Chpt.8 Statistical Inference: Hypothesis Testing

第八章 假设检验

pp. 1

南开大学计算机学院

目录——Content

- 假设检验的定义和基本概念

问题的提出

例1 设某种清漆的9个样品，其干燥时间（以小时计）分别为：

6.0 5.7 5.5 6.5 7.0 5.8 5.2 6.1 5.0

根据以往经验，干燥时间的总体服从正态分布N(6.0,0.36),现根据样本检验均值（$\bar{x}=$）是否与以往有显著差异？

换而言之：对于原有数据$X \sim N(\mu_0, \sigma_0^2)$,对于采集到的样本数据（是否有变化）

变化在这个场景下定义为样本均值$\bar{x}$（对应随机变量$\bar{X}$,）是否还和μ_0保持一致？

南开大学计算机学院

问题的提出

例2 一种摄影药品被其制造商声称其贮藏寿命是均值180天、标准差不多于10天的正态分布。

某位使用者担心标准差可能不超过10天。他随机选取12个样品并测试，得到样本标准差为8天。根据样本有充分证据证明标准差小于10天吗？

此时，对于样本方差s^2（对应随机变量S^2）是否小于σ_0？

南开大学计算机学院

问题的提出

例3 孟德尔遗传理论断言，当两个品种的豆杂交时，圆的和黄的、起皱的和黄的、圆的和绿的、起皱的和绿的豆的频数将以比例9：3：3：1发生。

在检验这个理论时，孟德尔分别得到频数315、101、108、32这些数据提供充分证据拒绝该理论吗？

即：经验数据的离散分布列，和孟德尔所判定的分布列一致吗？

问题的提出

上述三个问题可以看到：

我们统计实验获得的数据，隶属于某一个统计量（随机变量），具有一定的变化性。

我们需要判定一个研究的假设：

即某一个结果数值上的不一致是来自真实的差异，还是随机变量的变化；

或者我们需要判定，某一个结果的一致，是偶然发生的，还是具有极高概率发生的“几乎必然”发生

也就是，我们的假设成立吗？

假设检验的定义及基本概念

错误的例子：经过改进的A方法比B方法更好。
正确的例子：经过改进的A方法比B方法，均值提高了/方差变小了。

What Is

统计假设——通过实际观察或理论分析，对总体分布形式或对总体分布形式中的某些参数作出某种假设（可量化）

假设检验——根据抽取的样本观测值，构造适当的统计量，对上述假设的正确性进行判断，从而决定接受假设或拒绝假设，这一统计推断过程就是所谓的假设检验。

Notice 1：假设不是毫无根据的（通过实际观察或理论分析）

Notice 2：参数除了估计外，也可采用假设检验方法参与研究。

假设检验的定义及基本概念

[例1]：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，σ^2)。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？

注1：回顾第七章：参数估计，我们应该估计什么？$\hat{u} = \bar{x} = 0.511$

注2：我们无法进一步对$\hat{u}$ 与u 做出判定，且判定$\hat{u} \neq \mu$

注3*：频率派定义下，我们无法 $\hat{u}$为随机变量，只能归咎于$\bar{x}$不能准确预计$\hat{u}$，

注4：这种不准确就分为“有其他因素影响，$\bar{x}$实在无能为力啊”和“一不小心不准确了一下”

假设检验的定义及基本概念

一般情况下方差比较稳定，糖重服从正态分布N(μ，σ^2)，而判断机器是否正常，就相当于判断μ是否为0.5公斤。

为此，提出两个相对对立的假设：

$$H_0: \ \mu = \mu_0 = 0.5 \qquad H_1: \ \mu \neq \mu_0$$

然后给出一个合理的法则，根据法则，利用已知样本做出决策，是接受假设H_0（即拒绝假设H_1），还是拒绝假设H_0(即接受假设H_1)。

如果作出的决策是接受H_0，则认为$\mu=\mu_0=0.5$，即认为机器工作是正常的。

假设检验的定义及基本概念

假设的设立

通常称H_0假设为原假设或零假设，称H_1假设为备择假设

检验的目的就是要在原假设H_0与备择假设H_1之间

选择其中之一：

1）若认为原假设H_0是正确的，则接受H_0；

2）若认为原假设H_0是不正确的，则拒绝H_0，而接受备择假设H_1.

