北京航空航天大学
——经济管理学院——
BEIHANG UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT

第三讲：假设检验

康雁飞

数量经济与商务统计系

Outline

统计推断

- 统计推断的任务：根据样本信息，推断总体的统计规律
- 统计推断的基本问题
  1. 估计问题：根据样本信息，对总体分布中未知参数进行估计
  2. 假设检验问题：假设的建立及其各种检验
- 假设检验（Hypothesis Test）：K. Pearson 于 20 世纪初提出，之后由 Fisher 进行细化，并最终由 Neyman 和 E. Pearson 提出完整的假设检验理论
- 学习基本要求：理解并熟悉假设检验的基本思想、基本步骤、概念及其各种假设检验方法的构造方法、统计性质及其在不同分布情形下的计算、R 实现

Outline

1 假设检验的基本思想与概念

2 正态总体参数假设检验

3 其他分布参数的假设检验

4 总体分布检验

假设检验的基本思想与概念

1. 假设检验问题
2. 假设检验的基本步骤
3. 检验的 p 值

假设检验问题案例：女士品茶

一种奶茶有牛奶与茶按一定比例混合而成，分为两种：TM（先倒茶后倒奶），MT（先倒奶后倒茶）。某女士声称可以鉴别是 TM 还会 MT，周围品茶的人对此产生疑问。请思考：如何判断该论证正确?

- **Fisher 的做法:** Fisher 提出做一项试验来检验如下命题是否可以接受。

  假设 H: 该女士无此种鉴别能力

假设检验问题

- 随机调制 10 杯奶茶（TM 和 MT 均有），该女士一一品茶并说出是 TM 还是 MT。结果该女士竟然正确分辨出 10 杯奶茶中的每一杯。**请回答：此时该如何做出判断?**
- **Fisher 的想法:** 假设 H 成立, 10 次都猜对的概率为 $2^{-10} < 0.001$. 这是一个很小的概率, 在一次试验中几乎不会发生, 但如今该事件竟然发生了, 这有理由认为假设 H 不成立, 即拒绝假设 H, 认为该女士有辨别奶茶中 TM 和 MT 的能力。
- **小概率原理:** 若事件 A 发生是一个很小的概率 (比如 0.001), 则认为在一次试验中事件 A 几乎不会发生。

假设检验问题

假设检验可研究的问题举例:

1. 一种药品是否比另一种药品更有效?
2. 几种不同的肥料哪一种更有效?
3. 大学生的就业率与城市失业率之间是否存在关系?

假设检验问题

- 这些没有被数据验证的猜测就是假设检验问题，通过样本求证的过程就是假设检验（Hypothesis testing）。
- 上述研究问题都可以视为对不同分布总体的选择问题：分布间差异是本质的还是随机偶然因素引起的?
- 检验或检验法则：希望能够找到一个合理地做出可靠性判断的临界点，从而产生判别的条件和准则。

假设检验

假设检验就是一种统计推断方法，根据样本提供的信息对所提出的假设作出判断。

假设检验基本原理

- 小概率事件不是不可能事件，但在一次试验中出现的可能性很小，以至于实际上可以看成是不可能发生的。
- 在统计学中，这是小概率事件实际不可能原理，即小概率原理，是假设检验的基本依据。

假设检验基本原理

先提出假设 H_0，再根据一次抽样所得到的样本值进行计算。

1. 若导致小概率事件发生，则拒绝假设 H_0；
2. 否则，没有足够的理由拒绝假设 H_0。

假设检验：例 1

某厂有一批产品，共有 10000 件，需检验合格方能出厂。按规定次品率不能超过 2%。现从中任取 100 件，发现有 5 件次品，问这批产品能否出厂?

假设检验：例 1

- 提出假设：$H_0: p \leq 0.02$，其中 p 为总体次品率。
- $X_i \sim B(1, p), i = 1, 2, \cdots, 100.$
- $Y = X_1 + X_2 + \cdots + X_{100}$ 也就是抽取的 100 件产品中的次品数，$Y \sim B(100, p).$
- H_0 成立的条件下，
$$P(Y = 5; p) = C_{100}^5 p^5 (1 - p)^{95} \leq 0.035.$$
- 小概率事件发生，不能出厂。

假设检验：例 2

某食品厂生产猪肉罐头，按规定每瓶的标准重量为 500g，由以往经验知，该厂生产的猪肉罐头质量服从正态分布 $N(\mu, 4)$, 随机抽取 5 瓶, 其重量分别为 (单位: g):

$$501, 507, 498, 502, 504$$

能否认为该厂猪肉罐头的标准重量为 500g? 分析设该厂生产的猪肉罐头平均重量 $\mu = 500\ \mathrm{g}$ 则问题变为检验假设 $H_0 : \mu = 500$ 是否成立?

假设检验：例 2

1. 提出两个对立假设：

$$H_0 : \mu = \mu_0 = 500,\ H_1 : \mu \neq \mu_0.$$

2. 我们知道，$\bar{X}$ 是 μ 的无偏估计量，所以若 H_0 为真，$|\bar{x} - \mu_0|$ 不会太大，那么 $\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}}$ 不会太大。

假设检验：例 2

- 当 H_0 为真时，
$$u = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$
- $P\left(|u| \geq u_{1-\alpha/2}\right) = \alpha$，当 α 很小时，我们认为 $|u| \geq u_{1-\alpha/2}$ 是个小概率事件。
- 根据小概率原理，如果 H_0 为真，由一次试验得到了满足 $|u| = \frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} \geq u_{1-\alpha/2}$ 的观测值 $\bar{x}$，这几乎不可能发生。
- 我们有理由怀疑原来的假设，拒绝 H_0。

假设检验：例 2

- 拒绝域

$$\left\{x_1, \cdots, x_n : \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \geq u_{1-\frac{\alpha}{2}}\right\}$$

- 接收域

$$\left\{x_1, \cdots, x_n : \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \leq u_{1-\frac{\alpha}{2}}\right\}$$

假设检验：例 2

在例 2 中，$n = 5, \sigma = 2, \bar{x} = 502.4, \mu_0 = 500.$

- Q1: 若取 $\alpha = 0.05$，则 $u_{\alpha/2} = ?$
- Q2: $|u| = ?$
- Q3: 决策?

假设检验的一般步骤

1. 建立假设
2. 选择检验统计量, 给出拒绝域形式
3. 选择显著性水平
4. 给出拒绝域，做出判断

假设检验的一般步骤

一、建立假设

- 在假设检验中，常把一个被检验的假设称为原假设/零假设（null hypothesis），用 H_0 表示.
- 当 H_0 被拒绝时而接受的假设称为备择假设（alternative hypothesis），用 H_1 表示，它们常常成对出现.
- 设 θ 是参数，Θ 是参数空间，$\Theta_0 \subset \Theta$，$\Theta_1 \subset \Theta$ 且 $\Theta_0 \cap \Theta_1 = \varnothing$，则我们感兴趣的一对假设就是

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

如何选取零假设

零假设的选择没有很强的要求，以下列出一些指导原则：

- 通常将不应轻易加以否定的假设作为原假设。
- 通常将研究者想收集证据予以支持的假设作为备择假设。
- 信息准则：
  1. "保护原假设"原则：存在普遍成立的先验信息，这样只有样本表现出足够的说服力来推翻先验信息时，我们才认为原假设被拒绝，新结论成立。
  2. 样本先验信息原则：通常将样本显示出的特点作为对总体的猜想，并优先被选作备择假设。

假设检验的一般步骤

二、选择检验统计量，给出拒绝域形式

- 由样本对原假设进行判断总是通过一个统计量完成的，该统计量称为**检验统计量**，记为 $T(X_1, \cdots, X_n)$.
- 而检验等价于把样本空间化为成两个互不相交的部分：拒绝域 W 和接受域 $\bar{W}$.
  1. 如果 $x_1, \cdots, x_n \in W$，拒绝 H_0；
  2. 否则，没有足够的理由，拒绝 H_0。

例题回顾

- 例 1：$T=\sum\limits_{i=1}^{n} X_i$，拒绝域 $W=\left\{x_1,\cdots,x_n : T\left(x_1,\cdots,x_n\right)\geq c\right\}$.
- 例 2：$T=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$，拒绝域 $W=\left\{x_1,\cdots,x_n : \left|T\left(x_1,\cdots,x_n\right)\right|\geq c\right\}$.

