2) 若 σ_1^2 和 σ_2^2未知, 但已知 $\sigma_1^2=\sigma_2^2$, 设

$$S_W=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2},$$

则考虑枢轴变量

$$W=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2).$$

于是有

$$\Pr\left[-t_{\alpha/2}(n+m-2)<\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n}+\frac{1}{m}}}<t_{\alpha/2}(n+m-2)\right]=1-\alpha.$$

3) 求方差比 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的置信区间. 设枢轴变量

$$W=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n-1,m-1),$$

根据 F 分布的不对称性, 采用概率对称的区间

$$\Pr[W\leq a]=\Pr[W\geq b]=\alpha/2\quad\Rightarrow\quad b=F_{\frac{\alpha}{2}}(n-1,m-1),\quad a=F_{1-\alpha/2}(n-1,m-1).$$

由此可得置信区间

$$\Pr\left[\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n-1,m-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n-1,m-1)}\right]=1-\alpha.$$

9.3.5 单侧置信区间

对某些实际问题, 我们往往只关心置信区间的上限或下限, 例如, 次品率只关心上限, 产品的寿命只关心下限, 由此引入单侧置信区间及其估计.

定义9.5 (单侧置信区间)**.** *给定 $\alpha\in(0,1)$, 若样本 $X_1,\cdots,X_n$ 的统计量 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 满足*

$$\Pr\left[\theta>\hat{\theta}_1\right]\geq 1-\alpha,$$

则称 $(\hat{\theta}_1,+\infty)$ 为 θ 的置信度为 $1-\alpha$ 的单侧置信区间, $\hat{\theta}_1$ 称为单侧置信下限.

同理定义单侧置信上限. 对正态总体, 可以将相关置信区间的估计都扩展到单侧置信估计, 枢轴变量的定理类似, 我们将不再重复讨论, 下面仅举两个实例:

例9.16. *设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim\mathcal{N}(\mu,\sigma^2)$ 的样本, 若方差 σ^2 已知, 求 μ 的置信度为 $1-\alpha$ 的单侧置信下限和上限.*

解. 设样本均值 $\bar{X}=\sum_{i=1}^n X_i/n$, 根据正态分布的性质考虑枢轴变量

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

于是有

$$\Pr\left[\frac{\bar{X}-\mu}{S/\sqrt{n}} < \mu_\alpha\right] = 1-\alpha, \quad \Pr\left[\frac{\bar{X}-\mu}{S/\sqrt{n}} > -\mu_\alpha\right] = 1-\alpha,$$

整理计算完成估计. □

例9.17. *从一批出厂的灯泡中随机抽取 10 盏灯泡, 测试其寿命分别为: 1000, 1500, 1250, 1050, 950, 1000, 1150, 1050, 950, 1000, (单位: 小时). 假设这批灯泡的寿命服从正态分布, 求这批灯泡平均寿命的置信度为 95% 的单侧置信下限.*

解. 首先计算样本均值和样本修正方差分别为

$$\bar{X} = \sum_{i=1}^{10} X_i/10 = 1090 \quad \text{和} \quad S^2 = \sum_{i=1}^{10}(X_i-\bar{X})^2/9 = 8800/3.$$

根据正态分布的性质考虑枢轴变量

$$W = \frac{\bar{X}-\mu}{S/3} \sim t(9),$$

于是有

$$\Pr\left[\frac{\bar{X}-\mu}{S/3} < t_{0.05}(9)\right] = 0.95,$$

查表 $t_{0.05}(9) = 1.833$ 可得

$$\mu > \bar{X} - t_{0.05}(9)S/3 = 1090 - \sqrt{8800/3} \times 1.833/3 > 1056.$$

□

9.3.6 非正态分布的区间估计

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 若总体 X 的分布未知或非正态分布, 我们可以给出总体期望 $\mu = E[X]$ 的区间估计, 方法分为两种: 利用 Concentration 不等式和中心极限定理.

1. 首先考虑 Concentration 不等式, 若总体 $X \in [a,b]$, 设 $\bar{X} = \sum_{i=1}^n X_i/n$, 根据 Concentration 不等式有

   $$\Pr\left[|\mu - \bar{X}| \geq \epsilon\right] \leq 2\exp(-2n\epsilon^2/(b-a)^2).$$

   令 $\alpha = 2\exp(-2n\epsilon^2/(b-a)^2)$ 求解 $\epsilon = \sqrt{(b-a)^2\ln(2/\alpha)/n}$, 于是有

   $$\Pr\left[\bar{X} - \sqrt{(b-a)^2\ln(2/\alpha)/n} < \mu < \bar{X} + \sqrt{(b-a)^2\ln(2/\alpha)/n}\right] > 1-\alpha.$$

   可基于其它 Concentration 不等式给出类似的置信区间估计, 以及其它 sub-Gaussian 型随机变量的期望的置信区间估计.

