第8章 假设检验

Chpt. 8 Hypothesis Testing

统计推断

观察某工厂生成的产品，在100个产品中，发现6个次品

1）次品率是多少？→ 点估计

2）次品率的范围是多少？→ 区间估计（给定置信度）

3）次品率不会超过多少？→ 单侧区间估计（给定置信度）

4）次品率是不是5%？

5）次品率是否超过6%？

6）次品率是否不低于5%？

假设检验

参数问题：当已知总体（随机变量）的分布形式已知时，如何确定其未知的参数？

方法之一：采用估计（点估计或区间估计）的方法。

分布问题：我们不知道一个总体（随机变量）的分布，那么如何通过实际观察或理论分析，确定总体分布形式？

问题分析：参数估计的方法不可用。但是，可以假设总体服从某种分布，而后根据抽取的样本观测值，对假设的正确性进行判断、检验。

8.1 假设检验

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

假设检验——根据抽取的样本观测值，构造适当的统计量，对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（二者回答不同的统计推断问题）

8.1 假设检验

[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$N(\mu, \sigma^2)$。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？

8.1 假设检验

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：

$$H_0\text{：}\ \mu = \mu_0 = 0.5$$

$$H_1\text{：}\ \mu \neq \mu_0$$

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。

8.1 假设检验

通常称H_0假设为原假设，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.

假设检验基本思想

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。

假设检验基本思想

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；

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。

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

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

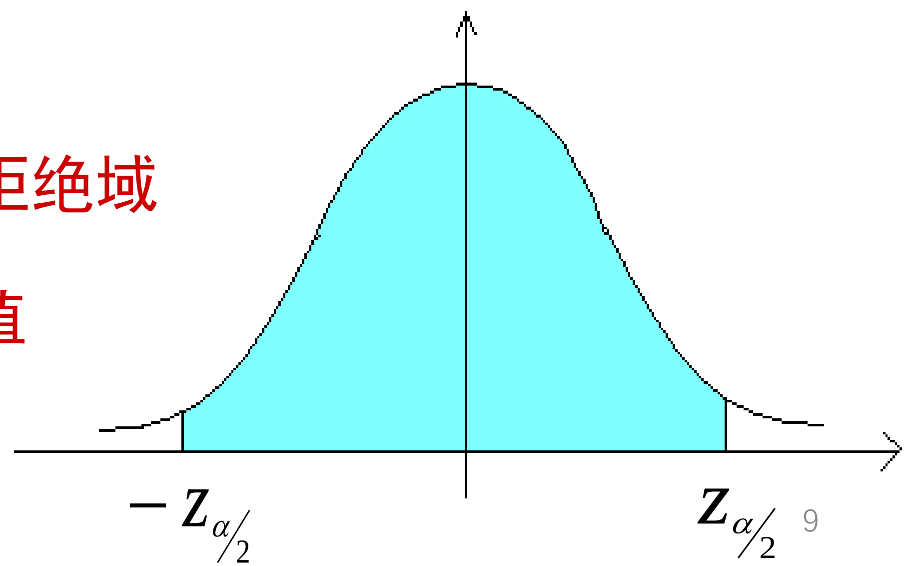
考虑一个小概率事件（取某个较小的正数α）

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

其中α称为显著性水平

$|Z| > z_{\alpha/2}$称为统计量Z的拒绝域

$z_{\alpha/2}$称为统计量 Z 的临界值

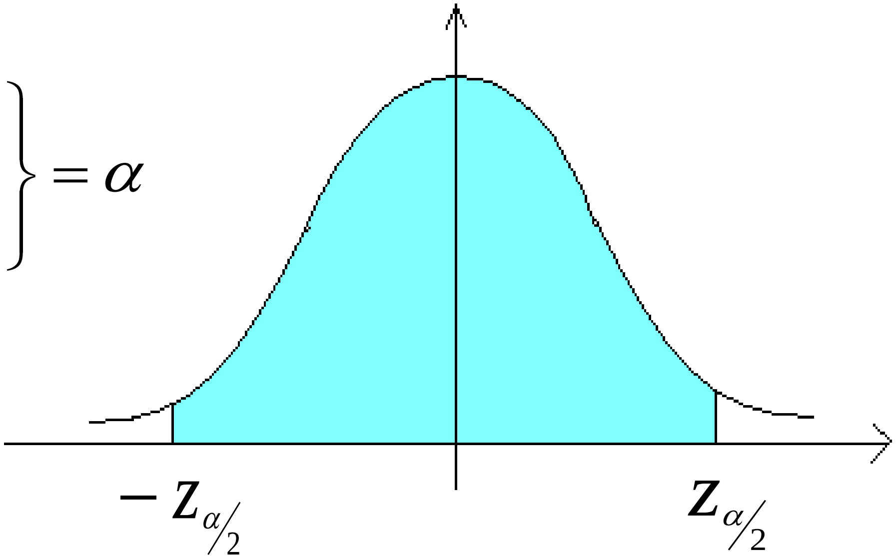


显著性水平的意义

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。

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



[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

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

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

为小概率事件，满足

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

小概率事件的实际不可能性原理

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

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

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，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，所以应当拒绝原假设H_0，接受备择假设H_1，即认为包装机今天工作不正常。

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 $P\left\{ |Z| > 2.58 \right\} = 0.01$

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，$\mu=0.5$（公斤）.