假设检验的一般步骤

三、选择显著性水平 α，并根据检验统计量在原假设下的分布，求出临界值 c; 或者在原假设下计算利用样本观测值能够做出拒绝原假设的概率。

两类错误

第一类错误 (Type I error)

H_0 为真，样本落在拒绝域中，从而拒绝了 H_0。第一类错误（拒真）概率记为 $\alpha(\theta) = P_\theta\{X \in \mathbf{W}\}, \theta \in \Theta_0$，或 $P_\theta\{X \in \mathbf{W}|H_0\}$。

第二类错误 (Type II error)

H_1 为真，但样本观测值落在接受域中，从而接受了 H_0。第二类错误（取伪）的概率记为 $\beta(\theta) = P_\theta\{X \in \bar{\mathbf{W}}\}, \theta \in \Theta_1$，或 $P_\theta\{X \in \bar{\mathbf{W}}|H_1\}$。

势函数 (power function)

- 任何一个假设检验都无法避免犯上述两种错误。能否使得一个检验犯两类错误的概率都尽可能小? 不能!
- 我们引入势函数的概念说明此问题.
- 犯第一类错误的概率 α 和犯第二类错误的概率 β 可以用同一个函数表示, 即所谓的势函数.

势函数

设检验问题

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

的拒绝域为 **W**, 则样本观测值落在拒绝域内的概率称为该检验的势函数, 记为

$$g(\theta) = P_\theta(X \in W), \quad \theta \in \Theta = \Theta_0 \cup \Theta_1$$

势函数 (power function)

势函数 $g(\theta)$ 是定义在参数空间 Θ 上的一个函数. 犯两类错误的概率都是参数 θ 的函数, 并可由势函数算得, 即:

$$g(\theta) = \begin{cases} \alpha(\theta), & \theta \in \Theta_0 \\ 1 - \beta(\theta), & \theta \in \Theta_1 \end{cases}$$

势函数：例 3

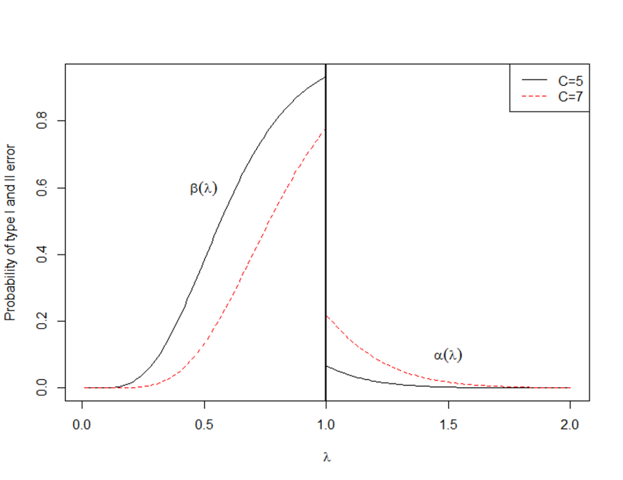
- 假设总体 X 来自 Poisson 分布 $\mathcal{P}(\lambda)$, 简单随机抽样 $X_1, X_2, \cdots, X_n$, 假设检验问题 $H_0: \lambda \geq 1 \quad \text{vs} \quad H_1: \lambda < 1$ 根据假设检验的步骤，可以选取

$$T = \sum_{i=1}^{n} X_i \Rightarrow W = \left\{X_1, \cdots, X_n : \sum_{i=1}^{n} X_i \leq C\right\}$$

- $n = 10$ 时, 对 $C = 5$ 和 $C = 7$ 我们考虑检验势函数随 λ 从 0 变化到 2 的情况.

■ 在零假设下, 我们注意到检验

$$\alpha(\lambda) = P(|) = P\left(\sum_{i=1}^{n} X_i \le C \mid \lambda \in H_0\right),$$
$$\beta(\lambda) = 1 - P(|) = 1 - P\left(\sum_{i=1}^{n} X_i \le C \mid \lambda \in H_1\right).$$


C=5
C=7
β(λ)
α(λ)
Probability of type I and II error
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
λ

显著性检验

折中：通常仅限制犯第一类错误的概率，这就是显著性检验。

显著性检验

对检验问题

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

如果一个检验满足对任意的 $\theta \in \Theta_0$, 都有

$$g(\theta) \leq \alpha,$$

则称该检验是显著性水平 (level of significance) 为 α 的显著性检验, 简称水平为 α 的检验。

显著性检验

- 不否定原假设并不是肯定原假设一定对，只是说差异还不够显著，还没有达到足以拒绝原假设的程度。
- 因此假设检验又称为显著性检验。

假设检验的一般步骤

三、选择显著性水平 α，并根据检验统计量在原假设下的分布，求出临界值 c; 或者在原假设下计算利用样本观测值能够做出拒绝原假设的概率。

给定显著性水平 α，水平为 α 的显著性检验要求

1. 例 1：对任意 $p \leq 0.02$ 有 $g(p) = P(Y \geq c | Y \sim B(100, p)) \leq \alpha$
   - $g(p)$ 关于 p 单调递增，只需要 $g(0.02) \leq \alpha$
   - $\alpha = 0.05$ 时可以计算临界值 $c = 5$，拒绝域 $W = \{x_1, \cdots, x_n : T(x_1, \cdots, x_n) \geq 5\}$

Lets go to R

```r
g <- function(c){
  # calculate g(0.02)
  p = rep(NA, c + 1)
  for (i in 0:c){
    p[i+1] = choose(100, i)*0.02^i*0.98^(100-i)
  }
  return(1 - sum(p))
}
print(c(g(3), g(4), g(5), g(6), g(7)))
```

```
## [1] 0.141038437 0.050830445 0.015483641 0.004062054 0.000931940
```

假设检验的一般步骤

2. 例 2：对于 $H_0 : \mu = \mu_0 = 500$ 有

$$g(500) = 2 * (1 - \Phi(c)) = \alpha, \; c = u_{1-\alpha/2}.$$

- $\alpha = 0.05$ 时可以计算临界值 $c = 1.96$，拒绝域 $W = \{x_1, \cdots, x_n : |T| \geq 1.96\}$

假设检验的一般步骤

四、在给定显著性水平 α 下，根据拒绝域的具体形式，根据样本观测值作出判断

给定显著性水平 $\alpha = 0.05$，

- 例 1：拒绝域 $W = \{x_1, \cdots, x_n : T(x_1, \cdots, x_n) \geq 5\}$，样本观测值 $T(x_1, \cdots, x_n) = 5$，因此 $(x_1, \cdots, x_n) \in W \to$ 拒绝H_0.
- 例 2：拒绝域 $W = \{x_1, \cdots, x_n : T(x_1, \cdots, x_n) \geq 1.96\}$，样本观测值 $T(x_1, \cdots, x_n) = 2.68$，因此 $(x_1, \cdots, x_n) \in W \to$ 拒绝H_0.

假设检验的一般步骤

1. 建立假设
2. 选择检验统计量，给出拒绝域形式
3. 选择显著性水平
4. 给出拒绝域，做出判断

假设检验：p 值

以例 1 拒绝域 $W=\{x_1,\cdots,x_n : T(x_1,\cdots,x_n)\geq c\}$ 为例，拒绝域的大小很显然受选择的显著性水平 α 影响。

- α 较大时, 比如 $\alpha=0.05$, 拒绝域较大: $W=T\geq 5$. 此时在 $\alpha=0.05$ 水平下，拒绝原假设
- α 较小时, 比如 $\alpha=0.01$, 拒绝域较小: $W=T\geq 7$. 此时在 $\alpha=0.01$ 水平下，不能拒绝原假设
- 不同的检验水平的选择导致接受或拒绝原假设的判断不同

假设检验：p 值

- p 值：利用样本观测值能够做出“拒绝原假设”这一判断的最小显著性水平
- 或者说在原假设成立的前提下，出现与样本相同或者更极端的情况的概率（尾概率）
- p 值越小，我们拒绝原假设的理由越充分，结果越显著

假设检验：p 值

引进检验的 p 值的概念有明显的好处:

1. 客观, 避免了事先确定显著水平;
2. 由检验的 p 值与人们心目中的显著性水平 α 进行比较可以很容易作出检验的结论：
   - 如果 $\alpha \geqslant p$, 则在显著性水平 α 下拒绝 H_0;
   - 如果 $\alpha < p$, 则在显著性水平 α 下保留 H_0.