假设检验的定义及基本概念

从抽样检查的结果知样本均值 $\bar{x}=0.511$，显然样本均值 $\bar{x}$ 与假设的总体均值 $\mu_0=0.5$ 之间存在差异. 但是，我们不能就此简单否定 $\mu=\mu_0=0.5$

对 $\bar{x}$ 与 μ_0 之间出现的差异可以有两种不同的解释：

(1) 原假设H_0是正确的

即总体均值$\mu=\mu_0=0.5$，由于抽样$\bar{X}$的随机性，与μ_0之间出现某些差异是完全可以接受的；

(2) 原假设H_0是不正确的

即总体均值$\mu\neq\mu_0$，因此 μ 与 μ_0 之间出现的差异不是随机的，即与μ_0之间存在实质性、显著性的差异。

假设检验的定义及基本概念

上述两种解释哪一种较合理呢？

回答这个问题的依据是**小概率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H_0正确下的小概率事件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发生的概率是小的，一次抽样一般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。

即所谓的假设-检验：

先研究题目要求对应的参数

再找到要研究的随机变量（含参数，已知分布）

该随机变量现有的样本数据是小概率事件吗？

然后进行等价的该假设发生与否的检验

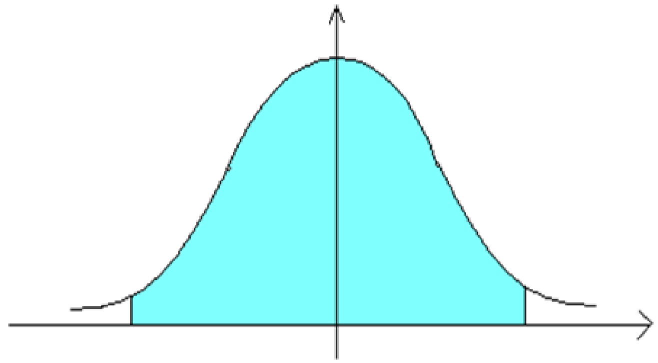
假设检验的定义及基本概念

[例1]：如原假设H_0正确，即$X \sim N(\mu_0, \sigma_0^2)$，则统计量

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

该统计量既包含统计量$\overline{X}$,也包含研究参数μ_0，考虑假设对应的一个小概率事件（取某个较小的正数α）

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



其中α称为**显著水平**，**Z称为检验统计量**

$z_{\alpha/2}$ 称为统计量 Z 的**临界值**.

假设检验的定义及基本概念

[例1]：通常α取较小的值，如0.05或0.01，当显著水平α＝0.05时，查表得$Z_{\alpha/2}$

$$z_{\alpha/2}=u_{0.025}=1.96$$

假如H_0正确，即 $X\sim N(\mu_0,\sigma_0^2)$，则 $\left\{\left|\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right|>1.96\right\}$

为小概率事件，满足

$$P\{|Z|>1.96\}=P\left\{\left|\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right|>1.96\right\}=0.05$$

假设检验的定义及基本概念

因为$\alpha=0.05$很小，所以事件$|Z|>1.96$是小概率事件，根据小概率事件的实际不可能性原理，可以认为在原假设H_0正确的条件下这样的事件实际上是不可能发生的。

但现在[例1]抽样检查的结果是

$$|z|=\frac{|0.511-0.5|}{0.015/\sqrt{9}}=2.2\quad>1.96$$

理解左边的式子：
样本数据$\bar{x}$为0.511,它是由μ_0=0.5,且受到σ_0浮动影响造成的概率<0.05,是小概率事件

小概率事件$|Z|>1.96$竟然发生，这只能表明：

抽样检查的结果与原假设H_0不相符合，即样本均值$\bar{x}$与假设的总体均值μ_0之间存在显著差异，因此，应当拒绝原假设H_0，接受备择假设H_1，即认为包装机今天工作不正常。

假设检验的定义及基本概念

Remark 1：假设检验的结论与选取的显著水平α密切相关

上述结论拒绝原假设H_0，接受备择假设H_1，是取显著水平α＝0.05时得到的；若改取显著水平α＝0.01，则 $z_{\alpha/2}=2.58$

从而有 $P\{\,|Z|>2.58\,\}=0.01$

因为抽样检查的结果是$|z|=2.2<2.58$，可见小概率事件$|Z|>2.58$ 没有发生，所以没有理由拒绝原假设H_0，就应当接受H_0，即可以认为该日包装机工作正常，μ=0.5（公斤）.