小概率事件的实际不可能性原理

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

双侧假设检验与单侧假设检验

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

这里备择假设H_1包含了$\mu<\mu_0$和$\mu>\mu_0$，因此称这类假设检验为双侧假设检验；

备择假设只包含只包含单侧区间时，这样的假设检验被称为单侧假设检验；

对应地，双侧假设检验的拒绝域也是双侧的，单侧假设检验的拒绝域是单侧的。

单侧假设检验

备择假设只包含只包含单侧区间时，这样的假设检验被称为

单侧假设检验

如：试验新工艺以提高材料的强度，这时所考虑的总体均值应该越大越好，如果我们能判断在新工艺下总体均值较以往正常生产的大，则可考虑采用新工艺，此时，我们需要如下假设检验：

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

上式称为右边检验。

有时，我们也需要左边检验，即：

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

[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 α＝0.05。

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。

单边检验的问题

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。

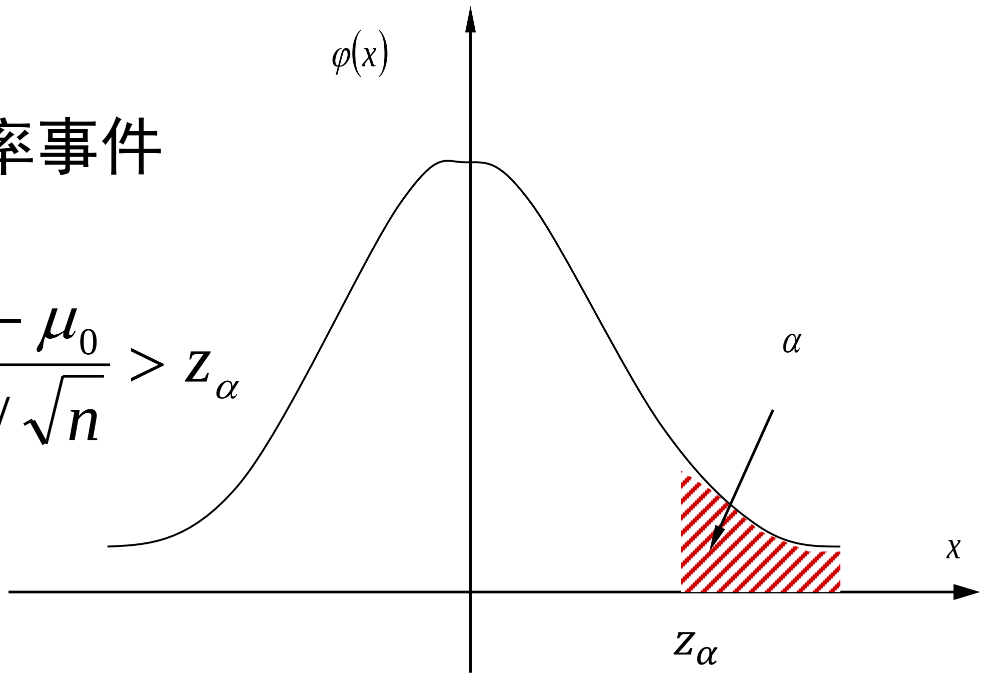
[1] 右边检验问题H_0：$\mu \le \mu_0$，H_1：$\mu > \mu_0$

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

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：$\mu \le \mu_0$



例2：燃烧率有无显著提高

现在 $n=25$，总体的标准方差$\sigma= 2cm/s$

样本均值为 $\bar{x}=41.25cm/s$

显著性水平$\alpha=0.05$。 $z_{\alpha}=1.65$

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

$$=3.125$$

$$>z_{\alpha}$$

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。

单边检验的问题

[2] 左边检测问题

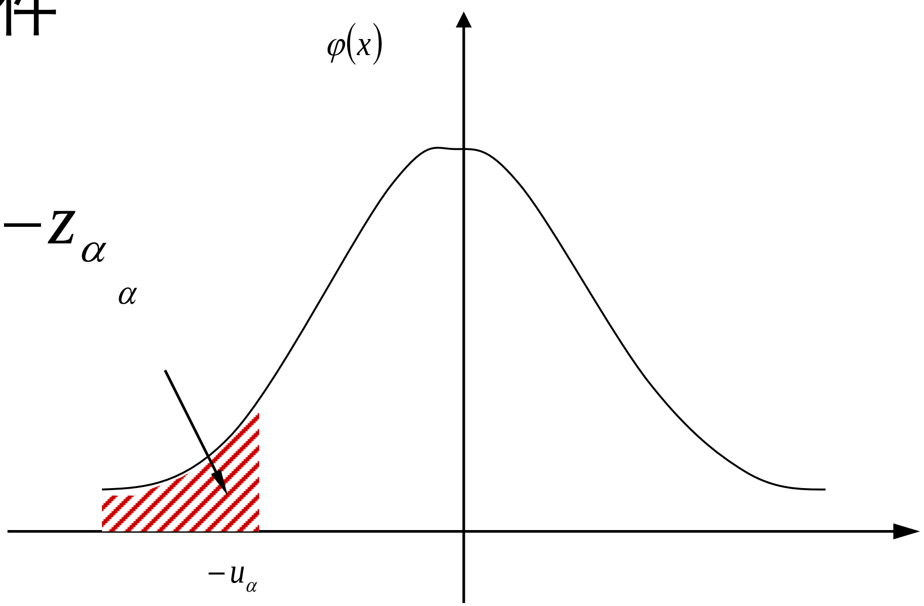
假设　　H_0：$\mu \geq \mu_0$，H_1：$\mu < \mu_0$

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 $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$

则拒绝 H_0：$\mu \geq \mu_0$



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.

假设检验可能犯的两类错误

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$P\{\overline{A}|H_1\} \le \beta$.

