概率统计 B
第六章 假设检验

根据李东风老师课件修改

2017 春季学期

本节目录

假设检验的问题

- 估计是统计学的一个重要问题，本章讨论另一个重要问题："假设检验"。
- 本节用一些例子给出假设检验的典型问题。
- **例 1.1** 某厂有一批产品，共 200 件，须经检验合格才能出厂。按国家标准，次品率不得超过 1%，今在其中抽取 5 件，发现这 5 件中含有次品。
- 问这批产品能否出厂？
- 直观看不能出厂，但理由何在？
- 设这批产品的次品率为 p。
- 问题化为：如何根据抽样的结果来判断不等式"$p \leq 0.01$"是否成立？

例 1.2

- **例 1.2** 用某仪器间接测量温度，重复 5 次，所得数据如下：

$$1250,\ 1265,\ 1245,\ 1260,\ 1275$$

  而用别的精确办法测得温度为 1277(可以看作温度的真值)。
- 试问此仪器间接的温度有无系统偏差？
- 用 X 表示这个仪器测得的数值，X 是随机变量。
- 得到的 5 个数据值是 X 的一个样本。
- 问题化为：如何判断等式“$E(X)=1277$”成立与否？

例 1.3

- **例 1.3** 某工厂近 5 年来发生了 63 次事故，这些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：事故发生是否与星期几有关？
- 用 X 表示这样的随机变量：若事故发生在星期 i，则 $X = i$。
- X 的可能取值集合为 $\{1, 2, \ldots, 6\}$（星期日是该厂厂休日）。
- 问题化为：如何判断

$$P(X = i) \equiv \frac{1}{6} \quad (i = 1, 2, \ldots, 6)$$

是否成立？

例 1.4

- **例 1.4** 在针织品的漂白工艺过程中，要考察温度对针织品的断裂强力（主要质量指标）的影响。
- 为了比较 70°C 与 80°C 的影响有无差别，在这两个温度下，分别重复做了 8 次试验，得数据如下（单位：千克力）

$$
\begin{aligned}
70^\circ\text{C时的强力：} & 20.5,\ 18.8,\ 19.8,\ 20.9, \\
& 21.5,\ 19.5,\ 21.0,\ 21.2 \\
80^\circ\text{C时的强力：} & 17.7,\ 20.3,\ 20.0,\ 18.8, \\
& 19.0,\ 20.1,\ 20.2,\ 19.1
\end{aligned}
$$

- 从试验数据看，两种温度下的强度有无区别？

- 用 X 表示 70°C 下的强力，Y 表示 80°C 下的强力，问题变成：
- 如何判断等式

$$E(X) = E(Y)$$

成立与否？
- 还可以进一步问等式

$$D(X) = D(Y)$$

成立与否？

例 1.5

- 比较新型吹风机与原吹风机。原吹风机是成功，但面临竞争压力。
- 新型吹风机成本减少了 15%，如果产品可靠性也不比原产品差则可以上市取代原产品。
- 否则不采纳新产品，因为产品有 1 年保质期，保质期内损坏需要免费更换。
- 公司进行了可靠性试验，将新产品和原产品各取 250 件在模拟一年使用的条件下进行试验。
- 发现新产品中有 11 个失效，而原产品中有 20 个失效。
- 问：新产品的可靠性是否不比原产品的差？
- 用 p_1 表示新产品的失效率，p_2 表示原产品的失效率。
- 问题化为：判断

$$p_1 \le p_2$$

例 1.6

- 怎样根据一个随机变量的样本值，判断该随机变量是否服从正态分布 N(μ, σ^2)?
- 更一般地，如何根据样本的特性去判断随机变量是否以给定的函数 $F(x)$ 为其分布函数？

假设检验问题

- 例 1.1—例 1.6 代表了一类广泛的统计学问题。
- 共同点是要从样本值出发去判断一个“看法”是否成立。
- 例 1.1 的“看法”是“次品率 $p \leq 0.01$”；
- 例 1.2 的看法是“$E(X) = 1277$”；
- 例 1.3 的看法是“$P(X = i) \equiv \frac{1}{6}$ $(i = 1, 2, \ldots, 6)$”；
- 例 1.4 的看法是“$E(X) = E(Y)$”；
- 例 1.5 的看法是“$p_1 \leq p_2$”；
- 例 1.6 的看法是“X 的分布函数是 $F(x)$”。

- “看法”又叫“假设”。这些例子就是所谓**假设检验**问题。
- 本章介绍一些常用的检验方法，判断所关心的“假设”是否成立。
- 例 1.1、例 1.2 和例 1.3 中的“假设”都是关于一个随机变量的参数的判断，这叫做**一个总体的参数检验问题**。
- 例 1.6 也是一个总体的检验问题，但不是参数检验，而是概率分布的检验问题。
- 例 1.4 和例 1.5 的“假设”是关于两个随机变量的判断，叫做**二总体的检验问题**。

假设检验的思想

- 假设检验的思想是“带概率的反证法”。
- 以例 1.1 为例说明。
- 要检验假设 “$p \leq 0.01$”。
- 在假设成立的条件下进行分析。如果假设成立，则总体中至多有 2 件次品。
- 任抽取 5 件产品，先来求这 5 件中“无次品”的概率。

- 用古典概型：

$$P(\text{无次品}) = \begin{cases} \frac{C_{198}^5}{C_{200}^5} & \text{当 200 件中有 2 件次品时} \\ \frac{C_{199}^5}{C_{200}^5} & \text{当 200 件中有 1 件次品时} \\ \frac{C_{200}^5}{C_{200}^5} & \text{当 200 件中没有次品时} \end{cases}$$

$$\geq \frac{C_{198}^5}{C_{200}^5} = \frac{198 \times 197 \times \cdots \times 194}{200 \times 199 \times \cdots \times 196} \geq 0.95$$

- 于是

$$P(\text{任取的 5 件中有次品}) \leq 1 - 0.95 = 0.05$$

- 计算说明：如果次品率真的 ≤ 0.01，那么抽取 5 件样品，出现次品的机会是很少的，平均在 100 次抽样中，出现不到 5 次。
- 如果 $p \le 0.01$ 成立，则在一次抽样中，人们实际上很少遇见出现次品的情况。
- 现在的实际情况是遇到了次品，这是“不合理”的。
- 产生这种不合理现象的根源在于假设 $p \le 0.01$；
- 因此假设“$p \le 0.01$”是不能接受的。
- 所以按照国家标准，这批产品不能出厂。

概率性质的反证法

- （1）反证法的思想。为了检验一个假设（如"$p \leq 0.01$"），先假定这个假设是成立的，如果实际从样本中观察到的情况在这个假设下是不合理的，就认为原来的假设是不正确的，**拒绝**原来的假设。
- 如果样本没有不合理现象，就**不能拒绝**原来的假设。
- （2）这不是纯粹的反证法。这里的"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小概率还是有可能发生的，一旦出现这种情况，我们拒绝原假设就是错误的。
- 所以在观察到小概率事件后拒绝原假设是有可能犯错的，只不过这种可能性比较小而且可以控制。
- 在原假设下概率小到什么程度算是"小概率事件"？通常取界限为 0.05，有时也取为 0.01 或 0.10。

本章内容

- §2 先讲一个正态总体的检验问题。
- §3 介绍假设检验的一般概念和数学描述，如功效、p 值等。
- §4 讲两个正态总体的检验问题。
- §5 介绍比率的假设检验（一个总体和两个总体）。
- §6 为概率分布的检验。

本节目录

一个正态总体的假设检验

- 设总体 $X \sim \mathrm{N}(\mu, \sigma^2)$, 关于一个正态总体有四种假设检验问题:
- 已知方差 σ^2, 检验假设 $H_0 : \mu = \mu_0$;
- 未知方差 σ^2, 检验假设 $H_0 : \mu = \mu_0$;
- 未知期望 μ，检验假设 $H_0 : \sigma^2 = \sigma_0^2$;
- 未知期望 μ，检验假设 $H_0 : \sigma^2 \leq \sigma_0^2$。

已知方差检验均值

- **例 2.1** 某车间生产铜丝，主要质量指标是折断力大小。
- 用 X 表示该车间生产的铜丝的折断力。
- 由以往经验，可以认为 X 服从正态分布，期望为 570 千克力，标准差是 8 千克力。
- 换了一批原料后，认为方差不会有什么变化，但需要检验折断力是否和原来一样。
- 即：已知方差 $\sigma^2 = 8^2$, 检验假设

$$H_0: \ \mu = 570$$

- 抽取了 10 个样品，测得折断力值为（单位：千克力）：

$$578, 572, 570, 568, 572, 570, 570, 570, 572, 596, 584$$

- 如何检验 H_0?

- 用概率性质的反证法。考虑在 H_0 成立的假设下，观测到的样本是否有不合理现象。
- 在 H_0 下，$X \sim \mathsf{N}(570, 8^2)$。
- 设 X 的一个样本为 $X_1, X_2, \ldots, X_n$（看成随机变量），则

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathsf{N}(570, \frac{8^2}{10}) \qquad U = \frac{\bar{X} - 570}{\sqrt{8^2/10}} \sim \mathsf{N}(0,1)$$

(正态分布的线性组合仍服从正态分布)

- 由正态分布的经验规则，$P(|U| > 1.96) = 0.05$ 所以在 H_0 成立时 $\{|U| > 1.96\}$ 是一个小概率事件。
- 计算得 $\bar{x} = 575.2$, $u = 2.05$，说明小概率事件发生了，认为是不合理的，故此拒绝 H_0，习惯上说“折断力的大小和原来**有显著差异**”。

例 2.2

- **例 2.2** 根据历史经验和资料分析，某砖瓦厂所生产的砖的“抗断强度”X 服从正态分布，方差 $\sigma^2 = 1.21$。
- 从一批产品中抽取 $n = 6$ 的样本，测得抗断强度（单位：kg · cm^{-2}）：32.56，29.66，31.64，30.00，31.87，31.03
- 问：这批砖的平均抗断强度可否认为是 32.80?
- **解** 待检验的假设是 $H_0 : \mu = 32.50$。
- 计算统计量 U:

$$U = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{31.13 - 32.50}{\sqrt{1.21/6}} = -3.05$$

- 于是 $|U| = 3.05 > 1.96$，小概率事件发生。拒绝 H_0，这批砖的平均抗断强度不能认为是 32.50 (或称这批砖的平均抗断强度与 32.50**有显著差异**)。

检验水平

- 以 H_0 下计算的概率小于 0.05 作为拒绝 H_0 的标准，这样 $\alpha = 0.05$ 叫做**检验水平**。
- 有些问题中需要把检验水平取得更小一些，如取 $\alpha = 0.01$。这时查表得 H_0 下

$$P(|U| > 2.58) = 0.01$$

- 在例 2.1 中，$U = 2.05$，不超过 2.58，在 0.01 水平下"小概率事件"没有发生，假设 H_0 与数据是**相容**的，简称 H_0 是相容的。
- 这与 $\alpha = 0.05$ 时结论不同，可见检验的结果与水平 α 的选择有关。
- 检验水平 α 的直观意义：把 H_0 成立时概率不超过 α 的事件当作一次观察时不会发生的"小概率事件"。

第一类错误

- 对于一般的 α（无论如何可以假设 $\alpha < 0.5$），取**临界值**λ 使得

$$P(|U| > \lambda) = \alpha$$

  从样本中计算统计量 U 的值，当 $|U| > \lambda$ 时就拒绝 H_0。
- 这样下结论不能绝对不犯错误（但是从部分（样本）推断整体（总体）本来就不能保证绝对正确，管中窥豹和盲人摸象是典型例子）。
- 即使 H_0 成立，仍有 α 的概率我们会拒绝 H_0，这种错误叫做**第一类错误**，检验水平 α 是犯第一类错误的概率的上界。

第二类错误

- 那么，是不是第一类错误越小越好？
- 不是。比如，我们取临界值 λ 接近于无穷大，这时 α 几乎等于零。
- 但是，这样检验相当于不论 H_0 成立与否都不拒绝 H_0，于是当 H_0 不成立时一定会犯错误。
- 当 H_0 不成立时，如果按照检验规则，数据与 H_0 相容，不能拒绝 H_0，就犯了**第二类错误**。
- 用 β 表示第二类错误的概率。
- 两类错误概率越小越好，但两者互相矛盾。
- 经典的统计假设检验做法是固定第一类错误概率的水平 α，然后尽可能设法减小第二类错误概率，如设计更好的检验方案，增大样本量。

