

统计学 Statistics

陈灯塔

Econ, XJTUCC

CDT.WISE@G

2021 年 5 月 9 日

Free, non profit classroom only

内容梗概

- 1 假设检验：思想
- 2 假设检验：均值
- 3 假设检验： p 值

人与猪

当我们知道物种的区别时，人与猪是不同的物种。当用统计方法时，通常的 5% 显著水平下，人与猪没有区别

小概率原理，是行为准则，不是数学定律，不是客观规律

- 0.0001% (百万分之一) 的区别就可以判定为赝品
- 人不是猪，哪怕 DNA 差别很小 (有说 99% 同，或者 96% 相同)

1 假设检验：思想

- 概念
- 一般步骤
- 茶余饭后
- 总结

猜硬币

你和小明猜硬币，10 次中小明猜中了 9 次。您认为如下哪种情况最可能

- ① 小明有特殊能力，猜硬币水平高
- ② 小明走狗屎运 (猜中至少 9 次的概率为 $\frac{1}{2^{10}} \left(\binom{10}{9} + \binom{10}{10} \right) = 0.0107$)
- ③ 小明涉嫌做假，魔术戏法或者硬币并不均匀等

Lady Tasting Tea

The experiment provides a subject with 8 randomly ordered cups of tea – 4 prepared by first pouring the tea, then adding milk, 4 prepared by first pouring the milk, then adding the tea. The subject has to select 4 cups prepared by one method.

The null hypothesis is that the subject has no ability to distinguish the teas.

wiki: Lady tasting tea

the number of successes is distributed according to the hypergeometric distribution.

if and only if the lady properly categorized all 8 cups was Fisher willing to reject the null hypothesis – effectively acknowledging the lady's ability at a 1.4% significance level (but without quantifying her ability)

概念

什么是假设

“我认为这种新药的疗效比原有的药物更有效!” (康复更快, 或者比例更高)

“硬币是均匀的” (出现正反面的概率都为 $1/2$)

对总体参数的具体数值所作的陈述 (总体参数包括总体均值、比例、方差等)

假设检验: 对总体参数提出一个假设, 用样本信息计算检验统计量, 以决定是否拒绝该假设

- 参数检验 parametric tests
- 非参数检验 nonparametric test (课本称为分布检验不恰当)

小概率原理

一次猜中密码，人们自然怀疑泄密或者有设置时依照某种偏好，不相信密码是保密且随机的

小概率事件在一次试验中不应该发生：小概率事件在很多次试验中才可能发生一次，实际决策时通常认为它在一次试验中是不会发生的；反之，如果在一次试验中小概率事件居然发生了，人们宁愿相信该事件的前提条件有错误。

提醒：小概率原理认为人生不能靠运气 (可遇不可求)，提升能力把小概率变成大概率

- 小概率事件改变世界。 小概率事件在一次试验下发生的概率小，并不是不可能发生。
- 小概率原理是信念，应称为**小概率信念**：是行为模式，非科学和非理性的行为习性
- 小概率原理有犯错误的可能！不是数学的反证法

原假设和备择假设

null hypothesis: 研究者想收集证据予以反对的假设; 又称“0 假设”, 表示为 H_0

之所以用零来修饰原假设, 其原因是原假设的内容总是表示总体参数没有差异或没有改变, 或变量间没有关系等等

- 零假设总是一个与总体有关的问题, 不是关于样本的
- 参数检验的零假设通常包含等号, 如 $=$, \geq 或 \leq

备择假设: alternative hypothesis, 表示为 H_1 或者 H_A , 也称研究假设、对立假设, 它是原假设的对立陈述, 研究者想收集证据予以支持的假设。原假设和备择假设必有一个成立, 而且只有一个成立

提出假设: 例子 ppt.15-7; 双侧检验与单侧检验 ppt.21

The type of test (左、右和双侧) is determined by the Alternative Hypothesis

因研究目的不同, 对同一问题可能提出不同的假设 (也可能得出不同的结论)

两类错误

犯错是小概率原理的必然，假设检验中的两类错误 ppt.24

- ① 第 I 类错误 (Type I: false positive, 弃真错误): 原假设为正确时拒绝原假设; 第 I 类错误的概率记为 α , 被称为显著性水平 $\alpha = \Pr(\text{拒绝 } H_0 \mid H_0 \text{ 为真})$
- ② 第 II 类错误 (Type II: false negative, 取伪错误): 原假设为错误时未拒绝原假设; 第 II 类错误的概率记为 β . 拒绝一个错误的原假设的概率

$$1 - \beta = \Pr(\text{拒绝 } H_0 \mid H_0 \text{ 为假})$$

这个概率被称为检验的势 (power, 也翻译为功效), 代表拒绝错误原假设的能力

type1-2 error.jpg (H_0 : 未孕)

影响第 II 类错误的因素: 显著性水平, 样本容量, 总体的分布 (均值、方差)

