第七章 假设检验

Instructor: 郝壮

haozhuang@buaa.edu.cn
School of Economics and Management
Beihang University

March 29, 2022

§7.1 假设检验的基本思想与概念

关于总体参数的统计推断有

- **参数估计:** 点估计; 区间估计
- **假设检验:** 利用样本, 求证对于总体中的某个参数或某种性质的推断(猜测)正确与否的过程.

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, 统计软件实现.

本节内容较为理论, 概念较多, 逻辑较长, 同学们应有心理准备, 仔细学习理解假设检验的推理步骤.

案例: Fisher的"女士品茶"和假设检验

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.

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.

假设检验可研究的问题举例

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

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**(critical value), 从而产生判别的条件和准则.

7.1.2 假设检验问题

例7.1.2 某厂生产的合金强度服从 $N(\theta, 16)$, 其中的设计值θ为不低于110(Pa). 即生产正常的标准是平均强度不低于110(Pa). 从产品中随机抽取25块合金, 测得强度值为$x_1, x_2, ..., x_{25}$, 其均值为$\bar{x} = 108.2$(Pa), 问当日生产是否正常?

- (1) 是参数估计问题吗?
- (2) 回答"是"还是"否", 假设检验问题.
- (3) 命题"合金平均强度不低于110Pa"正确与否仅涉及如下两个参数集合:

$$\Theta_0 = \{\theta : \theta \geq 110\} \quad \Theta_1 = \{\theta : \theta < 110\}$$

命题成立对应$\Theta_0 = \{\theta : \theta \geq 110\}$, 不成立对应$\Theta_1 = \{\theta : \theta < 110\}$.

- 这两个非空参数集合都称作**统计假设**(statistical hypothesis), 简称**假设**.
- (4) 我们的任务是利用样本去判断假设(命题) "$\Theta_0 = \{\theta : \theta \geq 110\}$"是否成立. 这里的"判断"在统计学中称为检验或检验法则.

7.1.2 假设检验的基本步骤

- 一. 建立假设
- 二. 选择检验统计量(test statistics), 给出拒绝域(rejection region)形式
- 三. 选择显著性水平
- 四. 给出拒绝域
- 五. 做出判断

7.1.2 假设检验的基本步骤

由于本节新的理论概念很多, 在给出严格概念之前, 我们先从一个完整的例题对假设检验的最基本的应用有所熟悉, 用实际应用增加理论概念的理解.

引例

某炼铁厂铁水含碳量X服从正态分布: $X \sim N(4.55, 0.108^2)$. 现改变工艺, 检测五炉铁水, 其含碳量为:

$$4.28, \quad 4.40, \quad 4.42, \quad 4.35, \quad 4.37$$

求得

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 4.364 \neq 4.55$$

假设改变工艺并未改变含碳量分布和含碳量的方差, 所以新的铁水含碳服从$N(\mu, 0.108^2)$. 问工艺的改变是否改变了含碳量的期望?

这是要回答如下的问题: 上述的$4.364 \neq 4.55$是因为测量造成的 (偶然因素——随机误差), 还是由于工艺条件改变造成的(系统性因素)?

引例

假设检验的基本思路:

1. 建立零假设(Null hypothesis) H_0 : 新工艺对平均含碳量μ无影响 [即: X依然服从$N(4.55, 0.108^2)$]

记$\mu_0 = 4.55$

则零假设/原假设 $H_0 : \mu = \mu_0$ (所观察到的现象是随机误差造成的)

备择假设(Alternative hypothesis $H_1 : \mu \neq \mu_0$ (所观察到的现象是真实的)

2. 构造“**检验统计量**”

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

思考：Z服从什么分布？(标准正态？)

思考: Z在什么情况下服从标准正态分布？

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

引例

思考: Z在什么情况下服从标准正态分布？

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

由于新的X服从 $N(\mu, 0.108^2)$ 所以只有当**零假设为真**时$H_0 : \mu = \mu_0 = 4.55$, Z服从标准正态

$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

3. 下一步, 选择"**检验水平**": $\alpha = 0.05$.

由于$P\{|Z| \geq 1.96\} = 0.05$, 可知:在 H_0为真的情况下, 事件$P\{|Z| \geq 1.96\}$是一个小概率事件.

4. 判断方法: 计算$|Z|$的观测值

$$|Z| = \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| = \left|\frac{4.364 - 4.55}{0.108/\sqrt{5}}\right| = 3.85 > 1.96$$

一般认为, 小概率事件是不经常发生的.

在原假设成立的条件下, 在一次实验中, 小概率事件居然发生了! 因此怀疑H_0假设的正确性, 而认为**样本可能来源于另一个分布!** 即工艺条件的改变使总体均值发生了显著变化. 因此拒绝 H_0, 认为$\mu \neq \mu_0$.

- 以上给出了假设检验的一个最基础的例子, 下面给出一般化的假设检验的基本步骤, 并解释前面提到的理论概念.

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:
  - **“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
  - **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

练习:(如何选择原假设)

1. 公共场所禁烟对降低吸烟率有效. 检验这一结论是否成立.

2. 采用新技术生产后, 将会使产品的使用寿命明显延长到1500小时以上. 检验这一结论是否成立.

如果Θ_0只包含一个点, 我们称之为**简单原假设**(simple null hypothesis), 包括三种常见检验

- 双侧检验

$$H_0: \theta = \theta_0, \quad H_1: \theta \neq \theta_0$$

- 单侧检验(右尾检验)

$$H_0: \theta = \theta_0, \quad H_1: \theta > \theta_0$$

- 单侧检验(左尾检验)

$$H_0: \theta = \theta_0, \quad H_1: \theta < \theta_0$$

如果Θ_0不止包含一个点, 我们称之为**复合(composit)原假设**

回顾**例7.1.2** 某厂生产的合金强度服从 $N(\theta, 16)$, 其中的设计值θ为不低于110(Pa). 即生产正常的标准是平均强度不低于110(Pa). 从产品中随机抽取25块合金, 测得强度值为$x_1, x_2, ..., x_{25}$, 其均值为$\bar{x} = 108.2$(Pa), 问当日生产是否正常?

在**例7.1.2**中, 我们可建立如下两个假设(Θ_0包含所有大于等于110的参数空间):

$$H_0 : \theta \in \Theta_0 = \{\theta | \theta \geqslant 110\} \quad \text{vs} \quad H_1 : \theta \in \Theta_1 = \{\theta | \theta < 110\}$$

即为复合假设. 这两个假设可以简写为

$$H_0 : \theta \geqslant 110 \quad vs \quad H_1 : \theta < 110$$

二. 选择检验统计量(test statistic), 给出拒绝域(rejection region)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**.
- 使原假设被拒绝的样本观测值所在区域称为**拒绝域**, 一般用$\mathbf{W}$表示. 相对地, 用 $\bar{\mathbf{W}}$表示**接受域**. 注意: 拒绝域和接受域都是就H_0而言的.
- 在例7.1.2中, 样本均值$\bar{x}$愈大, 意味着总体均值θ也大, 因此, 合理的拒绝域形如

$$\mathbf{W} = \{(x_1, \cdots, x_n) : \bar{x} \le c\} = \{\bar{x} \le c\}$$

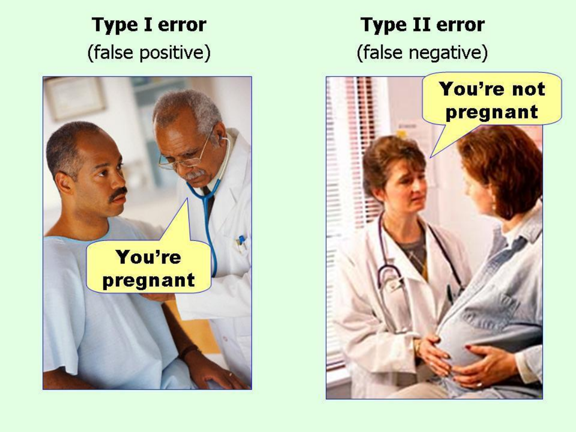
c是待定的**临界值**(critical value).

- 正如在数学上我们不能用一个例子去证明一个结论一样, 用一个样本(例子)不能证明一个命题(假设)是成立的, 但可以用一个例子(样本)推翻一个命题. 因此, 从逻辑上看, 注重拒绝域是适当的.
- 事实上, 在“拒绝原假设”和“拒绝备择假设(从而接收原假设)”之间还有一个模糊域, 如今我们把它并入接收域, 所以接收域是复杂的, 将之称为保留域也许更恰当, 但习惯上已把它称为接收域, 没有必要再进行改变, 只是应注意它的含义.

三. 选择显著性水平

检验可能犯以下**两类错误**:

- **第一类错误(Type I error)**: H_0为真, 但**样本观测值**落在拒绝域中, 从而拒绝原假设H_0.
  发生的概率称为犯第一类错误的概率, 或称**拒真概率**, 记为$\alpha = P_\theta\{X \in \mathbf{W}\}, \theta \in \Theta_0$, 或$P_\theta\{X \in \mathbf{W}|H_0\}$.
- **第二类错误(Type II error)**: H_0不真(即 H_1为真), 但**样本观测值**落在接受域中, 从而接受原假设H_0.
  其发生的概率称为犯第二类错误的概率, 或称**取伪概率**, 记为$\beta = P_\theta\{X \in \bar{\mathbf{W}}\}, \theta \in \Theta_1$, 或$P_\theta\{X \in \bar{\mathbf{W}}|H_1\}$.
- 任何一个假设检验都无法避免犯上述两种错误.



