Chpt.8 Statistical Inference: Hypothesis Testing

第八章 假设检验

问题引入

体重指数BMI是目前国际上常用的衡量人体胖瘦程度以及是否健康的一个标准. 专家指出, 健康成年人的BMI 取值应在 18.55- 24.99 之间.

某种减肥药广告宣称，连续使用该种减肥药一个星期便可达到减肥的效果.

问题引入

为了检验其说法是否可靠,随机抽取9位试验者(要求BMI 指数超过25、年龄在20-25岁女生), 先让每位女生记录没有服用减肥药前的体重,

然后让每位女生服用该减肥药, 服药期间, 要求每位女生保持正常的饮食习惯, 连续服用该减肥药1周后, 再次记录各自的体重.

测得服减肥药前后的体重差值(服药前体重-服药后体重) (单位: kg):

1.5， 0.6， -0.3， 1.1， -0.8， 0， 2.2， -1.0， 1.4

问题：根据目前的样本资料能否认为该减肥药广告中的宣称是可靠的?

问题分析

- 本质：用X代表服药前后体重的差值，一般可假设$X \sim N(\mu, \sigma^2)$，问题可归结为“判断是否有$\mu > 0$成立”？
- 思路：参数估计，μ的矩估计量和最大似然估计量都是$\bar{X}$，根据数据求得$\bar{x} = 0.522$，
- 问题：据此可以得出$\mu > 0$的结论吗？
- 不能！实验获得的数据，隶属于某一随机变量，具有一定的随机性（波动性）。我们要甄别来自真实的差异，还是随机变量的波动性。

假设检验

1. 建立假设

- 令服用减肥药前后体重差值$X \sim N(\mu, 0.36)$
- 建立假设

$$H_0: \mu \le 0, \quad H_1: \mu > 0$$

其中，H_0称为原假设（零假设），H_1称为备择假设（对立假设）。

Remark：原假设与备择假设是不对称的！

决定谁是原假设，依赖于立场、惯例、方便性

原假设的选择原则

I.保护原假设

如果错误地拒绝假设A比错误地拒绝假设B带来更严重的后果，则将A作为原假设！

例如，假设A: 新药有某种毒副作用

假设B: 新药无某种毒副作用

“有毒副作用”错误地当成“无毒副作用”比“无毒副作用”错误地当成“有毒副作用”带来的后果更严重。

将A选作原假设！

原假设的选择原则

II.原假设设为维持现状

为解释某些现象或效果的存在性，原假设常取为“无效果”、“无改进”、“无差异”等，拒绝原假设表示有较强的理由支持备择假设.

例如，刚才减肥药的例子中，

H_0: 药物无减肥效果，即$\mu \leq 0$

H_1: 药物有减肥效果，即$\mu > 0$

原假设的选择原则

III.原假设取简单假设

只有一个参数(或分布)的假设称为简单假设.

如果只有一个假设是简单假设，将其取为原假设。

假设的形式

设μ是反映总体某方面特征的量，是我们感兴趣的参数.

参数μ的假设有三种情形:

- H_0: $\mu \le \mu_0$, H_1: $\mu > \mu_0$--右边检验 $\Longleftrightarrow$ H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$
- H_0: $\mu \ge \mu_0$, H_1: $\mu < \mu_0$--左边检验 $\Longleftrightarrow$ H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$
- H_0: $\mu = \mu_0$, H_1: $\mu \ne \mu_0$--双边检验

判断依据

- 假设检验的思路：根据实验数据，在 H_0与 H_1中选择一个

1）若有充分理由说明原假设H_0是不正确的，则拒绝H_0，而接受备择假设H_1；

2）若没有充分理由说明原假设H_0是不正确的，则接受H_0.

法官的立场基于“疑罪从无”：
法官宣告被告“有罪”是需要充分的证据来推翻被告是“无罪”的假设；而宣判“无罪”，是由于没有充分的证据支持被告“有罪”，并不是有充分的证据支持被告“无罪”.