假设检验可能犯的两类错误

真实情况 检验决策	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) 在给定样本容量为n的条件下，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？

- 犯第一、二类错误，哪个的影响更大？

假设检验可能犯的两类错误

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

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

当$\dfrac{C}{\sigma/\sqrt{n}} = z_{\alpha/2}$时，$\alpha(C) = \alpha$

Nankai University

假设检验可能犯的两类错误

例1中，犯第二类错误的概率

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

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

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

因为 $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 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

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

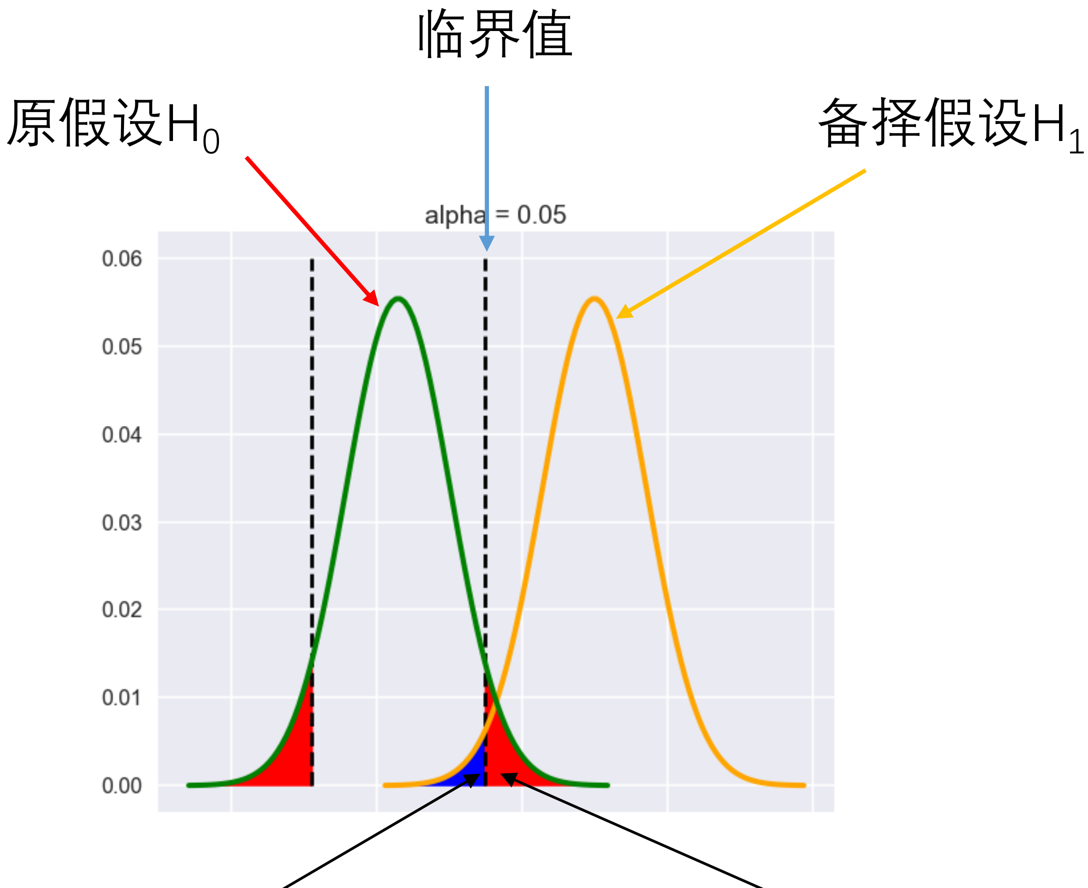
关于C的单调递增的函数

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！

假设检验可能犯的两类错误

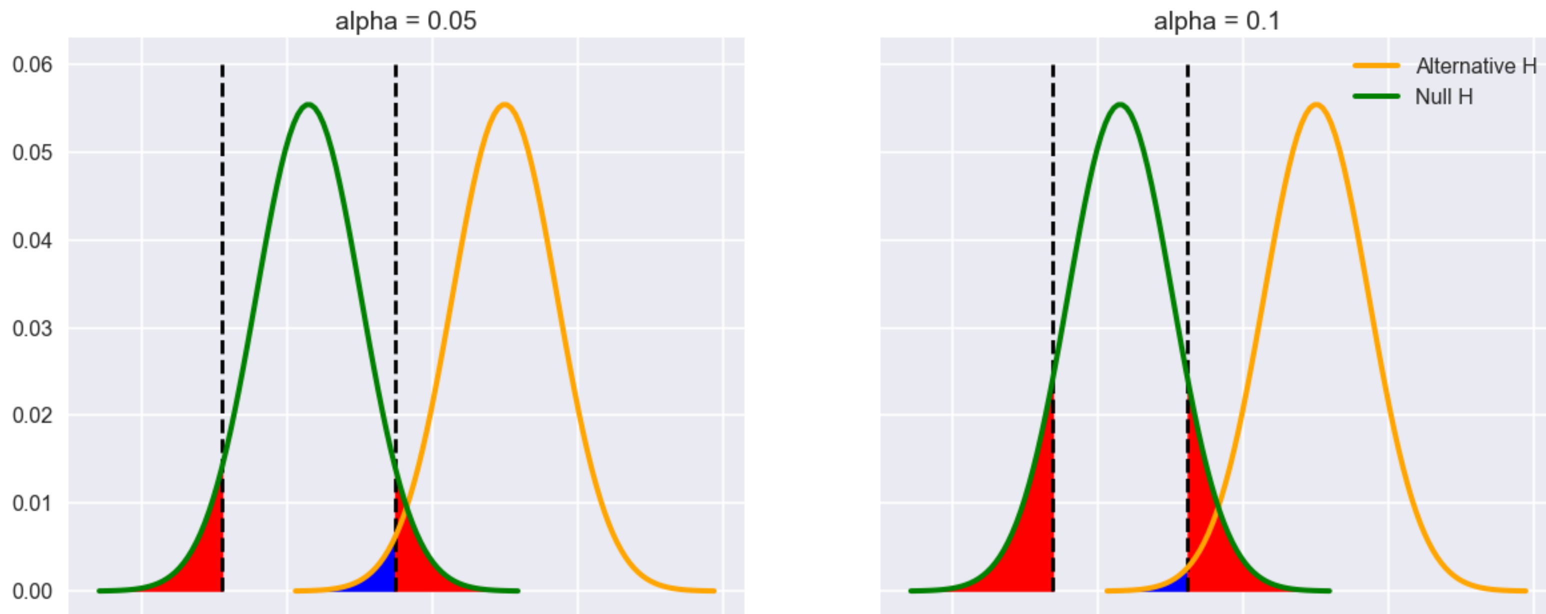


第二类错误：原假设为假，备择假设为真，但根据样本接受了原假设

第一类错误：原假设为真，但根据样本拒绝了原假设

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

假设检验可能犯的两类错误

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。

<u>假设检验可能犯的两类错误</u>

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，那么犯第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。犯第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>) 比如疫情期间，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第一类错误的概率，而尽量减少第二类错误。因为犯第二类错误的代价太大，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。

假设检验可能犯的两类错误

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。

8.2 正态总体均值的假设检验

(一) 关于单个正态总体$N(\mu, \sigma^2)$均值的假设检验

(1) σ^2已知，关于μ的检验 – Z检验

(2) σ^2未知，关于μ的检验 – t检验

(二) 关于两个正态总体均值差的假设检验

(1) 已知σ_1，σ_2

检验 H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

(2) 未知σ_1，σ_2，但是知道$\sigma_1=\sigma_2$

检验H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

回顾一下常用的统计量及其观测值：

(1) 样本均值 $\overline{X}=\dfrac{1}{n}\sum_{k=1}^{n}X_k$

(2) 样本方差 $S^2=\dfrac{1}{n-1}\sum_{k=1}^{n}(X_i-\overline{X})^2$

(3) 样本标准差 $S=\sqrt{S^2}=\sqrt{\dfrac{1}{n-1}\sum_{k=1}^{n}(X_i-\overline{X})^2}$

回顾一下几类抽样分布

单个正态总体

(1) 方差已知

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