Null hypothesis H_0: not pregnant

Type 1: falsely rejecting a null hypothesis

Type 2: falsely accepting a null hypothesis

势函数(power function)

能否使得一个检验犯两类错误的概率都尽可能小?

不能! 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**. 势函数是假设检验中最重要的概念之一.

定义7.1.1 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$\mathbf{W}$, **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:

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

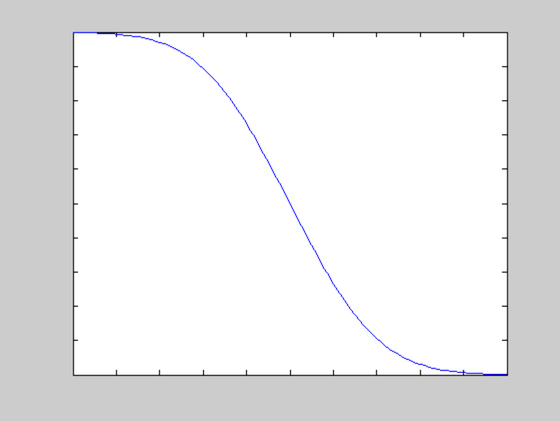
或

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

对**例7.1.2**, 其拒绝域为$W=\{\bar{x}\le c\}$, 注意到$\bar{x}\sim N(\theta,16/25)$, 由势函数定义可以算出该检验的势函数

$$\begin{aligned}g(\theta)&=P_\theta(X\in W)\\&=P_\theta(\bar{x}\le c)=P_\theta\left(\frac{\bar{x}-\theta}{4/5}\le\frac{c-\theta}{4/5}\right)=\Phi\left(\frac{c-\theta}{4/5}\right)\end{aligned}$$

- 这个势函数是θ的减函数



利用这个势函数容易写出犯两类错误的概率分别为

$$\alpha(\theta) = \Phi\left(\frac{c-\theta}{4/5}\right), \quad \theta \in \Theta_0$$

(原假设正确但拒绝了原假设的概率)

和

$$\beta(\theta) = 1 - \Phi\left(\frac{c-\theta}{4/5}\right), \quad \theta \in \Theta_1$$

(原假设错误但接受了原假设的概率)

由此可得如下结论:

- 当α减小时, c 也随之减小, 必导致β的增大;
- 当β减小时, c 会增大, 必导致α的增大;

- 说明:在样本量一定的条件下不可能找到一个使α和β都小的检验.
- 如必要, 怎样才能同时减少弃真概率和取伪概率? 增加样本容量!

- 英国统计学家 Neyman 和 Pearson 提出水平为α的显著性检验的概念.
- 通常做法是仅限制第一类错误的概率. 因为由检验假设的设定, 一般第一类错误更严重.

水平为α的检验

定义7.1.2 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

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

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.**

α为犯第一类错误的概率, 不能过小(会导致犯第二类错误的概率β过大). 通常的选择是$\alpha = 0.05$, 有时也选择$\alpha = 0.01$或$\alpha = 0.1$.

四. 给出拒绝域

确定显著性水平后, 可以定出检验的拒绝域W.

在例7.1.2中, 对于给定的α, **水平为α的检验**要求对任意的$\theta \geqslant 110$有

$$g(\theta) = \Phi\left(\frac{5(c-\theta)}{4}\right) \leq \alpha$$

由于$g(\theta)$关于θ单调减, 只需要

$$g(110) = \Phi\left(\frac{5(c-110)}{4}\right) = \alpha$$

上式可以改写为

$$\frac{5(c-110)}{4} = u_\alpha$$

从而可算出c的值

$$c = 110 + 0.8u_\alpha$$

因为$c = 110 + 0.8u_\alpha$, 所以检验的拒绝域为

$$W = \{\bar{x} \leqslant 110 + 0.8u_\alpha\}$$

若取$\alpha = 0.05$, 则$u_{0.05} = -u_{0.95}$. 这给出c的值为

$$c = 110 + 0.8u_{0.05} = 110 - 0.8 \times 1.645 = 108.684$$

检验的拒绝域为

$$W = \{\bar{x} \leq 108.684\}$$

若令

$$u=\frac{\bar{x}-110}{4/5}$$

则拒绝域有另一种表示:

$$W=\{u\le u_{0.05}\}=\{u\le -1.645\}$$

通常用检验统计量u表示拒绝域.

五. 作出判断

在有了明确的拒绝域后, 根据样本观测值我们可以做出判断:

- 当$u \leq -1.645$ 或 $\bar{x} \leq 108.684$ 时, 则拒绝H_0, 接受H_1
- 当$u > -1.645$ 或 $\bar{x} > 108.684$时, 则接受H_0

在**例7.1.2**中, 由于

$$u = \frac{108.2 - 110}{4/5} = -2.25 < -1.645 \text{或} \bar{x} = 108 < 108.684$$

因此拒绝原假设, 即认为该日生产不正常.

小结: 统计检验的步骤

1. 根据实际问题, 建立统计假设$H_0 \quad vs \quad H_1$.

2. 选取合适的检验统计量$T(X)$, 使该统计量当H_0成立时, T的分布完全已知, 并根据H_0, H_1的特点, 初步确定拒绝域W形状.

3. 确定显著性水平α, 确定具体的拒绝域W.

4. 由样本观测值$x_1, ..., x_n$, 计算检验统计量$T(x_1, ..., x_n)$, 由$T(x_1, ..., x_n)$判断是否拒绝原假设.

特别注意: 不拒绝H_0并不是肯定H_0一定对, 有可能是由于样本量小导致差异不够显著, 没有达到足以否定H_0的程度. 所以假设检验又叫"显著性检验".

p值概念的引出：显著水平选择问题

以例7.1.2检验的拒绝域为例

$$W = \{\bar{x} \leqslant 110 + 0.8u_\alpha\}$$

拒绝域的大小很显然受选择的检验水平α影响

- 当α较大, 如$\alpha = 0.05$, 则拒绝域较大
- 当α较小, 如$\alpha = 0.01$, 则拒绝域较小

相同的观测值$\bar{x}$可能会因为不同检验水平的选择导致接受或拒绝原假设的判断不同.

为此, 引入**检验的p值**(p-value)的概念, 表示利用样本观测值做出拒绝原假设的最小显著水平.

检验的p值

定义7.3.1 在一个假设检验问题中, 利用观测值能够做出拒绝原假设的最小显著性水平称为**检验的p值.**

引进检验的p值的概念有明显的好处:

- 第一, 客观, 避免了事先确定显著水平;
- 其次, 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.
- p 值在应用中很方便, 如今的统计软件中对检验问题一般都会给出检验的p值.

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5

7.1课后习题: 1-5

7.2课后习题: 1,2,5,6,9

上机实验3

使用Stata内置数据集auto.dta (sysuse auto). 再如下两种情况下对变量mpg做总体均值的假设检验.

a. 假设总体方差已知为36, 检验mpg均值是否小于20 (H0: 总体均值小于20).

b. 假设总体方差未知, 检验mpg均值是否等于18 (H0: 总体均值等于18).

7.2.1 单个正态总体均值的检验

设$x_1, ..., x_n$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

- (1)$H_0 : \mu \leqslant \mu_0$ vs $H_1 : \mu > \mu_0$
- (2)$H_0 : \mu \geqslant \mu_0$ vs $H_1 : \mu < \mu_0$
- (3)$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

- 当备择假设H_1在原假设 H_0 一侧时的检验称为**单侧检验**;
- 当备择假设H_1在原假设 H_0 两侧时的检验称为称为**双侧检验**.

一. 已知σ 时的u检验

设$x_1, \ldots, x_n$是来自总体$N(\mu, \sigma^2)$的样本, 考虑关μ的检验问题

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

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).

一. 已知σ 时的u检验

设$x_1, \ldots, x_n$是来自总体$N(\mu, \sigma^2)$的样本, 考虑关μ的检验问题

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

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.

综上, 对于

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

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$\{u \geqslant c\}$.

下一步, 选择显著性水平. 若要求检验的显著性水平为α, **则对于使原假设**$\{\mu \leqslant \mu_0\}$**成立的中任意的**μ, c应满足检验统计量落在拒绝域的概率不超过α, 即

$$P_{\mu_0}(u \geqslant c) \leqslant \alpha$$

由于在$\mu = \mu_0$时$u \sim N(0, 1)$, 若使上式中等号成立, 就有$c = u_{1-\alpha}$,

在H_0当中, 还存在$\mu < \mu_0$的情况, 但考虑到$\bar{x} \sim N(\mu, \sigma^2/n)$且

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

既然$\mu = \mu_0$的情况下

$$P_{\mu_0}(u \geqslant u_{1-\alpha}) = \alpha,$$

在$\mu < \mu_0$的情况, 如果 $c = u_{1-\alpha}$就一定能保证

$$P_{\mu_0}(u \geqslant u_{1-\alpha}) \leqslant \alpha.$$

所以, 所以最后拒绝域为

$$W = \{u \geqslant u_{1-\alpha}\}.$$

以上的分析隐含了一个重要的结论:

说明

$$\text{问题1:}H_0:\mu=\mu_0 \quad \text{vs} \quad H_1:\mu>\mu_0$$

和

$$\text{问题2:}H_0:\mu\leqslant\mu_0 \quad \text{vs} \quad H_1:\mu>\mu_0$$

两类检验是相同的, 其根本原因是这两类检验问题的拒绝域是相同的.

