mathrm{N}(\mu_1, \frac{1}{n_1}\sigma^2), \quad \bar{Y} \sim \mathrm{N}(\mu_2, \frac{1}{n_2}\sigma^2),$$

$$\frac{\star}{\sigma^2} \sim \chi^2(n_1 - 1), \quad \frac{\star}{\sigma^2} \sim \chi^2(n_2 - 1).$$

- 区别1、分子的方差.

$$(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2) \sim \mathrm{N}\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

- 于是, 若$\mu_1 = \mu_2$, 则$\frac{\bar{X}-\bar{Y}}{\star} \sim \mathrm{N}(0, 1)$.
- t 检验的思想: σ^2 未知时, 用其无偏估计S^2 代替.

- **区别2、分母的表达式**. 分别用$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2$ 和$S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$ 代替σ_1^2 和σ_2^2.
- 检验统计量与“枢轴量”：

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}},\quad T(\theta)=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}.$$

- 注: $T(\theta)$ 的精确分布复杂, 且依赖于σ_1^2/σ_2^2 的值.
- 可以证明: $T(\theta)$ **近似**服从t(m),
  其中, m 为最接近如下m^* 的整数.

$$m^*=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}. \tag{4.8}$$

- 检验统计量:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

- **双边.** $H_0 : \mu_1 = \mu_2$.
  否定域: $\mathcal{W} = \{(\vec{x}, \vec{y}) : |T(\vec{x}, \vec{y})| > \lambda\}$, 其中, $\lambda = t_{1-\alpha/2}(m)$.
- **单边.** $H_0 : \mu_1 \leqslant \mu_2$.
  否定域: $\mathcal{W} = \{(\vec{x}, \vec{y}) : T(\vec{x}, \vec{y}) > \lambda\}$, 其中, $\lambda = t_{1-\alpha}(m)$.

例4.7. 研究父亲患心脏病的家庭中, 子女的胆固醇水平是否偏高的问题. 数据如下, 问: μ_1 与μ_2 是否有显著差异. ($\alpha = 0.05$.)

X (父患病), n =100, $\bar{x} = 207.3$, s_X =35.6;

Y (父不患病), $n_2 = 74$, $\bar{y} = 193.4$, s_Y =17.3.

- 先检验$H_0: \sigma_1^2 = \sigma_2^2$. 计算: $F = 4.23$, 查表: $\lambda_1 = 0.6548$, $\lambda_2 = 1.5491$, $F = 4.23 > \lambda_2$, 拒绝H_0. (若接受H_0, 则用(1).)
- 计算统计量: $T = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = 3.40$. 计算$m^* = 151.4$, 得到$m = 151$. (没有t(151), 使用t(120) 的临界值作为近似.)
- **双边.** $H_0: \mu_1 = \mu_2$, 查表得$\lambda = t_{1-\alpha/2}(120) = 1.980$.
  下结论: $|T| = 3.40 > \lambda$, 拒绝H_0.
- **单边.** $H_0: \mu_1 \leqslant \mu_2$, 查表得$\lambda = t_{1-\alpha}(120) = 1.658$.
  下结论: $T = 3.40 > \lambda$, 拒绝H_0.
  认为父亲死于心脏病的孩子的胆固醇水平显著地更高.

t分布与F分布的关系:

设$X \sim \mathrm{t}(n)$, 则$Y = X^2 \sim \mathrm{F}(1, n)$.

- 证: 设$Z_0, Z_1, \cdots, Z_n$ 相互独立, 都$\sim \mathrm{N}(0, 1)$.
  则$X \overset{d}{=} \dfrac{Z_0}{\sqrt{\frac{1}{n}(Z_1^2+\cdots+Z_n^2)}}$. 于是$X^2 \overset{d}{=} \dfrac{Z_0^2}{\frac{1}{n}(Z_1^2+\cdots+Z_n^2)} \sim \mathrm{F}(1, n)$.
- 另证: $F_Y(y) = P(|X| \leqslant \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$, 故

$$\begin{aligned} p_Y(y) =& F_Y'(y) = p_X(\sqrt{y})\frac{1}{2\sqrt{y}} + p_X(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ =& \frac{1}{2\sqrt{y}}[p_X(\sqrt{y}) + p_X(-\sqrt{y})], \quad y > 0. \end{aligned}$$

将$t(n)$ 的密度代入$p_X(x)$, 便知$p_Y(y)$ 是$\mathrm{F}(1, n)$ 的密度.