(2) 方差未知

$$t=\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

两个正态总体

(1) 方差已知

$$Z=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

(2) 方差未知，但$\sigma_1=\sigma_2=\sigma$

$$T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

（一）关于单个正态总体N(μ，σ²)均值的假设检验

设总体X~ N(μ，σ²)，从中抽取容量为n的样本 $X_1, X_2, \cdots, X_n$

样本均值及样本方差分别为

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

考虑关于未知参数μ的假设检验.

(1) σ²已知，关于μ的检验－Z检验

采用统计量 $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ 来确定拒绝域

不同情况 $\left\{ \begin{array}{l} \text{确定关于}\mu\text{的不同的假设检验中的原假设}H_0\text{与备择假设}H_1 \\ \text{显著水平}\alpha\text{下关于原假设}H_0\text{的拒绝域} \\ \text{统计量的观测值落在拒绝域内的概率} \end{array} \right.$

（一）关于单个正态总体N(μ，σ²)均值的假设检验

原假设H_0	备择假设H_1	在显著水平α下关于原假设H_0的拒绝域	统计量的观测值落在拒绝域内的概率
$\mu = \mu_0$	$\mu \neq \mu_0$		
$\mu \leq \mu_0$	$\mu > \mu_0$		
$\mu \geq \mu_0$	$\mu < \mu_0$		

[1] 对于假设 H_0 ： $\mu = \mu_0$ H_1 ： $\mu \neq \mu_0$

我们用统计量Z，确定了相对于显著水平α的临界值，得到拒绝域 $|Z| > z_{\alpha/2}$

[2] 对于假设 H_0 ： $\mu \leq \mu_0$ $H_1 : \mu > \mu_0$

我们用统计量Z，确定了相对于显著水平α的临界值，得到拒绝域 $Z > z_{\alpha}$

[2] 对于假设 $H_0: \mu \le \mu_0 \quad H_1: \mu > \mu_0$

假设H_0正确，我们用统计量 $Z = \dfrac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$

相对于显著水平α 确定一个小概率事件

假设H_0正确，相当于H_1：$\mu > \mu_0$不成立

那么 $\bar{X}$ 的观察值 $\bar{x}$ 大于μ_0就是小概率事件

$$P\left\{\frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} > z_\alpha\right\} \le \alpha$$

拒绝域确定为

$$Z > z_\alpha$$

（一）关于单个正态总体N(μ，σ²)均值的假设检验

[3] 对于假设　$H_0\colon \mu \geq \mu_0$　$H_1\colon \mu < \mu_0$

我们用统计量Z，确定了相对于显著水平α的临界值，得到拒绝域 $Z < -z_\alpha$

原假设H_0	备择假设H_1	在显著水平α下关于原假设H_0的拒绝域	统计量的观测值落在拒绝域内的概率
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert Z\rvert > z_{\alpha/2}$	$P\{\lvert Z\rvert > z_{\alpha/2}\} = \alpha$
$\mu \leq \mu_0$	$\mu > \mu_0$	$Z > z_\alpha$	$P\{Z > z_\alpha\} \leq \alpha$
$\mu \geq \mu_0$	$\mu < \mu_0$	$Z < -z_\alpha$	$P\{Z < -z_\alpha\} \leq \alpha$

与备择假设方向一致

（一）关于单个正态总体N(μ，σ²)均值的假设检验

(2) σ^2未知，关于μ的检验 – t检验

在此前提下，关于的检验存在以下诸多情况

原假设H_0	备择假设H_1
$\mu = \mu_0$	$\mu \neq \mu_0$
$\mu \leq \mu_0$	$\mu > \mu_0$
$\mu \geq \mu_0$	$\mu < \mu_0$