- $X \sim N(\mu, 0.36)$，我们的假设是 $H_0: \mu \leq 0$， $H_1: \mu > 0$
- 根据样本数据，求得$\bar{x} = 0.522$。对$\bar{x} > 0$有两种解释：

(1) 原假设H_0是正确的

即总体均值$\mu \leq 0$，由于抽样的随机性，$\bar{x}$出现某些偏差是完全可以接受的；

(2) 原假设H_0是不正确的

即总体均值$\mu > 0$，因此 $\bar{x}$大于0并不是随机的，$\mu > 0$是显著成立的。

这两种解释哪个更合理呢？

判断依据是**小概率事件的实际不可能性原理**：

① 在承认原假设H_0正确的条件下，选取小概率事件A(对应随机变量含有假设中的参数）；

② 由抽样的结果考察A是否出现；

③ 如果H_0是正确的，那么事件A发生的概率很小，在一次抽样中一般A不会出现。由此，形成如下准则：

若A出现，则说明H_0不正确

若A没有出现，则认为H_0正确。

如何构造小概率事件？

回忆：$X \sim N(\mu, 0.36)$，检验假设是 $H_0: \mu \le 0, \quad H_1: \mu > 0$

当 $\overline{X} \ge C$ 时，拒绝原假设 H_0，
当 $\overline{X} < C$ 时，接受原假设 H_0，
其中 C 是待定的常数.

由于样本的随机性，任一检验规则在应用时，都有可能发生错误的判断
两类错误

检验决策 \ 真实情况	H_0为真	H_0非真
拒绝H_0	第I类错误(α)	正确
接受H_0	正确	第II类错误(β)

第I类错误：拒绝真实的原假设(弃真).
第II类错误：接受错误的原假设(取伪).

$\alpha = P\{$第I类错误$\} = P\{$拒绝$H_0 \mid H_0$是真实的$\}$,

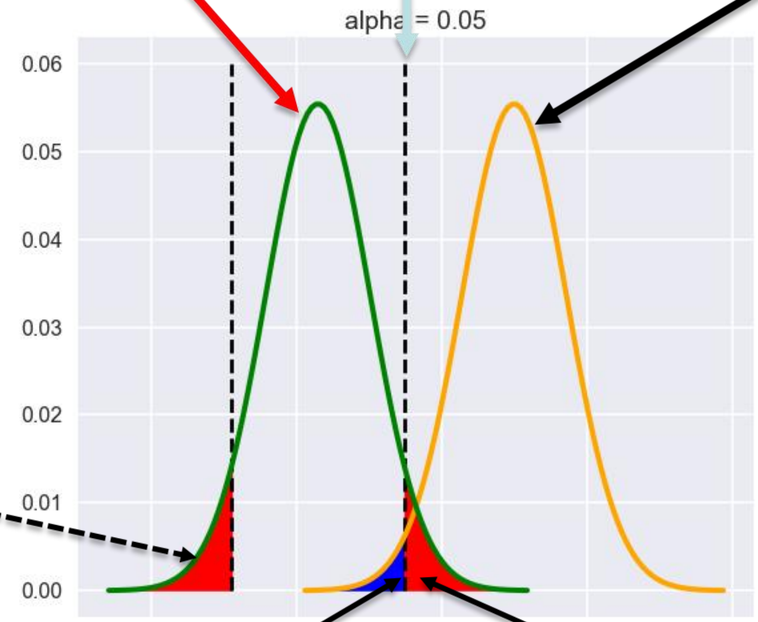
$\beta = P\{$第II类错误$\} = P\{$接受$H_0 \mid H_0$是错误的$\}$