假设检验的定义及基本概念

Remark2：假设检验中使用的推理方法是一种“反证法”，但这种“反证法”使用的不是纯数学中的逻辑推理，而仅仅是根据小概率事件的实际不可能性原理来推断的

假设检验的定义及基本概念

关于假设 H_0：$\mu = \mu_0 = 0.5$；　H_1：$\mu \neq \mu_0$ 的检验

当统计量 Z 的观测值的绝对值大于临界值 $z_{\alpha/2}$ 时，则拒绝原假设H_0，

这时Z的观测值落在区间$(-\infty, -z_{\alpha/2})$ 或 $(z_{\alpha/2}, \infty)$ 内

称上述这样的区间为关于原假设H_0拒绝域

拒绝域就是统计量只能以小概率发生的，对应的最外侧的定义域

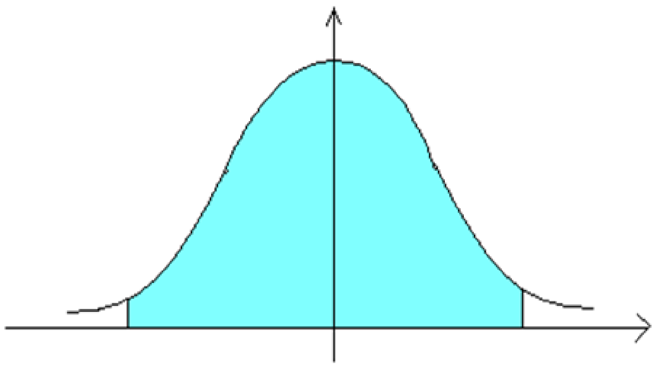
拒绝域的边界点称为临界点

假设检验的定义及基本概念

这里的拒绝域分别位于两侧，因此称这类假设检验为**双侧假设检验（双边假设检验或双边备择假设）**。

双边假设检验的假设形式

$$H_0: \ \mu = \mu_0 \qquad\qquad H_1: \ \mu \neq \mu_0$$



假设检验的定义及基本概念

具有形如

$$H_0: \ \mu \le \mu_0 \qquad\qquad H_1: \ \mu > \mu_0$$

的假设检验称为右边检验

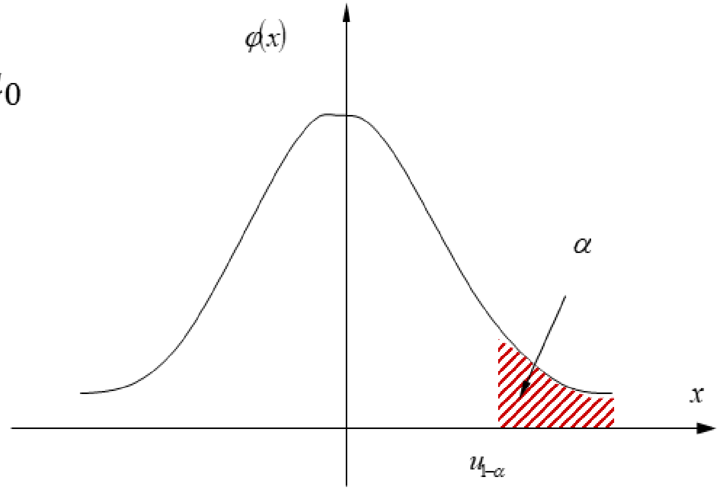
具有形如

$$H_0: \ \mu \ge \mu_0 \qquad\qquad H_1: \ \mu < \mu_0$$

的假设检验称为左边检验

右边检验和左边检验统称为单侧假设检验

单侧假设检验的拒绝域只位于一侧



假设检验的定义及基本概念

由于假设检验是根据样本做出的，样本本身就是随机变量，具有随机性，因此总有可能做出错误的判断。

在假设H_0实际上为真时，依然有可能犯拒绝H_0的错误，称这类“弃真”错误为第I类错误
在假设H_0实际上为不真时，依然有可能犯接受H_0的错误，称这类“取伪”错误为第II类错误