这可以简单理解为:

- 如果在问题1中拒绝了H_0, 即拒绝$\mu=\mu_0$, 则在问题2中也应拒绝$H_0:\mu\leqslant\mu_0$ (由于拒绝域相同)
- 如果在问题1中接受了H_0, 即接受$\mu=\mu_0$, 则在问题2中也应接受$H_0:\mu\leqslant\mu_0$

我们通过势函数的分析方法可以证明上述分析的正确性.

注意: 在此势函数推导方法与教材中不同, 而是从最基本的样本均值$\bar{x}$出发(7.1.2中描述的步骤), 更加便于理解.

在上文拒绝域用检验统计量u给出$W = \{u \geqslant c\}$. 相同地, 拒绝域可以通过样本均值给出$W = \{\bar{x} \geqslant c'\}$ 该检验的势函数是μ的函数, 可以用正态分布写出. 即

$$\begin{aligned} g(\mu) &= P_\mu\left(X \in W\right) = P_\mu\left(\bar{x} \geqslant c'\right) \\ &= P_\mu\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geqslant \frac{c'-\mu}{\sigma/\sqrt{n}}\right) = 1 - \Phi(\frac{c'-\mu}{\sigma/\sqrt{n}}) \end{aligned}$$

检验的显著水平为α即对于所有的$H_0: \mu \leqslant \mu_0$有

$$g(\mu) = 1 - \Phi(\frac{c'-\mu}{\sigma/\sqrt{n}}) \leqslant \alpha$$

由于$g(\mu)$是 μ的增函数, 根据**7.1.2(四)**的分析, 仅需

$$g(\mu_0) = 1 - \Phi(\frac{c' - \mu_0}{\sigma/\sqrt{n}}) = \alpha$$

即$\frac{c'-\mu_0}{\sigma/\sqrt{n}} = u_{1-\alpha}$, 所以

$$c' = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}$$

故拒绝域为$\{\bar{x} \geqslant c' = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\}$

用原检验统计量u表示, 则拒绝域为

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geqslant \frac{c' - \mu_0}{\sigma/\sqrt{n}} = \frac{\mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} - \mu_0}{\sigma/\sqrt{n}} = u_{1-\alpha}$$

计算检验的p值

最后, 我们还可以计算该检验的p值(拒绝原假设的最小显著性水平). 我们已推出用检验统计量u表示的拒绝域为

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geqslant u_{1-\alpha}.$$

则最小的使上式成立的检验水平α即为所求p值 (α越小, $u_{1-\alpha}$越大). 故

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = u_{1-p}.$$

即

$$\Phi(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}) = 1 - p.$$

解得: $p = 1 - \Phi(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}})$

若记检验统计量的**观测值**是 (注意: 此时$\bar{x}$表示的是样本观测值.)

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

则 $p = 1 - \Phi(u_0)$

- 当$p > \alpha$, 则观察值不在拒绝域里, 接受原假设
- 当$p \leqslant \alpha$, 则观察值落在拒绝域里, 拒绝原假设

对第二个检验问题

- (2)$H_0: \mu \geqslant \mu_0$ vs $H_1: \mu < \mu_0$

分析方法完全一致, 拒绝域为

$$W = \{u \leqslant u_\alpha\}.$$

检验的p值为 $p = \Phi(u_0)$

同样地,

$$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0$$

和该检验检验是相同的, 其根本原因是这两类检验问题的拒绝域是相同的.

对第三个检验问题

- (3)$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

分析方法类似, 由于为双侧检验, 拒绝域应该在两侧

$$W = \{|u| \geqslant c\}.$$

通过相同逻辑计算可得$c = u_{1-\alpha/2}$

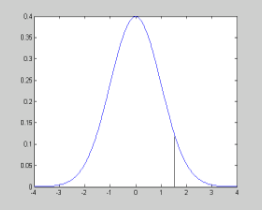
最后拒绝域为

$$W = \{|u| \geqslant u_{1-\alpha/2}\}.$$

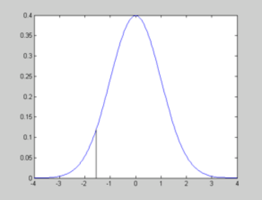
检验的p值为 $p = 2(1 - \Phi(|u_0|))$

三种假设的拒绝域形式分别见下图:

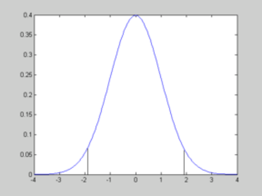
- $(1)H_0: \mu \leqslant \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0$
- $(2)H_0: \mu \geqslant \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0$
- $(3)H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0$



$W = \{u > c\}$



$W = \{u < c'\}$



$W = \{|u| > c''\}$

例7.2.1 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?

解:

例7.2.1 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?

解: 这是一个假设检验的问题, 总体$X \sim N(\mu, 0.2^2)$. 检验假设:$H_0 : \mu = 8 \quad v.s. \quad H_1 : \mu \neq 8$

这个双侧检验问题的拒绝域为$\{|u| \geq u_{1-\alpha/2}\}$

例7.2.1

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

$$\bar{x} = 8.15, \quad u_0 = \sqrt{5}(8.15 - 8)/0.2 = 1.68$$

u_0 值未落入拒绝域内, 故不能拒绝原假设, 可认为猜测成立.

也可以用p值完成假设检验. 由于 $u_0 = 1.68$, 所以 $p = 2(1 - \Phi(u_0)) = 0.09$, 在0.05水平下无法拒绝原假设. (注意：但若预先给定的水平是0.1, 则可以拒绝原假设).

上例的p值的Stata计算

上例的p值是多少?

解:

$$p = 2(1 - \Phi(|u_0|)) = 2(1 - \Phi((8.15 - 8)\sqrt{5}/0.2)) = .09353251$$

```
. di 2*(1-normal((8.15-8) * sqrt(5)/0.2))
.09353251
```

课堂练习

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).

- (A) 达到
- (B) 没有达到

课堂练习

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).

- (A) 达到
- (B) 没有达到

答案: B

解: 单侧检验问题:

$$H_0: \mu = 32 \quad vs \quad H_1: \mu < 32$$

拒绝域

$$W = \{u \leqslant u_\alpha\}.$$

选择检验水平$\alpha = 0.05$, $u_\alpha = u_{0.05} = -u_{0.95} = -1.65$

u的观测值为

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{29.5 - 32}{4/\sqrt{25}} = -3.125 < -1.65$$

所以, 拒绝原假设. 判断钻头寿命没有达到32.

```
. di invnormal(0.05)
-1.6448536
. di (29.5-32)/4*5
-3.125
```

由于σ未知,

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

不是统计量(所以也无法作为检验统计量). 此时将上式中未知的总体标准差σ替换成样本标准差s, 这就形成t**检验统计量:**

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

在$\mu = \mu_0$时,

$$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

此时, 由**推论5.4.2**,

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$$

用完全类似上小节的分析方法, 可得三种假设的拒绝域形式和p值为:

- (1) $H_0 : \mu \leqslant \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$
  拒绝域: $W = \{t \geqslant t_{1-\alpha}(n-1)\}$ (教材表4)
  p值: $p = P(t \geqslant t_0)$ (可用计算机软件求得, 后文详述)
- (2) $H_0 : \mu \geqslant \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$
  拒绝域: $W = \{t \leqslant t_{\alpha}(n-1)\}$
  p值: $p = P(t \leqslant t_0)$
- (3) $H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$
  拒绝域: $W = \left\{|t| \geqslant t_{1-\alpha/2}(n-1)\right\}$
  p值: $p = P(|t| \geqslant |t_0|)$

其中, $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t(n-1)$为$t$检验统计量的样本观测值.

由于实际工作中大多数情况总体方差是未知的, 所以t检验更常用.

例7.2.2 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

$$239.7 \quad 239.6 \quad 239 \quad 240 \quad 239.2$$

试判断该厂此类铝材的长度是否满足设定要求?

解: 这是一个关于正态均值的双侧假设检验问题, 且σ未知, 采用t检验, 拒绝域为:

$$\{|t| \geq t_{1-\alpha/2}(n-1)\}$$

若取 $\alpha = 0.05$, 则 $t_{0.975}(4) = 2.776$ (查表4, 或用stata计算, 见下页).

现由样本计算得到: $\bar{x} = 239.5, \quad s = 0.4$. 故

$$|t| = \sqrt{5} \cdot |239.5 - 240|/0.4 = 2.795$$

由于$2.7951 > 2.776$, 故拒绝原假设, 认为该厂生产的铝材的长度不满足设定要求. (**注意:**第二版教材错误t应为$|t|$, 第三版已修改.)

