2021年数理统计上机课

–假设检验

黄启岳

北京师范大学统计学院

2021 年 5 月 12 日

目录

基本概念

假设检验本质上是一种分类问题。

Θ_0与Θ_1是对参数空间Θ的一个划分，关注的问题为：

$$H_0 : \theta \in \Theta_0 \longleftrightarrow H_1 : \theta \in \Theta_1$$

与分类问题不同在于：假设检验保护原假设。

假设检验的逻辑

假设检验的核心逻辑可以概括为：小概率反证法。

- 考虑数学证明中的反证法，即假定结论不成立，通过逻辑推理得出与条件相悖的结论，即可证伪原命题的否命题，也就是原命题是真命题。
- 假设检验的逻辑和反证法是类似的。假设检验假定原假设H_0成立，若在这个被肯定的原假设下观测到了原假设下的“小概率事件”，统计学家们就要“怀疑”原假设的真实性了。

- 和反证法比较，类似在于：通过说明原假设下一个“不太可能”(数学中为不可能)发生的事件在假设下“确实发生了”，进而证伪原假设。
- 区别在于：数学中需要导出严格矛盾才能证伪原假设，统计学中则使用小概率事件代替了严格矛盾。
- 使用小概率事件代替严格矛盾意味着统计检验会“犯错”，因为小概率事件不是不可能事件。

基本步骤

- 确定原假设H_0和备择假设H_1，划分参数θ的取值空间；
- 基于原假设H_0以及数据构造合适的检验统计量W_n，W_n需要有确定的分布；
- 在给定显著水平α下，根据W_n的分布确定拒绝域，即“不太可能”发生事件的数学表达；
- 根据数据计算W_n的具体值，若拒绝域包含了W_n，则认为观测到了“不太可能”的事件，拒绝H_0，否则接受H_0。

两类错误

假设检验不同于反证法，是有可能犯错误的，分为第一类错误和第二类错误。第一类错误即H_0事实上成立但被假设检验否定；第二类错误即H_0事实上不成立但被假设检验接受。

Example (案例：审判)

张三被捕，被法院审判是否有罪。根据无罪推定原则，未经审判证明有罪确定前，推定被控告者无罪。则第一类错误为“张三事实上无罪却被捕”，第二类错误为“张三事实上有罪却被放”。

显著性水平

为了控制第一类错误和第二类错误的概率，使得假设检验"犯错"现象得到控制，

Definition (显著性水平)

犯第一类错误的最大概率α称为假设检验的显著性水平(Significant level)。即：

$$\sup P(H_0 \textit{ is rejected} | H_0 \textit{ is true}) = \alpha$$

仍然以审判为例，我们无法做到完全不冤枉好人，但会控制这种事情出现的概率维持在一个指定的水平。如$\alpha = 0.05$，表示审判(假设检验)发生冤枉好人事件(第一类错误)的概率不超过0.05。

在假设检验中，显著性水平α一般会直接给定。常用***表示显著性水平0.01，**表示显著性水平0.05，*表示显著性水平0.1。

检验功效

功效这一概念与显著性水平一同提出，遏制了显著性水平的“无序增长”。

以审判为例，如何做到不错抓好人？一种极端的方法是**把所有人都释放**，但坏人也被放了，这显然不是我们愿意看到的。这一现象即为第二类错误的概率高。最好的结果是：在控制错抓好人概率的前提下尽可能抓住坏人。

功效函数就是对上述问题的数学表达。

DEFINITION (功效函数(POWER FUNCTION))

功效函数定义为假设检验ϕ拒绝原假设$H_0 : \theta \in \Theta_0$的概率。

$$\beta_\phi(\theta) = \begin{cases} P(H_0 \textit{ is rejected} | H_0 \textit{ is true}), \textit{if } \theta \in \Theta_0 \\ 1 - P(H_0 \textit{ is accepted} | H_0 \textit{ is false}), \textit{if } \theta \in \Theta_1 \end{cases}$$

当H_0成立时，拒绝原假设相当于犯第一类错误，即功效等于第一类错误发生概率。

当H_0不成立时，拒绝原假设是正确的，也就是第二类错误的反面，即功效函数为1减去第二类错误的概率。

具体操作中，我们一般控制第一类错误的上确界不能超过给定值α，在此基础上让第二类错误概率β最小。即：

$$\max_{\sup P(\mathit{Type\ I\ error}) \leq \alpha} (1 - P(\mathit{Type\ II\ error}))$$