当样本容量固定时，若减少犯某一类错误的概率，则另一类错误的概率就会增大。

在给定样本容量的情况下，我们控制犯第I类错误的概率不大于α，这种只对犯第I类错误的概率加以控制，而不考虑犯第II类错误的概率的假设检验，称为显著性检验

假设检验的定义及基本概念

假设检验的基本方法：

将检验问题转化为一个变量问题

我们将希望证明的检验条件放到备择检验部分(H1)

除非条件里面有等号，则放入H0部分

我们目标检查备择假设是否是小概率

基于这个变量构造一个已知分布的随机变量

判断采样数据在现有分布下发生概率，是否是小概率事件

⇔采样数据对应的分位点与 小概率分位点的比较；拒绝域

假设检验的定义及基本概念

如何构造H0,H1

待验证部分为何放入H1?/如何选取H0 H1?

为什么这个“=”情况特殊？*

选择随机变量

非正态分布时、多种比对目标时，选那些随机变量？*

分位点与拒绝域

单尾检验的选择

拒绝域与置信区间的关系？*

假设检验的定义及基本概念

[例2] 某厂生产的固体燃料推进器的燃烧率服从正态分布

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

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$\alpha=0.05$。

假设检验的定义及基本概念

解：按题意，需检验假设

H_0：$\mu \le \mu_0 = 40$（即假设新方法没有提高燃烧率）

H_1：$\mu > \mu_0$（即假设新方法提高了燃烧率）

此可以成为右边检验问题。

类似地，有时我们需要检验假设

H_0：$\mu \ge \mu_0 = 40$，H_1：$\mu < \mu_0 = 40$

此称为左边检验。左边检验和右边检验简称为单边检验。

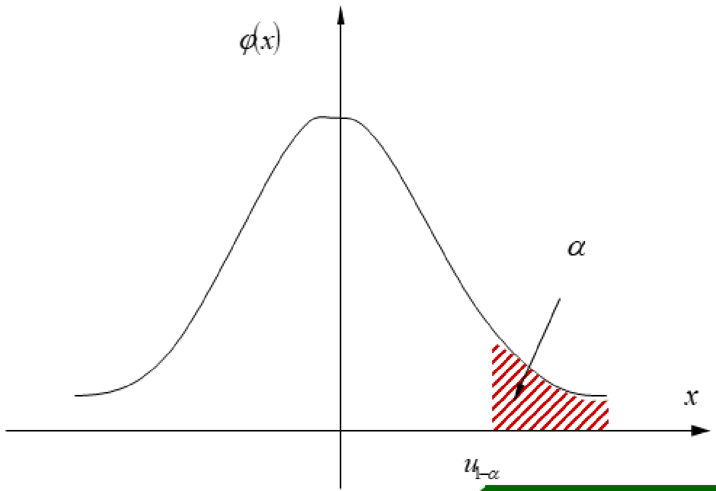
假设检验的定义及基本概念

设总体X~N(μ，σ²)，σ为已知， $X_1, X_2, \cdots, X_n$ 是来自 X的样本。给定显著性水平α，求解单边检验问题。

[1] 右边检验问题H_0：$\mu \le \mu_0$，H_1：$\mu > \mu_0$

如果H_0：$\mu \le \mu_0$成立，那么 $\bar{X}$ 的观察值 $\bar{x}$ 一般就要不大于μ_0 ,定义Z_α

而 $\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$ 是小概率事件



如果检测到它发生，即 $\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$

则拒绝 H_0：$\mu \le \mu_0$

假设检验的定义及基本概念

例2：燃烧率有无显著提高

现在 $n=25$，总体的标准方差σ= 2cm/s

样本均值为 $\bar{x} = 41.25cm/s$

显著性水平α=0.05。 $z_\alpha = 1.65$

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{41.25-40}{2/5} = 3.125 > z_\alpha$$