小结（一）

1. 理解并熟悉掌握假设检验的基本思想及一般原理
2. 理解假设检验中的基本概念：原假设、备择假设；检验统计量、拒绝域、接受域；两类错误、势函数、显著性水平；p 值等统计含义
3. 掌握假设检验的一般步骤，尤其是检验问题的提出、构造拒绝域形式及确定临界点等方法

作业

- 7.1 课后习题：1-5

Outline

1 假设检验的基本思想与概念

2 正态总体参数假设检验

3 其他分布参数的假设检验

4 总体分布检验

单个正态总体均值的检验

设 $X_1, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 有关参数 μ 假设检验常见的有三种基本形式（含单侧检验和双侧检验）

- I: $H_0 : \mu \leqslant \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$
- II: $H_0 : \mu \geqslant \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$
- III: $H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$

一、已知 σ 时的 μ 检验

设 $X_1, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 考虑关 μ 的单侧检验问题

$$H_0 : \mu \leqslant \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

由于 μ 的点估计是 $\bar{x}$, 且 $\bar{x} \sim N(\mu, \sigma^2/n)$, 故检验统计量可选为

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- 该检验用 u 检验统计量, 故称为 u 检验（也有称为 Z 检验)。
- 直觉上, $\bar{x}$ 不超过 μ_0 且与 μ_0 距离越远，越倾向接受原假设，反之则倾向拒绝。
- 但在有随机性存在的场合，$\bar{x}$ 超过或不超过 μ_0 可能由不确定性造成，所以距离 μ_0 要足够大（大过阈值 c）才有足够的信心拒绝或接受原假设。

对于检验

$$H_0 : \mu \leqslant \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

拒绝域应该大于临界值 c,

$$W = \{(x_1, \cdots, x_n) : u \geqslant c\}$$

或简化记为 $\{u \geqslant c\}$.

若要求检验的显著性水平为 α，则当原假设 $\{\mu \leqslant \mu_0\}$ 成立时，c 应满足

$$P(u \geqslant c) \leqslant \alpha$$

因此 $c = u_{1-\alpha}$.（想一想为什么?）

势函数

$$
\begin{aligned}
g(\mu) &= P(u \geq u_{1-\alpha}) \\
&= P(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq u_{1-\alpha}) \\
&= P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq u_{1-\alpha} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}) \\
&= 1-\Phi(u_{1-\alpha} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}).
\end{aligned}
$$

这是关于 μ 的增函数，所以当 $\mu \leq \mu_0$ 时，$g(\mu) \leq g(\mu_0) = \alpha$.

一、已知 σ 时的 μ 检验

检验统计量

$$u = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

I: $H_0 : \mu \leqslant \mu_0$ **vs** $H_1 : \mu > \mu_0$

- $W_I = \{(x_1, \cdots, x_n) : u \geq u_{1-\alpha}\}$
- $p_I = P(u \geq u_0) = 1 - \Phi(u_0)$

一、已知 σ 时的 μ 检验

II: $H_0 : \mu \geq \mu_0$ **vs** $H_1 : \mu < \mu_0$

- $W_{II} = \left\{(x_1, \cdots, x_n) : u \leq u_\alpha\right\}$
- $p_{II} = P\left(u \leq u_0\right) = \Phi\left(u_0\right)$

III: $H_0 : \mu = \mu_0$ **vs** $H_1 : \mu \neq \mu_0$

- $W_{III} = \left\{(x_1, \cdots, x_n) : |u| \geq u_{1-\alpha/2}\right\}$
- $p_{III} = P\left(|u| \geq |u_0|\right) = 2\left(1 - \Phi\left(|u_0|\right)\right)$

举例

从甲地发送一个讯号到乙地. 设乙地接收到的讯号值服从正态分布 $N(\mu, 0.2^2)$ 其中 μ 为甲发送的真实讯号. 现甲地重复发送同一讯号 5 次, 乙地接收到的讯号值为

$$8.05 \quad 8.15 \quad 8.2 \quad 8.1 \quad 8.25$$

设接收方有理由猜测甲地发送的讯号值为 8, 问能否接受这猜测?

- 这是一个假设检验的问题, 总体 $X \sim N(\mu, 0.2^2)$. 检验假设:$H_0 : \mu = 8 \quad \text{vs} \quad H_1 : \mu \neq 8$
- 这个双侧检验问题的拒绝域为 $\left\{|u| \geq u_{1-\alpha/2}\right\}$
- 取置信水平 $\alpha = 0.05$, 则查表知 $u_{0.975} = 1.96$. 用观测值可计算得

$$\bar{x} = 8.15, \quad u = \sqrt{5}(8.15 - 8)/0.2 = 1.6771$$

u 值未落入拒绝域内, 故不能拒绝原假设, 可认为猜测成立.

R

```r
x <- c(8.05, 8.15, 8.2, 8.1, 8.25)
mu0 <- 8; sigma <- 0.2; n <- 5
u.crit <- qnorm(0.975)
u.crit
```

```
## [1] 1.959964
```

```r
u.obs <- (mean(x) - mu0)/(sigma/sqrt(n))
u.obs
```

```
## [1] 1.677051
```

```r
p.value <- 2 * (1 - pnorm(u.obs))
p.value
```

```
## [1] 0.09353251
```

课堂练习一

某公司声称其钻头寿命服从正态分布 $X \sim N(32, 16)$，抽取一个容量为 $n = 25$ 的样本，$\bar{x} = 29.5$，问在 $\alpha = 0.05$ 的显著性水平下，其钻头期望寿命是否达到了公司声称的 32（设总体方差为 16 是正确的)?

R

```r
xbar <- 29.5
mu0 <- 32; sigma <- 4; n <- 25
u.crit <- qnorm(0.05)
u.crit
```

```
## [1] -1.644854
```

```r
u.obs <- (xbar - mu0)/(sigma/sqrt(n))
u.obs
```

```
## [1] -3.125
```

```r
p.value <- pnorm(u.obs)
p.value
```

```
## [1] 0.0008890253
```

二、未知 σ 时的 μ 检验

检验统计量

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

拒绝域（未知 σ 时的 μ 检验）

- $W_I = \{(x_1, \cdots, x_n) : t \geq t_{1-\alpha}\}$
- $W_{II} = \{(x_1, \cdots, x_n) : t \leq t_{\alpha}\}$
- $W_{III} = \left\{(x_1, \cdots, x_n) : |t| \geq t_{1-\alpha/2}\right\}$

二、未知 σ 时的 μ 检验

p 值（未知 σ 时的 μ 检验）

- $p_I = P\left(t \geq t_0\right)$
- $p_{II} = P\left(t \leq t_0\right)$
- $p_{III} = P\left(|t| \geq |t_0|\right)$

课堂练习二

某厂生产的某种铝材的长度服从正态分布, 其均值设定为 240 厘米。现从该厂抽取 5 件产品，测得其长度为

239.7　239.6　239　240　239.2

试判断该厂此类铝材的长度是否满足设定要求?

R

```r
x <- c(239.7, 239.6, 239, 240, 239.2)
mu0 <- 240
t.test(x, mu = mu0)
```

```
##
##  One Sample t-test
##
## data:  x
## t = -2.7951, df = 4, p-value = 0.04906
## alternative hypothesis: true mean is not equal to 240
## 95 percent confidence interval:
##  239.0033 239.9967
## sample estimates:
## mean of x
##     239.5
```

小结

单个正态总体均值的假设检验

检验法	H_0	H_1	检验统计量	拒绝域	p 值
u 检验 (σ 已知)	$\mu\leqslant\mu_0$	$\mu>\mu_0$	$u=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$\{u\geqslant u_{1-\alpha}\}$	$1-\Phi(u_0)$
	$\mu\geqslant\mu_0$	$\mu<\mu_0$		$\{u\leqslant u_\alpha\}$	$\Phi(u_0)$
	$\mu=\mu_0$	$\mu\neq\mu_0$		$\{\lvert u\rvert\geqslant u_{1-\alpha/2}\}$	$2(1-\Phi(\lvert u_0\rvert))$
t 检验 (σ 未知)	$\mu\leqslant\mu_0$	$\mu>\mu_0$	$t=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$\{t\geqslant t_{1-\alpha}(n-1)\}$	$P(t\geqslant t_0)$
	$\mu\geqslant\mu_0$	$\mu<\mu_0$		$\{t\leqslant t_\alpha(n-1)\}$	$P(t\leqslant t_0)$
	$\mu=\mu_0$	$\mu\neq\mu_0$		$\{\lvert t\rvert\geqslant t_{1-\alpha/2}(n-1)\}$	$P(\lvert t\rvert\geqslant\lvert t_0\rvert)$

注：$u_0=\sqrt{n}(\bar{x}-\mu_0)/\sigma$，$t_0=\sqrt{n}(\bar{x}-\mu_0)/s$，$t$ 是服从 $t(n-1)$ 的随机变量.