假设检验可能犯的两类错误

临界值C

原假设
$H_0: \mu \leq 0$

备择假设
$H_1: \mu > 0$



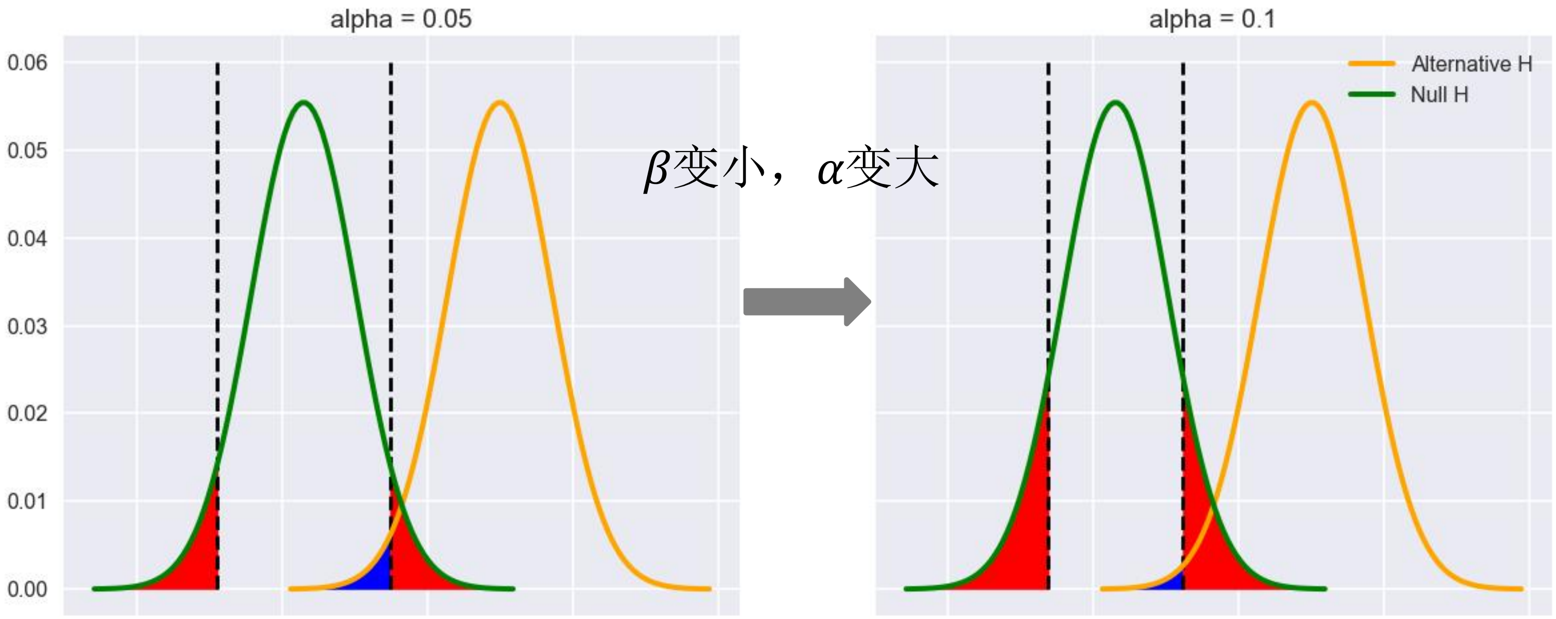
左边的红色区域
请忽略

第二类错误：原假设为假，备择假设为真，但根据样本接受了原假设，
即$\mu > 0$，且$\bar{X} < C$

第一类错误：原假设为真，但根据样本拒绝了原假设，
即$\mu \leq 0$，且$\bar{X} \geq C$

假设检验可能犯的两类错误

犯两类错误的概率相互制约！



犯两类错误的概率当然是越小越好，但当样本容量固定时，两类错误的概率不可能同时减小，而是其中一个减小，另一个就会增大。

如何构造小概率事件？

Neyman-Pearson原则：

首先控制犯第I类错误的概率不超过某个常数 $\alpha \in (0,1)$，再寻找检验，使得犯第II类错误的概率尽可能小. α称为**显著水平**.

常取 $\alpha = 0.01,\ 0.05,\ 0.1$ 等.

2.确定拒绝域的形式和检验统计量

回忆：$X \sim N(\mu, 0.36)$，检验假设是 $H_0: \mu \le 0$， $H_1: \mu > 0$

当$\overline{X} \ge C$时，拒绝原假设H_0，

当$\overline{X} < C$时，接受原假设H_0，

其中C是待定的常数.

拒绝域

取显著水平$\alpha = 0.05$. 考察第一类错误的概率

$$P(\overline{X} \ge C|\mu \le 0) = P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu}{\sigma/\sqrt{n}}|\mu \le 0)$$

$$\le P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C}{\sigma/\sqrt{n}}|\mu \le 0) \le \alpha$$

由于$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，可知$\frac{C}{\sigma/\sqrt{n}} \ge z_\alpha$，即$C \ge \frac{\sigma}{\sqrt{n}} z_\alpha = \frac{0.6}{\sqrt{9}} z_{0.05} = 0.329$

检验统计量

3.确定拒绝域临界值

根据Neyman-Pearson原则，为使犯第II类错误的概率尽可能小，取C=0.329.