2. 利用中心极限定理, 求枢轴变量的近似分布, 再给出置信区间估计. 设总体 X 的期望 $E(X) = \mu$, 方差 $\mathrm{Var}(X) = \sigma^2$, 利用中心极限定理设枢轴变量

   $$W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}.$$

枢轴变量 W 的分布近似于标准正态分布 $\mathcal{N}(0,1)$. 当方差 σ^2 已知时有

$$\Pr\left[-\mu_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$

当方差 σ^2 未时, 用修正样本方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ 代替方差 σ^2, 于是有

$$\Pr\left[-\mu_{\alpha/2} < \frac{\bar{X}-\mu}{S^2/\sqrt{n}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$

例9.18. *设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim Ber(p)$ 的样本, 求 p 的置信度为 $1-\alpha$ 的区间估计.*

解. 根据 Bernoulli 分布的性质有 $X_i \in \{0,1\}$ 以及 $p = E[X]$, 根据 Chernoff 不等式有

$$\Pr[|\bar{X}-p| > \epsilon p] \leq 2\exp(-np\epsilon^2/3),$$

设 $\alpha = 2\exp(-np\epsilon^2/3)$, 于是有

$$\Pr\left[\bar{X} - \sqrt{3p\ln(2/\alpha)/n} < p < \bar{X} + \sqrt{3p\ln(2/\alpha)/n}\right] \geq 1-\alpha,$$

最后求解 p 的置信区间.

方法二: 根据 Bernoulli 分布的性质有 $E[X] = p$ 和 $\mathrm{Var}(X) = p(1-p)$, 设枢轴变量

$$W = \frac{n\bar{X} - np}{\sqrt{np(1-p)}}$$

根据中心极限定理可知 W 近似于标准正态分布 $\mathcal{N}(0,1)$. 于是有

$$\Pr\left[-\mu_{\alpha/2} < \frac{n\bar{X}-np}{\sqrt{np(1-p)}} < \mu_{\alpha/2}\right] \approx 1-\alpha.$$

最后求解 p 的近似置信区间. □

10 假设检验(Hypothesis Testing)

根据样本信息来检验关于总体的某个假设是否正确, 此类问题称为 **假设检验问题**, 可分为两类:

- 参数检验问题: 总体分布已知, 检验某未知参数的假设;
- 非参数检验问题: 总体分布未知时的假设检验问题.

假设检验的方法: 先假设所做的假设 H_0 成立, 然后从总体中取样, 根据样本的取值来判断是否有'不合理'的现象出现, 最后做出接受或者拒绝所做假设的决定. '不合理'的现象指小概率事件在一次事件中几乎不会发生.

例10.1. *某产品出厂检验规定次品率 $p \leq 0.04$ 才能出厂, 现从 10000 件产品中任抽取 12 件, 发现 3 件是次品, 问该批产品是否该出厂; 若抽样结果有 1 件次品, 问该批产品是否该出厂?*

解. 首先做出假设 $H_0 : p \leq 0.04$. 若假设 H_0 成立, 设随机变量 $X \sim \mathrm{B}(12, p)$,

$$\Pr[X = 3] = \binom{12}{3} p^3 (1-p)^9 \leq 0.0097.$$

由此可知这是一个小概率事件, 一次试验不应该发生, 但却发生了, 故不合理, 原假设 $H_0 : p \leq 0.04$ 不成立, 即 $p > 0.04$, 该批产品不能出厂.

若 $X = 1$ 则

$$\Pr[X = 1] = p(1-p)^{11} \binom{12}{1} \geq 0.306.$$

这不是小概率事件, 没理由拒绝原假设H, 产品可以出厂. □

注: 当 $X = 1$ 情况下, 若直接利用参数估计

$$p = 1/12 = 0.083 > 0.04.$$

若仅仅采用参数估计而不用假设检验, 则不能出厂, 因此参数估计与假设检验是两回事.

在假设检验中, 需要对'不合理'的小事件给出一个定性描述, 通常给出一上界 α, 当一事件发生的概率小于 α 时则成为小概率事件. 通常取 $\alpha = 0.05, 0.1, 0.01$, 其具体取值根据实际问题而定. 在假定 H_0 成立下, 根据样本提供的信息判断出不合理的现象 (概率小于 α 的事件发生), 则认为假设 H_0 不显著, α 被称为显著水平.

注意: 不否定假设 H_0 并不是肯定假设 H_0 一定成立, 而只能说差异不够显著, 没达到否定的程度, 所以假设检验被称为"显著性检验".