§6.5 比率的假设检验

(1) 单总体($X \sim B(1,p)$), 正态近似法(大样本);

(2) 单总体, p值检验法(小样本);

(3) 两总体($X \sim B(1,p_1)$, $Y \sim B(1,p_2)$), 正态近似法(大样本).

(4) 两总体: Fisher精确估计法(不要求大样本).

- 适用范围: n 很大, 一般要求成功和失败个数都超过5个.
- 统计量: $S(\vec{X}) = S_n = X_1 + \cdots + X_n \sim B(n,p)$, $\hat{p} = S_n/n$.
- **双边.** $H_0 : p = p_0$. 在H_0 下, 统计量

$$\zeta(\vec{X}) = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \text{ 近似地} \sim \mathrm{N}(0,1).$$

否定域: $\mathcal{W} = \{\vec{x} : |\zeta(\vec{x})| \geqslant z_{1-\alpha/2}\}$.

- **单边.** 统计量与"枢轴量" (近似$\sim$ N(0,1)).

$$\eta(\vec{x}) = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}, \quad \xi := \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}}.$$

- $H_0 : p \leqslant p_0$. 否定域: $\mathcal{W} = \{\vec{x} : \eta(\vec{x}) \geqslant z_{1-\alpha}\}$.
- $H_0 : p \geqslant p_0$. 否定域: $\mathcal{W} = \{\vec{x} : \eta(\vec{x}) \leqslant -z_{1-\alpha}\}$.

例. 收藏家一年中购入了98 幅名画, 经鉴定其中26 幅是赝品. 问, 该收藏家的鉴定准确率$\geqslant 75\%$?

- $\hat{p} = \frac{98-26}{98} = 0.7347 \leqslant 75\%$.
- 按题意选择$H_0 : p \leqslant 0.75$(肯定不能否定该H_0), 或者, (更应该)选择$H_0 : p \geqslant 0.75$.
- 否定域: $\alpha = 0.05$,

$$\mathcal{W}_+ = \left\{ \vec{x} : \eta(\vec{x}) = \frac{\hat{p} - 0.75}{\sqrt{\hat{p}(1-\hat{p})/n}} > 1.645 \right\}$$

$$\mathcal{W}_- = \{ \vec{x} : \eta(\vec{x}) < -1.645 \} .$$

- 计算得到$\eta(\vec{x}) = -0.3432$. 下结论: 接受H_0 和H_0. (检验法也有局限性!)

(2) 单总体, p值检验法(小样本)

- **单边.** $H_0 : p \leqslant p_0 \longleftrightarrow H_a : p > p_0$.
- 统计量: $S(\vec{X}) = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$.
  否定域: $\mathcal{W} = \{\vec{x} : S(\vec{x}) \geqslant c\}$.
- 临界值c 为满足$\sup\limits_{p \leqslant p_0} P_p(S \geqslant k) \leqslant \alpha$ 的最小整数k.
- 固定k. $f(p) := P_p(S \geqslant k)$ 关于p 单调上升:

$$f(p) = \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i}$$
$$= \frac{n!}{(k-1)!(n-k)!} \int_0^p u^{k-1}(1-u)^{n-k} du.$$

- 因此, 临界值c 为满足下式的最小整数k.

$$P_{p_0}(S \geqslant k) \leqslant \alpha. \qquad (5.2)$$

- 否定域: $\mathcal{W}=\{\vec{x}: S(\vec{x})\geqslant c\}$, 其中$S=S(\vec{X})=\sum_{i=1}^{n}X_i \sim \mathrm{B}(n,p)$, c 为满足$P_{p_0}(S\geqslant k)\leqslant\alpha$ 的最小整数k_0.
- 求临界值c比较麻烦, 我们用p值来表示否定域. 记$S(\vec{x})=s_0$

$$p(\vec{x}) := \sup_{p\leqslant p_0} P_p(S\geqslant s_0) = P_{p_0}(S\geqslant s_0)$$
$$= \sum_{i=s_0}^{n} C_n^i p_0^i (1-p_0)^{n-i}.$$