检验p值

p值直观体现对原假设H_0的“怀疑程度”，一般p值越小，我们越怀疑原假设不成立。统计软件多使用p值反映假设检验的结果。

Definition (p值(p-value))

拒绝原假设H_0的最小显著性水平称为检验的p值。

参数型假设检验

- 我们主要研究参数的假设检验问题，即对总体的某个位置参数做出假设，抽取样本信息构造统计量，通过统计量反向推断是否接受原假设。
- 与之相对，非参数假设检验往往是对总体分布情况进行假设，不针对某个具体的参数。这里不再系统展开，相关内容可以在课程**非参数统计**中进一步学习。

单样本正态检验

首先介绍单样本正态检验，检验的总体已知服从正态分布。

根据中心极限定理(Central Limit Theorem,CLT)，很多分布最终可以近似处理为正态分布。在实际的生产生活中，我们同样很关注正态分布参数的一些性质，因此引入正态总体参数的假设检验。

基本假定

- 通常假定数据$X=(X_1,\ldots,X_n)$独立同分布服从正态分布$N(\mu,\sigma^2)$，是来自同一正态总体的样本。
- 均值函数记为：

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

- 方差记为：

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

均值方差均已知的均值检验

考虑$X_1, \ldots, X_n \sim N(\mu, \sigma_0^2)$，各个$X_i$相互独立。考虑假设检验问题：

$$H_0 : \mu = \mu_0 \longleftrightarrow H_1 : \mu > \mu_0$$

根据正态分布在线性运算下仍保持正态分布的性质，原假设成立的条件下成立：

$$Z = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma_0} \sim N(0, 1)$$

则除非$\overline{X}$比μ_0“大了太多”，否则我们接受原假设。$\overline{X}$比μ_0“大了太多”在原假设成立的前提下为小概率事件，至于“概率多小”，需要使用已有分布进行刻画。

例如我们对“小概率”的界定为$\alpha = 0.05$，即概率不到0.05的事件被认为“小概率”。则$\overline{X}$比μ_0“大了太多”是小概率事件表述为：

$$\sup P(\overline{X} > \mu_0) = \alpha = 0.05 \tag{1}$$

(1)式等价于：

$$\sup P(\overline{X} - \mu_0 > 0) = \alpha = 0.05 \tag{2}$$

(2)进一步变化为：

$$\sup P\left(\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma_0} = Z > 0\right) = \alpha = 0.05 \tag{3}$$

问题被转化为Z与0比大小，若Z显著大于0则$\overline{X}$比μ_0“大了太多”。Z显然服从标准正态分布，即可通过正态分布分位数评价Z是否“太大”。

针对单侧检验，Z“太大”代表Z超过了标准正态分布的0.95分位数。因此拒绝域构造为：

$$C_{0.05} = \{Z - 0 = Z > qnorm(0.95, 0, 1)\}$$

或采取构造p值的方式进行评价，即：

$$p - value = 1 - pnorm(Z, 0, 1)$$

p值代表原假设成立下Z>0的概率，也就是$\overline{X} > \mu_0$的概率。

R语言中没有类似检验的函数，编写了一个函数实现功能，即程序中的函数ztest1和ztest2，分别通过拒绝域和p值表明假设检验的结果。

方差未知的均值检验

针对同样的原假设，方差未知的情形下可以使用样本方差代替总体方差进行插入估计，即使用S^2代替σ_0^2，这样做的合理性是根据大数定律S^2收敛到σ_0^2。且：

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \tag{4}$$

即得到：

$$\frac{\sqrt{n}(\overline{X}-\mu_0)}{S} = \frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma_0} \times \left(\sqrt{\frac{1}{n-1} \times \frac{(n-1)S^2}{\sigma_0^2}}\right)^{-1} \tag{5}$$

根据(4)式和(5)式得到：

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \sim t(n-1)$$

拒绝域为：

$$C_\alpha = \{T > t_{(1-\alpha)}(n-1)\}$$

R语言中函数为t.test(x = , y = NULL, mu = 0, alternative = c("two.sided","greater","less"),paired = FALSE, var.equal = FALSE, conf.level = 0.05)，这里选择alternative = "greater"即可。