未知方差时期望的检验

- 设总体为 $\mathrm{N}(\mu, \sigma^2)$, μ 和 σ^2 都未知，检验 $H_0: \mu = \mu_0$(μ_0 已知)。
- 例如，例 1.2，测量值 X 服从正态分布，μ 和 σ^2 都未知，$\mu_0 = 1277$，检验 $H_0: \mu = 1277$。
- 如果 σ^2 已知，检验使用的统计量为

$$U = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

- 在 σ^2 未知时，在 U 的公式中用 σ^2 的估计量 S^2 (样本方差) 代替 σ^2 得到

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$$

- 由 §5.5(P151) 知 H_0 成立时 $T \sim \mathrm{t}(n-1)$ 分布。

- 对例 1.2，$n=5$，查自由度为 4 的 t 分布表得

$$P(|T|>2.776)=0.05$$

- 计算得 $\bar{x}=1259$, $S^2=142.5$，$T=-3.37$, $|T|>2.776$，小概率事件发生，所以拒绝 H_0，认为间接测温的平均值不等于精确测量值，即间接测温有系统误差。

正态总体方差未知时关于期望的检验步骤

(1) 提出带检验的假设 $H_0 : \mu = \mu_0 (\mu_0$ 已知)。

(2) 根据样本值 $x_1, x_2, \ldots, x_n$ 计算统计量

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

的值。

(3) 对于检验水平 α，自由度 $n-1$，查 t 分布临界值表（附表 2），得临界值 λ。

(4) 若 $|T| > \lambda$，拒绝 H_0；否则 H_0 相容。

两类错误的比较

- 在经典的统计假设检验中，第一类错误由检验水平 α 控制，第二类错误只能尽量设法减小但不可控制。
- 所以，设计检验 H_0 时，尽可能取 H_0 使得 H_0 代表原来的、标准的情况或做法，从而接受 H_0 即使犯错（发生第二类错误）也没有严重后果，而拒绝 H_0 则需要更为谨慎（可能发生第一类错误）。
- 这样，一旦拒绝 H_0，我们是比较有把握结论是可信的。
- 实践中，人们一般希望得到拒绝 H_0 的结论。
- 这样也会发生报告偏差：不同的研究人员对同一问题做了试验，接受 H_0 的时候就不报告了，拒绝 H_0 就报告出来。

例 2.3

- **例 2.3** 根据长期资料分析，知道某种钢生产出的钢筋的强度服从正态分布。今随机抽取 6 根钢筋进行强度试验，测得强度为（单位：$\mathrm{kg}\cdot\mathrm{mm}^{-2}$）：

$$48.5,\ 49.0,\ 53.5,\ 49.5,\ 56.0,\ 52.5$$

- 问：能否认为该种钢生产的钢筋的平均强度为 52.0?

- **解** 用 X 表示钢筋强度，$X \sim \mathsf{N}(\mu, \sigma^2)$.
- （1）要检验的假设是 $H_0 : \mu = 52.0$。
- （2）计算统计量 T 的值。$\bar{x} = 51.5, S^2 = 8.9,$

$$T = \frac{51.5 - 52.0}{\sqrt{8.9/6}} = -0.411$$

- （3）查附表 2，$\alpha = 0.05$，自由度 $n - 1 = 5$，得 $\lambda = 2.571$。
- （4）下判断。现在 $|T| = 0.411 < 2.571$，故 H_0 是相容的，不能否定钢筋的平均强度为 $52.0\mathrm{kg} \cdot \mathrm{mm}^{-2}$。

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。
- $\mu = E(X)$ 的置信水平为 $1 - \alpha$ 的置信区间是满足

$$\left| \frac{\bar{x} - \mu}{\sqrt{S^2/n}} \right| \le \lambda$$

  的 μ 的集合。
- 检验 $H_0 : \mu = \mu_0$ 的规则是：当且仅当

$$\left| \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}} \right| \le \lambda$$

  时不拒绝 H_0。
- 所以检验法也可以说是：当且仅当 μ_0 属于 μ 的置信水平为 $1 - \alpha$ 的置信区间时不拒绝 H_0。
- 反过来以可以由检验法构造置信区间。

方差的双边检验

- 对正态总体，设期望和方差未知，检验：

$$H_0 : \sigma^2 = \sigma_0^2$$

(σ_0^2 已知)

- **例 2.4** 某车间生产铜丝，生产一直比较稳定。
- 今从产品中随机抽出 10 根检查折断力，得数据如下（单位：千克力）：

$$578, 572, 570, 568, 572, 570, 570, 570, 572, 596, 584$$

- 问：是否可相信该车间的铜丝的折断力的方差为 64?

- 用 X 表示铜丝的折断力，已知 $X \sim \mathsf{N}(\mu, \sigma^2)$，要根据样本检验

$$H_0 : \sigma^2 = \sigma_0^2 \qquad\qquad (\sigma_0^2 = 64)$$

- 自然想到用 S^2 与 σ_0^2 比较。如果 S^2/σ_0^2 很大或很小，则应拒绝 H_0。如果 S^2/σ_0^2 的值接近于 1，应该接受 H_0。
- 取统计量

$$W = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}$$

在 W 很大或很小时拒绝 H_0。

- 由 §5.6, 当 H_0 成立时 $W \sim \chi^2(n-1)$。查 χ^2 分布临界值表找到 λ_1, λ_2 使得

$$P(W < \lambda_1) = 0.025$$

$$P(W > \lambda_2) = 0.025$$

- 则 H_0 成立时 $\{W < \lambda_1 或 W > \lambda_2\}$ 是小概率事件。

- 对例 2.4，从样本中计算得 $\bar{x} = 575.2$, $S^2 = 75.73$, 故

$$W = \frac{9 \times 75.73}{64} = 10.65$$

- 查 χ^2 分布表，自由度 $n - 1 = 9$，得 $\lambda_1 = 2.70$, $\lambda_2 = 19.0$。现在

$$\lambda_1 < W < \lambda_2$$

- 所以 H_0 是相容的，可以相信铜丝折断力方差为 64。

方差的单侧检验

- 设正态总体的期望和方差都未知，检验

$$H_0 : \sigma^2 \le \sigma_0^2 \qquad\qquad (\sigma_0^2\text{已知})$$

- 这种检验也是常用的。H_0 代表生产的加工精度没有变差，如果拒绝了 H_0，就说明加工精度变差了，要检查原因。
- 设总体 $X \sim \mathrm{N}(\mu, \sigma^2)$，$X_1, X_2, \ldots, X_n$ 为样本。
- 自然想到如果 S^2/σ_0^2 远大于 1 则应该拒绝 H_0。
- 仍采用统计量

$$W = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}$$

  当 W 很大时拒绝 H_0。
- 在 $H_0 : \sigma^2 \le \sigma_0^2$ 的情况下，W 的分布依赖于未知的 σ^2。
- 用不等式推导中的放大法。

- 令

$$Y=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}=\frac{\sigma_0^2}{\sigma^2}W$$

- 则 H_0 成立时 $Y\geq W$。
- $Y\sim\chi^2(n-1)$，设 λ 是 $\chi^2(n-1)$ 分布右侧 α 临界值，即

$$P(Y>\lambda)=\alpha$$

- 则 $Y>\lambda$ 是小概率事件。
- 但是 Y 不是统计量，其计算公式涉及未知参数 σ^2。
- 然而，在 H_0 下

$$P(W>\lambda)\leq P(\frac{\sigma_0^2}{\sigma^2}W>\lambda)=P(Y>\lambda)=\alpha$$

- 也就是说，H_0 下 $\{W>\lambda\}$ 也是小概率事件，而 W 是统计量。
- 从样本值 $x_1,x_2,\ldots,x_n$ 中计算得到 W，如果 $W>\lambda$，就应该拒绝 H_0。

方差检验步骤

- 检验步骤：
- (1) 提出待检验的假设 $H_0 : \sigma^2 = \sigma_0^2$(或 $H_0 : \sigma^2 \le \sigma_0^2$)。
- (2) 计算统计量

$$W = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}$$

- (3) 查 χ^2 分布临界值表（附表 3），自由度为 $n-1$，得 λ_1,λ_2 (或 λ) 满足

$$P(\chi^2 < \lambda_1) = (\chi^2 > \lambda_2) = \frac{\alpha}{2}$$

$$(\text{或} P(\chi^2 > \lambda) = \alpha)$$

- (4) 比较 W 与 λ_1, λ_2（或 λ），作出判断。

单边检验假设选取

- 对于单边的问题，假设可以是问题本身或者问题的反面。
- 一般取假设为原来的、标准的情况或做法，这样接受假设即使犯错后果也不严重。
- 如果希望回答可信，应该把希望得到的回答的反面作为假设（拒绝了 H_0 时的结论是比较可靠的）。

例 2.5

- **例 2.5** 已知罐头番茄汁中，维生素 C 含量服从正态分布。
- 按照规定，维生素 C 含量不得少于 21 毫克。
- 现从一批罐头中抽了 17 罐，算得维生素 C 含量统计量为 $\bar{x} = 23$, $S^2 = 3.98^2$。
- 为这批罐头的维生素 C 的含量是否合格？
- **解** 设这批罐头中维生素 C 含量 $X \sim \mathrm{N}(\mu, \sigma^2)$。
- 我们希望作出合格的判断时比较有把握，所以设 H_0 为合格的反面：

$$H_0 : \mu < 21$$

- 参考 $H_0 : \mu = \mu_0$ 时的做法，考虑统计量

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

当 T 很大时拒绝 H_0。

- 在 H_0 成立时 T 的分布依赖于未知参数 μ，取

$$Y = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t(n-1)$$

- 则 H_0 成立时 $Y > T$。
- 取 λ 使

$$P(Y > \lambda) = \alpha$$

- 则 H_0 成立时

$$P(T > \lambda) \le P(Y > \lambda) = \alpha$$

所以 $\{T > \lambda\}$ 是小概率事件。

- 为了从附表 2 中查 λ，注意自由度为 $n-1$，另外

$$P(Y > \lambda) = \alpha \Longleftrightarrow P(|Y| > \lambda) = 2\alpha$$

所以取 $\alpha = 0.05$ 时只要从附表 2 中查 0.10 对应的 λ。

- 对例 2.5，查附表 2，自由度 $n-1=16$，0.10 对应的临界值为 $\lambda = 1.746$。
- 计算得

$$T = \frac{\bar{X} - 21}{\sqrt{\frac{S^2}{n}}} = 2.07 > 1.746$$

- 所以小概率事件发生了，拒绝 H_0，认为该批罐头合格。

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 T 统计量近似标准正态分布。
- 所以 $\alpha = 0.05$ 时可以用 $\{|T| > 1.96\}$ 作为拒绝 $H_0: \mu = \mu_0$ 的标准。
- 为了使得这种近似有意义，样本量 n 一般不小于 30，最好是 50 以上，或 100 以上。

本节目录

假设检验的某些概念和数学描述

- 检验法；
- 两类错误；
- 功效函数；
- 临界值与 p 值；
- 假设检验与置信区间的联系。

零假设和对立假设

- 要检验的“假设”记作 H_0，叫做**零假设或原假设**。
- H_0 是关于随机变量 X(总体) 的一个“看法”。数学描述：
- 设 X 的分布函数为 $F(x,\theta)$，其中 $\theta \in \Theta$，这里 Θ 是实数（或向量、或其它符号）组成的已知集合。
- “看法” H_0 表示成：$\theta \in \Theta_0$，这里 Θ_0 是 Θ 的非空真子集。
- 把 $\theta \in \Theta - \Theta_0$ 叫做**对立假设或 备择假设**，记作 H_a。

例 2.1 的假设

- 例如，在例 2.1 中，考虑铜丝折断力的总体，要检验的假设是 $\mu = 570$。
- 这里

$$
\begin{aligned}
X \sim & \mathrm{N}(\mu, 8^2) \\
\theta = & \mu \\
\Theta = & (-\infty, \infty) \\
\Theta_0 = & \{570\} \\
\Theta - \Theta_0 = & (-\infty, 570) \cup (570, \infty) \\
H_0 : & \theta \in \Theta_0 \\
H_a : & \theta \in \Theta - \Theta_0
\end{aligned}
$$

检验法

- 检验法就是给出一个规则，对给定的样本值 $x_1, x_2, \ldots, x_n$，进行明确表态：不拒绝假设 H_0 还是拒绝假设 H_0。
- 设 S 是所有可能的样本值 $(x_1, x_2, \ldots, x_n)(n$ 固定) 组成的集合（样本空间）。很多情况下 $S = R^n$。
- 检验法就是指空间 S 的一个划分：$S = S_1 \cup S_2$，$S_1 \cap S_2 = \emptyset$。当 $(x_1, x_2, \ldots, x_n) \in S_1$ 时，不接受假设 H_0；当 $(x_1, x_2, \ldots, x_n) \in S_2$ 时，拒绝 H_0。
- S_1 叫做**接受域**, S_2 叫做**否定域或拒绝域**。
- 否定域与检验法互相唯一确定。

假设检验的两类错误

- 在解决假设检验的问题时，无论作出否定还是接受原假设 H_0 的决定，都有可能犯错误.
- 我们称否定 H_0 时犯的错误为第一类错误，接受 H_0 时犯的错误为第二类错误.
-