Stata 计算$t_{0.975}(4) = 2.776$

```
help invt

\\invt(df,p)
\\Description:  the inverse cumulative
\\Student's t distribution: if t(df,t) = p,
\\ then invt(df,p) = t
\\     Domain df:     2e-10 to 2e+17 (may be noninteg
\\     Domain p:      0 to 1
\\      Range:          -8e+307 to 8e+307

. di invt(4, 0.975)
2.7764451
```

下面用p值再做一次检验.

$$|t_0| = \sqrt{5} \cdot |239.5 - 240|/0.4 = 2.795$$

(注意:第二版教材错误t_0应为$|t_0|$, 第三版已修正.)

t服从自由度为4的t分布, 所以

$$p = P(|t| \geqslant |t_0|) = 2P(t \geqslant 2.795)$$

用计算机软件可计算$2P(t \geqslant 2.795) = 0.0491$, 由于$p$小于事先给定的显著水平, 故拒绝原假设.

用STATA计算$P(t \geqslant 2.795)$:

```
help t
//t(df,t)
//Description: the cumulative Student's
//t distribution with df degrees of freedom
//Range: 0 to 1
. di 2*(1-t(4, 2.795))
.04906106
```

表7.2.1 单个正态总体的均值的检验问题

检验法	H_0	H_1	检验统计量	拒绝域	p值
u检验 (σ已知)	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$	$\mu \ge \mu_0$ $\mu \le \mu_0$ $\mu \ne \mu_0$	$u = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$\{u \ge u_{1-\alpha}\}$ $\{u \le u_\alpha\}$ $\{\|u\| \ge u_{1-\alpha/2}\}$	$1-\Phi(u_0)$ $\Phi(u_0)$ $2(1-\Phi(\|u_0\|))$
t检验 (σ未知)	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$	$\mu \ge \mu_0$ $\mu \le \mu_0$ $\mu \ne \mu_0$	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$\{t \ge t_{1-\alpha}(n-1)\}$ $\{t \le t_\alpha(n-1)\}$ $\{\|t\| \ge t_{1-\alpha/2}(n-1)\}$	$P(t \ge t_0)$ $P(t \le t_0)$ $P(\|t\| \ge \|t_0\|)$

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

7.2.2 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 $x_1, \ldots, x_n$ 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

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

可以改写为

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

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

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

所以, 区间估计中的置信区间和和假设检验中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.

设$x_1,\ldots,x_m$是来自总体$N(\mu_1,\sigma_1^2)$的样本, $y_1,\ldots,y_n$是来自总体$N(\mu_2,\sigma_2^2)$的样本, 两样本**相互独立**. 考虑如下三类关于μ的检验问题

$$
\begin{aligned}
H_0: \mu_1-\mu_2 \leqslant 0 \quad &\text{vs} \quad H_1: \mu_1-\mu_2 > 0 \\
H_0: \mu_1-\mu_2 \geqslant 0 \quad &\text{vs} \quad H_1: \mu_1-\mu_2 < 0 \\
H_0: \mu_1-\mu_2 = 0 \quad &\text{vs} \quad H_1: \mu_1-\mu_2 \neq 0
\end{aligned}
$$

我们对常用的两中情况进行讨论: σ_1,σ_2已知的情况; $\sigma_1=\sigma_2=\sigma$未知的情况.

一. σ_1, σ_2已知, 两样本均值差$\mu_1 - \mu_2$的u检验

若σ_1, σ_2已知, 则$\mu_1 - \mu_2$的点估计的分布已知

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

这是因为$\bar{x} \sim N(\mu_1, \sigma_1^2/m)$, $\bar{y} \sim N(\mu_2, \sigma_2^2/n)$.

由此可采用u检验统计量,

$$u = (\bar{x} - \bar{y}) / \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

在$\mu_1 = \mu_2$时, $u \sim N(0, 1)$.

用类似7.2.1小结的的分析方法, 可得三种假设的拒绝域形式和p值为:

- (1) $H_0: \mu_1 - \mu_2 \leqslant 0 \quad$ vs $\quad H_1: \mu_1 - \mu_2 > 0$
  拒绝域: $W = \{u \geqslant u_{1-\alpha}\}$
  p值: $p = 1 - \Phi(u_0)$

- (2) $H_0: \mu_1 - \mu_2 \geqslant 0 \quad$ vs $\quad H_1: \mu_1 - \mu_2 < 0$
  拒绝域: $W = \{u \leqslant u_\alpha\}$
  p值: $p = \Phi(u_0)$

- (3) $H_0: \mu_1 - \mu_2 = 0 \quad$ vs $\quad H_1: \mu_1 - \mu_2 \neq 0$
  拒绝域: $W = \{|u| \geqslant u_{1-\alpha/2}\}$
  p值: $p = 2(1 - \Phi(|u_0|))$

其中, $u_0 = (\bar{x} - \bar{y}) / \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$为$u$的样本观测值.

二. $\sigma_1 = \sigma_2 = \sigma$但未知时, 两样本均值差$\mu_1 - \mu_2$的$t$检验

若$\sigma_1 = \sigma_2 = \sigma$未知, 则首先$\mu_1 - \mu_2$的点估计的分布为

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, (\frac{1}{m} + \frac{1}{n})\sigma^2\right)$$

这是因为$\bar{x} \sim N(\mu_1, \sigma^2/m)$, $\bar{y} \sim N(\mu_2, \sigma^2/n)$.

所以,

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, 1)$$

进一步, 由定理5.4.1,

$$\frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i-\bar{x})^2 \sim \chi^2(m-1), \quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})^2 \sim \chi^2(n-1)$$

所以

$$\frac{1}{\sigma^2}\left(\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2\right) \sim \chi^2(m+n-2)$$

综上, 有

$$t=\frac{\dfrac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{\dfrac{\frac{1}{\sigma^2}\left(\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2\right)}{m+n-2}}} \sim t(m+n-2)$$

记

$$s_w^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2\right]$$

则

$$t=\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$$

于是给出检验统计量

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

用类似7.2.1小结的的分析方法, 可得三种假设的拒绝域形式和p值为:

- (1) $H_0: \mu_1 - \mu_2 \leqslant 0$ vs $H_1: \mu_1 - \mu_2 > 0$
  拒绝域: $W = \{t \geqslant t_{1-\alpha}(m+n-2)\}$
  p值: $p = P(t \geqslant t_0)$
- (2) $H_0: \mu_1 - \mu_2 \geqslant 0$ vs $H_1: \mu_1 - \mu_2 < 0$
  拒绝域: $W = \{t \leqslant t_{\alpha}(m+n-2)\}$
  p值: $p = P(t \leqslant t_0)$
- (3) $H_0: \mu_1 - \mu_2 = 0$ vs $H_1: \mu_1 - \mu_2 \neq 0$
  拒绝域: $W = \left\{|t| \geqslant t_{1-\alpha/2}(m+n-2)\right\}$
  p值: $p = P(|t| \geqslant |t_0|)$

其中, $t_0 = \frac{(\bar{x}-\bar{y})}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$的样本观测值.

例7.2.3 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:

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

根据经验, 硬度服从正态分布, 且方差保持不变. 试在显著性水平α下判断镍合金的硬度是否有明显提高.

解: 用X 表示镍合金的硬度, Y 表示铜合金的硬度, 则由假定, $X \sim N\left(\mu_1, \sigma^2\right)$, $Y \sim N\left(\mu_2, \sigma^2\right)$.

要检验的假设是:$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 > \mu_2$.

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,

$$t = \frac{73.39 - 68.2756}{4.4494 \cdot \sqrt{\frac{1}{8} + \frac{1}{9}}} = 2.4233$$

查表知

$$t_{0.95}(15) = 1.7531$$

由于

$$t > t_{0.95}(15)$$

故拒绝原假设, 可判断镍合金硬度有显著提高.

成对数据(paired data)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. 通过下例解释, 两样本t检验结论对成对数据结构可能是错误的, 在此基础上, 提出成对数据t检验.

例 7.2.4 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?

一. 两样本t检验(错误方法)

假定$x \sim N(\mu_1, \sigma_1^2)$, $y \sim N(\mu_2, \sigma_2^2)$, x与y独立, 假定方差相等是合理的假设$\sigma_1^2 = \sigma_2^2 = \sigma^2$. 记两种种子样本均值和样本方差分别为 $\bar{x}, \bar{y}, s_x^2, s_y^2$. 则检验问题为

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2$$

两样本t检验的统计量t_1和拒绝域W_1分别为

$$t_1 = \frac{\bar{x}-\bar{y}}{s_w/\sqrt{n/2}}$$
$$W_1 = \left\{|t_1| > t_{1-a/2}(2n-2)\right|$$

其中$s_w^2 = \left(s_x^2 + s_y^2\right)/2$.

由数据

$$\bar{x} = 33.1, \bar{y} = 35.7, s_x^2 = 33,2110, s_y^2 = 14.2333, s_w^2 = 23.7222$$

所以

$$t_1 = \frac{33.1 - 35.7}{4.8705/\sqrt{10/2}} = -1.1937$$

若取$\alpha = 0.05$, 查表可得$t_{0.975}(18) = 2.1009$. 由于$|t_1| < t_{0.975}(18)$, 无法拒绝原假设. 认为两种种子没有显著差别.

然而, 上述的检验没有考虑到不同的土地质量对于谷物产量的影响. 假设土地质量对不同种子产量影响相同, 成对数据可以通过做差$d_i = x_i - y_i, i = 1, 2, ..., 10$去除这种影响.