(2) σ^2未知，关于μ的检验－t检验

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

给定显著水平为α；我们选择一个统计量，它应该与假设相关，但是不应包含σ。

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

假定H_0成立，构造一个小概率事件，来确定拒绝域。

H_0：$\mu \geq \mu_0$成立，那么 $\bar{X}$ 的观察值 $\bar{X}$ 小于μ_0就是小概率事件

$$P\left\{\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -t_\alpha\right\} \leq \alpha$$

确定拒绝域为 $t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} < -t_\alpha$

（一）关于单个正态总体N(μ，σ²)均值的假设检验

(2) σ²未知，关于μ的检验 – t检验

原假设H_0	备择假设H_1	在显著水平α下关于原假设H_0的拒绝域	统计量的观测值落在拒绝域内的概率
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t \rvert > t_{\alpha/2}$	$P\{\lvert t \rvert > t_{\alpha/2}\} = \alpha$
$\mu \le \mu_0$	$\mu > \mu_0$	$t > t_\alpha$	$P\{t > t_\alpha\} \le \alpha$
$\mu \ge \mu_0$	$\mu < \mu_0$	$t < -t_\alpha$	$P\{t < -t_\alpha\} \le \alpha$

例6：化肥厂用自动包装机包装化肥，某日测得9包化肥的重量(单位：kg) 如下：

49.7 49.8 50.3 50.5 49.7 50.1 49.9 50.5 50.4

设每包化肥的重量服从正态分布

问：是否可以认为每包化肥的平均重量为50 kg (取显著水平 $\alpha=0.05$)

解：设每包化肥的质量 $X\sim N(\mu, \sigma^2)$，可给出要检验的假设如下：

H_0：$\mu=\mu_0=50$；H_1：$\mu\neq\mu_0$.

因为未知σ，考虑统计量 $t=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$

其中$\mu_0=50$，$n=9$

则样本均值及修正样本标准差分别为

$$\bar{x} = 50.1, \qquad s \approx 0.335$$

统计量 t 的观测值 $t = \dfrac{50.1 - 50}{0.335 / \sqrt{9}} \approx 0.896$

查表得 $t_{\alpha/2}(n-1) = t_{0.025}(8) = 2.31$

因为 $|t| = 0.896 < t_{0.025}(8)$，所以接受原假设$H_0$

即在显著水平α=0.05下，可以认为每包化肥的平均质量

为50kg.

（二）关于两个正态总体均值的假设检验

有时，我们需要**比较两总体的参数（均值）是否存在显著差异**

- 两个农作物品种的产量
- 两种电子元件的使用寿命
- 两种加工工艺对产品质量的影响
- 两地区的气候差异等等

（二）关于两个正态总体均值的假设检验

总体$X \sim N(\mu_1, \sigma_1^2)$，从中抽取容量为$n_1$的样本$X_1, X_2, \cdots, X_{n_1}$，

总体 $Y \sim N(\mu_2, \sigma_2^2)$，从中抽取容量为$n_2$的样本 $Y_1, Y_2, \cdots, Y_{n_2}$

X, Y的样本均值及样本方差分别为 $\overline{X}$，S_1^2；$\overline{Y}$，S_2^2 。考虑

关于未知参数μ_1，μ_2，σ_1，σ_2的假设检验问题

先讨论均值μ_1、μ_2的假设检验

（二）关于两个正态总体均值的假设检验

(1) 已知σ_1，σ_2，检验 H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

针对以上假设，选取一个统计量，它涉及假设，但不含有未知参数。

使用统计量 $$Z=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

假设H_0正确，$\mu_1-\mu_2=\delta$，相对于显著水平α构造小概率事件

$$P\left(|Z|>z_{\alpha/2}\right)=\alpha$$

得到拒绝域为：$$|z|=\frac{\left|\overline{x}-\overline{y}-\delta\right|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\geq z_{\alpha/2}$$

为什么把假设做成如下形式：

H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

（1） $$Z=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

（2）这种形式覆盖了多种关于μ_1、μ_2的关系，如：

[1] $\delta=0$， $\mu_1=\mu_2$

[2] $\delta>0$， $\mu_1>\mu_2$

[3] $\delta<0$， $\mu_1<\mu_2$

南開大學 Nankai University

例7 据资料，已知某品种小麦每平方米产量(千克)的方差为 $\sigma^2 = 0.2$. 今在一块地上用A，B 两法，A法设12个样本点，得平均产量 $\bar{x} = 1.5$ ；B 法设8个样本点，得平均产量 $\bar{y} = 1.6$ ，试比较A、B两法的平均产量是否有显著差别。 $(\alpha = 0.05)$

解 假设： $H_0 : \mu_X = \mu_Y$， $H_1 : \mu_X \neq \mu_Y$

因为：

$$\left| \frac{\bar{x} - \bar{y}}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}} \right| = \left| \frac{1.5 - 1.6}{\sqrt{0.2/12 + 0.2/8}} \right| \approx 0.49 < 1.96 = z_{0.025}$$

所以接受H_0假设，即认为 A、B两法的平均产量无显著差别

（二）关于两个正态总体均值的假设检验

(2) 未知σ_1，σ_2，但是知道$\sigma_1=\sigma_2$

检验H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

使用统计量

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

（二）关于两个正态总体均值的假设检验

(2) 未知σ_1，σ_2，但是知道$\sigma_1 = \sigma_2$

检验H_0：$\mu_1 - \mu_2 = \delta$，H_1：$\mu_1 - \mu_2 \neq \delta$

提出关于μ_1，μ_2的不同的假设检验中的原假设H_0与备择假设H_1，在显著水平a下关于原假设H_0的拒绝域以及统计量的观测值落在拒绝域内的概率：

$$|t| = \frac{|\bar{x} - \bar{y} - \delta|}{s_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{\alpha/2}(n_1 + n_2 - 2)$$