$$1-\phi(3.125) = 1-0.999 = 0.001$$

小概率事件发生了，那说明原假设H_0不成立，

即新方法显著提高了燃烧率。

假设检验的定义及基本概念

[2] 左边检测问题

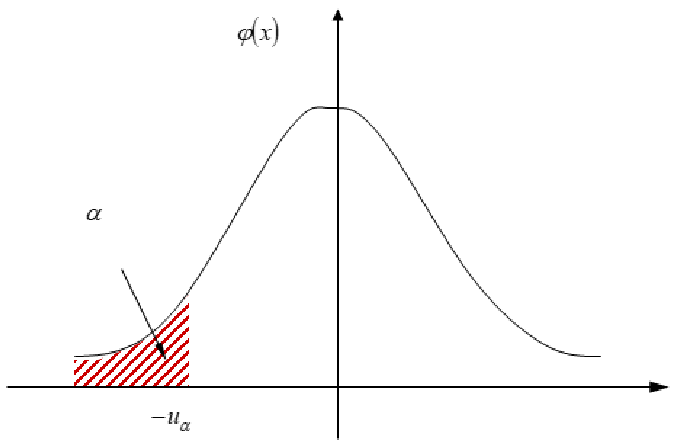
假设　H_0：$\mu \geq \mu_0$，H_1：$\mu < \mu_0$

如果H_0：$\mu \geq \mu_0$成立，那么 $\overline{X}$ 的观察值 $\overline{x}$ 一般就不小于μ_0

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$$ 是小概率事件

如果它发生，即 $\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$

则拒绝 H_0：$\mu \geq \mu_0$



假设检验的定义及基本概念

若将原假设和备择假设互换，即考虑左边检验

H_0： $\mu>\mu_0=40$ （即假设新方法提高了燃烧率）

H_1： $\mu\leq\mu_0$（即假设新方法没有提高燃烧率）

此可以成为右边检验问题。

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{41.25-40}{2/5}=3.125\ ?\ Z_{-\alpha}$$

$$3.125>Z_{-\alpha}=-1.65$$

所以上述事件不是小概率事件，我们不拒绝H_0，即新方法提高了燃烧率，结果和之前一样

左面例子只是为了展示单纯的交换，
正确例子应为
$\mathrm{H}_0: \mu \geq \mu_0$
$\mathrm{H}_1: \mu < \mu_0$

假设检验的定义及基本概念

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？

左边检验是比较3.125与1.65，右边检验比较的是3.125和-1.65。直觉上那个更“严格”？

在第四节我们会了解假设检验与置信区间的关系*，同时讨论“=”为什么必须在左边原假设。

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），从而得到想要的情况。

假设检验的定义及基本概念

假设检验的一般步骤

根据实际问题提出原假设H_0与备择假设H_1，即，指出需要检验的假设的具体内容；

↓

选取适当的统计量(涉及假设的问题），并在假定原假设H_0成立的条件下确定该统计量的分布；

↓

根据问题的需要，适当选取显著水平α(α的值一般比较小)，构建统计量的概率为α的事件，也就是据统计量的分布查表确定α的临界值；

↓

根据样本观测值计算统计量的观测值，与临界值比较，作出拒绝或接受原假设H_0的判断.

假设检验的定义及基本概念

由于假设检验使用的是根据小概率的实际不可能性原理作出判断的一种“反证法”，而无论小概率事件A发生的概率如何小，它还是有可能发生的。

假设检验可能作出以下两类错误的判断：

(1)第一类错误“弃真”：

即原假设H_0实际上是正确的，但却错误地拒绝了H_0.

由于小概率事件A发生时才会拒绝H_0，所以犯第一类错误的概率为

$P\{A\,|\,H_0\}\le\alpha$.