- 当且仅当$\star\leqslant\alpha$ 时, 拒绝H_0.
- 将s_0 视为固定值, $f(p)=\sum_{i=s_0}^{n} C_n^i p^i(1-p)^{n-i}$ 关于p 严格单调递增. 存在唯一的$p_\alpha(s_0)$ 使得$f(p_\alpha(s_0))=\alpha$.
- 补充定义: $s_0=0$ 时, $f(p)\equiv 1$. 规定$p_\alpha(0)=0$.
- 因此, 当且仅当$p_0\leqslant p_\alpha(s_0)$ 时, 拒绝H_0.

- 否定域: $\mathcal{W} = \{\vec{x} : p_0 \leqslant p_\alpha(s_0)\}$, 其中$s_0 = \sum_{i=1}^n x_i$.

  $p_\alpha(s_0)$ 为$f(p) := \sum_{i=s_0}^n C_n^i p^i (1-p)^{n-i} = \alpha$ 的(唯一)解.

- **可以证明**如下的式成立: 记$n_1 = 2(n-k+1)$, $n_2 = 2k$,

$$p_\alpha = p_\alpha(k) = \left\{1 + \frac{n-k+1}{k} F_{1-\alpha}(n_1, n_2)\right\}^{-1}.$$

- 于是, $p_\alpha(S_0)$ 有如下表达式:

$$\left\{1 + \frac{n-S_0+1}{S_0} F_{1-\alpha}(2(n-S_0+1), 2S_0)\right\}^{-1}. \tag{5.5}$$

记$n_1 = 2(n-k+1)$, $n_2 = 2k$, 往证$f(p) := \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} = \alpha$ 的(唯一)解有如下表达式:

$$p_\alpha = p_\alpha(k) = \left\{1 + \frac{n-k+1}{k} F_{1-\alpha}(n_1, n_2)\right\}^{-1}.$$

- 作变量替换: 令$u = (1 + \frac{n-k+1}{k}x)^{-1} = (1 + \frac{n_1}{n_2}x)^{-1}$, 即$x = x(u) = \frac{k}{n-k+1}(\frac{1}{u} - 1)$. 则$x$ 关于u 严格单调下降. 当$u \to 0$ 时, $x(u) \to \infty$.

$$\begin{aligned}
f(p) &= \frac{n!}{(k-1)!(n-k)!}\int_0^p u^{k-1}(1-u)^{n-k}du \qquad \left(记\ C_1 = \frac{n!}{(k-1)!(n-k)!}\right)\\
&= C_1 \int_{\infty}^{x(p)} \left(\frac{1}{1+\frac{n_1}{n_2}x}\right)^{k-1} \left(\frac{\frac{n_1}{n_2}x}{1+\frac{n_1}{n_2}x}\right)^{n-k} d\frac{1}{1+\frac{n_1}{n_2}x}\\
&= C_1 \int_{x(p)}^{\infty} \frac{\left(\frac{n_1}{n_2}x\right)^{n-k}}{\left(1+\frac{n_1}{n_2}x\right)^{n-1}} \cdot \frac{\frac{n_1}{n_2}}{\left(1+\frac{n_1}{n_2}x\right)^2} dx \qquad \left(记\ C_2 = C_1\left(\frac{n_1}{n_2}\right)^{n-k+1}\right)\\
&= \int_{x(p)}^{\infty} C_2 x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} dx. \qquad \left(n-k = \frac{n_1}{2}-1,\ n+1 = \frac{n_1+n_2}{2}\right)
\end{aligned}$$

- 上式中的被积函数是$F(n_1, n_2)$ 的密度(见定义4.1). (注: 令$p \to 1$, 则$f(p) \to 1$, $x(p) \to 0$.) 因此, $P(F_{n_1,n_2} > x(p)) = f(p)$, 其中$F_{n_1,n_2} \sim F(n_1, n_2)$.
- 取$\lambda = x(p_\alpha)$, 即$p_\alpha = (1 + \frac{n-k+1}{k}\lambda)^{-1}$. 则$P(F_{n_1,n_2} > \lambda) = f(p_\alpha) = \alpha$. 因此$\lambda = F_{1-\alpha}(n_1, n_2)$.