		检验结果	
		H_0	H_1
真实	H_0	✓	X(第 I 类)
情况	H_1	X(第 II 类)	✓

- 假设检验一般控制第一类错误在检验水平 α 以下，所以否定 H_0 时结论比较可靠。
- 如果承认 H_0，可能犯第二类错误，错误概率可能会比较大。
- 要同时减小两类错误是不可能的，一种错误减小另一种错误必然增

功效函数

- 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本，W 是否定域，当事件 $\{(X_1, X_2, \ldots, X_n) \in W\}$ 发生时拒绝零假设 H_0。
- 设 $X \sim F(x, \theta)$，此事件的概率记作

$$M_W(\theta) = P_\theta((X_1, X_2, \ldots, X_n) \in W), \quad \theta \in \Theta$$

称作否定域 W(或对应的检验法) 的**功效函数**。

- 功效函数在两种情况下的意义:

		意义	记号
真实	$H_0(\theta \in \Theta_0)$	第 I 类错误概率	$\alpha_W(\theta)$
情况	$H_1(\theta \in \Theta_1)$	$1-$ 第二类错误概率	$1-\beta_W(\theta)$

检验水平

- 在 H_0 成立的时候，功效函数越小越好；
- 当 H_1 成立的时候，功效函数越大越好。
- **定义 3.1** 给定 $\alpha(0 < \alpha < 1)$，如果

$$\alpha_W(\theta) \leq \alpha, \quad \forall \theta \in \Theta_0$$

则称 W 是**检验水平 (或显著性水平)** 为 α 的否定域（检验法）。

- 如果

$$\sup_{\theta \in \Theta_0} \alpha_W(\theta) = \alpha$$

则称 α 为 W 的精确检验水平。

检验法讨论

- 统计学中假设检验一般都是固定一个检验水平 α，找到 H_1 下第二类错误尽可能小（功效尽可能大）的检验法进行检验。
- 这样，如果最后结论是拒绝 H_0，可能犯的是第 I 类错误，此错误概率收到检验水平的控制，所以结论是比较可信的。
- 如果最后结论是接受 H_0，可能犯的是第 II 类错误，此错误概率虽然已经尽可能控制但不像第 I 类错误那样有明确界限。
- 所以接受 H_0 的结论是一种“维持原状、不改变原来结论”之类的做法，一般不能把接受 H_0 作为一个新的发现使用。
- 这和“带概率的反证法”是一致的，反证法不能推出矛盾，并不意味着假设一定成立。

临界值

- 否定域 W 一般用一个直观上有明确意义的统计量（称为**检验统计量**）$\varphi(X_1, X_2, \ldots, X_n)$ 来确定。
- 单边情形的否定域：

$$W = \{(x_1, x_2, \ldots, x_n) : \varphi(x_1, x_2, \ldots, x_n) > \lambda\} \tag{3.1}$$

- 双边情形的否定域：

$$\begin{aligned} W = \Big\{&(x_1, x_2, \ldots, x_n) : \varphi(x_1, x_2, \ldots, x_n) < \lambda_1 \text{或} \\ &\varphi(x_1, x_2, \ldots, x_n) > \lambda_2\Big\} \end{aligned} \tag{3.2}$$

- (3.1) 中 λ 叫做单边情形的临界值，(3.2) 中 λ_1 和 λ_2 叫做双边情形的临界值。
- 临界值根据检验水平确定。

单侧临界值确定

- 为了使得 H_0 成立时第一类错误概率不超过检验水平 α，应找 λ 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) = \alpha \tag{3.3}$$

- 如果检验统计量服从连续型分布，这样的 λ 存在。
- 如果检验统计量服从离散型分布，这样的 λ 不一定存在，这时找 λ 使得

$$\begin{aligned} &\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) \le \alpha \\ <&\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \ge \lambda) \end{aligned} \tag{3.4}$$

(这是水平为 α 的最小的 λ)

- 注意：第一类错误概率不是越小越好！所以 λ 不是越大越好。

双边情形的临界值

- 对于双边情形，应找 λ_1 和 λ_2 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) < \lambda_1) = \frac{\alpha}{2} \tag{3.5}$$

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda_2) = \frac{\alpha}{2} \tag{3.6}$$

- 如果检验统计量服从连续型分布，这样的 λ_1 和 λ_2 存在。

- 如果检验统计量服从离散型分布，这样的 λ_1 和 λ_2 不一定存在，这时找 λ_1 和 λ_2 使得

$$
\begin{aligned}
& \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) < \lambda_1) \le \frac{\alpha}{2} \\
< & \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \le \lambda_1) \qquad (3.7) \\
& \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda_2) \le \frac{\alpha}{2} \\
< & \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \ge \lambda_2) \qquad (3.8)
\end{aligned}
$$

- 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

- **例 3.1** 设 $X \sim \mathrm{N}(\mu, \sigma^2)$，未知 σ，检验假设

$$H_0 : \mu \leq \mu_0 \longleftrightarrow H_a : \mu > \mu_0$$

- 令 $\theta = (\mu, \sigma^2)$，及

$$\begin{aligned}
\Theta =& \{(\mu, \sigma^2) : \mu \in (-\infty, \infty), \sigma^2 > 0\} \\
\Theta_0 =& \{(\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\} \\
\Theta_1 =& \Theta - \Theta_0 = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}
\end{aligned}$$

- 检验可表示为

$$H_0 : \theta \in \Theta_0 \longleftrightarrow H_a : \theta \in \Theta_1$$

- 设样本为 $X_1, X_2, \ldots, X_n$，取检验统计量

$$\varphi(X_1, X_2, \ldots, X_n) = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}$$

- φ 值越大，数据越倾向于否定 H_0 而接受 H_1。取单边否定域

$$\begin{aligned} W =& \{(x_1, x_2, \ldots, x_n) : \varphi(x_1, x_2, \ldots, x_n) > \lambda\} \\ =& \left\{(x_1, x_2, \ldots, x_n) : \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}} > \lambda\right\} \end{aligned}$$

- 对给定检验水平 α，临界值 λ 应满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) = \alpha$$

- 当 $\mu \leq \mu_0$ 时，

$$\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} \leq \frac{\bar{X}-\mu}{\sqrt{S^2/n}}$$

- 于是

$$\sup_{\theta\in\Theta_0} P_\theta\left(\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} > \lambda\right) = \sup_{\theta\in\Theta_0} P_\theta\left(\frac{\bar{X}-\mu}{\sqrt{S^2/n}} > \lambda\right)$$

- 其中 $\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$ 对任意 (μ, σ^2) 都服从 $\mathrm{t}(n-1)$ 分布，所以可取 λ 为 $\mathrm{t}(n-1)$ 分布的 $1-\alpha$ 分位数，则上式右边为 α。
- 否定域求得。

p 值方法

- p 值方法是确定否定域的另一方法，p 值可以直观表示数据中否定 H_0 的倾向的强烈程度。
- 以单边否定域 (3.1) 为例:

$$W = \{(x_1, x_2, \ldots, x_n) : \varphi(x_1, x_2, \ldots, x_n) > \lambda\} \tag{3.1}$$

- 设 $x_1, x_2, \ldots, x_n$ 是样本值，$X_1, X_2, \ldots, X_n$ 是总体分布参数为 θ 时的样本。

- 定义

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n) \right)
\end{aligned}
$$

- $p(x_1, x_2, \ldots, x_n)$ 是 H_0 成立的条件下统计量 $\varphi(X_1, X_2, \ldots, X_n)$ 取到和观测到的统计量值 $\varphi(x_1, x_2, \ldots, x_n)$ 一样大或更大的概率的最大值。
- 这个概率是数据取值倾向于否定 H_0 的情况的概率。
- 如果 H_0 成立，这个概率应该很小。
- 注意：

  (1) $p(x_1, x_2, \ldots, x_n)$ 取值于 $[0, 1]$;

  (2) $p(X_1, X_2, \ldots, X_n)$ 是统计量。

p 值与否定域

- 用 p 值可以表示否定域：
- **引理 3.1** 设对给定的 $\alpha \in (0,1)$，恰有一个 λ 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) = \alpha \tag{3.10}$$

则

$$\varphi(x_1, x_2, \ldots, x_n) > \lambda \Longleftrightarrow p(x_1, x_2, \ldots, x_n) < \alpha$$

- 即：当且仅当 p 值小于检验水平 α 时否定零假设 H_0:

$$W = \{(x_1, x_2, \ldots, x_n) : p(x_1, x_2, \ldots, x_n) < \alpha\}$$

- 这种确定否定域的方法叫做 **p 值方法**。
- 这个否定域的精确检验水平为 α。

p 值特点

- p 值越小，从数据看否定 H_0 的倾向越强烈。
- p 值大小代表了数据与 H_0 的相容程度，当 p 值小于 α 时就不相容了。
- p 值总是在 $[0,1]$ 取值的。
- p 值是能够否定 H_0 可取的最小的检验水平 α_0，取比 p 值更小的检验水平 $\alpha < \alpha_0$ 就不能否定 H_0 了。

引理 3.1 的证明

- 设 $p(x_1, x_2, \ldots, x_n) < \alpha$, 即

$$\sup_{\theta \in \Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n)\right) < \alpha \qquad (*)$$

- 而 λ 满足

$$\sup_{\theta \in \Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) > \lambda\right) = \alpha$$

- 所以一定有

$$\varphi(x_1, x_2, \ldots, x_n) > \lambda$$

- 这是因为如果

$$\varphi(x_1, x_2, \ldots, x_n) \leq \lambda$$

则

$$\begin{aligned} & \varphi(X_1, X_2, \ldots, X_n) > \lambda \\ \Longrightarrow & \varphi(X_1, X_2, \ldots, X_n) > \varphi(x_1, x_2, \ldots, x_n) \\ \Longrightarrow & \varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n) \end{aligned}$$

从而

$$\sup_{\theta \in \Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n)\right) \geq \alpha$$

- 反过来，设 $\varphi(x_1, x_2, \ldots, x_n) > \lambda$, 则有 $\varepsilon > 0$ 使得

$$\varphi(x_1, x_2, \ldots, x_n) - \varepsilon > \lambda$$

- 于是

$$\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n) \right) \\
\leq & \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(X_1, X_2, \ldots, X_n) > \varphi(x_1, x_2, \ldots, x_n) - \varepsilon \right) \\
\leq & \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(X_1, X_2, \ldots, X_n) > \lambda \right) = \alpha \qquad (**)
\end{aligned}$$

- 且 (**) 式的小于等于号必为严格小于号，否则令

$$\lambda' = \varphi(x_1, x_2, \ldots, x_n) - \varepsilon$$

则 $\lambda' > \lambda$ 也是

$$\sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) = \alpha$$

- 即：$\varphi(x_1, x_2, \ldots, x_n) > \lambda$ 时有

$$p(x_1, x_2, \ldots, x_n) < \alpha$$

- 引理 3.1 证毕。

精确水平不等于 α 的情况

- 给定 $\alpha \in (0,1)$ 不一定有临界值 λ 使得检验的精确水平为 α。
- 这时求满足 (3.4) 的 λ。
- p 值定义不变。
- **引理 3.2** 设对给定的 $\alpha \in (0,1)$，有 λ 满足

$$
\begin{aligned}
&\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda) \leq \alpha \\
<&\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \geq \lambda)
\end{aligned} \tag{3.11}
$$

则

$$\varphi(x_1, x_2, \ldots, x_n) > \lambda \Longleftrightarrow p(x_1, x_2, \ldots, x_n) \leq \alpha$$

引理 3.2 的证明

- **引理 3.2 的证明** 设 $\varphi(x_1,x_2,\ldots,x_n)>\lambda$，则

$$
\begin{aligned}
&P_\theta\big(\varphi(X_1,X_2,\ldots,X_n)\geq\varphi(x_1,x_2,\ldots,x_n)\big)\\
\leq&P_\theta(\varphi(X_1,X_2,\ldots,X_n)>\lambda)
\end{aligned}
$$

- 两边取 $\sup_{\theta\in\Theta_0}$，有

$$
\begin{aligned}
&p(x_1,x_2,\ldots,x_n)\\
=&\sup_{\theta\in\Theta_0}P_\theta(\varphi(X_1,X_2,\ldots,X_n)\geq\varphi(x_1,x_2,\ldots,x_n))\\
\leq&\sup_{\theta\in\Theta_0}P_\theta(\varphi(X_1,X_2,\ldots,X_n)>\lambda)\leq\alpha\quad(\text{由 }(3.11))
\end{aligned}
$$

- 反过来，如果 $\varphi(x_1, x_2, \ldots, x_n) \le \lambda$, 则

$$
\begin{aligned}
& P_\theta\big(\varphi(X_1, X_2, \ldots, X_n) \ge \varphi(x_1, x_2, \ldots, x_n)\big) \\
\ge & P_\theta(\varphi(X_1, X_2, \ldots, X_n) \ge \lambda)
\end{aligned}
$$