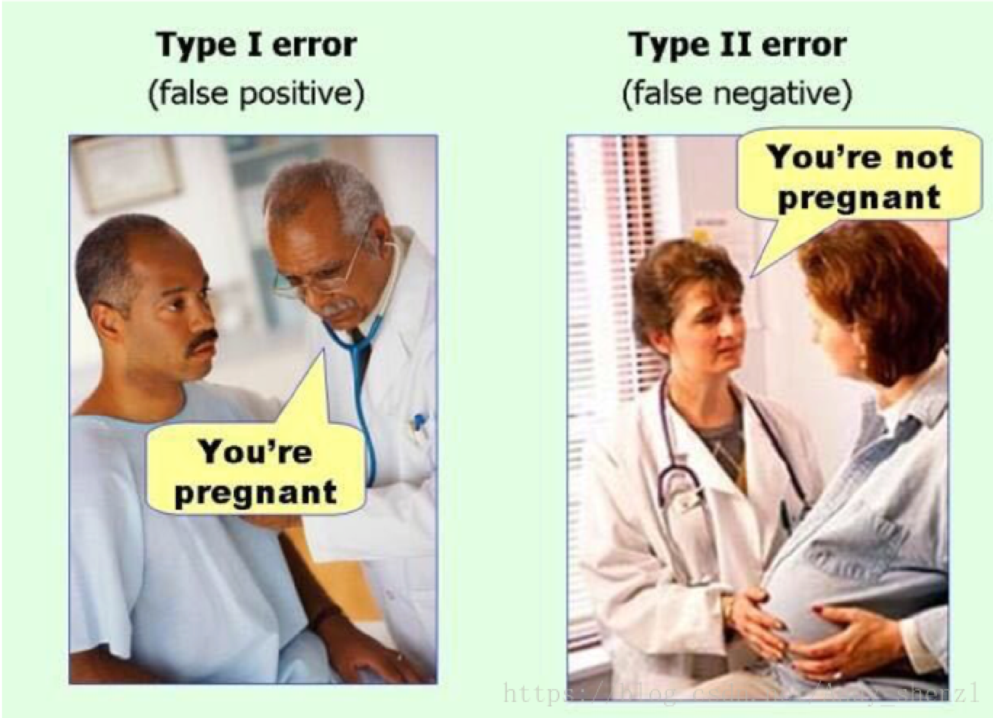
(2) 第二类错误“取伪”：

即原假设H_0实际上是不正确的，但却错误地接受了H_0，　犯第二类错误的概率记为$P(\overline{A}|H_1)\le\beta$.

南开大学计算机学院

H_0: 男人不会怀孕
H_1: 图中男人怀孕了

H_0: 图中女人没怀孕
H_1: 图中女人怀孕了



假设检验的定义及基本概念

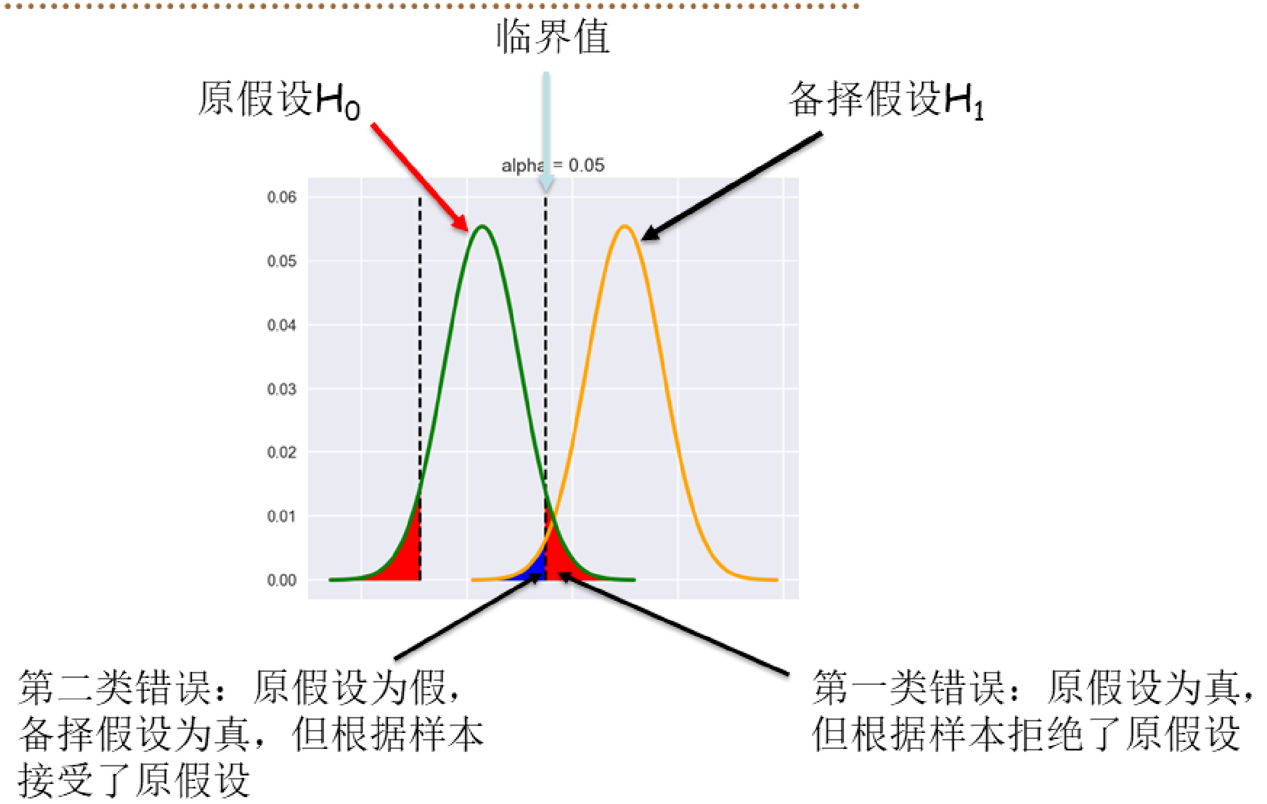
检验决策 \ 真实情况	H_0为真	H_0非真
拒绝H_0	第一类错误(α)	正确
接受H_0	正确	第二类错误(β)