假设检验与置信区间的关系

- 检验统计量与 6.6 节中置信区间所用的枢轴量是相似的, 两者之间存在非常密切的关系.
- 设 $x_1, \cdots, x_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, 我们以 σ^2 未知场合为例, 讨论关于均值 μ 的检验问题.
- 考虑双侧检验问题:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

- 则水平为 α 的检验拒绝域为

$$W=\left\{|t| \geqslant t_{1-\alpha/2}(n-1)\right\}.$$

- 接受域为

$$\bar{W}=\left\{|t| < t_{1-\alpha/2}(n-1)\right\}.$$

其中

$$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

- 接受域

$$\bar{W} = \left\{ |\frac{\bar{x} - \mu_0}{s/\sqrt{n}}| < t_{1-\alpha/2}(n-1) \right\}.$$

可以改写为

$$\bar{W} = \left\{ \bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \le \mu_0 \le \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right\}.$$

- 而原假设 $H_0 : \mu = \mu_0$ 中 μ_0 的取值没有限制, 可取 $(-\infty, \infty)$.

- 回顾 σ^2 未知场合 μ 的 $1-\alpha$ 置信区间为

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$

- 这和原假设的接受域是一一对应的.

反之若有一个如上的 $1-\alpha$ 置信区间, 也可获得关于 $H_0: \mu = \mu_0$ 的水平为 α 的显著性检验.

- 所以: “正态均值 μ 的 $1-\alpha$ 置信区间” 与” 关于 $H_0: \mu = \mu_0$ 的双侧检验问题的水平 α 的检验” 是一一对应的.

类似地,

- “参数 μ 的 $1-\alpha$ 置信上限” 与” 关于 $H_0: \mu \leqslant \mu_0$ 的单侧检验问题的水平 α 的检验” 是一一对应的.
- “参数 μ 的 $1-\alpha$ 置信下限” 与” 关于 $H_0: \mu \geqslant \mu_0$ 的单侧检验问题的水平 α 的检验” 是一一对应的.

作业

- 7.2 课后习题：1、2、6、7、10

两个正态总体均值差的检验

- 设 $X_1, \cdots, X_m$ 是来自 $N(\mu_1, \sigma_1^2)$ 的样本，$Y_1, \cdots, Y_n$ 是来自 $N(\mu_2, \sigma_2^2)$ 的样本，且两个样本相互独立。
- 考虑三种关于 $\mu_1 - \mu_2$ 的检验问题

- I: $H_0 : \mu_1 - \mu_2 \leq 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$
- II: $H_0 : \mu_1 - \mu_2 \geq 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 < 0$
- III: $H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0$

一、σ_1，σ_2 已知时的 u 检验

检验统计量

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

拒绝域

- $W_I = \left\{(x_1, \cdots, x_n) : u \geq u_{1-\alpha}\right\}$
- $W_{II} = \left\{(x_1, \cdots, x_n) : u \leq u_{\alpha}\right\}$
- $W_{III} = \left\{(x_1, \cdots, x_n) : |u| \geq u_{1-\alpha/2}\right\}$

二、$\sigma_1 = \sigma_2 = \sigma$ 未知时的 t 检验

检验统计量

$$t = \frac{\bar{x} - \bar{y}}{s_w\sqrt{\frac{1}{m} + \frac{1}{n}}},$$

其中 $s_w^2 = \frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}$

拒绝域

- $W_I = \{(x_1, \cdots, x_n) : t \geq t_{1-\alpha}(m+n-2)\}$
- $W_{II} = \{(x_1, \cdots, x_n) : t \leq t_{\alpha}(m+n-2)\}$
- $W_{III} = \left\{(x_1, \cdots, x_n) : |t| \geq t_{1-\alpha/2}(m+n-2)\right\}$

举例

某厂铸造车间为提高铸件的耐磨性而试制了一种镍合金铸件以取代铜合金铸件，为此，从两种铸件中各抽取一个容量分别为 8 和 9 的样本，测得其硬度为

镍合金：

$$\begin{array}{cccc} 76.43 & 76.21 & 73.58 & 69.69 \\ 65.29 & 70.83 & 82.75 & 72.34 \end{array}$$

铜合金：

$$\begin{array}{ccccc} 73.66 & 64.27 & 69.34 & 71.37 & \\ 69.77 & 68.12 & 67.27 & 68.07 & 62.61 \end{array}$$

设正态分布且方差不变，在水平 α 下判断镍合金的硬度是否有明显提高。

举例

用 X 表示镍合金的硬度, Y 表示铜合金的硬度, 则由假定,

$$X \sim N\left(\mu_1, \sigma^2\right), \quad Y \sim N\left(\mu_2, \sigma^2\right).$$

要检验的假设是： $H_0 : \mu_1 \leq \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2.$

举例

经计算，

$$\bar{x} = 73.39, \quad \bar{y} = 68.2756,$$

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 191.7958, \quad \sum_{i=1}^{9} (y_i - \bar{y})^2 = 91.1552$$

从而，

$$s_w = \sqrt{\frac{1}{8+9-2}(191.7958 + 91.1552)} = 4.3432$$

所以，

$$t = \frac{73.39 - 68.2756}{4.3432 \cdot \sqrt{\frac{1}{8} + \frac{1}{9}}} = 2.4234$$

问题

1. 拒绝域?
2. p 值?

R

```r
x <- c(76.43, 76.21, 73.58, 69.69, 65.29, 70.83, 82.75, 72.34)
y <- c(73.66, 64.27, 69.34, 71.37, 69.77, 68.12, 67.27, 68.07, 62.61)
t.test(x, y, alternative = 'greater', var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = 2.4234, df = 15, p-value = 0.01424
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.41478     Inf
## sample estimates:
## mean of x mean of y
##  73.39000  68.27556
```

成对数据检验

- 成对数据 (paired data) 在进行两总体均值比较中经常出现, 若直接使用两样本 t 检验结论对成对数据结构可能是错误的（思考：为什么?)
- 成对数据检验提供了一个运用固定效应 (fixed effects) 去除不可观测因素的基本思想.

举例

为了比较两种种子的优劣, 选取 10 块土质不全相同的土地, 将每块土地分为面积相同的两部分, 分别种植两种种子, 其他条件相同, 得到产量如下:

土地	1	2	3	4	5	6	7	8	9	10
种子 1 的产量 x	23	35	29	42	39	29	37	34	35	28
种子 2 的产量 y	30	39	35	40	38	34	36	33	41	31
差 d=x-y	-7	-4	-6	2	1	-5	1	1	-6	-3

假定单位产量服从正态分布, 问两种种子的平均单位产量在显著性水平 $\alpha = 0.05$ 上有无显著差异?

成对数据 t 检验

- 假定 $x \sim N(\mu_1, \sigma_1^2)$, $y \sim N(\mu_2, \sigma_2^2)$. 则 $d = x - y \sim N(\mu, \sigma_d^2)$, 其中 $\mu = \mu_1 - \mu_2$, $\sigma_d^2 = \sigma_1^2 + \sigma_2^2$. 检验问题转化为

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0$$

- 即单样本 t 检验问题. 此时 t 统计量为

$$t_2 = \bar{d} / \left(s_d / \sqrt{n} \right)$$

其中

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i, \quad s_d = \left(\frac{1}{n-1} \sum_{i=1}^{n} \left(d_i - \bar{d} \right)^2 \right)^{1/2}$$

- 拒绝域为

$$W_1 = \left\{ |t_2| \geqslant t_{1-\alpha/2}(n-1) \right|$$

- 由数据

$$n = 10, \quad \bar{d} = -2.6, s_d = 3.5024$$

所以

$$t_2 = \frac{-2.6}{3.5024/\sqrt{10}} = \frac{-2.6}{1.1076} = -2.3475$$

- 若取 $\alpha = 0.05$, 查表可得 $t_{0.975}(9) = 2.2622$. 由于 $|t_2| > 2.2622$, 故拒绝原假设. 认为两种种子有显著差别.