- 两边取 $\sup_{\theta\in\Theta_0}$，有

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \ge \varphi(x_1, x_2, \ldots, x_n)) \\
\ge & \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \ge \lambda) > \alpha \quad (\text{由 } (3.11))
\end{aligned}
$$

- 因此 $p(x_1, x_2, \ldots, x_n) \le \alpha$ 时一定有 $\varphi(x_1, x_2, \ldots, x_n) > \lambda$。
- 引理 3.2 证毕。

- 在引理 3.2 条件下，若 λ 满足 (3.11)(不一定唯一)，则否定域

$$W = \{(x_1, x_2, \ldots, x_n) : \varphi(x_1, x_2, \ldots, x_n) > \lambda\}$$

  的精确检验水平不超过 α，是一个检验水平为 α 的否定域。
- 样本值落入否定域 W 的充分必要条件是 p 值小于等于 α，与引理 3.1 的做法只有微小的的差别（引理 3.1 要求 p 值严格小于 α）。
- 这也叫做 p 值方法。

例 3.2

- **例 3.2** 设 $X \sim \mathrm{N}(\mu, \sigma^2)$。未知 σ，检验 $H_0 : \mu \leq \mu_0$。
- 否定域为

$$W = \{(x_1, x_2, \dots, x_n) : \varphi(x_1, x_2, \dots, x_n) > \lambda_0\}$$

$$\varphi(x_1, x_2, \dots, x_n) = \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}}$$

λ_0 为 $\mathrm{t}(n-1)$ 的 $1-\alpha$ 分位数。

- 计算 p 值（是统计量）

$$\begin{aligned} & p(x_1, x_2, \dots, x_n) \\ = & \sup_{\theta \in \Theta_0} P_\theta \left(\varphi(X_1, X_2, \dots, X_n) \geq \varphi(x_1, x_2, \dots, x_n)\right) \end{aligned}$$

- 这里 $\theta = (\mu, \sigma^2)$, $\Theta_0 = \{(\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}$。

- 当 $\mu \leq \mu_0$ 时

$$\frac{\bar{x}-\mu}{\sqrt{S^2/n}} \geq \frac{\bar{x}-\mu_0}{\sqrt{S^2/n}}$$

- 所以

$$\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \sup_{\theta \in \Theta_0} P_\theta \left(\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \geq \frac{\bar{x}-\mu_0}{\sqrt{S^2/n}} \right) \\
= & P_{(\mu_0, \sigma^2)} \left(\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \geq \frac{\bar{x}-\mu_0}{\sqrt{S^2/n}} \right) \\
= & P\left(T \geq \frac{\bar{x}-\mu_0}{\sqrt{S^2/n}} \right) \qquad (*)
\end{aligned}$$

(其中 T 表示 t$(n-1)$ 分布随机变量)

- 由 (*) 可见（或由引理 3.1 可知）

$$p(x_1, x_2, \ldots, x_n) < \alpha \Longleftrightarrow \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}} > \lambda$$

- 例如，为了检验 $H_0 : \mu \leq 25$，设某样本 $n = 64$，$\bar{x} = 25.9$, $S^2 = 17.3$，则

$$\varphi(x_1, x_2, \ldots, x_n) = \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}} = 1.731$$

$$p(x_1, x_2, \ldots, x_n) = P(T \geq 1.731) = 0.044 < 0.05$$

(这个 p 值可以在 R 软件中用 `1 - pt(1.731, 63)`计算)

- 在 $\alpha = 0.05$ 检验水平下应拒绝 H_0。

双边否定域的 p 值

- 考虑否定域 (3.2):

$$W = \{(x_1, x_2, \ldots, x_n):\ \varphi(x_1, x_2, \ldots, x_n) < \lambda_1 \text{或} \\ \varphi(x_1, x_2, \ldots, x_n) > \lambda_2\} \tag{3.2}$$

- 不去考虑由 α 确定 λ_1 和 λ_2 的问题，直接定义 p 值。
- 根据 “p 值是能够拒绝 H_0 的最小的可取检验水平”，可以找到特定的 λ_0，$\lambda_1 \leq \lambda_0 < \lambda_2$。

- 当 $\varphi(x_1, x_2, \ldots, x_n) \leq \lambda_0$ 时，令

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \min\left\{ \sup_{\theta\in\Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \leq \varphi(x_1, x_2, \ldots, x_n)\right),\ 1 \right\}
\end{aligned}
\tag{3.12}
$$

- 当 $\varphi(x_1, x_2, \ldots, x_n) > \lambda_0$ 时，令

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \min\left\{ \sup_{\theta\in\Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n)\right),\ 1 \right\}
\end{aligned}
\tag{3.13}
$$

- **定义 3.3** 由 (3.12) 和 (3.13) 定义的 $p(x_1, x_2, \ldots, x_n)$ 叫做双边情形下样本值 $(x_1, x_2, \ldots, x_n)$ 的 p 值。

引理 3.3

- **引理 3.3** 设对给定的 $\alpha \in (0,1)$，有唯一的 λ_1 和 λ_2 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) < \lambda_1) = \frac{\alpha}{2} \tag{3.14}$$

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda_2) = \frac{\alpha}{2} \tag{3.15}$$

则

$$\begin{aligned} &\varphi(x_1, x_2, \ldots, x_n) < \lambda_1 \text{ 或} \varphi(x_1, x_2, \ldots, x_n) > \lambda_2 \\ \Longleftrightarrow &p(x_1, x_2, \ldots, x_n) < \alpha. \end{aligned}$$

- 否定域为 p 值小于 α。

引理 3.4

- **引理 3.4** 设对给定的 $\alpha \in (0,1)$，有 λ_1 和 λ_2 满足

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1,X_2,\ldots,X_n) < \lambda_1) \le \frac{\alpha}{2} < \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1,X_2,\ldots,X_n) \le \lambda_1) \tag{3.16}$$

$$\sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1,X_2,\ldots,X_n) > \lambda_2) \le \frac{\alpha}{2} < \sup_{\theta\in\Theta_0} P_\theta(\varphi(X_1,X_2,\ldots,X_n) \ge \lambda_2) \tag{3.17}$$

则

$$\varphi(x_1,x_2,\ldots,x_n) < \lambda_1 \text{ 或} \varphi(x_1,x_2,\ldots,x_n) > \lambda_2$$
$$\Longleftrightarrow p(x_1,x_2,\ldots,x_n) \le \alpha.$$

- 与引理 3.3 的差别是否定域为 p 值小于等于 α。

引理 3.3 的证明

- **引理 3.3 的证明** 设 $\varphi(x_1, x_2, \ldots, x_n) < \lambda_1$，则 $\varphi(x_1, x_2, \ldots, x_n) < \lambda_0$ （只要 $\lambda_0 \in [\lambda_1, \lambda_2)$）。
- 而且存在 $\varepsilon > 0$ 使得 $\varphi(x_1, x_2, \ldots, x_n) < \lambda_1 - \varepsilon$。
- 由 (3.12)

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
\leq & 2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \leq \varphi(x_1, x_2, \ldots, x_n)) \\
\leq & 2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) < \lambda_1 - \varepsilon) < \alpha
\end{aligned}
$$

(由 (3.17) 对 λ_1 的要求及唯一性)

- 若 $\varphi(x_1, x_2, \ldots, x_n) > \lambda_2$, 则 $\varphi(x_1, x_2, \ldots, x_n) > \lambda_0$ （只要 $\lambda_0 \in [\lambda_1, \lambda_2)$）。
- 而且存在 $\varepsilon > 0$ 使得 $\varphi(x_1, x_2, \ldots, x_n) > \lambda_2 + \varepsilon$。
- 由 (3.13) 得

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
\leq & 2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n)) \\
\leq & 2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) > \lambda_2 + \varepsilon) < \alpha
\end{aligned}
$$

(由 (3.17) 对 λ_2 的要求及唯一性)

- 于是我们证明了引理 3.3 的 “$\Longrightarrow$”。

- 反过来，设 $p(x_1, x_2, \ldots, x_n) < \alpha$。
- 若 $\varphi(x_1, x_2, \ldots, x_n) \le \lambda_0$, 则

$$
\begin{aligned}
&p(x_1, x_2, \ldots, x_n) \\
=&2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \le \varphi(x_1, x_2, \ldots, x_n)) \\
&(\text{这是因为 } p(x_1, x_2, \ldots, x_n) < \alpha \text{ 时 (3.12) 不能取 1}) \\
<&\alpha
\end{aligned}
$$

- 由 (3.14) 中 λ_1 的唯一性用反证法可知这时

$$\varphi(x_1, x_2, \ldots, x_n) < \lambda_1$$

- 若 $\varphi(x_1, x_2, \ldots, x_n) > \lambda_0$, 则

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
=& 2 \sup_{\theta \in \Theta_0} P_\theta(\varphi(X_1, X_2, \ldots, X_n) \geq \varphi(x_1, x_2, \ldots, x_n)) \\
& (\text{这是因为 } p(x_1, x_2, \ldots, x_n) < \alpha \text{ 时 (3.13) 不能取 1}) \\
<& \alpha
\end{aligned}
$$

- 由 (3.15) 中 λ_2 的唯一性用反证法可知这时

$$
\varphi(x_1, x_2, \ldots, x_n) > \lambda_2
$$

- 这样，引理 3.3 的必要性和充分性都证明了。

- 从引理 3.3 证明看出，在取定了检验水平 α 以后，λ_0 可以取区间 $[\lambda_1, \lambda_2)$ 中的任何值。
- 这样当可以拒绝 H_0 时，p 值定义不受 λ_0 选取的影响。当 H_0 相容时，(3.12) 和 (3.13) 两个 p 值定义受到 λ_0 选取影响但都会做出相容的判断。
- 如果在引理 3.3 条件下定义

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
=& 2\min\left\{\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \le \varphi(x_1, x_2, \ldots, x_n)\right),\right. \\
& \left.\sup_{\theta\in\Theta_0} P_\theta\left(\varphi(X_1, X_2, \ldots, X_n) \ge \varphi(x_1, x_2, \ldots, x_n)\right)\right\}
\end{aligned}
$$

  则引理 3.3 仍成立，证明类似，但不再需要 λ_0。
- 引理 3.4 证明类似。

- 引理 3.3 和引理 3.4 给出了双边检验问题的 p 只方法。
- 在引理 3.3 条件下，否定域为 p 值小于 α。
- 在引理 3.4 条件下，否定域为 p 值小于等于 α。
- 两种情况的检验水平都不超过 α。
- 优点：适用于任何检验水平；p 值大小给出了拒绝零假设的强烈程度。
- 现代的统计软件中假设检验的结果一般都给出 p 值。

例 3.3

- **例 3.3** 设 $X \sim \mathrm{N}(\mu, \sigma^2)$，$\mu, \sigma^2$ 均未知。
- 检验

$$H_0 : \sigma^2 = \sigma_0^2 \longleftrightarrow H_a : \sigma^2 \neq \sigma_0^2$$

- 用检验统计量

$$\varphi(X_1, X_2, \dots, X_n) = \frac{1}{\sigma_0^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- 当 φ 太大或太小时拒绝 H_0。
- 使用双边否定域

$$\begin{aligned} W = \{ &(x_1, x_2, \dots, x_n) : \varphi(x_1, x_2, \dots, x_n) < \lambda_1 \\ &\text{或} \varphi(x_1, x_2, \dots, x_n) > \lambda_2 \} \end{aligned}$$

- 从直观看 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 可以作为 σ^2 的估计。
- H_0 成立时 $\frac{1}{n-1}\varphi(x_1, x_2, \ldots, x_n)$ 应该与 1 相差不大，$\varphi(x_1, x_2, \ldots, x_n)$ 应该与 $n-1$ 相差不大。
- λ_1 应该小于 $n-1$，λ_2 应该大于 $n-1$。取 $\lambda_0 = n-1$，则 $\lambda_1 < n-1 < \lambda_2$。
- $\varphi(x_1, x_2, \ldots, x_n) \leq n-1 \Leftrightarrow S^2 \leq \sigma_0^2$.