(1)能否设计一种假设检验的方法，同时减小第一类错误和第二类错误的概率？ **不能**

(2)如果不能同时减小犯第一、二类错误的概率，那么假设检验的原则是什么？

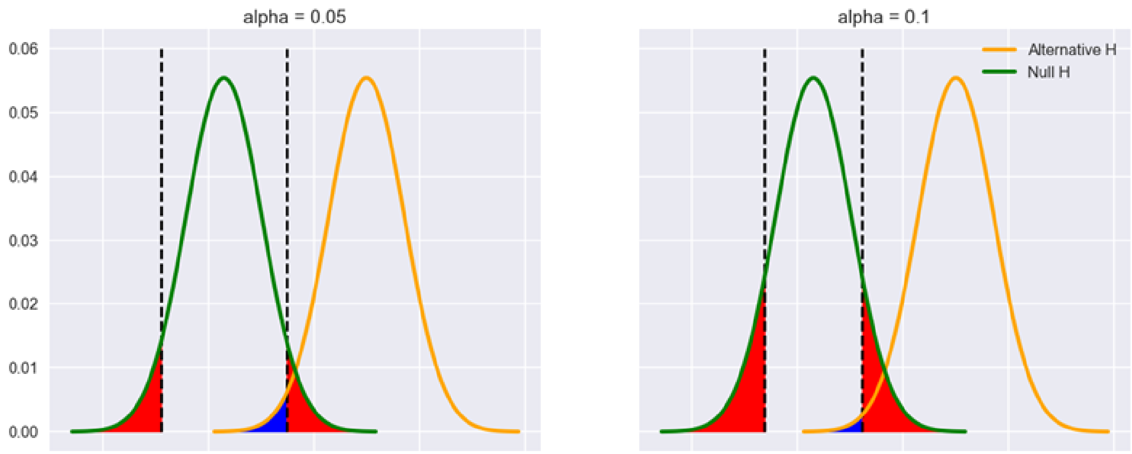
- 优先控制一种错误的概率
- 犯第一、二类错误，哪个的影响更大？

假设检验的定义及基本概念



假设检验可能犯的两类错误

犯两类错误的概率相互制约！



正态分布的曲线是固定的，增大α，β会减小

假设检验的定义及基本概念

例1中，犯第一类错误的概率

$$\alpha(C) = P\{A|H_0\} = P\{|\overline{X} - \mu_0| \geq C|\mu = \mu_0 = 0.5\}$$

$$= P\left\{\frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} \geq \frac{C}{\sigma/\sqrt{n}}|\mu = \mu_0 = 0.5\right\}$$