二. 成对数据t检验(正确方法)

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$\mu = \mu_1 - \mu_2$. 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.

根本原因: 两个总体不相互独立.

特别注意

教材错误: 教材中所写的总体方差$\sigma_d^2 = \sigma_1^2 + \sigma_2^2$ 仅在对两总体分布的独立性做某些假设时成立:

- 1. x_i, y_i独立独立时成立, 但如果土地质量对来自同一块土地的两种种子产量有相同影响, 则 x_i, y_i独立一般不成立.
- 2. 若对 x_i, y_i做如下结构化假设: $x_i = s_{1i} + \epsilon_i$, $y_i = s_{2i} + \epsilon_i$, 其中, ϵ_i表示第i块土地对种子1, 种子2 的相同影响; s_{1i}, s_{2i}独立. 则$d = s_{1i} - s_{2i}$.
  $\sigma_d^2 = \sigma_{s1}^2 + \sigma_{s2}^2 < \sigma_1^2 + \sigma_2^2$

如不对总体分布独立性做假设则$\sigma_d^2 = \sigma_1^2 + \sigma_2^2$不成立!

但并不影响我们用样本方差s_d^2对未知的总体方差做一致估计! 所以教材中推断过程和结论依然正确.

单或双样本u检验与t检验

```
//u test (z tests)
// (mean-comparison tests, known variance)
help ztest


// ttest (mean-comparison tests),
// One-sample t test/ Two-sample t test
help ttest
```

单样本正态均值u检验(总体方差已知)

```
//单样本u test (z tests), 假设正态分布, 假设总体方差已知
. sysuse auto
(1978 Automobile Data)

. ztest mpg==20, sd(6)

One-sample z test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     mpg |      74     21.2973    .6974858           6    19.93025    22.66434
------------------------------------------------------------------------------
    mean = mean(mpg)                                              z =   1.8600
Ho: mean = 20

    Ha: mean < 20              Ha: mean != 20              Ha: mean > 20
 Pr(Z < z) = 0.9686        Pr(|Z| > |z|) = 0.0629        Pr(Z > z) = 0.0314
```

单样本t检验

```
//单样本t test, 假设正态分布, 假设总体方差未知
. sysuse auto
(1978 Automobile Data)

. ttest mpg==20

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     mpg |      74     21.2973    .6725511    5.785503     19.9569    22.63769
------------------------------------------------------------------------------
    mean = mean(mpg)                                              t =   1.9289
Ho: mean = 20                                    degrees of freedom =       73

    Ha: mean < 20                Ha: mean != 20                 Ha: mean > 20
 Pr(T < t) = 0.9712         Pr(|T| > |t|) = 0.0576          Pr(T > t) = 0.0288
```

双样本u检验，非成对

```
//双样本u test (z tests)，假设正态分布，假设总体方差已知
. webuse fuel

Two-sample unpaired z test using variables
. ztest mpg1==mpg2, unpaired sd(3)

Two-sample z test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .8660254           3    19.30262    22.69738
    mpg2 |      12       22.75    .8660254           3    21.05262    24.44738
---------+--------------------------------------------------------------------
    diff |               -1.75    1.224745               -4.150456    .6504558
------------------------------------------------------------------------------
    diff = mean(mpg1) - mean(mpg2)                                z =  -1.4289
Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0765         Pr(|Z| > |z|) = 0.1530          Pr(Z > z) = 0.9235
```

双样本t检验, 非成对

```
//双样本t test (z tests), 假设正态分布
. webuse fuel

. ttest mpg1==mpg2, unpaired

Two-sample t test with equal variances
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .7881701    2.730301    19.26525    22.73475
    mpg2 |      12       22.75    .9384465    3.250874    20.68449    24.81551
---------+--------------------------------------------------------------------
combined |      24      21.875    .6264476    3.068954    20.57909    23.17091
---------+--------------------------------------------------------------------
    diff |               -1.75    1.225518               -4.291568    .7915684
------------------------------------------------------------------------------
    diff = mean(mpg1) - mean(mpg2)                                t =  -1.4280
Ho: diff = 0                                     degrees of freedom =       22

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0837         Pr(|T| > |t|) = 0.1673          Pr(T > t) = 0.9163
```

双样本t检验，成对

```
//双样本t test (z tests), 假设正态分布
. webuse fuel
. ttest mpg1==mpg2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .7881701    2.730301    19.26525    22.73475
    mpg2 |      12       22.75    .9384465    3.250874    20.68449    24.81551
---------+--------------------------------------------------------------------
    diff |      12       -1.75    .7797144     2.70101    -3.46614   -.0338602
------------------------------------------------------------------------------
     mean(diff) = mean(mpg1 - mpg2)                               t =  -2.2444
 Ho: mean(diff) = 0                              degrees of freedom =       11

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0232         Pr(|T| > |t|) = 0.0463          Pr(T > t) = 0.9768
```

一, 单个正态总体方差的χ^2检验

设$x_1, ..., x_n$是来自总体$N(\mu, \sigma^2)$的样本, 有关参数σ假设检验常见的有三种基本形式

- (1)$H_0 : \sigma^2 \leqslant \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$
- (2)$H_0 : \sigma^2 \geqslant \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$
- (3)$H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$

通常假定μ未知, 它们采用的检验统计量都是

$$\chi^2 = (n-1)s^2/\sigma_0^2$$

由定理5.4.1, 当$\sigma^2 = \sigma_0^2$时, $\chi^2 \sim \chi^2(n-1)$.

取显著性水平为α, 则对应三个检验问题的拒绝域和p值依次分别为

- (1) $H_0: \sigma^2 \leqslant \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$
  拒绝域: $W = \left\{\chi^2 \geq \chi^2_{1-\alpha}(n-1)\right\}$
  p值: $p = P(\chi^2 \geqslant \chi_0^2)$
- (2) $H_0: \sigma^2 \geqslant \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$
  拒绝域: $W = \left\{\chi^2 \leq \chi^2_{\alpha}(n-1)\right\}$
  p值: $p = P(\chi^2 \leqslant \chi_0^2)$
- (3) $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$
  拒绝域:
  $W = \left\{\chi^2 \leq \chi^2_{\alpha/2}(n-1) \quad \text{或} \quad \chi^2 \geq \chi^2_{1-\alpha/2}(n-1)\right\}$
  p值: $p = 2min\{P(\chi^2 \geqslant \chi_0^2), P(\chi^2 \leqslant \chi_0^2)\}$

其中, $\chi_0^2 = (n-1)s^2/\sigma_0^2$为$\chi^2$的样本观测值.

拓展知识

单个总体**标准差检验**和**方差检验**是同一问题. 为什么?

- (1)$H_0 : \sigma \leqslant \sigma_0$ vs $H_1 : \sigma > \sigma_0$
- (2)$H_0 : \sigma \geqslant \sigma_0$ vs $H_1 : \sigma < \sigma_0$
- (3)$H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma \neq \sigma_0$

故标准差检验的检验统计量和方差检验相同, 拒绝域, p值也相同.

例7.2.5 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.

解:

例7.2.5 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.

解: 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:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \times 0.025}{0.016} = 37.5 > 36.415$$

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.

二, 两个正态总体方差比的F 检验

设$x_1, \dots, x_m$是来自总体$N(\mu_1, \sigma_1^2)$的样本, $y_1, \dots, y_n$是来自总体$N(\mu_2, \sigma_2^2)$的样本, 两样本相互独立. 考虑如下三类关于σ_1^2, σ_2^2的检验问题

$$\begin{array}{lcl} H_0 : \sigma_1^2 \leqslant \sigma_2^2 & \text{vs} & H_1 : \sigma_1^2 > \sigma_2^2 \\ H_0 : \sigma_1^2 \geqslant \sigma_2^2 & \text{vs} & H_1 : \sigma_1^2 < \sigma_2^2 \\ H_0 : \sigma_1^2 = \sigma_2^2 & \text{vs} & H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}$$

μ_1, μ_2未知. 记样本方差 s_x^2, s_y^2 分别是σ_1^2和σ_2^2 的无偏估计. 可建立F检验统计量

$$F = \frac{s_x^2}{s_y^2}$$

由推论5.4.1, 当$\sigma_1^2 = \sigma_2^2$时, $F \sim F(m-1, n-1)$.

取显著性水平为α, 则对应三个检验问题的拒绝域和p值依次分别为

- (1) $H_0: \sigma_1^2 \leqslant \sigma_2^2$ vs $H_1: \sigma_1^2 > \sigma_2^2$
  拒绝域: $W = \{F \geq F_{1-\alpha}(m-1, n-1)\}$
  p值: $p = P(F \geqslant F_0)$
- (2) $H_0: \sigma_1^2 \geqslant \sigma_2^2$ vs $H_1: \sigma_1^2 < \sigma_2^2$
  拒绝域: $W = \{F \leq F_{\alpha}(m-1, n-1)$
  p值: $p = P(F \leqslant F_0)$
- (3) $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$
  拒绝域: $W =$
  $\{F \leq F_{\alpha/2}(m-1, n-1)$ 或 $F \geq F_{1-\alpha/2}(m-1, n-1)\}$
  p值: $p = 2\min\{P(F \geqslant F_0), P(F \leqslant F_0)\}$

其中, $F_0 = \frac{s_x^2}{s_y^2}$为$F$的样本观测值.