统计显著性

在假设检验中，如果样本提供的证据拒绝原假设，我们说检验的结果是**显著的**，如果不拒绝原假设，我们则说结果是不显著的

小概率原理默认样本结果并不是偶然得到的，无视运气、机遇的成分，“大势所趋”。
significant (显著的) 一词的意义在这里并不是“重要的”，而是指“非偶然的”

一般步骤

步骤

假设检验的一般步骤

- ① 提出原假设和备择假设 (单侧, 双侧)
- ② 选择适当的检验统计量, 明确其概率分布
- ③ 取得样本, 计算出检验统计量的观测值
- ④ 统计决策, 是否拒绝零假设: 给定显著性水平 α
 - ① 确定临界值和拒绝域, 是否落入拒绝域
 - ② 计算 p 值 (将单独讲), 是否小于显著水平 [现代方法, 推荐]
 - ③ 对于参数假设检验, 还可以计算置信区间, 是否包含参数值 (不包含真值是小概率)
- ⑤ 作出检验结论 (经济意义方面的)

茶余饭后

拒绝和不拒绝

假设检验的目的在于试图找到证据拒绝原假设，而不在于证明什么是正确的

- 不拒绝原假设时，并未给出明确的结论，不能说原假设是正确的，也不能说它不是正确的
- 不能拒绝时，未能提供有用信息，对决策没有帮助
- Hypothesis testing *emphasizes the rejection*, which is based on a probability, rather than the acceptance, which requires extra steps of logic

当没有足够证据拒绝原假设时，不采用“接受原假设”的表述，而是“不拒绝原假设”

- “接受”的说法有时会产生误导，因为这种说法似乎暗示着原假设已经被证明是正确的了。事实上，总体真值是未知的，由样本提供的信息也就自然无法证明 H_0 是否正确
- 我们可能不拒绝 $\mu = 1$ ，同时也不能拒绝 $\mu = 2$ ；我们不要说接受 $\mu = 1$ 同时也接受 $\mu = 2$
- The phrase “accept the null hypothesis” may suggest it has been proved simply because it has not been disproved, a logical fallacy known as the argument from ignorance.

第一类错误

为什么先控制第一类错误？而不是先去控制第二类错误？

- 假设检验的结果无非两种，拒绝原假设或者不拒绝原假设。不拒绝原假设的情况下对决策没有帮助，白忙活了。**人们希望的结果是拒绝原假设**，于是在拒绝原假设的情况下，错误率要小，即第一类错误率要小
- 两类错误的犯错成本不同，导致了 H_0 和 H_1 处在不平等的位置。假设检验在数学结构上，能完美控制第一类错误的概率。因此，特意地建立原假设/备择假设使得更加严重的错误恰好成为第一类错误 (因为严重才设定成第一类错误；而不是因为是第一类错误才严重)

风吹草动

- 儿童的 H_0 : 狼来了 (第一类错误: 狼真的来了以为没来)
- 猎人的 H_0 : 狼没有来 (第一类错误: 误以为狼来了)

显著水平 0.05

Fisher 没有任何高深的理由解释他为什么选择 0.05: The value for which $P = 0.05$, or 1 in 20, is 1.96 or nearly 2 ; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not

- Because of copyright restrictions he could only print an excerpt, and he decided to take 0.05 as something that should work in most cases.
 - If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 per cent point), or one in a hundred (the 1 per cent point).
- $\pm 3PE$ (central $\sim 95\%$ of values) was in common use in the late 1890s and this translates to 0.05. 1 in 20 is easier to interpret for most people than a z value of 2 or in terms of PE and thus explains why 0.05 became more popular. $\pm 3PE \rightarrow \pm 2\sigma \rightarrow \pm 1.96\sigma \rightarrow \alpha = 0.05$

如果读到 *size of test*, 本科生可以把它理解为显著水平 significance level of a test (In most cases, one uses tests whose size is equal to the significance level)

总结

- ① 假设检验的基本思想和原理 (理解)
- ② 两类错误，假设检验的基本概念 (掌握)
- ③ 假设检验的步骤 (理解)

Keyword: 假设检验, 小概率原理, 原假设, 备择假设, 第 I 类错误, 第 II 类错误, 显著性

Homework: (* by group)

- Time loss of smart phones, pads and computers in classroom
- 基本概念要牢记, 牢记, 牢记!
- 统计学中的假设检验是什么? 用纯文字表达, 受众为普通大众 (受过义务教育). 联系自己所学专业的实际, 提出一个需要假设检验的问题
- Read
 - ① Textbook: S6.6.4–5
 - ② Reference book: corresponding chapters