R

```r
x <- c(23, 35, 29, 42, 39, 29, 37, 34, 35, 28)
y <- c(30, 39, 35, 40, 38, 34, 36, 33, 41, 31)
t.test(x, y, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  x and y
## t = -2.3475, df = 9, p-value = 0.04348
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.10545182 -0.09454818
## sample estimates:
## mean of the differences
##                    -2.6
```

正态总体方差的检验

- 单个正态总体方差的 χ^2 检验
- 两个正态总体方差的 F 检验

单个正态总体方差的 χ^2 检验

- 设 $X_1, \cdots, X_m$ 是来自 $N(\mu, \sigma^2)$ 的样本，考虑三种关于 σ^2 的检验问题:

- I: $H_0 : \sigma^2 \leq \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 > \sigma_0^2$
- II: $H_0 : \sigma^2 \geq \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 < \sigma_0^2$
- III: $H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 \neq \sigma_0^2$

单个正态总体方差的 χ^2 检验

检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}.$$

拒绝域

- $W_I = \left\{(x_1, \cdots, x_n) : \chi^2 \geq \chi^2_{1-\alpha}(n-1)\right\}$
- $W_{II} = \left\{(x_1, \cdots, x_n) : \chi^2 \leq \chi^2_{\alpha}(n-1)\right\}$
- $W_{III} = \left\{(x_1, \cdots, x_n) : \chi^2 \leq \chi^2_{\alpha/2}(n-1), \chi^2 \geq \chi^2_{1-\alpha/2}(n-1)\right\}$

单个正态总体方差的 χ^2 检验

p 值

- $p_I = P\left(\chi^2 \geq \chi_0^2\right)$
- $p_{II} = P\left(\chi^2 \leq \chi_0^2\right)$
- $p_{III} = 2 * \min\{P\left(\chi^2 \leq \chi_0^2\right), P\left(\chi^2 \geq \chi_0^2\right)\}$ 其中 $\chi_0^2 = (n-1)s^2/\sigma_0^2$ 为检验统计量的样本观测值。

单个正态总体方差的 χ^2 检验：举例

某类钢板每块的重量 X 服从正态分布，其质量指标是重量的方差不得超过 0.016. 现从某天生产的钢板中随机抽取 25 块，得其样本方差 $s^2 = 0.025$，问该天生产的钢板是否满足要求？

单个正态总体方差的 χ^2 检验：举例

- 原假设为 $H_0: \sigma^2 \leq 0.016$, 备择假设为 $H_1: \sigma^2 > 0.016$. 此处 $n = 25$, 若取 $\alpha = 0.05$, 则查表知 $\chi^2_{0.95}(24) = 36.415$.
- 计算可得:
$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \times 0.025}{0.016} = 37.5 > 36.415$$
由此, 在显著性水平 0.05 下, 我们拒绝原假设, 认为该天生产的钢板重量不符合要求.
- 计算 p 值: $p = P(\chi^2 \geqslant \chi_0^2) = P(\chi^2 \geqslant 37.5) = 0.039 < 0.05$, 故在给定显著水平 $\alpha = 0.05$ 下拒绝原假设.

R

```r
n <- 25; s <- sqrt(0.025); sigma0 <- sqrt(0.016)
chisq.obs <- (n - 1)*s^2/sigma0^2
chisq.obs
```

```
## [1] 37.5
```

```r
chisq.crit <- qchisq(0.95, 24)
chisq.crit
```

```
## [1] 36.41503
```

```r
p.value <- 1-pchisq(chisq.obs, 24)
p.value
```

```
## [1] 0.0389818
```

两个正态总体方差的 F 检验

- 设 $X_1, \cdots, X_m$ 是来自 $N(\mu_1, \sigma_1^2)$ 的样本，$Y_1, \cdots, Y_n$ 是来自 $N(\mu_2, \sigma_2^2)$ 的样本，且两个样本相互独立。
- 考虑三种关于 σ_1^2, σ_2^2 的检验问题

- I: $H_0 : \sigma_1^2 \leq \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 > \sigma_2^2$
- II: $H_0 : \sigma_1^2 \geq \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 < \sigma_2^2$
- III: $H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$

两个正态总体方差的 F 检验

检验统计量

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

拒绝域

- $W_I = \{(x_1, \cdots, x_n) : F \geq F_{1-\alpha}(m-1, n-1)\}$
- $W_{II} = \{(x_1, \cdots, x_n) : F \leq F_{\alpha}(m-1, n-1)\}$
- $W_{III} = \left\{(x_1, \cdots, x_n) : F \leq F_{\alpha/2}(m-1, n-1),\ F \geq F_{1-\alpha/2}(m-1, n-1)\right\}$

两个正态总体方差的 F 检验：举例

甲、乙两台机床加工某种零件，零件的直径服从正态分布，总体方差反映了加工精度，为比较两台机床的加工精度有无差别，现从各自加工的零件中分别抽取 7 件产品和 8 件产品，测得其直径为

$$X: \quad 16.2 \quad 16.8 \quad 15.8 \quad 15.5 \quad 16.7 \quad 15.6 \quad 15.8$$

$$Y: \quad 15.9 \quad 16.0 \quad 16.4 \quad 16.1 \quad 16.5 \quad 15.8 \quad 15.7 \quad 15.0$$

两个正态总体方差的 F 检验：举例

- 这是一个双侧假设检验问题，原假设是 $H_0 : \sigma_1^2 = \sigma_2^2$，备择假设为 $H_1 : \sigma_1^2 \neq \sigma_2^2$.
- 此处 $m = 7, n = 8$，经计算 $s_x^2 = 0.2729, \quad s_y^2 = 0.2164$. 于是 $F_0 = \frac{0.2729}{0.2164} = 1.261$.
- 若取 $\alpha = 0.05$，查表知

$$F_{0.975}(6, 7) = 5.12$$

$$F_{0.025} = \frac{1}{F_{0.975}(7, 6)} = \frac{1}{5.70} = 0.175$$

两个正态总体方差的 F 检验：举例

- 其拒绝域为 $W = \{F \leqslant 0.175 或 F \geqslant 5.12\}$
- 由此可见, 样本统计量观测值 $F_0 = \frac{0.2729}{0.2164} = 1.261$ 未落入拒绝域, 即在 0.05 水平下可以认为两台机床的加工精度一致.
- 最后, 计算 p 值. 此处 $F_0 = 1.261$, 故

$$p = 2\min\{P(F \geqslant 1.261), P(F \leqslant 1.261)\}$$
$$= 2\min\{0.3803, 0.6197\} = 0.7606 > 0.05$$

不能拒绝原假设。

R

```r
x <- c(16.2, 16.8, 15.8, 15.5, 16.7, 15.6, 15.8)
y <- c(15.9, 16.0, 16.4, 16.1, 16.5, 15.8, 15.7, 15.0)
var.test(x, y)
```