南開大學 Nankai University

（二）关于两个正态总体均值的假设检验

两个正态总体均值差的单侧检验

左侧检验

$$H_0: \mu_1 - \mu_2 \geq \delta, \qquad H_1: \mu_1 - \mu_2 < \delta$$

右侧检验

$$H_0: \mu_1 - \mu_2 \leq \delta, \qquad H_1: \mu_1 - \mu_2 > \delta$$

和单个正态总体类似，可以分别得到两个正态总体均值差的左侧检验和右侧检验的拒绝域

[例8] 某灯泡厂在采用一种新工艺的前后，分别抽取10个灯泡进行寿命(单位：小时)检测，计算得到：

- 采用新工艺前，灯泡寿命的样本均值为2460，样本标准差为56；
- 采用新工艺后，灯泡寿命的样本均值为2550，样本标准差为48.

设灯泡的寿命服从正态分布

问题：是否可以认为采用新工艺后灯泡的平均寿命有显著提高(取显著水平α=0.01)?

解： 设采用新工艺前后的灯泡寿命分别为

$$X \sim N(\mu_1, \sigma_1^2) \qquad Y \sim N(\mu_2, \sigma_2^2)$$

因为未知σ_1，σ_2，暂时先假定$\sigma_1=\sigma_2$ (将在以后检验)，检验假设

$$H_0\text{：} \mu_1 \geq \mu_2 \qquad\qquad H_1\text{：} \mu_1 < \mu_2$$

取统计量

$$T = \bar{X} - \bar{Y} \Big/ S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sim t(n_1 + n_2 - 2)$$

其中$n_1=n_2=10$， $\bar{x} = 2460$， $\bar{y} = 2550$

$$s_1^2 = 56^2 \qquad s_2^2 = 48^2$$

$$s_w = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9\times 56^2+9\times 48^2}{20-2}} \approx 52.15$$

$$T \approx \frac{2460-2550}{52.15\sqrt{\frac{1}{10}+\frac{1}{10}}} \approx -3.86$$

查表得

$$t_\alpha(n_1+n_2-2) = t_{0.01}(18) = 2.55$$

因为$T=-3.86<-t_{0.01}(18)=-2.55$，所以拒绝原假设$H_0$，接受备择假设$H_1$

即在显著水平α=0.01下，认为采用新工艺后灯泡的平均寿命有显著提高.

例：(<u>成对数据检验</u>) 为了比较两种谷物种子的优劣，特选取10块土质不全相同的土地，并将每块土地分为面积相同的两部分，分别种植这两种种子，施肥与田间管理在20小块土地上都是一样的（假定两种种子产量的方差相等），下面是各小块上的单位产量：

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，试问：两种种子的平均单位产量在显著性水平$\alpha = 0.05$上有无显著性差异？

解：假定$X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$，且X，Y独立，这里假定两个总体的方差是相等的，因此可以应用两个总体的均值差的t检验来分析

记两种种子单位产量的样本均值为$\overline{X},\overline{Y}$，样本方差为$S_1^2,S_2^2$

需要做如下检验问题：

$$H_0: \mu_1-\mu_2=0,\ H_1: \mu_1-\mu_2\neq 0$$

取统计量

$$T=\overline{X}-\overline{Y}\Big/ S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sim t(n_1+n_2-2)$$

其中$n_1 = n_2 = n = 10$

$\overline{x} = 33.1, \overline{y} = 35.7, S_1^2 = 33.2110, S_2^2 = 14.2333$

$$s_w = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{s_1^2 + s_2^2}{2}} \approx 4.8705$$

所以统计量T的观测值 $t = \frac{\overline{x} - \overline{y}}{s_w\sqrt{\frac{2}{10}}} = -1.1937$

拒绝域为：$|T| \geq t_{\alpha/2}(2n - 2) = t_{0.025}(18) = 2.1009$

因为$|t| < 2.1009$，故不应该拒绝原假设，即认为两种种子的单位产量平均值没有显著差别。

我们换一个角度来讨论这个问题，本题中出现了成对数据，同一块土地上用两种种子的2个产量，其差$d_i = x_i - y_i$(i = 1, ⋯ , 10)排除了土质差异这个不可控因素的影响，主要反映两种种子的优劣。

我们知道 $D = X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

记$\mu = \mu_1 - \mu_2, \sigma^2 = \sigma_1^2 + \sigma_2^2$，我们要做的检验问题如下

$$H_0: \mu = 0, \qquad H_1: \mu \neq 0$$

即解决一个单样本t检验问题，其拒绝域为

$$|T'| \geq t_{\alpha/2}(n-1) = t_{0.025}(9) = 2.2622$$

本题中 $n = 10, \overline{d} = -2.6, s_d = 3.5024$

可计算检验统计量$t' = \frac{\overline{d} - \mu_0}{s_d/\sqrt{n}} = -2.3475$

因为$|t| > 2.2622$，故应该拒绝原假设，即认为两种种子的单位产量平均值有显著差别

进一步，平均单位产量差的估计量$\hat{\mu} = \overline{x} - \overline{y} = -2.6$，可见种子y要比种子x的平均单位产量高