拓展知识

两个正态总体方差比检验和标准差比检验也是同一问题, 有相同的检验统计量, 拒绝域, p值.

例7.2.6 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

$$\begin{array}{lcccccccc} X(\text{甲机床}): & 16.2 & 16.4 & 15.8 & 15.5 & 16.7 & 15.6 & 15.8 & \\ Y(\text{乙机床}): & 15.9 & 16.0 & 16.4 & 16.1 & 16.5 & 15.8 & 15.7 & 15.0 \end{array}$$

解: 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 $F = \frac{0.2729}{0.2164} = 1.261$.

若取$\alpha = 0.05$, 查表知

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

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

其拒绝域为

$$W = \{F \leqslant 0.175 \text{或} F \geqslant 5.12\}$$

由此可见, 样本统计量观测值$F = \frac{0.2729}{0.2164} = 1.261$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

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

不能拒绝原假设.

Stata做方差检验/标准差检验

```
//sdtest — Variance-comparison tests
help sdtest

sdtest performs tests on the equality of standard deviations (variances).

// In the first form, sdtest tests that the standard deviation of varname is #.

// In the second form, sdtest performs the same test,
//        using the standard deviations of the two groups defined by groupvar.

// In the third form, sdtest tests that varname1 and varname2 have the same standard deviatio
```

注意: sdtest做单总体检验的统计量是标准差, 可能需要做方差标准差变换; 对两总体方差比做检验无需变换.

单样本标准差检验. **思考: 如何做方差检验?**

```
. sysuse auto
(1978 Automobile Data)

. sum mpg

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         mpg |         74     21.2973    5.785503         12         41

. sdtest mpg==5

One-sample test of variance
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     mpg |      74     21.2973    .6725511    5.785503     19.9569    22.63769
------------------------------------------------------------------------------
    sd = sd(mpg)                                          c = chi2 =  97.7384
Ho: sd = 5                                   degrees of freedom =       73

     Ha: sd < 5                 Ha: sd != 5                 Ha: sd > 5
  Pr(C < c) = 0.9717       2*Pr(C > c) = 0.0565         Pr(C > c) = 0.0283
.
```

Stata做方差检验/标准差检验

两样本方差比检验.

```
. webuse fuel
. sum
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        mpg1 |        12          21    2.730301         17         25
        mpg2 |        12       22.75    3.250874         17         28

. sdtest mpg1 == mpg2

Variance ratio test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .7881701    2.730301    19.26525    22.73475
    mpg2 |      12       22.75    .9384465    3.250874    20.68449    24.81551
---------+--------------------------------------------------------------------
combined |      24      21.875    .6264476    3.068954    20.57909    23.17091
------------------------------------------------------------------------------
    ratio = sd(mpg1) / sd(mpg2)                                   f =   0.7054
Ho: ratio = 1                                    degrees of freedom =   11, 11

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.2862         2*Pr(F < f) = 0.5725           Pr(F > f) = 0.7138
```

作业6

7.2课后习题: 14,16,18,21,26

7.3课后习题: 1-6

7.3.1 指数分布参数的假设检验

指数分布假设检验的推导过程类似正态分布假设检验, 在此仅给出结论(思路见教材).

设 $x_1, x_2, \ldots, x_n$ 是来自指数分布$X \sim Exp(1/\theta)$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

$$\chi^2 = \frac{2n\bar{x}}{\theta_0}$$

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

取显著性水平为α, 则对应三个检验问题的拒绝域和p值依次分别为

- (1) $H_0: \theta \leqslant \theta_0 \quad \text{vs} \quad H_1: \theta > \theta_0$
  拒绝域: $W = \left\{\chi^2 \geqslant \chi^2_{1-\alpha}(2n)\right\}$
  p值: $p = P\left(\chi^2 \geqslant \chi^2_0\right)$
- (2) $H_0: \theta \geqslant \theta_0 \quad \text{vs} \quad H_1: \theta < \theta_0$
  拒绝域: $W = \left\{\chi^2 \leqslant \chi^2_{\alpha}(2n)\right\}$
  p值: $p = P\left(\chi^2 \leqslant \chi^2_0\right)$
- (3) $H_0: \theta = \theta_0 \quad \text{vs} \quad H_1: \theta \neq \theta_0$
  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: $p = 2\min\left\{P\left(\chi^2 \geqslant \chi^2_0\right), P\left(\chi^2 \leqslant \chi^2_0\right)\right\}$

其中, $\chi^2_0 = \frac{2n\bar{x}}{\theta_0}$为$\chi^2$的样本观测值.

例7.3.1 设我们要检验某种元件的平均寿命不小于6000小时, 假定元件寿命为指数分布, 现取5个元件投入试验, 观测到如下5个失效时间: 395, 4094, 119, 11572, 6133.

解: 由于待检验的假设为

$$H_0 : \theta \geqslant 6000 \quad \text{vs} \quad H_1 : \theta < 6000$$

若取 $\alpha = 0.05$, 则检验拒绝域为:

$$\{\chi^2 \leq \chi^2_{0.05}(10) = 3.94\}$$

经计算得

$$\chi^2 = \frac{10\bar{x}}{\theta_0} = \frac{10 \times 4462.6}{6000} = 7.4377 > 3.94$$

故不能拒绝原假设, 可以认为平均寿命不低于6000小时.

下面, 让我们调换检验的备择假设与原假设.

解: 设检验的假设为

$$H_0: \theta \leqslant 6000 \quad \text{vs} \quad H_1: \theta > 6000$$

若取 $\alpha = 0.05$, 则检验拒绝域为:

$$\{\chi^2 \geq \chi^2_{0.95}(10) = 18.307\}$$

经计算得

$$\chi^2 = \frac{10\bar{x}}{\theta_0} = \frac{10 \times 4462.6}{6000} = 7.4377$$

故不能拒绝原假设, 认为平均寿命低于6000小时.

思考: 这里发生了什么? 逻辑上两个检验有错误么? 为什么? 如何改善?

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:

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

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

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

这时, 可取$c=c_0+1$, 此时相当于把显著性水平由α降低到

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

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

$$p=P(x\geqslant x_0)$$

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: $p = P(x \geqslant x_0)$
- (2) $H_0: p \geq p_0 \quad vs \quad H_1: p < p_0$
  p值: $p = P(x \leqslant x_0)$
- (3) $H_0: p = p_0 \quad vs \quad H_1: p \neq p_0$
  p值: $p = 2\min\{P(x \leqslant x_0), \quad P(x \geqslant x_0)\}$

其中, x_0为x的样本观测值.

例7.3.2 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%?

解: 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$n = 20, t_0 = 7$,

$$\begin{aligned} p &= 2\min\{P(T \leqslant 7), P(T \geqslant 7)\} \\ &= 2\min\{0.4159, 0.7500\} = 0.8318 < 0.05 \end{aligned}$$

不能拒绝原假设, 可以认为优制品率保持在40%.

计算复杂! 如何解决?

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$x_1, \ldots, x_n$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$\theta$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

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

其中$\hat{\theta}$为θ的一致估计(证明用Slusky Theorem, 见第六章).

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.

7.3.3 大样本检验: 重要拓展

特别注意: 教材中为MLE, 实际上, 这里只要保证 $\hat{\theta}$为θ的一致估计即可. 所以, MLE 和矩估计都是可选的估计方法.

由于矩估计不需要知道总体分布, 所以即使我们不知道总体的分布, 利用中心极限定理也可以给出有关总体均值的假设检验.

由此,

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

亦可写为

$$u = \frac{\sqrt{n}\left(\bar{x} - \theta_0\right)}{s} \dot{\sim} N(0,1)$$

其中, s为样本标准差. 上式的成立在上一章已经证明.

例7.3.3 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不高于10%?

解: 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.

例 7.3.4 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$\alpha = 0.05$).

解: 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$W=\{u\geq 1.645\}$

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

(亦可不假设总体分布做假设检验. 如何实现? 作为课后练习.)

注意

大样本检验是近似的: 近似的含义是指检验的实际显著性水平与原先设定的显著性水平有差距, 这是由于诸如u 的分布与$N(0,1)$有差别.

如果n 很大, 则这种差异就很小. 实用中我们一般并不清楚对一定的n, u 的分布与$N(0,1)$ 的差异有多大, 因而也就不能确定检验的实际水平与设定水平究竟差多少.

因此, 使用大样本方法应要求样本容量尽可能大.