例5.1. 原药有效率为$p_0 = 0.80$. 制药公司声称新药有效率高于0.80. 新药的数据: $n = 30$, $s_0 = 27$. 检验$H_0 : p \leqslant p_0$.

- $n_1 = 2(n - s_0 + 1) = 8$, $n_2 = 2s - 0 = 54$.
- $\alpha = 0.05$. 查表得到$F(n_1, n_2) = F(8, 54)$ 的临界值$\lambda = 2.13$.
- 计算$p_\alpha(s_0)$, 并判断:

$$p_\alpha(s_0) = \left(1 + \frac{4}{27} \times 2.13\right)^{-1} = 0.76 < p_0 = 0.80.$$

- 不拒绝H_0, 无明显理由说新药比原来药物有更高的有效率.
- 注: 若数据为$s = 28$, 则可拒绝H_0.
- 注: 有可能增大样本量后有可以得到显著结果.

- **单边′.** $H_0 : p \geqslant p_0$. 否定域: $\mathcal{W} = \{\vec{x} : S(\vec{x}) \leqslant c\}$, 其中, $S(\vec{x}) = \sum_{i=1}^{n} x_i$, c 是满足$P_{p_0}(S(\vec{X}) \leqslant k) \leqslant \alpha$ 的最大的整数.
- p值:

$$p(\vec{x}) = P_{p_0}(S \leqslant s_0) = \sum_{i=0}^{s_0} C_n^i p_0^i (1-p_0)^{n-i}.$$

- 否定域: $\mathcal{W} = \{\vec{x} : p(\vec{x}) \leqslant \alpha\} = \{\vec{x} : p_0 \leqslant \tilde{p}_\alpha(S(\vec{x}))\}$, 其中$\tilde{p}_\alpha(s_0)$ 为如下方程的唯一解.

$$\sum_{i=0}^{s_0} C_n^i p_0^i (1-p_0)^{n-i} = \alpha. \tag{5.9}$$

- 注: 也可以将问题变为令$1-p=q$. 检验$H_0 : q \leqslant q_0$.

- **双边.** $H_0 : p = p_0$. 记$S(\vec{x}) = \sum_{i=1}^{n} x_i$,
  否定域: $\mathcal{W} = \{\vec{x} : S(\vec{x}) \leqslant c_1$ 或 $S(\vec{x}) \geqslant c_2\}$.
  c_1 是满足$P_{p_0}(S(\vec{X}) \geqslant k) \leqslant \frac{\alpha}{2}$ 的最小的整数.
  c_2 是满足$P_{p_0}(S(\vec{X}) \leqslant k) \leqslant \frac{\alpha}{2}$ 的最大的整数.
- 用p值方法. 记$s_0 = \sum_{i=1}^{n} x_i$. p值定义为

$$p(\vec{x}) = 2\min\left\{P_{p_0}(S \leqslant S_0), P_{p_0}(S \geqslant S_0)\right\}.$$

- 否定域: $\mathcal{W} = \{\vec{x} : p(\vec{x}) \leqslant \alpha\}$.

(3)两总体，正态近似法(大样本)

- **单边**. $H_0: p_1 \leqslant p_2$ 或 $H_0: p_1 \geqslant p_2$.
- p_i 的无偏估计: $\hat{p}_i = \frac{1}{n_i} S_i$, 其中$S_1 = \sum_{i=1}^{n_1} X_i$, $S_2 = \sum_{i=1}^{n_2} Y_i$.
- 方差: $\sigma_i^2 = D(\hat{p}_i) = \frac{p_i(1-p_i)}{n_i}$, 用$\frac{\hat{p}_i(1-\hat{p}_i)}{n_i}$ 代替.
- 统计量η 与"枢轴量" ξ:

$$\eta(\vec{x}, \vec{y}) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}, \quad \xi = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\star + \star}}.$$

- 当n_1 和n_2 相当大($n_i\hat{p}_i(1-\hat{p}_i) \geqslant 5$)时, ξ 近似服从$N(0,1)$.
- 否定域:

$\mathcal{W} = \{(\vec{x}, \vec{y}) : \eta(\vec{x}, \vec{y}) > z_{1-\alpha}\}$, $\mathcal{W} = \{(\vec{x}, \vec{y}) : \eta(\vec{x}, \vec{y}) < z_{\alpha}\}$.