2 假设检验：均值

- 总体均值
- 总体比率
- 总结

拒绝域

ppt.35-44

- 双侧检验

$$\mathbb{H}_0 : \mu = \mu_0 \quad \mathbb{H}_1 : \mu \neq \mu_0$$

- 单侧检验: 右侧

$$\mathbb{H}_0 : \mu \geq \mu_0 \quad \mathbb{H}_1 : \mu < \mu_0$$

左侧

$$\mathbb{H}_0 : \mu \leq \mu_0 \quad \mathbb{H}_1 : \mu > \mu_0$$

总体均值

总体均值的假设检验

基本假设：样本来自于正态总体 $N(\mu, \sigma^2)$, 样本均值为 \bar{x} , 样本方差为 s^2

- 总体方差 σ^2 已知: 给定常数 μ_0 , 建立统计假设为

$$\mathbb{H}_0 : \mu = \mu_0 \quad \mathbb{H}_1 : \mu \neq \mu_0$$

统计量

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

给定 α , 得上分位数 $z_{\frac{\alpha}{2}}$ 和 z_α : 拒绝域为 $|z| \geq z_{\frac{\alpha}{2}}$ (双侧), $z \geq z_\alpha$ (右侧) 和 $z \leq -z_\alpha$ (左侧)

- 总体方差未知: 与方差已知的区别为统计量

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

给定 α , 得到上分位数 $t_{\frac{\alpha}{2}}(n-1)$ 和 $t_\alpha(n-1)$: 拒绝域为 $|t| \geq t_{\frac{\alpha}{2}}(n-1)$ (双侧), $t \geq t_\alpha(n-1)$ (右侧) 和 $t \leq -t_\alpha(n-1)$ (左侧).

置信区间法

p165 E6-15

- 95% 的置信区间为 $\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$, 题中

$$\alpha = 0.05, \quad \bar{x} = 255.8, \quad z_{\frac{\alpha}{2}} = 1.96$$

代入数值得

$$\Delta = 1.96 * \frac{5}{\sqrt{40}} = 1.5495$$

即 95% 的置信区间为 $[\alpha$ 为是总体参数未在区间内的比例 \rightarrow 未包含真值是小概率事件]

$$[\bar{x} - \Delta, \bar{x} + \Delta] = [254.25, 257.35]$$

- 包含容量 255 毫升, 不能拒绝原假设 统计上认为不符合标准的证据不足

更多例子: p166-7, E6-16 17 18 19

总体比率

总体比率

检验

$$\mathbb{H}_0 : p = \pi \quad \mathbb{H}_1 : p \neq \pi$$

统计量

$$Z = \frac{p - \pi}{\sqrt{(1 - \pi)\pi/n}} \sim N(0, 1)$$

p170 E6-21

总结

- ① 总体均值的检验 (掌握)
- ② 总体成数的检验 (了解)

Keyword: 拒绝域, 总体均值的假设检验, 总体比率的假设检验

Homework: (* by group)

- Time loss of smart phones, pads and computers in classroom
- 基本概念要牢记, 牢记, 牢记!
- * Salary-stu.xlsx, 电脑里完成其中 sheet2 里的内容 (纸版或者电子版都不需要提交)
- Read
 - ① Textbook: S6.7.2 (关于假设检验的 p 值, 自己上网找, 不要多于 30 分钟)
 - ② Reference book: corresponding chapters

3 假设检验： p 值

- 计算
- 总结

p -value

什么是 p 值? (p -value): 如果原假设是正确的话, 我们得到得到目前这个样本数据的可能性有多大, 如果这个可能性很小, 就应该拒绝原假设

- 显著性检验的目的是要描述样本所提供不利于原假设的证据有多强。 p 值就在做这件事
- 统计量落在拒绝域不同的地方, 实际的显著性是不同的。 p 值比临界值提供更丰富的信息

the p -value is a function of the sample. New sample, new p -value.

It may be less than or greater than the the type I error rate, the size or the significance level.

好处: 0 到 1 之间标准化, 临界值则取自差别很大

计算

计算 p -value

得到统计量 T 后 [左、右和双侧看 \mathbb{H}_1]

- 左侧 $p = \Pr(T \leq t | \mathbb{H}_0)$ 注意含等于号, 离散分布时有区别
- 右侧 $p = \Pr(T \geq t | \mathbb{H}_0)$
- 双侧 $p = 2 \min\{\Pr(T \geq t | \mathbb{H}_0), \Pr(T \leq t | \mathbb{H}_0)\}$

Excel 函数: `NORM.S.DIST(T,TRUE)`

p165 E6-15: $Z = 1.01$, $p = 0.3125$ [$=2*(1-\text{NORM.S.DIST}(1.01,\text{TRUE}))$]; swp
 $(1 - \text{NormalDist}(1.01)) * 2$]

更多例子: p166-7, E6-16 17 18 19 [部分题目由学生当场演示]

总结

① p 值的概念和计算 (掌握)

Keyword: p 值

Homework: (* by group)

- Time loss of smart phones, pads and computers in classroom
- 基本概念要牢记, 牢记, 牢记!
- * Salary-stu.xlsx, 电脑里完成其中 sheet2 里的内容, 采用 p 值的方法 (电子版自己保留, 纸质版抄录结果提交)
- Read
 - ① Textbook: S7.1–3
 - ② Reference book: corresponding chapters