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 1.2607, num df = 6, denom df = 7, p-value = 0.7608
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2463031 7.1804281
## sample estimates:
## ratio of variances
##           1.260726
```

小结

正态总体方差的假设检验

检验法	H_0	H_1	检验统计量	拒绝域	p 值
χ^2 检验	$\sigma^2 \leqslant \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 \geqslant \chi_{1-\alpha}^2(n-1)$	$P(\chi^2 \geqslant \chi_0^2)$
	$\sigma^2 \geqslant \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$\chi^2 \leqslant \chi_{\alpha}^2(n-1)$	$P(\chi^2 \leqslant \chi_0^2)$
	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$		$\chi^2 \leqslant \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \geqslant \chi_{1-\alpha/2}^2(n-1)$	$2\min\{P(\chi^2 \leqslant \chi_0^2), P(\chi^2 \geqslant \chi_0^2)\}$
F 检验	$\sigma_1^2 \leqslant \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \dfrac{s_x^2}{s_y^2}$	$F \geqslant F_{1-\alpha}(m-1,n-1)$	$P(F \geqslant F_0)$
	$\sigma_1^2 \geqslant \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$		$F \leqslant F_{\alpha}(m-1,n-1)$	$P(F \leqslant F_0)$
	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$		$F \leqslant F_{\alpha/2}(m-1,n-1)$ 或 $F \geqslant F_{1-\alpha/2}(m-1,n-1)$	$2\min\{P(F \leqslant F_0), P(F \geqslant F_0)\}$

小结（二）

1. 掌握单个正态总体均值的检验（单样本 u 检验、单样本 t 检验）
2. 掌握两个正态总体均值差的检验（两样本 u 检验、两样本 t 检验）
3. 理解并掌握成对数据的 t 检验
4. 掌握正态总体方差的检验（χ^2 检验和 F 检验）
5. 熟悉上述假设检验的一般步骤，尤其是检验问题的提出、构造拒绝域形式及确定临界点等方法

作业

- 7.2 课后习题：12, 13, 14, 18, 21, 26

Outline

其他分布参数的假设检验

1. 指数分布参数的假设检验
2. 比例 p 的假设检验
3. 大样本下总体均值的假设检验
4. 似然比检验（拓展内容）

指数分布参数的假设检验

设 $X_1, X_2, \cdots, X_n$ 是来自指数分布 $X \sim \mathrm{Exp}(1/\theta)$ 的样本, 指数分布的期望 $E(X) = \theta$, 关于 θ 的如下检验问题:

$$H_0 : \theta \leqslant \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta \geqslant \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0$$

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

检验统计量 χ^2 为

$$\chi^2 = \frac{2n\bar{X}}{\theta}$$

在 $\theta = \theta_0$ 时, $\chi^2 = \frac{2n\bar{x}}{\theta_0} \sim \chi^2(2n)$.

指数分布、伽马分布和 χ^2 分布

- $\mathrm{Exp}(1/\theta) = \mathrm{Gamma}(1, 1/\theta)$.
- Gamma 分布具有可加性：$n\bar{X} \sim \mathrm{Gamma}(n, 1/\theta)$.
- $\frac{2n\bar{X}}{\theta} \sim \mathrm{Gamma}(n, 1/2) = \chi^2(2n)$.

取显著性水平为 α, 则对应三个检验问题:

1. $H_0 : \theta \leqslant \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$
    - $H_0 : \theta \leqslant \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$
    - 拒绝域: $W = \left\{\chi^2 \geqslant \chi^2_{1-\alpha}(2n)\right\}$
    - p 值: $p = P\left(\chi^2 \geqslant \chi_0^2\right)$
2. $H_0 : \theta \geqslant \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0$
    - $H_0 : \theta \geqslant \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0$
    - 拒绝域: $W = \{\chi^2 \leqslant \chi^2_{\alpha}(2n)\}$
    - p 值: $p = P\left(\chi^2 \leqslant \chi_0^2\right)$
3. $H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$
    - $H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$
    - 拒绝域: $W = \left\{\chi^2 \leqslant \chi^2_{\alpha/2}(2n)\text{或}\chi^2 \geqslant \chi^2_{1-\alpha/2}(2n)\right\}$
    - p 值: $p = 2\min\left\{P\left(\chi^2 \geqslant \chi_0^2\right), P\left(\chi^2 \leqslant \chi_0^2\right)\right\}$

其中, $\chi_0^2 = \frac{2n\bar{x}}{\theta}$ 为 χ^2 的样本观测值。

比例 p 的假设检验

- 比例 p 可看作某事件发生的概率. 作 n 次独立试验, 以 x 记该事件发生的次数, 则 $X \sim b(n, p)$. 我们可以根据 X 检验关于 p 的一些假设. 以如下检验问题为例:

$$H_0 : p \leq p_0 \quad vs \quad H_1 : p > p_0$$

- 直观上检验统计量可取 X 本身, 拒绝域为: $W = \{x \geqslant c\}$, 由于 x 只取整数值, 故 c 可限制在非负整数中.

- 一般情况下，对给定的 α, 不一定能正好取到一个正整数 c 使下式成立:

$$P\left(x \geq c; p_0\right) = \sum_{i=c}^{n} \binom{n}{i} p_0^i \left(1 - p_0\right)^{n-i} = \alpha$$

这是在对离散总体作假设检验中普遍会遇到的问题.

- 一般较常见的是找一个 c_0, 使得

$$\sum_{i=c_0+1}^{n} \binom{n}{i} p_0^i \left(1 - p_0\right)^{n-i} < \alpha < \sum_{i=c_0}^{n} \binom{n}{i} p_0^i \left(1 - p_0\right)^{n-i}$$

这时, 可取 $c = c_0 + 1$, 此时相当于把显著性水平由 α 降低到

$$\sum_{i=c_0+1}^{n} \binom{n}{i} p_0^i \left(1 - p_0\right)^{n-i}$$

- 事实上, 在离散场合使用 p 值更加简便, 只需根据观测值 $x = x_0$ 计算 p 值

$$p = P_{p_0}(x \geqslant x_0)$$

并与给定的 α 比较即可.

三个检验问题的 p 值依次分别为

1. $H_0: p \leq p_0 \quad vs \quad H_1: p > p_0$
   - p 值: $p = P_{p_0}(x \geqslant x_0)$
2. $H_0: p \geq p_0 \quad vs \quad H_1: p < p_0$
   - p 值: $p = P_{p_0}(x \leqslant x_0)$
3. $H_0: p = p_0 \quad vs \quad H_1: p \neq p_0$
   - p 值: $p = 2\min\{P_{p_0}(x \leqslant x_0), \quad P_{p_0}(x \geqslant x_0)\}$ 其中, x_0 为 x 的样本观测值.

举例

某厂生产的产品优质品率一直保持在 40%, 近期对该厂生产的该类产品抽检 20 件, 其中优质品 7 件, 在 α 下能否认为优质品率仍保持在 40%?

- 以 p 表示优质品率, T 表示 20 件产品中的优质品件数, 则 $T \sim b(20, p)$, 待检验的假设为

$$H_0 : p = 0.4 \quad \text{vs} \quad H_1 : p \neq 0.4$$

- 由 $n = 20, t_0 = 7$,

$$\begin{aligned} p &= 2\min\left\{P_{\theta_0}(T \leqslant 7), P_{\theta_0}(T \geqslant 7)\right\} \\ &= 2\min\{0.4159, 0.7500\} = 0.8318 > 0.05 \end{aligned}$$

- 不能拒绝原假设, 可以认为优制品率保持在 40%.

大样本下总体均值的假设检验

在二点分布参数 p 的检验问题中, 临界值的确定比较繁琐, 使用不太方便. 如果样本量较大, 我们可用近似的检验方法—大样本检验. 大样本检验一般思路如下。

- 设 $X_1, \cdots, X_n$ 是来自某总体的样本, 又设该总体均值为 θ, 方差为 θ 的函数, 记为 $\sigma^2(\theta)$. 譬如, 对二点分布 $b(1, \theta)$, 其方差 $\sigma^2(\theta) = \theta(1-\theta)$ 是均值 θ 的函数. 考虑如下三类假设检验问题
  1. $H_0 : \theta \le \theta_0 \quad vs \quad H_1 : \theta > \theta_0$
  2. $H_0 : \theta \ge \theta_0 \quad vs \quad H_1 : \theta < \theta_0$
  3. $H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \ne \theta_0$

- 在样本容量 n 充分大时, 由中心极限定理

$$\bar{X} \dot{\sim} N\left(\theta, \sigma^2(\theta)/n\right)$$

- 故可采取如下检验统计量

$$u = \frac{\sqrt{n}\left(\bar{x} - \theta_0\right)}{\sqrt{\sigma^2(\hat{\theta})}} \dot{\sim} N(0, 1)$$

其中 $\hat{\theta}$ 为 θ 的一致估计。

由此近似地确定拒绝域和 p 值 (与单样本正态场合完全相同):

1. $H_0 : \theta \le \theta_0 \quad vs \quad H_1 : \theta > \theta_0$
   - 拒绝域: $W = \{u \ge u_{1-\alpha}\}$; p 值: $1 - \Phi(u_0)$
2. $H_0 : \theta \ge \theta_0 \quad vs \quad H_1 : \theta < \theta_0$
   - 拒绝域: $W = \{u \le u_{\alpha}\}$; p 值: $\Phi(u_0)$
3. $H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \ne \theta_0$
   - 拒绝域: $W = \{|u| \ge u_{1-\alpha/2}\}$; p 值: $2(1 - \Phi(|u_0|))$ 其中, $u_0 = \frac{\sqrt{n}(\bar{x}-\theta_0)}{\sqrt{\sigma^2(\hat{\theta})}}$ 为 u 的观测值.

举例

某厂产品的不合格品率不高于 10%, 在一次例行检查中, 随机抽取 80 件, 发现有 11 件不合格品, 在 $\alpha = 0.05$ 下能否认为不合格品率仍为 10%?

- 这是关于不合格品率的检验, 假设为:

$$H_0 : \theta \leq 0.1 \quad \text{vs} \quad H_1 : \theta > 0.1$$

因为 $n = 80$ 比较大, 可采用大样本检验方法. $\bar{x} = 11/80$, $\theta_0 = 0.1$, $\hat{\theta}_{MLE} = \bar{x}$, $\sigma^2(\hat{\theta}) = \hat{\theta}(1 - \hat{\theta}) = 11/80 \times 69/80 = 0.1186$

- 检验统计量为

$$u = \frac{\sqrt{80}\left(\frac{11}{80} - 0.1\right)}{\sqrt{0.1186}} = 0.9739$$

若取 $\alpha = 0.05$, 则 $u_{0.95} = 1.645$, 故拒绝域为

$$W = \{u \geq 1.645\}$$

故不能拒绝原假设.

举例

某建筑公司宣称其麾下建筑工地平均每天发生事故数不超过 0.6 起, 现记录了该公司麾下建筑工地 200 天的安全生产情况, 事故数记录如下:

一天发生的事故数	0	1	2	3	4	5	$\geqslant 6$	合计
天数	102	59	30	8	0	1	0	200

试检验该建筑公司的宣称是否成立 (取 $\alpha = 0.05$).

- 以 X 记建筑工地一天发生的事故数, 可认为 X 服从泊松分布, $X \sim P(\lambda)$. 要检验的假设是:

$$H_0 : \lambda \leq 0.6 \quad \text{vs} \quad H_1 : \lambda > 0.6$$

- 由于 $n = 200$ 很大, 可以采用大样本检验, 泊松分布的均值和方差都是 λ, 且 $\hat{\lambda} = \bar{x} = 0.74$. 这里 $\bar{x} = 0.74$, 检验统计量为

$$u = \frac{\sqrt{n}(\bar{x} - \lambda_0)}{\sqrt{\hat{\lambda}}} = \frac{\sqrt{200}(0.74 - 0.6)}{\sqrt{0.74}} = 2.302$$

若取 $\alpha = 0.05$, 则 $u_{0.95} = 1.645$, 拒绝域为 $W = \{u \geq 1.645\}$

- $u = 2.302$ 落入拒绝域, 故拒绝原假设, 认为该建筑公司的宣称不成立.

似然比检验（拓展内容）

- 似然比检验是利用似然函数的比值来进行的假设检验方法.
- 其基本思想是: 要检测某个参数限制是否是正确的, 可以将加入附加限制条件的模型的似然函数最大值与无参数限制的模型的似然函数最大值进行比较. 如果参数限制是正确的, 那么加入这样一个参数应当不会造成似然函数最大值的大幅变动.
- 一般使用两者的比例来进行比较, 对数似然比检验统计量的 2 倍渐近服从卡方分布.

广义似然比

设 $x_1, \cdots, x_n$ 为来自密度函数为 $p(x;\theta), \theta \in \Theta$ 的总体的样本. 考虑如下检验问题:

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$$

令

$$\Lambda\left(x_1, \cdots, x_n\right) = \frac{p\left(x_1, \cdots, x_n; \hat{\theta}\right)}{p\left(x_1, \cdots, x_n; \hat{\theta}_0\right)}$$

其中 $\hat{\theta}$ 表示在全参数空间上的最大似然估计, $\hat{\theta}_0$ 表示在子参数空间 Θ_0 上 θ 的最大似然估计. 则我们称统计量 $\Lambda\left(x_1, \cdots, x_n\right)$ 为上述假设检验的似然比 **(likelihood ratio)**, 也称广义似然比.

- 分子:没有假设时似然函数的最大值
- 分母:原假设成立条件下似然函数的最大值
- 如果 $\Lambda(x_1, \cdots, x_n)$ 很大, 则说明 $\theta \in \theta_1$ 的可能性要远大于 $\theta \in \theta_0$ 的可能性. 于是, 有理由相信 H_0 不成立.

小结（三）

1. 指数分布参数的假设检验
2. 比例 p 的假设检验
3. 大样本下总体均值的假设检验（重点掌握内容）
4. 似然比检验（拓展内容）

作业

- 7.3 课后习题：3-6

- 参数假设检验问题：前面讨论的检验问题都是在总体分布形式已知的前提下对分布的参数建立假设并进行检验
- 思考：总体分布的形式是否正确？如何对总体分布的形式建立假设并进行检验呢？

Outline

1 假设检验的基本思想与概念

2 正态总体参数假设检验

3 其他分布参数的假设检验

4 总体分布检验

总体分布检验

1. 分布的 χ^2 拟合优度检验
2. 列联表的独立性检验
3. 正态性检验（拓展内容）

分布的 χ^2 拟合优度检验

(孟德尔遗传学试验) 孟德尔按颜色与形状把豌豆分成四类：黄圆, 绿圆, 黄皱, 绿皱. 孟德尔根据遗传学原理判断这四类豆子的比例应为 $9:3:3:1$. 为做验证, 孟德尔在一次豌豆实验中获得 $n=566$ 个豌豆, 其中四类豆子的个数分别为 $315, 108, 101, 32$. 问该实验数据是否与孟德尔提出的理论比例吻合?

孟德尔遗传学试验

这是一类分类数据检验问题 (总体分布只取有限个值的情况). 一般形式为: 根据某项指标, 总体被分成 k 类: $A_1, A_2, \cdots, A_k$. 现对该总体作了 n 次观测, k 个类出现的频数分别为:

$$n_1, \ldots, n_k, \text{且} \sum_{i=1}^{k} n_i = n$$

我们关心各类元素在总体中所占的比率的假设:

$$H_0 : P(A_i) = p_i, \quad i = 1, 2, \cdots, k$$

其中诸 $p_i \geq 0$, 且 $\sum_{i=1}^{k} p_i = 1$.

若诸 p_i 均已知

- 如果 H_0 成立, 则对每一类 A_i, 其频率 n_i/n 与概率 p_i 应较接近. 即观测频数n_i 与理论频数np_i 应相差不大. 据此, 英国统计学家 K.Pearson 提出如下检验统计量:

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

并证明在 H_0 成立时对充分大的 n, 上式给出的检验统计量近似服从自由度为 $k-1$ 的 χ^2 分布.

- 拒绝域为:

$$W = \left\{\chi^2 \geq \chi^2_{1-\alpha}(k-1)\right\}$$

孟德尔遗传学试验

- 现对孟德尔遗传学试验 χ^2 拟合优度检验. 注意到

$$k=4, n=556, n_1=315, n_2=108, n_3=101, n_4=32$$

- 待检验假设为:

$$H_0: p_1=\frac{9}{16}, p_2=\frac{3}{16}, p_3=\frac{3}{16}, p_4=\frac{1}{16}$$

$$\chi^2=\frac{(315-312.75)^2}{312.75}+\frac{(108-104.25)^2}{104.25}+\frac{(101-104.25)^2}{104.25}+\frac{(32-34.75)^2}{34.75}=0.47$$

取 $\alpha=0.05$, 则 $\chi^2_{0.95}(3)=7.81>0.47$, 无法拒绝 H_0, 认为孟德尔的结论可接受.

R 实现