- 当 $S^2 \le \sigma_0^2$ 时定义

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \min\left\{ 2 \sup_{\substack{\mu \in (-\infty, \infty) \\ \sigma^2 = \sigma_0^2}} P_{(\mu, \sigma^2)}(\varphi(X_1, X_2, \ldots, X_n) \le \varphi_0), 1 \right\} \\
= & \min\{2P(\xi \le \varphi_0), 1\}
\end{aligned}
$$

- 其中 $\varphi_0 = \varphi(x_1, x_2, \ldots, x_n)$, ξ 为 $\chi^2(n-1)$ 随机变量。
- 当 $S^2 > \sigma_0^2$ 时类似有

$$
p(x_1, x_2, \ldots, x_n) = \min\{2P(\xi \ge \varphi_0), 1\}
$$

- 回到例 2.4 的 10 个铜丝折断力数据，要检验 $\sigma^2 = 64$。
- $S^2 = 75.7 > 64$。
- $\varphi_0 = \frac{1}{\sigma_0^2}\sum_{i=1}^{10}(x_1 - \bar{x})^2 = 10.65.$
- $P(\xi \geq \varphi_0) = 0.30(\xi \sim \chi^2(9))$。
- p 值为 0.60。
- 只要检验水平 α 不大于 0.60，H_0 都是相容的。
- 但是注意：假设检验应该预先确定检验水平，而不能看到 p 值后再选检验水平。

假设检验与置信区间的联系

- 在 §6.2.2 讲关于 $\mu = \mu_0$ 的检验时，我们讨论了检验的接受域与 μ 的置信区间的联系。
- 这种联系是一般性的。
- 设 X 的分布函数为 $F(x, \theta)$，θ 是未知参数, $\theta \in \Theta$, $X_1, X_2, \ldots, X_n$ 是 X 的样本。
- 对任何 $\theta_0 \in \Theta$, 考虑假设问题

$$H_0 : \theta = \theta_0 \longleftrightarrow H_a : \theta \neq \theta_0$$

- 设 $A(\theta_0)$ 是 H_0 的检验水平为 α 的接受域（其补集是 H_0 的检验水平为 α 的否定域）。

- 即当且仅当 $(x_1, X_2, \ldots, x_n) \in A(\theta_0)$ 时接受 $H_0: \theta = \theta_0$，且

$$P_{\theta_0}((X_1, X_2, \ldots, X_n) \notin A(\theta_0)) \leq \alpha$$

$$P_{\theta_0}((X_1, X_2, \ldots, X_n) \in A(\theta_0)) \geq 1 - \alpha \qquad (*)$$

- 令

$$S(x_1, x_2, \ldots, x_n) = \{\theta : (x_1, x_2, \ldots, x_n) \in A(\theta)\} \qquad (3.18)$$

- 则

$$(x_1, x_2, \ldots, x_n) \in A(\theta_0) \Longleftrightarrow \theta_0 \in S(x_1, x_2, \ldots, x_n) \qquad (**)$$

- 由 (*) 和 (**)，得

$$P_{\theta_0}(\theta_0 \in S(X_1, X_2, \ldots, X_n)) \geq 1 - \alpha$$

- 其中的 θ_0 是任意的，所以

$$P_\theta(\theta \in S(x_1, x_2, \ldots, x_n)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

- 可见，θ 的集合 $S(x_1, x_2, \ldots, x_n)$ 如果是个区间，则它是 θ 的置信水平为 $1-\alpha$ 的置信区间。
- 我们可以用 $\theta = \theta_0$ 的检验接受域来构造 θ 的置信区间。
- 反过来，如果统计量 $\underline{\theta}(X_1, X_2, \ldots, X_n)$ 和 $\overline{\theta}(X_1, X_2, \ldots, X_n)$ 使得

$$P_\theta\left(\theta \in (\underline{\theta}, \overline{\theta})\right) \geq 1-\alpha, \quad \forall \theta \in \Theta$$

则 $\theta_0 \in (\underline{\theta}, \overline{\theta})$ 可以作为 $H_0 : \theta = \theta_0$ 的检验水平为 α 的接受域。

例 3.4

- **例 3.4** 设 $X \sim \mathrm{N}(\theta, 1)$，$\theta \in (-\infty, \infty)$。
- $(X_1, X_2, \ldots, X_n)$ 是 X 的样本。
- 对任何 θ_0，考虑检验问题

$$H_0 : \theta = \theta_0 \longleftrightarrow H_a : \theta \neq \theta_0$$

- 从 §1 知可用接受域

$$A(\theta_0) = \{(x_1, x_2, \ldots, x_n) : |\bar{x} - \theta_0| \leq c\}$$

- 其中

$$c = \sqrt{\frac{1}{n}} z_{1-\frac{\alpha}{2}}$$

- z_p 表示标准正态分布的 p 分位数：

$$\begin{aligned} z_p =& \Phi^{-1}(p) \\ \Phi(p) =& p \end{aligned}$$

- 这时检验的精确水平为 α。
- 有

$$P_{\theta_0}(|\bar{X} - \theta_0| \leq c) = 1 - \alpha$$

- 因 θ_0 任意，所以

$$P_{\theta}(|\bar{X} - \theta| \leq c) = 1 - \alpha, \quad \forall \theta \in (-\infty, \infty)$$

- 于是

$$P_\theta(\bar{X} - c \le \theta \le \bar{X} + c) = 1 - \alpha, \quad \forall \theta \in (-\infty, \infty)$$

- 从接受域得到了 θ 的置信水平为 $1-\alpha$ 的置信区间。
- 反过来，当且仅当 θ_0 属于这个置信区间时接受 $H_0 : \theta = \theta_0$，检验水平为 α。

本节目录

两个正态总体的假设检验

- 设 $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, 且 X, Y 相互独立。
- (1) 未知 σ_1^2, σ_2^2，但知道 $\sigma_1^2 = \sigma_2^2$，检验假设

$$H_0 : \mu_1 = \mu_2$$

- (2) 未知 μ_1, μ_2，检验假设

$$H_0 : \sigma_1^2 = \sigma_2^2$$

- (3) 未知 μ_1, μ_2，检验假设

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$

- (4) 未知 σ_1^2, σ_2^2，但知道 $\sigma_1^2 \neq \sigma_2^2$，检验假设

$$H_0 : \mu_1 = \mu_2$$

独立两样本 t 检验

- 在已知两个独立正态总体方差相等时检验两个总体的期望是否相等，使用“独立两样本 t 检验”。
- 在实际问题中，用来比较独立的两组的同一属性平均来说有无显著差异。
- **例 4.1(即例 1.4)** 在针织品的漂白工艺过程中，要考察温度对针织品的断裂强力（主要质量指标）的影响。
- 为了比较 70°C 与 80°C 的影响有无差别，在这两个温度下，分别重复做了 8 次试验，得数据如下（单位：千克力）

$$
\begin{aligned}
70^{\circ}\mathrm{C}\text{时的强力：} & 20.5,\ 18.8,\ 19.8,\ 20.9, \\
& 21.5,\ 19.5,\ 21.0,\ 21.2 \\
80^{\circ}\mathrm{C}\text{时的强力：} & 17.7,\ 20.3,\ 20.0,\ 18.8, \\
& 19.0,\ 20.1,\ 20.2,\ 19.1
\end{aligned}
$$

- 用 X, Y 分别表示 $70°\mathrm{C}$ 与 $80°\mathrm{C}$ 下的断裂强力，试验结果按常识判断是独立的。
- 根据过去的经验，可以认为 X, Y 分别服从正态分布且方差相等。
- 即 $X \sim \mathsf{N}(\mu_1, \sigma^2)$，$Y \sim \mathsf{N}(\mu_2, \sigma^2)$，要检验 $H_0 : E(X) = E(Y)$，即 $H_0 : \mu_1 = \mu_2$。
- 自然想到比较两个样本的平均值。$70°\mathrm{C}$ 的样本的平均强力为 20.4 千克力，$80°\mathrm{C}$ 的样本的平均强力为 19.4 千克力，$70°\mathrm{C}$ 的平均值高出 1 千克力。

- 能否就此断言 70°C 下的总体期望值 $E(X)$ 与 80°C 的总体期望值 $E(Y)$ 不同？
- 要注意的是，我们只是在一组样本中观测到了样本平均值有 1 千克力的差距。换一组样本，差距可能就变了。
- 样本得到的两个样本均值的差即使在 $\mu_1 = \mu_2$ 时也一般不会等于零。
- 如果这个差距很大，我们可以比较有把握地说两个总体期望不同。
- 如果差距很小，我们可以认为两个总体期望相同。
- 问题是，如何找这个临界值？
- 思路是找到问题中的随机性的分布，然后在 $\mu_1 = \mu_2$ 情况下产生较大差距为小概率事件时否定 H_0。

两样本 t 检验的推导

- 设 $x_1, x_2, \ldots, x_n$ 是来自 X 的样本值，$y_1, y_2, \ldots, y_n$ 是来自 Y 的样本值。
- 设法研究样本平均值的差

$$\bar{x} - \bar{y}$$

  的分布，当此差超出了 $\mu_1 = \mu_2$ 时随机变化的正常范围时否定 $H_0 : \mu_1 = \mu_2$。
- 还是“带概率的反证法”的思想。先假设 $\mu_1 = \mu_2$，看是否有小概率事件发生。
- 在假设 $\mu_1 = \mu_2$ 的条件下，$\bar{X} - \bar{Y}$ 的分布仍与未知的 σ^2 有关。其方差等于

$$\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n}$$

- 于是

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{2\sigma^2}{n}}} \sim \mathsf{N}(0,1)$$

- 这不是统计量。估计 σ^2 为

$$\hat{\sigma}^2 = \frac{1}{2n-2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]$$

  称为合并方差估计。

- 在 $H_0:\mu_1=\mu_2$ 成立时统计量

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{2\hat{\sigma}^2}{n}}} \sim \mathsf{t}(2n-2) \tag{4.2}$$

- 查 t$(2n-2)$ 分布临界值表可得 λ 满足

$$P(|T|>\lambda)=\alpha$$

- 从样本中计算统计量 T 的值，当 $|T|>\lambda$ 时拒绝 $H_0:\mu_1=\mu_2$，当 $|T|\leq\lambda$ 时 H_0 相容。
- 称为**两样本 t 检验**。

例 4.1（续）

- 把两样本 t 检验一般步骤应用于例 4.1。
- (1) 提出待检验的假设:

$$H_0 : E(X) = E(Y) \longleftrightarrow H_a : E(X) \neq E(Y)$$

- (2) 计算 t 统计量的值。

$$n = 8 \quad \bar{x} = 20.4 \quad \bar{y} = 19.4$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 6.20 \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 5.80$$

$$\hat{\sigma}^2 = \frac{1}{2 \times 8 - 2}(6.20 + 5.80) = 0.8571$$

$$T = \frac{20.4 - 19.4}{\sqrt{\frac{2*0.8571}{8}}} = 2.16$$

- (3) 查 t 分布表，自由度是 $2n-2=14$，取 $\alpha=0.05$，得 $\lambda=2.145$。
- (4) 下结论：$|T|=2.16>\lambda$ 所以否定零假设，认为 $70°\mathrm{C}$ 下的总体期望值 $E(X)$ 与 $80°\mathrm{C}$ 的总体期望值 $E(Y)$ 不等，而且是 $70°\mathrm{C}$ 下的强力更大。
- 第 (3)、(4) 步也可以计算 p 值:

$$p(x_1,x_2,\dots,x_n)=P(|\xi|\geq|2.16|)=0.0486<\alpha$$

其中 ξ 为服从 t$(2n-2)$ 分布的随机变量。

两样本的样本量不等情况

- 设 $X_1, X_2, \ldots, X_{n_1}$ 是来自 $\mathrm{N}(\mu_1, \sigma^2)$ 的样本，$Y_1, Y_2, \ldots, Y_{n_2}$ 是来自 $\mathrm{N}(\mu_2, \sigma^2)$ 的样本。
- 这时定义 σ^2 的估计为 (合并方差估计)

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2\right]$$

- 检验统计量为

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

- 在 $H_0 : \mu_1 = \mu_2$ 下 $T \sim \mathrm{t}(n_1 + n_2 - 2)$, 使用、$\mathrm{t}(n_1 + n_2 - 2)$ 分布临界值，当且仅当 $|T| > \lambda$ 时拒绝 H_0。

- 这个检验叫做**两样本 t 检验**，也叫平均数的显著性鉴定。
- 如果拒绝了 $H_0 : \mu_1 = \mu_2$，一般称（在 α 水平下）两个总体的平均数 **有显著 (性) 差异**。

例 4.2

- **例 4.2** 研究口服避孕药对妇女血压影响。
- 对某公司工作的 35 岁至 39 岁的非怀孕妇女，用抽查方法收集到如下数据。
- 有 8 人使用口服避孕药，其收缩压平均值为 132.86（mmHg），标准差为 15.35。
- 有 21 人未使用，收缩压平均值为 127.44(mmHg)，标准差为 18.23。
- 问：这两种血压的平均值的差异是否显著？

- **解** 假设使用口服避孕药的妇女的收缩压总体为 $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$，假设不使用的妇女的收缩压总体为 $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，并假定 $\sigma_1^2 = \sigma_2^2 = \sigma^2$。
- 计算得

$$\begin{aligned}\hat{\sigma}^2 =& \frac{1}{8+21-2}((8-1)\times 15.35^2 + (21-1)\times 18.23^2) \\ =& 294.95\end{aligned}$$