两种方案得到完全不同的结论，那种方案更加合理？为什么？

第二种方案更合理

因为成对数据的差d_i消除了试验单元（如土质）之间的差异，从而用于检验的标准差$s_d = 3.5024$已排除土质差异的影响，只保留了种子之间的差异；

而使用两样本的t检验所用的$s_w = 4.8705$还含有土质差异，从而使得标准差变大，导致要检验的因子不显著。

注1：一般而言，成对数据场合下，化为单样本的t检验所作的结论更可靠。

注2：如果上述问题中的10块土地的土质完全一样，即参加比较的试验单元完全一样，则用两个样本t检验会更好一些。

因为二样本t检验可提供更多的自由度去估计误差。

注3：成对数据的获取事先要做周密的试验设计，保证在获得成对数据时不能发生数据“错位”，从而获得准确的成对数据。

8.3　正态总体方差的假设检验

（一）单个正态总体标准差的假设检验

（二）两个正态总体标准差的假设检验

8.4　非正态总体参数的假设检验

8.5　分布律的假设检验

回顾一下几类抽样分布

χ^2分布

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

认为均值μ未知，关于方差σ^2检验的统计量选此

F分布

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

8.3 正态总体方差的假设检验

（一）单个正态总体标准差的假设检验

设总体X ~ N(μ，σ²)，σ未知，$X_1, X_2, \cdots, X_n$ 是来自X的样本，要求对σ做出各种检验。

假设　H_0：$\sigma = \sigma_0$　　　H_1：$\sigma \neq \sigma_0$

按照前面的步骤，选择蕴含假设， 但是不包含未知参数的统计量。

假设　$H_0: \sigma = \sigma_0$　$H_1: \sigma \neq \sigma_0$

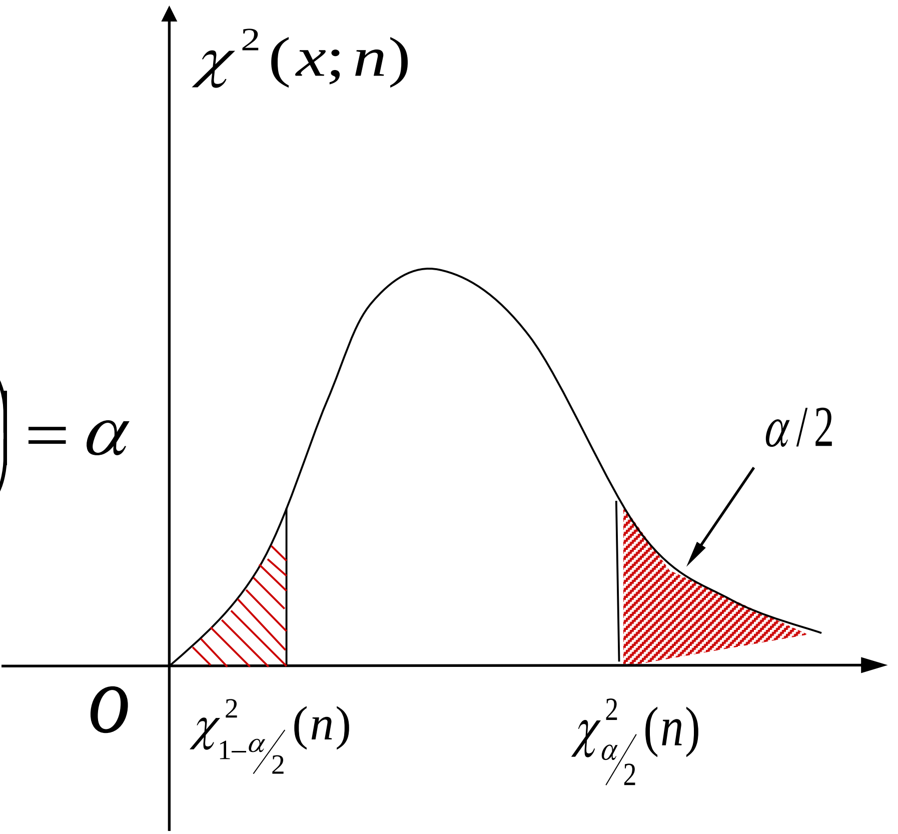
如果承认H_0，那么 $\dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$ ，其对应的小概率事件可以考虑为接近原点及远离原点时的区间。

对于给定的显著水平α，取两个临界点 $\chi^2_{1-\alpha/2}$ ，$\chi^2_{\alpha/2}$

$$P\left(\chi^2 < \chi^2_{1-\alpha/2}\right) + P\left(\chi^2 > \chi^2_{\alpha/2}\right) = \alpha$$

拒绝域为：

$$\chi^2 < \chi^2_{1-\alpha/2} \text{ 或 } \chi^2 > \chi^2_{\alpha/2}$$



原假设H_0	备择假设H_1	在显著水平α下关于原假设H_0的拒绝域	统计量的观测值落在拒绝域内的概率
$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\chi^2 < \chi^2_{1-\alpha/2}$ $\chi^2 > \chi^2_{\alpha/2}$	$P\left\{ \chi^2 < \chi^2_{1-\alpha/2} \right\}+$ $P\left\{ \chi^2 > \chi^2_{\alpha/2} \right\}=\alpha$
$\sigma \le \sigma_0$	$\sigma > \sigma_0$	$\chi^2 > \chi^2_{\alpha}$	$P\left\{\chi^2 > \chi^2_{\alpha}\right\} \le \alpha$
$\sigma \ge \sigma_0$	$\sigma < \sigma_0$	$\chi^2 < \chi^2_{1-\alpha}$	$P\{\chi^2 < \chi^2_{1-\alpha}\} \le \alpha$

例9：自动车床加工的某种零件的直径(单位：mm) 服从正态分布 N(μ，σ^2)

原来的加工精度 $\sigma^2 \le 0.09$，经过一段时间后，要检验是否保持原来的加工精度，为此，从该车床加工的零件中抽取 30 个，测得数据如下：

零件直径	9.2	9.4	9.6	9.8	10.0	10.2	10.4	10.6	10.8
频 数	1	1	3	6	7	5	4	2	1

问：现在的加工精度是否显著变差 (取显著水平 α=0.05) ?