因此，当$\bar{X} \geq C$=0.329时，拒绝原假设 H_0。

4.根据样本数据做出判断

根据样本数据，$\bar{x} = 0.522$>0.329，拒绝原假设，认为减肥药确实有效。

[理由]：$P(\bar{X} \geq 0.329|\mu \leq 0) \leq 0.05$，$\bar{x} = 0.522$>0.329，根据**小概率事件的实际不可能性原理**，从而拒绝原假设 H_0。

假设检验的一般步骤

（1）根据实际问题提出原假设和备择假设；

（2）提出检验统计量和拒绝域的形式；

（3）在给定的显著水平α下，根据Neyman-Pearson原则求出拒绝域的临界值;

（4）根据实际样本观测值作出判断.

单边检验

[例1]：某厂生产的固体燃料推进器的燃烧率服从正态分布

N(μ, σ^2)， $\mu = 40cm/s$ ， $\sigma = 2cm/s$。

现在用新方法生产了一批推进器，从中随机取 n＝25只，测得燃烧率的样本均值为 $\bar{x} = 41.25cm/s$。设在新方法下总体的标准差仍为 2cm/s

问：用新方法生产的推进器的燃烧率是否较以往生产的推进器的燃烧率有显著的提高？取显著性水平α＝0.05。

右边检验

解：设 $H_0: \mu \le \mu_0 = 40$ （无提升）

$H_1: \mu > \mu_0$

此为右边检验

若 $\bar{X} \ge C$，则拒绝 H_0

若 $\bar{X} < C$，则接受 H_0.

考察第一类错误的概率

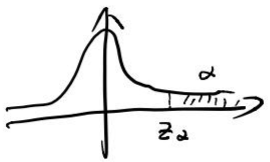
$$P(\bar{X} \ge C \mid \mu \le \mu_0) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu}{\sigma/\sqrt{n}} \,\middle|\, \mu \le \mu_0\right)$$

$$\le P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu \le \mu_0\right) \le \alpha$$

（因为 $\mu \le \mu_0$，则 $\frac{C-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu_0}{\sigma/\sqrt{n}}$，　$\frac{C-\mu_0}{\sigma/\sqrt{n}}$　$\frac{C-\mu}{\sigma/\sqrt{n}}$

则 $\left\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu}{\sigma/\sqrt{n}}\right\} \subseteq \left\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{C-\mu_0}{\sigma/\sqrt{n}}\right\}$）

由于 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，因此 $\frac{C-\mu_0}{\sigma/\sqrt{n}} \ge z_\alpha = z_{0.05} = 1.65$



$$C \ge \frac{\sigma}{\sqrt{n}} z_\alpha + \mu_0 = \frac{2}{\sqrt{25}} \times 1.65 + 40 = 40.66$$

根据 Neyman-Pearson 准则，$C = 40.66$

即 $\bar{X} \ge 40.66$，拒绝 H_0.

$\because \bar{X} = 41.25 > 40.66$

$\therefore$ 拒绝 H_0，即认为燃烧率确实有提升.

左边检验

若将原假设和备择假设互换，即考虑左边检验

把上面的H_0与H_1互换，则当前的

$H_0: \mu > \mu_0 = 40$（有提升）

$H_1: \mu \le \mu_0$（无提升）

此时变为左边检验

若 $\bar{x} \le c$，则拒绝 H_0

若 $\bar{x} > c$，则接受 H_0.

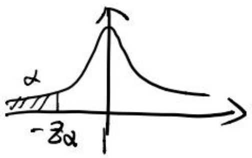
考察第I类错误的概率.

$$P(\bar{x} \le c \mid \mu > \mu_0) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{c-\mu}{\sigma/\sqrt{n}} \,\middle|\, \mu > \mu_0\right)$$

（由$\mu>\mu_0$，与前面类似推导）

$$\le P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{c-\mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu > \mu_0\right) \le \alpha.$$

由 $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，知



$$\frac{c-\mu_0}{\sigma/\sqrt{n}} \le -z_\alpha \qquad z_{0.05}=1.65$$

$$c \le -\frac{\sigma}{\sqrt{n}} z_\alpha + \mu_0 = -\frac{2}{\sqrt{25}} \times 1.65 + 40 = 39.34$$

由N-P准则，$c=39.34$.

即 $\bar{x} \le 39.34$，则拒绝H_0.

$\bar{x} = 41.25 > 39.34$，不能拒绝H_0，即接受H_0，

认为燃烧率有提升.

可以看到，在前面的例题中，交换了原假设和备择假设后，分别进行左边检验和右边检验，结果是一样的。

那么这两个选取，是否有好坏之分?