前面的例子初步介绍了假设检验的基本思想和方法, 下面再进一步说明假设检验的一般步骤:

例10.2. *假设某产品的重量服从 $\mathcal{N}(500, 16)$, 随机取出 5 件产品, 测得重量为 $509, 507, 498, 502, 508$, 问产品的期望是否正常? (显著性水平 $\alpha = 0.05$)*

解. 下面给出假设检验的一般步骤:

- 第一步: 提出原假设 $H_0: \mu = 500$ 和备择假设 $H_1: \mu \neq 500$;
- 第二步: 设计检验统计量, 在原假设 H_0 成立下的条件下求出其分布. 令样本均值 $\bar{X} = \sum_{i=1}^{5} X_i/5 = 504.8$, 设检验统计量为

  $$Z = \frac{\bar{X} - 500}{\sqrt{16/5}} \sim \mathcal{N}(0,1).$$

  检验统计量能衡量差异大小且分布已知.
- 第三步: 给定显著性水平 $\alpha = 0.05$, 查表得到临界值 $\mu_{0.025} = 1.96$, 使得

  $$\Pr[|Z| > 1.96] = 0.05$$

  成为一个小事件, 从而得到否定域 $\{Z\colon |Z| > 1.96\}$.
- 第四步: 将样本值代入计算统计量 Z 的实测值

  $$|Z| = \frac{|\bar{X} - 500|}{\sqrt{16/5}} = \frac{4.8}{4/\sqrt{5}} = 1.2 \times \sqrt{5} = 2.68 > 1.96.$$

  根据实测值 Z 落入否定域 $\{Z\colon |Z| > 1.96\}$, 从而拒绝原假设 H_0.

$\square$

由此归纳出假设检验的一般步骤:

1) 根据实际问题提出原假设 H_0 和备择假设 H_1;
2) 确定检验统计量 (分布已知);
3) 确定显著性水平 α, 并给出拒绝域;
4) 由样本计算统计量的实测值, 判断是否接受原假设 H_0.

假设检验可分为如下三类:

- 原假设 $H_0: \mu = \mu_0$ 和备选假设 $H_1: \mu \neq \mu_0$, 称为 **双边假设检验**;
- 原假设 $H_0: \mu \leq \mu_0$ 和备选假设 $H_1: \mu > \mu_0$, 称为 **右边检验**;
- 原假设 $H_0: \mu \geq \mu_0$ 和备选假设 $H_1: \mu < \mu_0$, 称为 **左边检验**.

右边检验和左边检验又被通称为双边检验.

下面研究假设检验是否会犯错, 假设检验的核心是先假设原判断假设 H_0 成立, 然后根据样本的取值来判断是否有'不合理'的现象出现, 即“小概率”原理, 然而小概率事件在一次试验中不发生并不意味着小概率事件不发生. 可能发生如下两种错误:

- 第 I 类错误: “弃真”, 即当 H_0 为真时, 我们仍可能拒绝 H_0.
- 第 II 类错误: “存伪”, 即当 H_0 不成立时, 我们仍可能接受 H_0.

两类错误如下表格所示

假设检验的决定	真实情况: H_0 为真	真实情况: H_0 为假
拒绝 H_0	第 I 类错误	正确
接受 H_0	正确	第 II 类错误

设犯第 I 类错误的概率为 α, 即显著性水平, 第 II 类错误的概率用 β 表示, 即

$$\alpha = \Pr\left[\text{拒绝 } H_0 | H_0 \text{ 为真}\right] \qquad \beta = \Pr\left[\text{接受 } H_0 | H_0 \text{ 为假}\right].$$

这两类错误互相关联, 当样本容量固定时, 一类错误概率的减少导致另一类错误概率的增加. Neymam-Pearson 原则: 在控制第 I 类错误的前提下, 尽可能减小第 II 类错误的概率.

10.1 正态总体期望的假设检验

10.1.1 方差已知的单个正态总体的期望检验 (Z 检验)

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 若方差 σ^2 已知, 检验原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$. 设样本均值为$\bar{X} = \sum_{i=1}^n X_i/n$, 根据正态分布的性质选择统计量

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

给定显著性水平 α, 得到拒绝域为 $|Z| \geq \mu_{\alpha/2}$, 这种检验方法称为 **Z 检验法**.

关于 Z 检验法的双边和单边检验有

i) 原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$, 拒绝域为 $\{Z : |Z| \geq \mu_{\alpha/2}\}$;

ii) 原假设 $H_0: \mu \geq \mu_0$ 和备择假设 $H_1: \mu < \mu_0$, 拒绝域为 $\{Z : Z \leq -\mu_\alpha\}$;

iii) 原假设 $H_0: \mu \leq \mu_0$ 和备择假设 $H_1: \mu > \mu_0$, 拒绝域为 $\{Z : Z \geq \mu_\alpha\}$.

例10.3. *已知某产品的重量 $X \sim \mathcal{N}(4.55, 0.108^2)$, 现随机抽取 5 个产品, 其质量分别为 4.28, 4.40, 4.42, 4.35, 4.27. 问产品的期望在 $\alpha = 0.05$ 下有无显著性变化. ($\mu_{0.025} = 1.96$)*

解. 首先提出原假设 $H_0: \mu = 4.55$ 和备择假设 $H_1: \mu \neq 4.55$. 若 H_0 成立, 选择检验量

$$Z = \frac{\bar{X} - 4.55}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

求得拒绝域为 $|Z| \geq \mu_{\alpha/2} = 1.96$. 计算样本均值可知$\bar{X} = 4.364$, 于是有

$$\frac{\bar{X} - 4.55}{0.108/\sqrt{5}} = 3.851 > 1.96,$$

由此可拒绝 H_0, 说明有显著变化. □