- **双边.** $H_0 : p_1 = p_2$.
- p_i 的无偏估计: $\hat{p}_i = \frac{1}{n_i} S_i$, 其中$S_1 = \sum_{i=1}^{n_1} X_i$, $S_2 = \sum_{i=1}^{n_2} Y_i$.
- $\hat{p}_1, \hat{p}_2$ 均近似服从正态, 方差: $\sigma_i^2 = D(\hat{p}_i) = \frac{p_i(1-p_i)}{n_i}$.
- 若$p_1 = p_2$ (记为p), 则
    - $\sigma^2 = D(\hat{p}_1 - \hat{p}_2) = D(\hat{p}_1) + D(\hat{p}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1-p)$.
    - 在σ^2 中, 用$\hat{p} := \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} = \frac{S_1+S_2}{n_1+n_2}$ 代替p.
- 当n_1 和n_2 相当大($n_i\hat{p}_i(1-\hat{p}_i) \geqslant 5$)时, ζ 近似服从$N(0,1)$.

$$\zeta(\vec{x}, \vec{y}) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \hat{p}(1-\hat{p})}}.$$

- 否定域: $\mathcal{W} = \{(\vec{x}, \vec{y}) : |\zeta(\vec{x}, \vec{y})| > z_{1-\frac{\alpha}{2}}\}$.

例5.2. 研究口服避孕药对年龄在40至44岁的妇女心脏的影响. 三年内出现心肌梗死的数据如下. 服药: $n_1 = 5000$, $s_1 = 13$ 人; 不服药: $n_2 = 10000$, $s_2 = 7$. 问: 心肌梗死比率是否有显著差异?

- 统计量: $\hat{p}_1 = \frac{13}{5000} = 0.0026$, $\hat{p}_2 = \frac{7}{10000} = 0.0007$.
- 近似条件成立: $n_i \hat{p}_i 1(1 - \hat{p}_i)$ 分别是6.66, 6.70, 均$\geqslant 5$.
- **双边.** $H_0 : p_1 = p_2$. 计算得$\zeta = 3.01$, $(\hat{p} = 0.00133.)$
  查表& 下结论: $\alpha = 0.01$, $z_{1-\frac{0.01}{2}} = 2.58 < |\zeta|$, 拒绝$H_0$.
- **单边.** $H_0 : p_1 \leqslant p_2$. 计算得$\eta = 2.48$.
  查表& 下结论: $\alpha = 0.01$, $z_{1-0.01} = 2.33 < \eta$, 拒绝H_0.
- 在0.01水平下, 两组的心肌梗死比率有显著差异, 口服避孕药对心肌梗死比率有显著增大.

(3)两总体, Fisher精确检验法

- $s_1 = \sum_{i=1}^{n_1} x_i$, $s_2 = \sum_{i=1}^{n_2} y_i$, $t = s_1 + s_2$.
- 记$p(i) = C_{n_1}^{i} C_{n_2}^{t-i} / C_{n_1+n_2}^{t}$. 特别地, $p(s_1) = C_{n_1}^{s_1} C_{n_2}^{s_2} / C_{n_1+n_2}^{s_1+s_2}$.
- **单边.** $H_0 : p_1 \leqslant p_2$, 或$H_0 : p_1 \geqslant p_2$. p值:

$$p_1(\vec{x}, \vec{y}) = \sum_{i \geqslant s_1} p(i), \quad p_2(\vec{x}, \vec{y}) = \sum_{i \leqslant s_1} p(i).$$

- **双边.** $H_0 : p_1 = p_2$. p值:

$$p_3(\vec{x}, \vec{y}) = 2 \min \left\{ p_1(\vec{x}, \vec{y}),\ p_2(\vec{x}, \vec{y}) \right\}.$$