```r
x <- c(315, 108, 101, 32)
chisq.test(x, p = c(9, 3, 3, 1)/16)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  x
## X-squared = 0.47002, df = 3, p-value = 0.9254
```

若诸 p_i 不完全已知

若诸 $p_i, i=1,\cdots,k$ 由 $r(r<k)$ 个未知参数 $\theta_1,\cdots,\theta_r$ 确定, 即

$$p_i = p_i\left(\theta_1,\cdots,\theta_r\right), \quad i=1,\cdots,k$$

- 首先给出 $\theta_1,\cdots,\theta_r$ 的极大似然估计, 然后给出诸 $p_i, i=1,\cdots,k$ 的极大似然估计 $\hat{p}_i = p_i\left(\hat{\theta}_1,\cdots,\hat{\theta}_r\right)$.
- Fisher 证明了

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n\hat{p}_i\right)^2}{n\hat{p}_i}$$

在 H_0 成立时近似服从自由度为 $k-r-1$ 的 χ^2 分布, 于是检验拒绝域为 $\left\{\chi^2 \geq \chi^2_{1-\alpha}(k-r-1)\right\}$.

举例

卢瑟福在 2608 个等时间间隔内观测一枚放射性物质放射的粒子数 X. 下表是观测结果的汇总, 其中 n_i 表示 2608 次观测中放射粒子数为 i 的次数.

i	0	1	2	3	4	5	6	7	8	9	10	$\geqslant 11$
n_i	57	203	383	525	532	408	273	139	45	27	10	6

试利用该组数据检验该放射物质在单位时间内放射出的粒子数是否服从泊松分布 $P(\lambda)$.

- 观测到 $0, 1, \cdots, 11$ 共 12 个不同取值, 这相当于把总体分成 12 类. 这里有一个未知参数 λ, 采用极大似然估计,
$$\hat{\lambda} = \bar{x} = 3.870.$$
- 将 $\hat{\lambda}$ 代入概率函数可以估计出诸 $\hat{p}_i$.
$$\hat{p}_k = \frac{\hat{\lambda}^k}{k!}\mathrm{e}^{-\hat{\lambda}}, k = 0, 1, 2, \cdots$$

- 于是可计算出

$$\chi^2 = \sum_{i=1}^{12} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 12.8967$$

- 若取 $\alpha = 0.05$, 则

$$\chi^2_{1-\alpha}(k - r - 1) = \chi^2_{0.95}(10) = 18.307$$

- 本例中 $\chi^2 = 12.8967 < 18.307$, 故没有足够理由拒绝原假设.
- 请使用 **R** 计算出此处检验的 p 值.

列联表的独立性检验

客户购物网站的整体满意度评价结果是否与地区有关?

客户购物网站的整体满意度评价结果

满意度	地区			
	东部	中部	西部	合计
满意	126	158	35	319
不满意	34	82	65	181
合计	160	240	100	500

- 这里涉及到两个类别变量：满意度和地区
- 如何检验两个类别变量之间是否存在关系?

列联表

1. 研究两个类别变量时，每个变量有多个类别，通常将两个变量多个类别的频数用交叉表的形式表示出来个变量放在行 (row) 的位置，称为行变量，其类别数 (行数) 用 r 衣示另一个变量放在列 (column) 的位置，称为列变量，其类别数 (列数) 用 c 表示
2. 这种由两个或两个以上类别变量交叉分类的频数分布表称为列联表 (contingency table)
3. 一个由 r 行和 c 列组成的列联表也称为 $r \times c$ 列联表

列联表

$r \times c$ 列联表

$A \backslash B$	1	$\cdots$	j	$\cdots$	c	行和
1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1c}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	n_{11}	$\cdots$	n_{ij}	$\cdots$	n_{ic}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	n_{r1}	$\cdots$	n_{rj}	$\cdots$	n_{rc}	$n_{r\cdot}$
列和	$n_{\cdot 1}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot c}$	n

列联表分析的基本问题是：考察各属性之间有无关联，即判别两属性是否独立.

列联表的独立性检验

如在前例中，问题是：客户购物网站的整体满意度评价结果是否与地区有关？在 $r \times c$ 列联表，若以 $p_{i\cdot}$, $p_{\cdot j}$ 和 p_{ij} 分别表示总体中的个体属于 A，属于 B 和同时属于 A 与 B 的概率，可得一个二维离散分布表，则 "A、B 两属性独立" 的假设可以表述为

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}, \quad i = 1, \cdots, r, \quad j = 1, \cdots, c$$

二维离散分布表 (注意我们不知道每个概率 p_{ij})

$A\backslash B$	1	$\cdots$	j	$\cdots$	c	行和
1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1c}	$p_{1\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	p_{11}	$\cdots$	p_{ij}	$\cdots$	p_{ic}	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	p_{r1}	$\cdots$	p_{rj}	$\cdots$	p_{rc}	$p_{r\cdot}$
列和	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot c}$	1

- 这里诸 p_{ij} 共有 rc 个参数, 在原假设 H_0 成立时, 这 rc 个参数 p_{ij} 由 $r+c$ 个参数 $p_{1\cdot}, \cdots, p_{r\cdot}$ 和 $p_{\cdot 1}, \cdots, p_{\cdot c}$ 决定. 在这 $r+c$ 个参数中存在两个约束条件:

$$\sum_{i=1}^{r} p_{i\cdot} = 1, \quad \sum_{j=1}^{c} p_{\cdot j} = 1$$

所以, 此时 p_{ij} 实际上由 $r+c-2$ 个独立参数所确定. 据此, 检验统计量为

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$$

- 其中诸 $\hat{p}_{ij}$ 是在 H_0 成立下得到的 p_{ij} 的极大似然估计, 其表达式为

$$\hat{p}_{ij} = \hat{p}_{i\cdot}\hat{p}_{\cdot j} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n}$$

- 在 H_0 成立时, 上式服从自由度为 $rc-(r+c-2)-1=(r-1)(c-1)$ 的 χ^2 分布. 对给定的显著性水平 α, 检验的拒绝域为:

$$W=\left\{\chi^2 \geq \chi^2_{1-\alpha}((r-1)(c-1))\right\}$$

- 对购物网站案例进行课堂练习

R 实现