方差的χ^2检验

原假设分为双边与单边，即:

- $H_0: \sigma = \sigma_0$
- $H_0: \sigma > \sigma_0$
- $H_0: \sigma < \sigma_0$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \tag{6}$$

问题：根据(6)式，如何构造拒绝域？如何编写χ^2检验的函数？(不幸的是，R语言中没有现成的函数)

两样本正态检验

针对单个样本的假设检验问题，我们可以提出合理的参数假设值，这在实际应用中很难做到。现实中常常取为“实验组”和“对照组”，在两样本之间做出比较。

基本假设

类似于一样本情形，仅仅变为：第一个总体$\{x_i\}_{i=1}^{m}$均值为μ_1，第二个总体$\{y_j\}_{j=1}^{n}$的均值为μ_2。μ_1和μ_2分别代表两个总体的均值，σ_1和σ_2分别为两总体标准差，S_1与S_2为样本标准差。

方差相等的两样本均值检验

这里讨论双边假设的情形。即$\sigma_1 = \sigma_2 = \sigma$。第一个总体$\{x_i\}_{i=1}^m$均值为$\mu_1$，第二个总体$\{y_j\}_{j=1}^n$的均值为$\mu_2$，原假设为：

$$H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 \neq \mu_2$$

统计量：

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

方差不等的均值检验

相比于方差相等的情形，方差不等的情形只是方差表达更加复杂。直接给出统计量：

$$T = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\left(\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}\right)}} \sim t(m+n-2)$$

这两种情况对应的R中函数仍为t.test()，需要同时输入x与y，并指定均值方差等信息。

方差的F检验

类似一样本情形，原假设为：

$$H_0 : \sigma_1^2 = \sigma_2^2 \longleftrightarrow H_1 : \sigma_1^2 \neq \sigma_2^2$$

检验统计量为：

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

拒绝域为：

$$C_\alpha = \left\{ F_{1-\frac{\alpha}{2}}(m-1, n-1) \leq F \leq F_{\frac{\alpha}{2}}(m-1, n-1) \right\}$$

R语言中对应函数为var.test()。

功效的模拟计算

最后简单介绍功效的模拟计算问题。有的检验例如t检验存在直接计算功效的函数power.t.test()函数，但并不是所有函数都有计算功效的现成的函数。

事实上不同检验相互比较的一个途径就是比较在同等数据下的功效高低，当然如果功效不容易直接计算也可以计算渐近相对效率。

这里采用的技术路线有别于其他两条，即通过生成相应分布随机数对实际的判断情况进行模拟。

算法：以一样本T检验为例

- 首先构造t检验的函数，确定显著性水平α
- 构造H_1成立的环境，则此时接受原假设为第二类错误
- 进行M次循环，其中第二类错误出现m次
- 功效函数近似为：$\hat{\beta}(T) = 1 - \frac{m}{M}$

应用

非参数统计中有一种检验为秩和检验，常用于和t检验比较功效，一样本与两样本的情况皆有，除了使用渐近相对效进行理论计算，还可以通过随机模拟的方式比较两种检验在不同分布以及不同分布参数假定下的功效。

选读：BOOTSTRAP T 检验

自助检验步骤如下：

- 首先将两总体X与Y的均值修正到相等，即

$$\widetilde{X}_i = X_i - \bar{X} + \bar{Z}, \widetilde{Y}_i = Y_i - \bar{Y} + \bar{Z}$$

- 之后进行自助抽样，获得自助样本$\{\widetilde{X_i^b}\}, \{\widetilde{Y_i^b}\}$，一共进行B次，针对每一组$\{\widetilde{X_i^b}\}, \{\widetilde{Y_i^b}\}$计算：

$$T_b = \frac{\overline{X_b} - \overline{Y_b}}{\sqrt{\frac{S_{X_b}^2}{m} + \frac{S_{Y_b}^2}{n}}}$$

- 基于原样本计算

$$T_{obs} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$$

- 最后计算p值，p值为：

$$p = \frac{1}{B}\sum_{i=1}^{B} \mathbf{I}_{\{T_b > T_{obs}\}}$$

若p值小于给定的显著性水平α，则拒绝原假设$H_0 : X = Y$。