- T 统计量为

$$T = \frac{132.86 - 127.44}{\sqrt{\left(\frac{1}{8} + \frac{1}{21}\right) 294.95}} = 0.760$$

- 查 27 个自由度的 t 分布的 $\alpha = 0.05$ 临界值得 $\lambda = 2.052$，$|T| < \lambda$ 所以 H_0 相容，两个平均值无显著差异。

成对数据的比较

- 有些实际问题中的数据是成对的，比如同一个人两次测量同一指标。
- 这时，两个变量 X 和 Y 一般是不独立的，不能使用上述的两样本 t 检验。
- 设 (X,Y) 的样本为 $(X_1,Y_1),(X_2,Y_2),\dots,(X_n,Y_n)$。
- 考虑新的总体 $Z=X-Y$，样本为 $Z_i=X_i-Y_i, i=1,2,\dots,n$。
- 如果 (X,Y) 服从联合正态分布则 Z 也服从正态分布。
- 要检验 $H_0: E(X)=E(Y)$，只要检验等价的假设 $E(Z)=0$。
- 问题化为单样本 t 检验问题。

例 4.3

- **例 4.3** 为了鉴定两种工艺方法生产的产品某性能指标有无显著差异，对于 9 批材料分别用两种工艺进行生产，得到该指标的 9 对数据如下：

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：根据以上数据，能否说两种工艺生产的产品性能指标有显著差异？（检验水平 $\alpha = 0.05$）

- 计算 9 对数据的差：

0.10	0.09	-0.12	0.18	-0.18	0.11	0.12	0.13	0.11

- 设差的总体为 $Z \sim \mathrm{N}(\mu, \sigma^2)$，检验 $H_0 : \mu = 0$。
- 用 §2 中的 t 检验，计算得

$$T = \frac{\bar{x}}{\sqrt{S^2/9}} = 1.467$$

- 查 t(8) 分布临界值表得 $\alpha = 0.05$ 对应临界值 $\lambda = 2.306$。
- 或计算 p 值:

$$p = P(|\xi| > 1.467) = 0.19$$

- 现在 $|T| < \lambda$（p 值小于 α），所以 H_0 相容，两种工艺方法生产的产品性能指标无显著差异。
- 也称这两种工艺方法对产品该性能指标**无显著影响**。

方差的双边检验

- 设 $X_1, X_2, \ldots, X_{n_1}$ 是来自总体 $\mathrm{N}(\mu_1, \sigma_1^2)$ 的样本，$Y_1, Y_2, \ldots, Y_{n_2}$ 是来自总体 $\mathrm{N}(\mu_2, \sigma_2^2)$ 的样本，两个总体独立。
- 要求检验

$$H_0 : \sigma_1^2 = \sigma_2^2$$

- 用“带概率的反证法”。先假设 H_0 成立。
- 要比较 σ_1^2 和 σ_2^2，想到比较其估计量：

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

- 取检验统计量

$$F = \frac{S_1^2}{S_2^2}$$

当 F 远大于 1 或远小于 1 时拒绝 H_0。

- 如果 F 的分布不依赖于未知参数，则可以取 λ_1 和 λ_2 使得

$$P(F < \lambda_1) = \frac{\alpha}{2}, \quad P(F > \lambda_2) = \frac{\alpha}{2}$$

当 $F < \lambda_1$ 或 $F > \lambda_2$ 时拒绝 H_0 （这样第 I 类错误概率等于 α，是水平 α 的检验法）

F 分布

- 如果随机变量 Z 有如下分布密度

$$
\begin{aligned}
& f_{n_1,n_2}(u) \\
=& \begin{cases} \dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} u^{\frac{n_1}{2}-1} \left(1+\dfrac{n_1}{n_2}u\right)^{-\frac{n_1+n_2}{2}}, & u>0 \\ 0 & u \le 0 \end{cases}
\end{aligned}
$$

  则称 Z 服从自由度为 n_1, n_2 的 F 分布，这里 n_1, n_2 是两个正整数，分别称为第一自由度和第二自由度（或分子自由度和分母自由度），记 $Z \sim \mathrm{F}(n_1, n_2)$。

- 上面的统计量 F 在 $H_0 : \sigma_1^2 = \sigma_2^2$ 成立时服从 $\mathrm{F}(n_1 - 1, n_2 - 1)$ 分布。

F 分布的临界值

- F 分布的临界值在第 434–439 页附表 4–附表 6 中，只给出了右侧临界值：

$$P(F > \lambda) = \alpha$$

- F 分布有性质：

$$F \sim \mathsf{F}(n_1 - 1, n_2 - 1) \Longrightarrow \frac{1}{F} \sim \mathsf{F}(n_2 - 1, n_1 - 1)$$

- 所以对于 $H_0 : \sigma_1^2 = \sigma_2^2$ 的检验，为求临界值 λ_1 和 λ_2，先查自由度为 $(n_1 - 1, n_2 - 1)$ 的水平 $\alpha/2$ 的临界值 λ_2。
- 再查自由度为 $(n_2 - 1, n_1 - 1)$ 的水平 $\alpha/2$ 的临界值 λ_1'，并令 $\lambda_1 = 1/\lambda_1'$。

- **例 4.1(续)** 考虑 70°C 和 80°C 下强力的方差的比较。
- 这里 $n_1 = n_2 = 8$，$s_1^2 = 6.20/7$, $s_2^2 = 5.80/7$,

$$F = \frac{s_1^2}{s_2^2} = \frac{6.20}{5.80} = 1.07$$

- 查 F 分布表，取 $\alpha = 0.05$，查自由度 $(7, 7)$ 的 0.025 临界值得

$$\lambda_2 = 4.99$$

对于 λ_1，这里自由度交换后仍为 $(7, 7)$，得

$$\lambda_1 = \frac{1}{4.99} = 0.200$$

- 现在 $\lambda_1 < F < \lambda_2$ 所以 H_0 相容，在 0.05 水平下不能否认 70°C 下和 80°C 下有相同的方差（或称：70°C 下和 80°C 下的方差**无显著差异**）。

例 4.4（例 4.2 续）

- **例 4.4（例 4.2 续）** 检验使用口服避孕药和不使用的妇女的血压的方差是否相等：

$$H_0: \sigma_1^2 = \sigma_2^2 \longleftrightarrow H_a: \sigma_1^2 \neq \sigma_2^2$$

- 这里 $n_1 = 8$, $n_2 = 21$, $S_1^2 = (15.35)^2$, $S_2^2 = (18.23)^2$，

$$F = \frac{S_1^2}{S_2^2} = 0.709$$

- 查自由度 $(7, 20)$ 的 0.025 水平 F 分布临界值得 $\lambda_2 = 3.01$, 查自由度 $(20, 7)$ 的 0.025 水平 F 分布临界值得 $\lambda_1 = 1/4.42 = 0.226$。
- 现在 $\lambda_1 < F < \lambda_2$，所以 H_0 相容，在 0.05 水平下不能否认两组人的血压有相同的方差（或称：两组人的方差**无显著差异**）。

关于查表

- 上例中 λ_2 在附表 5 中直接查到。而 $1/\lambda_1$ 需要找自由度 $(20, 7)$ 情况下的临界值，表中 $n_2 = 7$ 的行存在，但 $n_1 = 20$ 的列不存在。
- 怎么处理？
- 简化的做法是找最近的表格值：最近的自由度是 $(24, 7)$，表格值为 4.42。
- 更精细一点的做法是用线性插值近似：要求解的临界值和 $(12, 7)$ 自由度下的 4.67 以及 $(24, 7)$ 自由度下的 4.42 最接近，按照自由度接近程度进行插值：

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

- 如果两个自由度都不在表格中，这种方法要两次插值，而且精度也只是略有提高。
- 用统计软件计算更容易。如在 R 中

两总体方差单边检验

- 考虑未知 μ_1, μ_2，$H_0 : \sigma_1^2 \le \sigma_2^2$ 检验问题。
- **例 4.5** 有两台车床生产同一种型号的滚珠。可认为直径分别服从正态分布。
- 从这两台车床生产的产品中分别抽取 8 个和 9 个，测得滚珠直径如下（毫米）：

$$甲车床: 15.0, 14.5, 15.2, 15.5, 14.8, 15.1, 15.2, 14.8$$

$$乙车床: 15.2, 15.0, 14.8, 15.2, 15.0, 15.0, 14.8, 15.1, 14.8$$

- 问：乙车床产品直径的方差是否比甲车床的小？

- **解** 用 X, Y 分别表示甲、乙两车床的产品的直径。
- 设 $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，$X, Y$ 独立。
- 问题是想判断 $\sigma_1^2 > \sigma_2^2$ 是否成立，把它作为对立假设：

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \longleftrightarrow H_a : \sigma_1^2 > \sigma_2^2$$

- 这样在拒绝 H_0 后就可使断言乙车床的方差较小。

方差单边检验的推导

- 设 $X_1, X_2, \ldots, X_{n_1}$ 是来自总体 X 的样本，$Y_1, Y_2, \ldots, Y_{n_2}$ 是来自总体 Y 的样本，$X \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathsf{N}(\mu_2, \sigma_2^2)$，$X, Y$ 独立。
- 要检验

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \longleftrightarrow H_a : \sigma_1^2 > \sigma_2^2$$

- H_0 相当于 $\frac{\sigma_1^2}{\sigma_2^2} \leq 1$，用统计量

$$F = \frac{S_1^2}{S_2^2}$$

当 F 超过 1 且很大时拒绝 H_0。

- F 在 H_0 下的分布依赖于 σ_1^2, σ_2^2 的值。
- 令

$$\tilde{F} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

则 $\tilde{F} \sim \mathsf{F}(n_1 - 1, n_2 - 1)$, 且在 $H_0 : \sigma_1^2 \leq \sigma_2^2$ 下

$$\tilde{F} = \frac{\sigma_2^2}{\sigma_1^2} F \geq F$$

- 取 λ 使

$$P(\tilde{F} > \lambda) = \alpha$$

则

$$F > \lambda \Longrightarrow \tilde{F} > \lambda$$

$$P(F > \lambda) \leq P(\tilde{F} > \lambda) = \alpha$$

- 于是

- 回到例 4.5。
- 第一步：提出待检验的假设 $H_0 : \sigma_1^2 \leq \sigma_2^2$。
- 第二步：计算统计量 $F = S_1^2/S_2^2$ 的值。

$$
\begin{aligned}
& n_1 = 8, \quad n_2 = 9 \\
& S_1^2 = 0.09554, \quad S_2^2 = 0.02611 \\
& F = S_1^2/S_2^2 = 3.66
\end{aligned}
$$

- 查 $\alpha = 0.05$ 的 F$(7, 8)$ 分布右侧临界值得 $\lambda = 3.50$。
- 现在 $F = 3.66 > 3.50$，在 0.05 水平下否定零假设，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

- **例 4.6** 赈灾捐赠的男女差异。随机抽查了 25 个男士，平均捐赠 12.40 美元，标准差 2.50 美元；随机抽查了 25 个女士，平均捐赠 8.90 美元，标准差 1.34 美元。
- 问：男士捐赠额的方差是否大于女士捐赠额的方差？
- **解** 设一个男士的捐赠额 X 服从 $\mathrm{N}(\mu_1, \sigma_1^2)$，一个女士的捐赠额 Y 服从 $\mathrm{N}(\mu_2, \sigma_2^2)$，问题 $\sigma_1^2 > \sigma_2^2$ 作为对立假设：

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \longleftrightarrow H_a : \sigma_1^2 > \sigma_2^2$$

- 计算 F 统计量得 $F = (2.50)^2/(1.34)^2 = 3.48$，查自由度 $(24, 24)$ 的 F 分布水平 0.01 的临界值得 $\lambda = 2.66$。$F = 3.48 > 2.66$ 所以拒绝 H_0，在 0.01 水平下男士捐赠额的方差显著大于女士捐赠额的方差。

方差不等时均值的比较

- 如果未知 σ_1^2, σ_2^2 但知道 $\sigma_1^2 \neq \sigma_2^2$，检验 $H_0: \mu_1 = \mu_2$，这个问题称为 Behrens–Fisher 问题。
- 可以建立一个统计量分布在 H_0 下近似服从 t 分布的检验法。
- 设 $X_1, X_2, \ldots, X_{n_1}$ 是来自总体 X 的样本，$Y_1, Y_2, \ldots, Y_{n_2}$ 是来自总体 Y 的样本，$X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，$\sigma_1^2 \neq \sigma_2^2$, X, Y 独立。

- 令

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \qquad \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$$

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

- 易知

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

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0, 1)$$

- 在 $H_0: \mu_1 = \mu_2$ 下

$$\xi = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0, 1)$$

- 在 $|\xi|$ 太大时应该拒绝 H_0。但是 ξ 不是统计量。
- 用估计量代替 σ_1^2, σ_2^2，得统计量

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

- 在 H_0 下 T 的精确分布复杂且依赖于 σ_1^2/σ_2^2 的值，但 T 近似服从 $\mathrm{t}(m^*)$:

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

- 查表时可取 m^* 为最近的整数。查 t 分布临界值表找到 λ 使 $P(|T|>\lambda)=\alpha$，当且仅当 $|T|>\lambda$ 时拒绝 $H_0:\mu_1=\mu_2$。
- 类似可解决当 $\sigma_1^2\neq\sigma_2^2$ 时 $H_0:\mu_1\leq\mu_2$ 的检验问题。

例 4.7

- **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。
- 随机调查了 100 个 2 到 14 岁的孩子（父亲死于心脏病），其胆固醇水平平均值为 207.3，标准差为 35.6;
- 另外随机调查了父亲无心脏病史的 74 个 2 至 14 岁的孩子，其胆固醇水平平均值为 193.4，标准差为 17.3。
- 问：前者的胆固醇水平的平均值与后者的胆固醇水平的平均值是否有显著差异？

- **解** 设父亲死于心脏病的孩子的胆固醇水平 $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, 父亲无心脏病史的孩子的胆固醇水平 $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$，其中 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 都未知，要检验 $H_0 : \mu_1 = \mu_2 \longleftrightarrow H_a : \mu_1 \neq \mu_2$
- 首先判断 $\sigma_1^2 = \sigma_2^2$ 是否相等。设 $\alpha = 0.05$，计算得 $F = 4.23$，查表得 $\lambda_1 = 0.6548$, $\lambda_2 = 1.5491$，$F = 4.23 > \lambda_2$，所以方差有显著差异。
- 计算 T 统计量

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = 3.40$$

- 计算得 $m^* = 151.4$，$m = 151$，查 t(151) 的双侧 0.05 水平临界值得 $\lambda = 1.980$(使用 t(120) 的临界值)。现在 $|T| = 3.40 > \lambda$ 所以拒绝 H_0，认为在 0.05 水平下胆固醇水平的平均值有显著差异，父亲死于心脏病的孩子的胆固醇水平更高。

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 $X \sim f(x)$，则

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

- t(n) 分布的分布密度为

$$p_X(x) = \frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

- 所以

$$\begin{aligned}
p_X(\sqrt{y}) =& p_X(-\sqrt{y}) = \frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y}{n}\right)^{-\frac{n+1}{2}} \\
p_Y(y) =& \frac{1}{\sqrt{y}}\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y}{n}\right)^{-\frac{n+1}{2}} \\
=& \frac{\Gamma\left(\frac{1+n}{n}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{n}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1}\left(1+\frac{y}{n}\right)^{-\frac{n+1}{2}} \\
\sim & \mathsf{F}(1,n)
\end{aligned}$$

- 于是

$$P(|X|>\lambda) = P(Y>\lambda^2)$$

本节目录

比率的假设检验

- 设 $X \sim \mathrm{b}(1,p)$, $0<p<1$ 是未知参数，p 就是“比率”，如成功率、失败率、有效率等。
- 本节包括：
- 单总体比率的单边、双边假设检验；
- 两总体比率的单边、双边假设检验。

单总体比率右侧假设检验

- 设总体 $X \sim \mathrm{b}(1, p)$，$X_1, X_2, \ldots, X_n$ 为样本。
- 检验单边假设问题 $H_0 : p \leq p_0 \longleftrightarrow H_a : p > p_0$
- 用统计量

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

当 $\hat{p}$ 超过 p_0 很多时拒绝 H_0。

- 改用统计量 $S = \sum_{i=1}^{n} X_i$ 当 S 很大时拒绝 H_0。
- 否定域是 $W = \{S \geq c\}$

- 临界值 c 取为满足

$$\sup_{p \le p_0} P_p(S \ge c) \le \alpha \tag{5.1}$$

的最小整数。

- $S \sim B(n,p)$。有恒等式

$$\begin{aligned} P_p(S \ge k) &= \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} \\ &= \frac{n!}{(k-1)!(n-k)!} \int_0^p u^{k-1}(1-u)^{n-k} du \end{aligned}$$

- $P_p(S \ge c)$ 作为 p 的函数是增函数，所以 (5.1) 化为

$$P_{p_0}(S \ge c) \le \alpha \tag{5.2}$$

- 求临界值 c 比较麻烦，我们用 p 值来表示否定域。
- 设样本值为 $x_1, x_2, \ldots, x_n$，记 $S_0 = \sum_{i=1}^{n} x_i$，则 p 值为

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \sup_{p \le p_0} P_p(S \ge S_0) = P_{p_0}(S \ge S_0) \\
= & \sum_{i=S_0}^{n} C_n^i p_0^i (1-p_0)^{n-i}
\end{aligned}
$$

- 当且仅当 p 值小于等于 α：

$$
\sum_{i=S_0}^{n} C_n^i p_0^i (1-p_0)^{n-i} \le \alpha \tag{5.3}
$$

  时拒绝 H_0。
- 统计软件中可以计算这样的 p 值。

- (5.3) 左边是 p 的严格单调递增函数，为了判断 (5.3) 是否成立，还可以解方程

$$\sum_{i=S_0}^{n} C_n^i p_0^i (1-p_0)^{n-i} = \alpha$$

得唯一解 $p_\alpha(S_0)$($S_0 \geq 1$ 时)。

- $S_0 = 0$ 时规定 $p_\alpha(0) = 0$。
- 这样 $S_0 \geq c$ 当且仅当 $p_0 \leq p_\alpha(S_0)$。
- 设 $F_b(n_1, n_2)$ 表示 F(n_1, n_2) 分布的 b 分位数，则

$$\begin{aligned} & p_\alpha(S_0) \\ = & \left\{1 + \frac{n - S_0 + 1}{S_0} F_{1-\alpha}(2(n - S_0 + 1), 2S_0)\right\}^{-1} \end{aligned} \tag{5.5}$$

例 5.1

- 原药物有效率为 0.80，制药公司声称新药有效率高于 0.80，且药价更低。
- 收集了临床数据，使用新药的病人中随机抽查了 30 人，其中 27 人有效。
- 问：能否认为新药的有效率高于 0.80？

- **解** 用 X 表示新药的效果，$X=1$ 表示有效，$X=0$ 表示无效，$p=P(X=1)$。
- $n=30$，$S_0=\sum_{i=1}^n x_i=27$。
- 取检验水平 $\alpha=0.05$，从 (5.5) 得

$$\begin{aligned} p_\alpha(S_0) =& p_{0.05}(27) = \left\{1+\frac{4}{27}F_{0.95}(8,54)\right\}^{-1} \\ =& \left(1+\frac{4}{27}\times 2.13\right)^{-1} = 0.76 < p_0 = 0.80 \end{aligned}$$

- 所以不能拒绝 $H_0: p \le p_0$，没有理由说新药比原来药物有更高的有效率。
- 用 R 软件可以计算 p 值:

```
1 - pbinom(26, 30, 0.8)
```

结果为 0.1227，超过 α，结果不显著。

- 如果 30 人中有 28 人有效，则 p 值为 0.0442，在 0.05 水平下显著。
- 这个问题结果不显著，但是不能断言新药不如老药，有可能增大样本量后可以得到显著结果。

单总体比率左侧假设检验

- 再来考虑单总体假设检验

$$H_0 : p \geq p_0 \longleftrightarrow H_a : p < p_0$$

- 否定域为

$$W = \{S \leq c\}$$

- c 是满足

$$\sup_{p \geq p_0} P_p(S \leq c) \leq \alpha \qquad (5.6)$$

  的最大整数。

- $P_p(S \leq c)$ 是 p 的严格单调递减函数。(5.6) 化为

$$P_{p_0}(S \leq c) \leq \alpha \qquad (5.7)$$

- 用 p 值方法。p 值为

$$
\begin{aligned}
p(x_1, x_2, \ldots, x_n) =& P_{p_0}(S \le S_0) \\
=& \sum_{i=0}^{S_0} C_n^i p_0^i (1-p_0)^{n-i}
\end{aligned}
$$

- 当且仅当 p 值小于等于 α 时拒绝 H_0。
- 也可以解方程

$$
\sum_{i=0}^{S_0} C_n^i p_0^i (1-p_0)^{n-i} = \alpha \tag{5.9}
$$

得 $\tilde{p}_\alpha(S_0)$，当且仅当 $\tilde{p}_\alpha(S_0) \le p_0$ 时拒绝 H_0。

单总体比率的双侧检验

- 设 $X_1, X_2, \ldots, X_n$ 是 $X \sim \mathsf{B}(n,p)$ 的样本，$S = \sum_{i=1}^{n}$。
- 检验

$$H_0 : p = p_0 \longleftrightarrow H_a : p \neq p_0$$

- 当 S 太大或太小时拒绝 H_0。否定域为

$$W = \{S \leq c_1 \text{或} S \geq c_2\}$$

- 其中 c_1, c_2 取为最大的整数 c_1 和最小的整数 c_2 满足

$$P_{p_0}(S \leq c_1) \leq \frac{\alpha}{2}, \quad P_{p_0}(S \geq 2_1) \leq \frac{\alpha}{2}$$

- 用 p 值方法。设样本值为 $x_1, x_2, \ldots, x_n$， $S_0 = \sum_{i=1}^{n} x_i$。
- p 值为

$$\begin{aligned} & p(x_1, x_2, \ldots, x_n) \\ = & 2\min\left\{P_{p_0}(S \leq S_0), P_{p_0}(S \geq S_0)\right\} \end{aligned}$$

- 当且仅当 p 值小于等于 α 时拒绝 H_0。

两总体比率比较

- 设两个总体 X 与 Y 相互独立，$X \sim \mathrm{b}(1, p_1)$, $Y \sim \mathrm{b}(1, p_2)$, 样本分别为 $X_1, X_2, \ldots, X_{n_1}$, $Y_1, Y_2, \ldots, Y_{n_2}$。
- 考虑如下三个假设检验问题：

$$H_0 : p_1 \le p_2 \longleftrightarrow H_a : p_1 > p_2$$

$$H_0 : p_1 \ge p_2 \longleftrightarrow H_a : p_1 < p_2$$

$$H_0 : p_1 = p_2 \longleftrightarrow H_a : p_1 \ne p_2$$

- 令

$$S_1 = \sum_{i=1}^{n_1} X_i \qquad S_2 = \sum_{i=1}^{n_2} Y_i$$

- 则

$$\hat{p}_1 = \frac{S_1}{n_1} \qquad \hat{p}_2 = \frac{S_2}{n_2}$$

  分别是 p_1 和 p_2 的估计。
- 当 $\hat{p}_1$ 远大于 $\hat{p}_2$ 时拒绝 $H_0 : p_1 \le p_2$;
- 当 $\hat{p}_1$ 远小于 $\hat{p}_2$ 时拒绝 $H_0 : p_1 \ge p_2$;
- 当 $\hat{p}_1$ 和 $\hat{p}_2$ 相差很多时拒绝 $H_0 : p_1 = p_2$。
- 使用两种方法：大样本情形的正态近似方法和 Fisher 精确检验方法。

正态近似法

- 由注意到 $\hat{p}_1$ 和 $\hat{p}_2$ 分别是两个样本的样本平均值，所以

$$D(\hat{p}_1) = \frac{D(X)}{n_1} = \frac{p_1(1-p_1)}{n_1} \qquad D(\hat{p}_2) = \frac{D(Y)}{n_2} = \frac{p_2(1-p_2)}{n_2}$$

- 令

$$\xi = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \tag{5.11}$$

$$\eta = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \tag{5.12}$$

$$\zeta = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \tag{5.13}$$

- 其中

$$\hat{p} = \frac{1}{n_1 + n_2}(n_1\hat{p}_1 + n_2\hat{p}_2) = \frac{S_1 + S_2}{n_1 + n_2}$$