右边检验是比较 41.25与40.66， 左边检验比较的是 41.25 和39.34。 直觉上那个更“严格”？

一般地，在有关参数的假设检验中，备择假设是我们根据样本资料希望得到支持的假设。

我们希望数据，能够通过“较为严格”的统计计算，支持我们否定掉不想要的情况（原假设），从而得到备择假设，即想要的情况。

双边检验

[例2]：某车间用一台包装机包装葡萄糖，包装好的葡萄糖重量是一个随机变量X，它服从正态分布N(μ，σ²)。当机器正常时，其均值为0.5 公斤，标准差为0.015公斤。

某日开工后为检验包装机工作是否正常，随机抽取它所包装的9袋糖，称得净重为（公斤）：

0.497 / 0.506 / 0.518 / 0.524 / 0.498 / 0.511 / 0.520 / 0.515 / 0.512

假设产品的方差是已知并不变的，问机器工作是否正常？

取显著水平$\alpha = 0.05$.

解: $H_0: \mu=\mu_0$ (正常)　　$\alpha=0.05,\ n=9,\ \sigma=0.015,\ \mu_0=0.5$

$H_1: \mu\neq\mu_0$ (不正常)

此为双边检验.

设若 $|\bar{X}-\mu_0|\geq C$, 则拒绝 H_0

若 $|\bar{X}-\mu_0|<C$, 则接受 H_0.

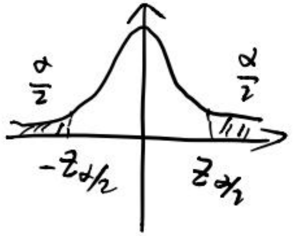
$$\left(\begin{array}{l} X\sim N(\mu,\sigma^2) \\ \text{当}\ \mu=\mu_0,\ X\sim N(\mu_0,\sigma^2) \\ \qquad\qquad \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1) \end{array}\right)$$

考察第I类错误的概率

$$P(|\bar{X}-\mu_0|\geq C \mid \mu=\mu_0)=P\left(\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq\frac{C}{\sigma/\sqrt{n}}\ \middle|\ \mu=\mu_0\right)\leq\alpha.$$

$$\frac{C}{\sigma/\sqrt{n}}\geq z_{\alpha/2}=z_{0.025}=1.96$$

$$C\geq\frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2}=\frac{0.015}{\sqrt{9}}\times1.96=0.0098$$



取 $C=0.0098$, 即 $|\bar{X}-\mu_0|\geq 0.0098$ 时, 拒绝 H_0.

$\because\ \bar{x}=0.511,\ |\bar{x}-\mu_0|=0.511-0.5=0.011>0.0098$

$\therefore$ 拒绝 H_0, 即认为当天机器不正常.

练习题

例 2 公司从生产商购买牛奶。公司怀疑生产商在牛奶中掺水以谋利。通过测定牛奶的冰点，可以检验出牛奶是否掺水. 天然牛奶的冰点温度近似服从正态分布，均值 $\mu_0=-0.545$ ℃，标准差 $\sigma=0.008$ ℃. 牛奶掺水可使冰点温度升高而接近于水的冰点温度(0 ℃). 测得生产商提交的 5 批牛奶的冰点温度，其均值为 $\bar{x}=-0.535$ ℃，问是否可以认为生产商在牛奶中掺了水? 取 $\alpha=0.05$.

解 按题意需检验假设

$$H_0: \mu \leqslant \mu_0 = -0.545(\text{即设牛奶未掺水}),$$

$$H_1: \mu > \mu_0(\text{即设牛奶已掺水})$$

这是右边检验问题，其拒绝域如(1.6)式所示，即为

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geqslant z_{0.05} = 1.645.$$

现在 $z = \dfrac{-0.535-(-0.545)}{0.008/\sqrt{5}} = 2.7951 > 1.645$，$z$ 的值落在拒绝域中，所以我们在显著性水平 $\alpha = 0.05$ 下拒绝 H_0，即认为牛奶商在牛奶中掺了水. □

显著性水平的意义

α越大，要拒绝H_0所需要的$|\bar{X}-\mu_0|$越小，$\bar{X}$和μ_0的差异越不显著，检验越不严格；

α越小，要拒绝H_0所需要的$|\bar{X}-\mu_0|$越大，$\bar{X}$和μ_0的差异越显著，检验越严格。

$$P\left\{\,|Z|>z_{\alpha/2}\right\}=P\left\{\left|\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right|>z_{\alpha/2}\right\}=\alpha$$