因为 $\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ 所以犯第一类错误的概率为

$$\alpha(C) = 2 - 2\Phi\left(\frac{C}{\sigma/\sqrt{n}}\right)$$

若$\frac{C}{\sigma/\sqrt{n}} = z_{\alpha/2}$，则$\alpha(C) = \alpha$

假设检验的定义及基本概念

例1中，犯第二类错误的概率

$$\alpha(C) = P\{\overline{A}|H_1\} = P\{|\overline{X} - \mu_0| < C|\mu \neq \mu_0\}$$

$$= P\{\mu_0 - C < \overline{X} < \mu_0 + C|\mu \neq \mu_0\}$$

$$= P\left\{\frac{\mu_0 - C - \mu}{\sigma/\sqrt{n}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{\mu_0 + C - \mu}{\sigma/\sqrt{n}} |\mu \neq \mu_0\right\}$$

因为 $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 所以犯第二类错误的概率为

$$\beta(C) = \Phi\left(\frac{\mu_0 + C - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu_0 - C - \mu}{\sigma/\sqrt{n}}\right) \quad \mu \neq \mu_0$$

假设检验可能犯的两类错误

$$\alpha(C) = 2 - 2\Phi\left(\frac{C}{\sigma/\sqrt{n}}\right)$$ 关于C的单调递减的函数

$$\beta(C) = \Phi\left(\frac{\mu_0 + C - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu_0 - C - \mu}{\sigma/\sqrt{n}}\right) \quad \mu \neq \mu_0$$

关于C的单调递增的函数

给定样本容量为n的样本前提下，不可能找到C，使得$\alpha(C)$和$\beta(C)$都尽可能小

- $\alpha(C)$减小时，C增大，$\beta(C)$增大
- $\beta(C)$减小时，C减小，$\alpha(C)$增大

犯两类错误的概率相互制约！

假设检验的定义及基本概念

犯一、二类错误的危害（公认观点）

- 犯第一类错误（弃真）的危害较大，由于报告了本不存在的现象，则因此现象而衍生出的后续研究、应用的危害将是不可估量的。
- 相对而言，犯第二类错误（取伪）的危害则相对较小，因为研究者如果对自己的假设很有信心，可能会重新设计实验，再次来过，直到得到自己满意的结果。

例3. (犯第一类错误影响巨大) 我们检查一批灯泡的寿命，原假设认为灯泡寿命是合格的，但是我们抽查的时候恰好查到了寿命比较低的，这样就拒绝了原假设，认为这一批灯泡是不合格的。

那么结果是什么呢？就需要把这一批灯泡全部返厂重新检测、返修，甚至是销毁。

假设检验的定义及基本概念

例4. (<u>疑罪从无</u>) 判定一个嫌疑人是不是犯了罪，原假设就是这个人没有犯罪，那么犯一类错误就是认为嫌疑人有罪，而事实上没有犯罪，也就是被冤枉。犯二类错误就是把有罪的人判定成无罪。

例5. (<u>对两类错误的控制需视具体情况而定</u>) 比如非典期间，为尽量减少病原的传播，不惜花费大量的精力来减少犯第二类错误的概率。

对所有人进行体温测量，只要发现发烧立即隔离。

这里面肯定会有很多没有携带病原体的人被认为是携带者，其实这里就是加大了犯一类错误的概率，而尽量减少二类错误。因为犯二类错误的成本太大，宁可错误的隔离3000人，也不能放过一个携带者，因为放过一个就会造成非常严重的后果。

假设检验的定义及基本概念

犯两类错误的概率当然是越小越好，但当样本容量固定时，不可能同时把α、β都减得很小，而是减少其中一个，另一个就会增大；

要使α、β都很小，只有通过增大样本容量。在实际问题中，一般总是控制犯第一类错误的概率α.

Neyman-Pearson原则：保护原假设，即限制α的前提下，使β尽可能的小。

注意：“接受H_0”，并不意味着H_0一定为真；
“拒绝H_0”，也不意味着H_0一定不真。

假设检验的定义及基本概念

显著性水平的意义

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

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