课堂练习

某公司拟开发新产品, 计算出盈亏平衡点为: 市场占有率=10%. 因此, 只有当市场占有率大于10%才可获利. 现抽取100个潜在用户, 其中有14%的用户表示有意购买此新产品. 问: 可否判断实际市场占有率大于10% ($\alpha = 0.05$)? (请先用经验判断)

解:

课堂练习

某公司拟开发新产品, 计算出盈亏平衡点为: 市场占有率=10%. 因此, 只有当市场占有率大于10%才可获利. 现抽取100个潜在用户, 其中有14%的用户表示有意购买此新产品. 问: 可否判断实际市场占有率大于10% ($\alpha = 0.05$)? (请先用经验判断)

解: 已知: 买和不买服从两点分布$b(1, \theta)$, 且

$$n = 100, \theta_0 = 10\%, \bar{x} = 14\%, \hat{\theta}_{MLE} = \bar{x} = 0.14,$$

$$\sigma^2(\hat{\theta}) = \hat{\theta}(1 - \hat{\theta}) = 0.14 \times (1 - 0.14) = 0.1204$$

提出假设:$H_0 : \theta \leq 10\% \quad vs \quad H_1 : \theta > 10\%$

检验统计量:

$$u = \frac{\sqrt{n}\,(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\hat{\theta})}} = \frac{\sqrt{100}\,(0.14 - 0.1)}{\sqrt{0.1204}} = 1.153$$

拒绝域: $W=\{u\geq u_{1-\alpha}\}=\{u\geq 1.65\}$

故无法拒绝原假设, 不能判断市场占有率大于10%.

思考并计算(3'):

1. 观察值$\bar{x}$至少多大才能拒绝原假设?

2. 在给定观察值$\bar{x}=14\%$下, n有多大才能拒绝原假设?

课堂练习

一种罐装饮料采用自动生产线, 每罐的容量是255ml. 为检验每罐容量是否符合要求, 质量检验人员在每天生产的饮料中随机抽取了40罐进行检验, 测得平均容量为255.8ml, 标准差是5ml. 取显著水平为0.05, 检验该天生产的饮料容量是否符合标准要求?

$$\mu_0 = 255, \quad n = 40, \quad \bar{x} = 255.8, \quad s = 5, , \quad \alpha = 0.05$$

解: 此题目总体分布未知, 可用样本标准差近似总体标准差(样本标准差s是总体标准差σ的一致估计, 矩估计性质)

检验问题为:$H_0 : \mu = 255 \quad vs \quad H_1 : \mu \neq 255$

拒绝域: $W = \{|u| \geq u_{1-\alpha/2}\} = \{|u| \geq 1.96\}$

检验统计量

$$u = \frac{\sqrt{n}\,(\bar{x} - \mu_0)}{s} = \frac{\sqrt{40}\,(255.8 - 255)}{5} = 1.01 < 1.96$$

不能拒绝H_0

课堂练习

例题 :某小麦品种的平均产量一直低于$5200kg/hm^2$. 为检验改良后新品种的平均产量是否有显著提高, 随机抽取了36个地块进行种植. 得到的样本平均产量为$5275kg/hm^2$, 标准差为$120kg/hm^2$. 问改良品种的产量是否有显著提高? $(\alpha = 0.05)$

解:

例题 :某小麦品种的平均产量一直低于$5200kg/hm^2$. 为检验改良后新品种的平均产量是否有显著提高, 随机抽取了36个地块进行种植. 得到的样本平均产量为$5275kg/hm^2$, 标准差为$120kg/hm^2$. 问改良品种的产量是否有显著提高? $(\alpha = 0.05)$

解: 检验问题为:$H_0 : \mu \leq 5200 \quad vs \quad H_1 : \mu > 5200$

拒绝域: $W = \{u \geq u_{1-\alpha}\} = \{u \geq 1.65\}$

检验统计量

$$u = \frac{\sqrt{n}\,(\bar{x} - \mu_0)}{s} = \frac{\sqrt{36}\,(5275 - 5200)}{120} = 3.75 > 1.645$$

拒绝H_0, 认为产量显著提高.

拓展阅读: 似然比检验(likelihood ratio test)的基本思想 (教材§ 7.4.1)

似然比检验(奈曼和皮尔逊, 1928)是利用似然函数的比值来进行的假设检验方法. 故在MLE中应用.

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**.

定义7.4.1: 设$x_1, ..., x_n$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

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

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$\Lambda(x_1, \cdots, x_n)$为上述假设检验的**似然比(likelihood ratio)**, 也称**广义似然比**.

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

分子: 没有假设时似然函数的最大值 (unrestricted)

分母: 原假设成立条件下似然函数的最大值 (restricted)

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.

由于**对数似然比检验统计量的2倍渐近服从卡方分布**, 所以实际应用时, 应用对数似然函数的最大值计算:

$$LR = 2\frac{LL_{unrestricted}}{LL_{restricted}}$$

拓展阅读: Wooldridge. Introductory Econometrics: A Modern Approach. 5th Edition. Chapter 17.

§7.4 分布拟合(拟合优度, goodness of fit)检验

前文均是介绍对总体分布的某个参数建立假设并进行检验的方法. **分布的拟合检验**是对总体分布形式的假设检验. 这是一类**非参数检验**问题.

分布拟合优度检验在总体分布为离散分布时应用较为广泛, 在总体分布为连续时有较强的应用局限性.

下面以一个例子介绍分类数据的总体分布拟合优度检验问题:

例7.4.2 (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?

解: 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: $A_1, A_2, \ldots, A_k$. 现对该总体作了n 次观测, k 个类出现的频数分别为:

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

我们关心各类元素在总体中所占的比率的假设:

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

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

一. 诸p_i均已知

定理7.4.1: 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:

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

证明略.

应用上述定理对**例7.4.2 (孟德尔遗传学试验)**做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.

例: 某彩票用摇大转盘的方法确定最后中奖金额. 大转盘均分为20份, 其中金额为5万, 10万, 20万, 30万, 50万, 100万的分别占2份, 4份, 6份, 4份, 2份, 2份. 假定大转盘是均匀的, 摇到每份是等可能的, 于是摇出各个奖项的概率如下:

额度	5	10	20	30	50	100
p	0.1	0.2	0.3	0.2	0.1	0.1

现20人参加摇奖, 摇得5万, 10万, 20万, 30万, 50万和100万的人数分别为2, 6, 6, 3, 3, 0, 由于没有一个人摇到100万, 于是有人怀疑大转盘是不均匀的, 那么该怀疑是否成立呢? 这就需要对转盘的均匀性作检验.

解: 这是一个典型的分布拟合优度检验, 总体共有6类, 其发生概率分别为0.1, 0.2, 0.3, 0.2, 0.1和0.1, 这里$k=6$, 检验拒绝域为:

$$\{\chi^2 \geq \chi^2_{1-\alpha}(5)\}$$

若取$\alpha=0.05$, 则查表知$\chi^2_{0.95}(5)=11.07$

由本例数据可以算出

$$\chi^2 = \frac{(2-2)^2}{2} + \frac{(6-4)^2}{4} + \frac{(6-6)^2}{6} + \frac{(3-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(0-2)^2}{2} = 3.75$$

由于χ^2未落入拒绝域, 故接受原假设, 没有理由认为转盘不均匀.

在分布拟合检验中使用p 值也是方便的. 本例中, 以T 记服从$\chi^2(5)$的随机变量, 则使用统计软件可以算出

$$p = P(T \geq 3.75) = 0.5859$$

这个p 值就反映了数据与假设的分布拟合程度的高低, p值越大, 拟合越好.

STATA χ^2拟合优度检验

```
ssc install tab_chi
help chitest

chitesti 315 108 101 32 \ 556*9/16 556*3/16 556*3/16 556*1/16
```

二. 诸 p_i 不完全已知 (教材7.4.3)

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

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

首先给出 $\theta_1, ..., \theta_r$ 的极大似然估计, 然后给出
诸$p_i, i = 1, ..., k$ 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

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

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

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

例7.4.3 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.

解: 本例中, 要检验总体是否服从泊松分布.

观测到 $0, 1, ..., 11$ 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,

$$\hat{\lambda} = \bar{x} = \frac{1}{2608}(1 \times 203 + 2 \times 383 + \ldots + 11 \times 6) = 3.870$$

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$\chi^2 = 12.8967 < 18.307$, 故接受原假设. 使用统计软件可以计算出此处检验的p 值是0.2295.

连续总体拟合优度检验不常用, 且存在应用局限, 略

7.4.4 列联表(contingency table)的独立性检验

列联表是将观测数据按两个或更多属性 (定性变量) 分类时所列出的频数表. 例如, 对随机抽取的1000人按性别(男或女)及色觉(正常或色盲) 两个属性分类, 得到如下二维列联表, 简称2×2表.

	视觉	
性别	正常	色盲
男	535	65
女	382	18

一般化地, 若总体中的个体可按两个属性A与B分类, A 有r个类 $A_1, \ldots, A_r$, B 有c个类 $B_1, \ldots, B_c$.

从总体中抽取大小为n的样本, 设其中有 n_{ij}个个体既属于A_i类又属于B_j 类, n_{ij} 称为频数, 将$r \times c$个 n_{ij} 排列为一个r行c列的二维列联表, 简称$r \times c$表

表7.4.3 $r \times c$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_{i1}	$\cdots$	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.**

如在前例中, 问题是:一个人是否色盲与其性别是否有关?
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(表7.4.4), 则"A, B两属性独立"的假设可以表述为

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

表7.4.4 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_{i1}	$\cdots$	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

这就变为上一小节中诸p_{ij}不完全已知时的分布拟合检验.
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$p_{1\cdot},\dots,p_{r\cdot}$和 $p_{\cdot 1},\dots,p_{\cdot c}$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

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

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:

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

例7.4.5 为研究儿童智力发展与营养的关系, 某研究机构调查了1436名儿童, 得到如表7.4.5的数据, 试在显著性水平0.05下判断智力发展与营养有无关系.