- 当 n_1 和 n_2 相当大（一般要求 $n_1\hat{p}_1(1-\hat{p}_1) \geq 5, n_2\hat{p}_2(1-\hat{p}_2) \geq 5$）时，$\xi$ 近似服从标准正态分布。
- 有

$$P(\xi > z_{1-\alpha}) \approx \alpha$$

两总体比率右侧检验的正态近似

- 在 $H_0: p_1 \le p_2$ 成立时，$\xi \ge \eta$，

$$P(\eta > z_{1-\alpha}) \le P(\xi > z_{1-\alpha}) \approx \alpha$$

- 所以

$$H_0: p_1 \le p_2 \longleftrightarrow H_a: p_1 > p_2$$

的否定域可取为

$$W = \{\eta > z_{1-\alpha}\}$$

两总体比率左侧检验的正态近似

- 类似地，对

$$H_0 : p_1 \geq p_2 \longleftrightarrow H_a : p_1 < p_2$$

- 在 H_0 下 $\xi \leq \eta$。
- 从而

$$P(\eta < z_\alpha) \leq P(\xi < z_\alpha) \approx \alpha$$

- 否定域可取为

$$W = \{\eta < z_\alpha\}$$

两总体比率双侧检验的正态近似

- 考虑检验问题

$$H_0: p_1 = p_2 \longleftrightarrow H_a: p_1 \neq p_2$$

- 当 $H_0: p_1 = p_2$ 成立时，只要 n_1, n_2 相当大（一般要求 $n_1\hat{p}_1(1-\hat{p}_1) \geq 5,\ n_2\hat{p}_2(1-\hat{p}_2) \geq 5$），统计量 ζ 近似服从标准正态分布，

$$P(|\zeta| > z_{1-\frac{\alpha}{2}}) \approx \alpha$$

- 否定域可取为

$$W = \{|\zeta| > z_{1-\frac{\alpha}{2}}\}$$

例 5.2

- **例 5.2** 研究口服避孕药对年龄在 40 至 44 岁的妇女的心脏的影响。
- 5000 个使用口服避孕药的妇女中三年内出现心肌梗死的有 13 人；
- 在 10000 个不服用口服避孕药的妇女中出现心肌梗死的有 7 人。
- 问：两组人的心肌梗死比率是否有显著差异？

- **解** 用 p_1 表示使用口服避孕药的妇女中三年内出现心肌梗死的比率，p_2 表示不使用口服避孕药的妇女中三年内出现心肌梗死的比率。
- 要检验

$$H_0 : p_1 = p_2 \longleftrightarrow H_a : p_1 \neq p_2$$

- 统计量

$$\begin{aligned}\hat{p}_1 =& \frac{13}{5000} = 0.0026 \\ \hat{p}_2 =& \frac{7}{10000} = 0.0007 \\ \hat{p} =& \frac{13+7}{5000+10000} = 0.00133\end{aligned}$$

- 近似条件：$n_1\hat{p}_1(1-\hat{p}_1)=6.66\geq 5$, $n_2\hat{p}_2(1-\hat{p}_2)=6.70\geq 5$，可以用统计量 ζ。
- $\zeta=3.01$，水平取为 $\alpha=0.01$，否定域为 $W=\{|\zeta|>z_{1-\frac{0.01}{2}}\}=\{|\zeta|>2.58\}$。
- $|\zeta|>2.58$，拒绝 H_0，在 0.01 水平下两组的心肌梗死比率有显著差异。
- 或称口服避孕药对心肌梗死比率有显著影响。

两样本比率比较的 Fisher 精确检验

- Fisher 精确检验计算精确 p 值，不要求大样本。
- 令 $S_1 = \sum_{i=1}^{n_1} X_i$，$S_2 = \sum_{i=1}^{n_2} Y_i$。
- 令

$$S_1^0 = \sum_{i=1}^{n_1} x_i,\ S_2^0 = \sum_{i=1}^{n_2} y_i,\ t = S_1^0 + S_2^0 \tag{5.14}$$

- 考虑检验问题

$$H_0 : p_1 \le p_2 \longleftrightarrow H_a : p_1 > p_2$$

- p 值为

$$p_1(S_1^0) = \sum_{i \ge S_1^0} p(i) \tag{5.15}$$

- 其中

$$p(i) = \frac{\binom{n_1}{i}\binom{n_2}{t-i}}{\binom{n_1+n_2}{t}} \quad (i = 0, 1, \ldots)$$

- 当且仅当 p 值小于等于 α 时拒绝 $H_0 : p_1 \le p_2$。

- 对于假设

$$H_0: p_1 \geq p_2 \longleftrightarrow H_a: p_1 < p_2$$

- p 值为

$$p_1(S_1^0) = \sum_{i \leq S_1^0} p(i) \tag{5.17}$$

- 对于双边假设

$$H_0 : p_1 = p_2 \longleftrightarrow H_a : p_1 \neq p_2$$

- p 值为

$$p_3(S_1^0) = 2\min\left\{\sum_{i \leq S_1^0} p(i), \sum_{i \geq S_1^0} p(i)\right\} \tag{5.18}$$

- 计算 $p(i)$ 的递推公式

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

例 5.3

- **例 5.3** 某公安局有两个专案组，在过去一年内一组接手 25 件人命案，侦破了 23 件；二组接手 35 件人命案，侦破了 30 件。
- 问：两个组的侦破能力有无差别？
- **解** 比较两个组的侦破率。
- 设两个组的侦破率分别为 p_1, p_2。要检验 $H_0 : p_1 = p_2$。

- 用 Fisher 精确检验。
- $n_1 = 25,\ n_2 = 35,\ S_1^0 = 23,\ S_2^0 = 30,\ t = 53$。
- p 值为

$$p_3(S_1^0) = 2\min\left\{\sum_{i=0}^{23} p(i),\ \sum_{i=23}^{25} p(i)\right\}$$

- 其中

$$\begin{aligned}
p(23) =& 0.252 \\
p(24) =& p(23)\frac{2\times 30}{24\times 6} = 0.105 \\
p(25) =& p(24)\frac{1\times 29}{25\times 7} = 0.017
\end{aligned}$$

- 于是

$$\sum_{i=23}^{25} p(i) = 0.374$$

$$\sum_{i=0}^{23} p(i) = 1 - \sum_{i=23}^{25} p(i) + p(23) = 0.878$$

$$p_3(S_1^0) = 2 \times 0.374 = 0.748 > 0.05$$

- 在 0.05 水平下不应拒绝 $H_0 : p_1 = p_2$。
- 两个专案组在破案能力上没有显著差异。

本节目录

总体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。

拟合优度卡方检验

- 检验

$$H_0: X\text{的分布函数为}F(x)$$

- 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本。
- 类似于画直方图，在数轴上取 m 个点：$t_1, t_2, \ldots, t_m$（$t_1 < t_2 < \cdots < t_m$），把数轴 $(-\infty, +\infty)$ 分成 $m+1$ 段：

$$(-\infty, t_1],\ (t_1, t_2],\ \ldots,\ (t_{m-1}, t_m],\ (t_m, +\infty)$$

- 用 ν_i 表示 $X_1, X_2, \ldots, X_n$ 中落入第 i 段的个数（$i = 1, 2, \ldots, m+1$）。这里 ν_i 是频数，$\frac{\nu_i}{n}$ 是频率。
- 在 H_0 下 X 落入第 i 段的概率 p_i 可计算：

$$\begin{aligned} p_1 =& P(X \le t_1) = F(t_1) \\ p_i =& P(t_{i-1} < X \le t_i) = F(t_i) - F(t_{i-1}), \end{aligned}$$

- 根据概率与频率的关系（大数定律），如果 H_0 成立，频率 $\frac{\nu_i}{n}$ 应该接近于概率 p_i。
- 定义

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

- 当 V 很大时拒绝 H_0。
- 其中多乘的因子 $\frac{n}{p_i}$ 的作用是放大 p_i 很小的项的作用，分布密度曲线的形状恰好是更多由两侧 (对应 p_i 很小) 的形状决定，如果不乘这个因子则 p_i 很小的项作用被严重低估。

- 另一解释：记

$$\xi_i = \frac{\nu_i}{n} - p_i$$

- 则 H_0 成立时 $\nu_i \sim \mathsf{B}(n, p_i)$，

$$\begin{aligned} E(\xi_i) =& 0 \\ D(\xi_i) =& \frac{p_i(1-p_i)}{n} \end{aligned}$$

- 为标准化 $\left(\frac{\nu_i}{n} - p_i\right)^2$, 应该用

$$\frac{\left(\frac{\nu_i}{n} - p_i\right)^2}{\frac{p_i(1-p_i)}{n}}$$

- 但其中的 p_i 一般都比较小 (尤其是样本量很大时)，$1 - p_i \approx 1$，就变成了

$$\frac{\left(\frac{\nu_i}{n} - p_i\right)^2}{\frac{p_i}{n}} = \left(\frac{\nu_i}{n} - p_i\right)^2 \frac{n}{p_i}$$

- 否定域为

$$W = \{V > \lambda\}$$

- 要找临界值 λ 使 W 的检验水平为 α。
- 需要 V 在 H_0 下的分布。在 H_0 成立的条件下，样本量较大时 V 近似服从 $\chi^2(m)$。
- 查 $\chi^2(m)$ 的右侧 α 水平临界值就可以得到 λ。
- 这种检验叫做**拟合优度（卡方）检验**，或分布的**卡方检验**。
- 当零假设下的分布函数包含从样本中估计的未知参数时，V 的自由度改为 m 减去估计的未知参数的个数。

例 6.1

- **例 6.1** 某车间生产滚珠，随机抽取了 50 个产品，直径数据（mm）：

15.0	15.8	15.2	15.1	15.9	14.7	14.8	15.5	15.6
15.3	15.1	15.3	15.0	15.6	15.7	14.8	14.5	14.2
14.9	14.9	15.2	15.0	15.3	15.6	15.1	14.9	14.2
14.6	15.8	15.2	15.9	15.2	15.0	14.9	14.8	14.5
15.1	15.5	15.5	15.1	15.1	15.0	15.3	14.7	14.5
15.5	15.0	14.7	14.6	14.2				

- 直方图：演示。
- 计算得 $\bar{x} = 15.1$, $S^2 = (0.4325)^2$。
- 问：滚珠直径是否服从 $\mathsf{N}(15.1, (0.4325)^2)$?

- **解** 分组与直方图法类似。找到样本值中最小与最大值，取比最小数略小的 a，比最大数略大的 b，将区间 $[a,b]$ 做 $m+1$ 等分，得分点

$$a < t_1 < t_2 < \cdots < t_m < b$$

- m 的个数选取参考直方图方法。
- 这 50 个数据最小 14.2，最大 15.9，取 $a = 14.05$, $b = 16.15$, $b - a = 2.1 = 0.3 * 7$, $m = 6$，分成 7 段。
- 在 $F(x)$ 为 $\mathrm{N}(15.1, (0.4325)^2)$ 分布函数时求得各个 p_i。
- 为求 $F(x)$，用

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\Phi(-x) = 1 - \Phi(x) \quad (x > 0)$$

$(\mu, \sigma^2$ 是正态分布的期望和方差$)$

- 计算得

i	p_i	np_i	ν_i	$(np_i-\nu_i)^2$	$\frac{(np_i-\nu_i)^2}{np_i}$
1	0.0414	2.070	3	0.8649	0.4178
2	0.1077	5.385	5	0.1482	0.0275
3	0.2154	10.770	10	0.5925	0.0551
4	0.2710	13.550	16	6.0025	0.4430
5	0.2154	10.770	8	7.6729	0.7124
6	0.1077	5.385	6	0.3782	0.0702
7	0.0414	2.070	2	0.0049	0.0024

- 检验统计量

$$V = \sum_{i=1}^{7} \frac{(\nu_i - np_i)^2}{np_i} = 1.7284$$

- 自由度用 $7-1-2$（额外扣除两个估计的未知参数的自由度）。取 $\alpha = 0.05$，$\lambda = 9.49$。
- $V < \lambda$，H_0 相容，在 0.05 水平下不拒绝总体服从正态分布的假设。

- 注意：如果最后的结论是不拒绝 H_0，可能会有较大的第二类错误概率。
- 有可能会以威布尔分布作为零假设，检验不拒绝；再以对数正态作为零假设，检验还是不拒绝。
- 但是，只要不拒绝零假设，就可以说明数据与该分布差距不大。

离散分布的卡方检验

- 连续型分布用卡方检验需要分组，离散型分布不需要分组。
- 概率特别小的组可以合并。
- 设 X 的分布是

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

$x_1, x_2, \ldots, x_n$ 是样本值。

- 取统计量

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

- 其中 ν_i 表示 n 个样品中 a_i 出现的频数。
- 在 H_0 下 V 近似服从 $\chi^2(m)$; 如果分布中含有从样本值估计的未知参数，则自由度要扣除未知参数个数。

例 6.2（例 1.3 续）

- **例 6.2（例 1.3 续）** 某工厂近 5 年来发生了 63 次事故，这些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：事故发生是否与星期几有关？
- **解** 用 X 表示这样的随机变量：若事故发生在星期 i，则 $X = i$。
- X 的可能取值集合为 $\{1, 2, \ldots, 6\}$（星期日是该厂厂休日）。
- 检验

$$H_0 : P(X = i) \equiv \frac{1}{6} \quad (i = 1, 2, \ldots, 6)$$

- 使用统计量 (6.1)。$m=5$, H_0 成立时 $p_i=1/6, i=1,2,\ldots,6$。
- 计算得 $V=1.67$。
- 查 5 个自由度卡方分布右侧 $\alpha=0.05$ 临界值得 $\lambda=11.07$。
- $V<\lambda$，H_0 相容，不能认为出事故与星期几有关。