- 拒绝域: $\mathcal{W} = \{(\vec{x}, \vec{y}) : p_i(\vec{x}, \vec{y}) \leqslant \alpha\}$.
- 实践中, $p(i)$ 用如下递推公式计算.

$$p(i+1) = p(i) \frac{(n_1 - i)(t - i)}{(i+1)(n_2 - t + i + 1)}. \tag{5.19}$$

例5.3. 某公安局有两个专案组. 在过去一年内一组接手$n_1 = 25$ 件人命案, 侦破了$s_1 = 23$ 件; 二组接手$n_2 = 35$ 件人命案，侦破了$s_2 = 30$ 件. 问: 两个组的侦破能力有无差别?

- 设两个组的侦破率分别为p_1, p_2. 检验$H_0 : p_1 = p_2$.
- $t = 53$. $p(23) = C_{25}^{23} C_{35}^{30} / C_{23+25}^{25+35} = 0.252 > \alpha (= 0.05)$,
  $\cancel{p(24) = p(23)\frac{2\times 30}{24\times 6} = 0.105, \quad p(25) = p(24)\frac{1\times 29}{25\times 7} = 0.017.}$
- p值: $p_i(\vec{x}, \vec{y}) \geqslant p(23) > \alpha$.

$$\cancel{p_1(\vec{x}, \vec{y}) = \sum_{i=23}^{25} p(i) = 0.374,}$$

$$\cancel{p_2(\vec{x}, \vec{y}) = \sum_{i=0}^{23} p(i) = 1 - p_1(\vec{x}, \vec{y}) + p(23) = 0.878,}$$

$$\cancel{p_3(\vec{x}, \vec{y}) = 2 \times 0.374 = 0.748.}$$

- 在$\alpha = 0.05$ 水平下, $H_0 : p_1 \leqslant p_2$, $H_0 : p_1 \geqslant p_2$, $H_0 : p_1 = p_2$ 均不拒绝. 两个专案组在破案能力上没有显著差异.

§6.6 总体的分布函数的假设检验

- 假设检验的参数方法先假定总体服从某种带有未知参数的分布(常用正态分布), 然后回答针对总体参数的问题.
- 还可以不假定总体分布类型, 直接回答分布有关的问题, 如总体是否来自正态分布.
- 如何判断一个总体X 是否分布函数为$F(x)$?
- 有时候从学科知识可以建模得到, 如前面放射性粒子数服从泊松分布的模型推导.
- 很多情况下只能从观测数据判断.
- 一般先作直方图(对连续型总体), 推测可能的分布类型, 再进行检验.

- **连续型.** 检验$H_0: X$的分布函数为$F(x)$.
- 类似于画直方图. 选$t_1 < \cdots < t_m$ 将实数轴分成$m+1$ 段,

$$I_1 = (-\infty, t_1],\ I_2 = (t_1, t_2],\ \ldots,\ I_m = (t_{m-1}, t_m],\ I_{m+1} = (t_m, +\infty).$$

  对应的概率$p_i = P(X \in I_i)$ 可用$F(x)$ 表达.
- 注: 一般选$t_1, \cdots, t_m$ 使得p_i's 比较小, ν_i's 不太小.
- $\nu_i =$ 落入I_i 的数据个数, 即, 频数. $X \in I_i$ 的频率$= \frac{\nu_i}{n}$.
- 若H_0 成立, 则$\frac{\nu_i}{n} \approx p_i$.
- 否定域: $\mathcal{W} = \{\vec{x} : V(\vec{x}) > \lambda\}$, 其中, 检验统计量

$$V(\vec{x}) = \sum_{i=1}^{m+1} \left(\frac{\nu_i}{n} - p_i\right)^2 \cdot \frac{n}{p_i} = \sum_{i=1}^{m+1} \frac{(\nu_i - np_i)^2}{np_i}.$$