表7.4.5 儿童智力与营养的调查数据

	智商				合计
	< 80	$80 \sim 89$	$90 \sim 99$	≥ 100	
营养不良	367	342	266	329	1304
营养良好	56	40	20	16	132
合计	423	382	286	345	1436

解: 用A表示营养状况, 它有两个水平:A_1 表示营养良好, A_2表示营养不良； B表示儿童智商, 它有四个水平, B_1, B_2, B_3, B_4分别表示表中四种情况. 沿用前面的记号, 首先建立假设

H_0: 营养状况与智商无关联, 即A与B独立的.

统计表示如下:

$$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, 2, \quad j = 1, 2, 3, 4$$

在原假设H_0成立下, 可以计算诸参数的极大似然估计值:

$$\begin{aligned}
\hat{p}_{1\cdot} &= 1304/1436 = 0.9081, \quad & \hat{p}_{2\cdot} &= 132/1436 = 0.0919 \\
\hat{p}_{\cdot 1} &= 423/1436 = 0.2946, \quad & \hat{p}_{\cdot 2} &= 382/1436 = 0.2660 \\
\hat{p}_{\cdot 3} &= 286/1436 = 0.1992, \quad & \hat{p}_{\cdot 4} &= 345/1436 = 0.2403
\end{aligned}$$

进而可给出诸$n\hat{p}_{ij}=n\hat{p}_{i\cdot}\hat{p}_{\cdot j}$, 如

$$n\hat{p}_{11}=1436\times 0.9081\times 0.2496=384.1677$$

其它结果见表7.4.6

表7.4.6 诸$n\hat{p}_{ij}=n\hat{p}_{i\cdot}\hat{p}_{\cdot j}$ 的计算结果

	<80	$80\sim 89$	$90\sim 99$	≥ 100	$\hat{p}_{i\cdot}$
营养不良	384.1677	346.8724	259.7631	313.3588	0.9081
营养良好	38.8779	35.1036	26.2881	31.7120	0.0919
$\hat{p}_{\cdot j}$	0.2946	0.2660	0.1992	0.2403	

由表7.4.5和表7.4.6可以计算检验统计量的值

$$\chi^2 = \frac{(367-384.1677)^2}{384.1677} + \frac{(342-346.8724)^2}{346.8724} + \cdots + \frac{(16-31.712}{31.7120}$$

$$= 19.2785$$

此处$r=2, c=4, (r-1)(c-1)=3$, 若取$\alpha=0.05$, 查表有$\chi^2_{0.95}(3)=7.815$, 由于$19.2785>7.815$, 故拒绝原假设, 认为营养状况对智商有影响.

本例中检验的p 值为0.0002.

STATA χ^2列联表独立性检验

```
help tabulate twoway
sysuse auto
tabulate length weight, chi
```

7.4.3 正态性检验(拓展内容)

正态分布是最常用的分布, 用来判断总体分布是否为正态分布的检验方法称为正态性检验, 它在实际问题中大量使用.

一, 正态概率纸(Q-Q Plot)

正态概率纸可用来作正态性检验, 方法如下:利用样本数据在概率纸上描点, 用目测方法看这些点是否在一条直线附近, 若是的话, 可以认为该数据来自正态总体, 若明显不在一条直线附近, 则认为该数据来自非正态总体.

正态概率纸可以作为正态性检验的根本原因:用经验分布函数逼近总体分布函数.

例7.4.4 随机选取10个零件, 测得其直径与标准尺寸的偏差如下:(单位:丝)

$$9.4 \quad 8.8 \quad 9.6 \quad 10.2 \quad 10.1 \quad 7.2 \quad 11.1 \quad 8.2 \quad 8.6 \quad 9.6$$

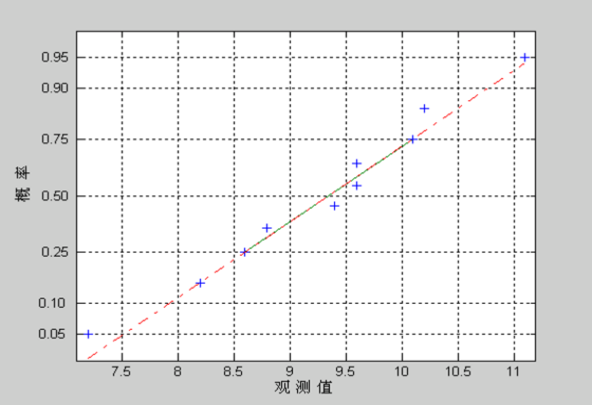
在正态概率纸上作图步骤如下:

- (1) 首先将数据排序:
  $7.2 \quad 8.2 \quad 8.6 \quad 8.8 \quad 9.4 \quad 9.6 \quad 9.8 \quad 10.1 \quad 10.2 \quad 11.1;$
- (2) 对每一个i, 计算修正频率:
  $(i-0.375)/(n+0.25), i=1,2,...,n$
- (3) 将点逐一点在正态概率纸上,
- (4) 观察上述n个点的分布:

若诸点在一条直线附近, 则认为该批数据来自正态总体;

若诸点明显不在一条直线附近, 则认为该批数据的总体不是正态分布.

从图7.4.2可以看到, 10个点基本在一条直线附近, 故可认为直径与标准尺寸的偏差服从正态分布.



如果从正态概率纸上确认总体是非正态分布时, 可对原始数据进行变换后再在正态概率纸上描点, 若变换后的点在正态概率纸上近似在一条直线附近, 则可以认为变换后的数据来自正态分布, 这样的变换称为正态性变换.

常用的正态性变换有如下三个:**对数变换**$y = ln(x)$, **倒数变换** $y = 1/x$ 和**根号变换**$y = \sqrt{x}$

例7.5.2 随机抽取某种电子元件10个，测得其寿命数据如下:

110.47,　99.16,　97.04,　77.60,　4269.82,

539.35,　179.49,　782.93,　561.10,　286.80.

图7.5.3 例7.5.2 的正态概率纸

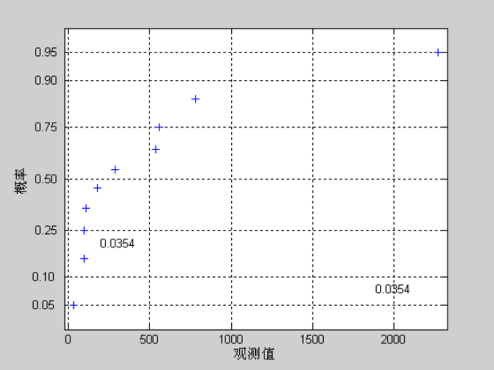


图7.5.3 给出这10个点在正态概率纸上的图形, 这10个点明显不在一条直线附近, 所以可以认为该电子元件的寿命的分布不是正态分布.

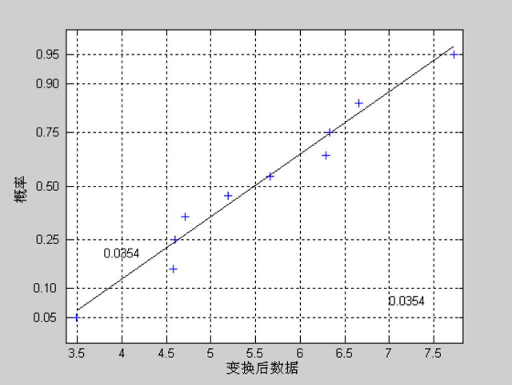
对该10个寿命数据作对数变换,结果见下表

表7.5.2 对数变换后的数据

i	$x_{(i)}$	$\ln x_{(i)}$	$\frac{i-0.375}{n+0.375}$	i	$x_{(i)}$	$\ln x_{(i)}$	$\frac{i-0.375}{n+0.375}$
1	32.62	3.4849	0.061	6	286.80	5.6588	0.549
2	97.04	4.5752	0.159	7	539.35	6.2904	0.646
3	99.16	4.5967	0.256	8	561.10	6.3299	0.743
4	110.47	4.7048	0.354	9	782.93	6.6630	0.841
5	179.49	5.1901	0.451	10	2269.82	7.7275	0.939

利用表7.5.2 中最后两列上的数据在正态概率纸上描点, 结果见图7.5.4, 从图上可以看到10个点近似在一条直线附近, 说明对数变换后的数据可以看成来自正态分布. 这也意味着, 原始数据服从对数正态分布

图7.4.4 变换后数据的正态概率纸



STATA Q-Q plot

```
sysuse nlsw88.dta
help pnorm
help qnorm

pnorm wage

gen lnwage=ln(wage)

pnorm lnwage
```

二, W 检验 (略).

三, EP 检验 (略).

7.6 非参检验(略).

作业7

7.4课后习题： 5,6,7,13,15

上机实验4

1. 利用Stata对书中例7.4.2做拟合优度检验.

2. (QQ Plot)选取任意可得数据, 利用Stata对该数据做正态性检验Q-Q plot. 讨论数据是否可认为是正态分布, 如不符合, 对数据做某种正态变换, 并讨论变换后的数据后是否可认为是正态分布.

复习思考题

假设检验的基本思想是什么? 其中使用了一条什么原理?

检验的显著性水平α的意义是什么?

比较双边, 左边和右边检验的拒绝域

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?

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?

成对数据差的t检验适用于哪些特殊场合?

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?