```r
eg1 <- as.table(matrix(c(126, 34, 158, 82, 35, 65), 2, 3))
row.names(eg1) <- c('Satisfied', 'Unsatisfied')
colnames(eg1) <- c('East', 'Middle', 'West')
chisq.test(eg1)
```

```
##
##  Pearson's Chi-squared test
##
## data:  eg1
## X-squared = 51.827, df = 2, p-value = 5.572e-12
```

正态性检验（拓展内容）

1. 图检验法: 正态概率图（Quantile-to-Quantile plot; QQ 图; R 函数: `qqplot()`)
2. 正态性检验
   - Kolmogorov-Smirnov（KS）正态性检验
   - Liliefor 正态分布检验
   - Shapiro-Wilk（SW）正态性检验
3. 正态性检验的 R 函数
   - `ks.test(x, pnorm, mean, sd, alternative)` (已知的正态分布)
   - `ks.test(x, pnorm, sample mean, sample sd, alternative)` (未知正态分布，样本均值和样本标准差替代 μ 和 σ)
   - `shapiro.test(x, alternative)`

QQ 图

- 是一种通过两组数据的分位数大小比较数据分布的图形工具
- 一般用于数据与已知分布的比较（基本原理是将数据的样本分位数和已知分布的总体分位数相比较，观察是否在一条直线上. 最常用的是正态 QQ 图, 可以直观反映数据是否服从正态分布)
- 也可以比较两组数据的分布 (基本原理是将两组数据分别从小到大排序后, 组成数据对 $x_{(i)}, y_{(i)}$, 描绘二者的散点图. 如果两组数据的分布相近,QQ 图近似成直线; 反之, 则认为两组数据的分布有较大差异)
- 在 R 中调用 `qqplot()`

小结（四）

1. 理解并熟悉分布的 χ^2 拟合优度检验的原理、一般步骤等
2. 熟悉掌握列联表的独立性检验的方法（一般步骤）
3. 熟悉掌握分布 χ^2 拟合优度检验和列联表独立性检验的 R 实现
4. 掌握利用图检验法进行分布检验的原理（尤其是正态分布）
5. 拓展内容：正态性检验（自学）

作业

- 7.4 课后习题：5-7

假设检验模块复习思考题

1. 假设检验的基本思想是什么？其中使用了一条什么原理？
2. 检验的显著性水平 α 的意义是什么？
3. 比较双边、左边和右边检验的拒绝域
4. 使用 u 检验法，t 检验法，χ^2 检验法，F 检验法，分别可以进行哪些假设检验？
5. 正态总体期望与方差的区间估计和假设检验两者之间有什么相似之处？
6. 成对数据差的 t 检验适用于哪些特殊场合？
7. 分布拟合的 χ^2 检验和列联表独立性检验的基本步骤是什么？