- $H_0: X \sim F(x)$. 否定域: $\mathcal{W} = \{\vec{x} : V(\vec{x}) > \lambda\}$, 其中,

$$V(\vec{x}) = \sum_{i=1}^{m+1} \left(\frac{\nu_i}{n} - p_i\right)^2 \cdot \frac{n}{p_i} = \sum_{i=1}^{m+1} \frac{(\nu_i - np_i)^2}{np_i}.$$

- $\frac{\nu_i}{n}$ 近似服从$\mathrm{N}(p_i, \frac{1}{n}p_i(1-p_i)) \approx \mathrm{N}(p_i, \frac{1}{n}p_i)$ (p_i 比较小). 因此$(\frac{\nu_i}{n} - p_i)\sqrt{\frac{n}{p_i}}$ 近似服从$\mathrm{N}(0,1)$.
- 可以证明: 在H_0 下, $V(\vec{X})$ 近似服从$\chi^2(m)$.
- **临界值**: $\lambda = \chi^2_{1-\alpha}(m)$.
- 注: 若$V(\vec{x})$ 中没有因子$\frac{n}{p_i}$, 则p_i 较小的项会被严重低估.
- 注: 若零假设下的分布函数包含r 个未知参数$\theta_1, \cdots, \theta_r$, 则先用点估计$\hat{\theta}_k$'s 代替$\theta_k$'s. 用$\hat{p}_i = F_{\hat{\theta}}(X \in I_i)$ 代替p_i. 此时, $V(\vec{x})$ 近似服从$\chi^2(m-k)$.

例6.1. 滚珠直径X. $n = 50$, 数据: $\cdots$. 检验X 的分布.

- 观察数据, 提出$H_0 : X \sim N(\mu, \sigma^2)$. 2 个未知参数$\mu$, σ^2.
- 用无偏估计$\hat{\mu} = \bar{x} = 15.1$, $\hat{\sigma}^2 = S^2 = 0.4325^2$ 代替μ, σ^2.
- 数据最小14.2, 最大15.9. 取$a = 14.05$, $b = 16.15$.
  $b - a = 2.1 = 0.3 * 7$. 取$m = 6$, 将$[a, b]$ 等分为7 段.
- 统计出ν_i's. 查N(0, 1) 的临界值表可得$p_i = P(Y \in I_i)$,
  其中$Y \sim N(\hat{\mu}, \hat{\sigma}^2)$. 计算得$V(\vec{x}) = \sum_{i=1}^{7} \frac{(\nu_i - np_i)^2}{np_i} = 1.7284$.
- 查$\chi^2(m - 2) = \chi^2(4)$ 的临界值表, $\alpha = 0.05$, 得$\lambda = 9.49$.
- **下结论**: $V(\vec{x}) < \lambda$. 接受H_0.
- 注: 可能会有较大的第二类错误概率. 有可能以威布尔分布、对数正态等作为零假设, 检验均不拒绝. 只要不拒绝, 就可以说明数据与该分布差距不大.

- **离散型**. 一般不需要分组, (概率特别小的组可以合并).
- 设X 的分布是

$$P(X=a_i)=p_i,\quad (i=1,2,\ldots,m+1).$$

- $\nu_i=$ 数据中a_i 出现的次数. 统计量:

$$V=V(\vec{x})=\sum_{i=1}^{m+1}\frac{(\nu_i-np_i)^2}{np_i}. \tag{6.1}$$

- 在H_0 下, $V(\vec{X})$ "还是" 近似服从$\chi^2(m)$.
- 注: 若分布中含有r 个未知参数, 处理办法同连续型情形.
- 否定域及临界值同连续型情形.

例6.2. (例1.3续). 某工厂近5年来发生了$n = 63$ 次事故, 在工作日的分布如下. 问: 事故发生是否与星期几有关?

星期	一	二	三	四	五	六
次数	9	10	11	8	13	12

- $H_0 : P(X = i) \equiv \frac{1}{6}$. $i = 1, \cdots, 6 = m + 1$. 故$m = 5$.
- 计算$V(\vec{x}) = 1.67$.
- 查$\chi^2(5)$ 的临界值表, $\alpha = 0.05$, 得$\lambda = 11.07$.

  下结论: $K(\vec{x}) < \lambda$, 数据与H_0 相容.

  不能认为出事故与星期几有关.