解：要检验的假设是: H_0：$\sigma \le \sigma_0 = 0.3$; H_1：$\sigma > \sigma_0 = 0.3$

因未知μ，考虑统计量 $\dfrac{(n-1)S^2}{\sigma_0^2} \boxed{\sim \chi^2(n-1)}$。

因为$\sigma \le \sigma_0$，所以检验统计量不服从χ^2分布

已知 $\sigma_0^2 = 0.09$，$n=30$，且样本方差 $s^2 \approx 0.1344$

由此得统计量χ^2的观测值 $\chi^2 = \dfrac{(30-1)\times 0.1344}{0.09} \approx 43.3$

查表得 $\chi_\alpha^2(n-1) = \chi_{0.05}^2(29) = 42.6$

因 $\chi^2 = 43.3 > \chi_{0.05}^2(29) = 42.6$

所以拒绝原假设 H_0：$\sigma \le \sigma_0 = 0.3$，接受备择假设$H_1$：$\sigma > \sigma_0 = 0.3$

即在显著水平$\alpha = 0.05$下，认为该车床的加工精度变差了。

（二）两个正态总体标准差的假设检验

总体 $X \sim N(\mu_1, \sigma_1^2)$，从中抽取$n_1$个样本 $X_1, X_2, \cdots, X_{n_1}$

总体$Y \sim N(\mu_2, \sigma_2^2)$，从中抽取$n_2$个样本 $Y_1, Y_2, \cdots, Y_{n_2}$

X，Y的样本均值及样本方差分别为 $\overline{X}, S_1^2; \overline{Y}, S_2^2$

问题：考虑μ_1，μ_2，σ_1，σ_2未知时，对于方差的假设检验：

$$H_0\text{：}\ \sigma_1^2 \le \sigma_2^2 \qquad H_1\text{：}\ \sigma_1^2 > \sigma_2^2$$

注意要体现σ_1与σ_2关系，不能使用 $S_1^2 \big/ S_2^2 \sim F(n_1 - 1, n_2 - 1)$

改用 $$\frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

（二）两个正态总体标准差的假设检验--F 检验法

假设H_0成立，寻求一个小概率事件，进一步根据抽样值进行判断。这是一个比较特殊的，取上述分布的α分位点 F_α

注意到S_1^2, S_2^2是σ_1^2, σ_2^2的无偏估计，当H_0为真时 $\dfrac{\sigma_2^2}{\sigma_1^2} \geq 1$

$$\left\{\frac{S_1^2}{S_2^2} \geq F_\alpha\right\} \subset \left\{\frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \geq F_\alpha\right\}$$

判断 $\dfrac{s_1^2}{s_2^2} \geq F_\alpha$成立否，如成立，那必然有 $\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \geq F_\alpha$

$$P\left\{\frac{S_1^2}{S_2^2} \geq F_\alpha\right\} < P\left\{\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \geq F_\alpha\right\} = \alpha$$

即小概率事件发生了，就可以否定 H_0：$\sigma_1^2 \leq \sigma_2^2$

例10：测得两批电子器材的样本的电阻为（单位：Ω）

第一批：0.140 0.138 0.143 0.142 0.144 0.137

第二批：0.135 0.140 0.142 0.136 0.138 0.140

设这两批器材的电阻均服从正态分布

问：这两批电子器件电阻的方差是否存在显著差异（显著度 $\alpha=0.05$）

解 这是一个两正态总体的方差检验问题，用 F 检验法

假设 $H_0:\sigma_1^2=\sigma_2^2$; $H_1:\sigma_1^2\neq\sigma_2^2$ 由样本观测数据得

$S_1^2=0.0028^2; S_2^2=0.0027^2$ $F=1.108$ 而 $F_{\alpha/2}(5,\ 5)=7.15$

$F_{1-\alpha/2}(5,\ 5)=\dfrac{1}{7.15}\approx 0.14$ 得 $F_{1-\alpha/2}(5,\ 5)<F<F_{\alpha/2}(5,\ 5)$

所以，接受原假设，即可认为两批电子器材的方差相等

例11：对甲、乙两种玉米进行评比试验，得如下产量资料：

甲：	951	966	1008	1082	983
乙：	730	864	742	774	990

问：这两种玉米的产量差异有没有统计意义？$(\alpha = 0.05)$

解：理解这个问题，就是要检验两种玉米的产量均值是否有差异

此处，可以假设玉米产量服从正态分布

甲：$X \sim N(\mu_1,\ \sigma_1^2)$

乙：$Y \sim N(\mu_2,\ \sigma_2^2)$

检验玉米产量的差异可以有多种方法：

[1] 直接估计方差，而后直接检验 $\mu_1 = \mu_2$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

该统计量中的未知数是?

$\hat{\sigma}_1^2 = S_1^2, \hat{\sigma}_2^2 = S_2^2$

[2] 也可以先看看二者的方差是否相等 $\sigma_1 = \sigma_2$ ，如相等则可以用另外的检验

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

我们下面用第二种方法，先检验二者的方差是否相等 $\sigma_1 = \sigma_2$

如相等再检验二者均值是否有差异 $\mu_1 = \mu_2$

解：先对方差作检验：$H_{10}: \sigma_1 = \sigma_2$， $H_{11}: \sigma_1 \neq \sigma_2$

$$\bar{x} = 998.0, \quad \boxed{s_1^2 = 51.5^2}; \qquad \bar{y} = 820.0, \quad \boxed{s_2^2 = 108.6^2}$$

$$F = \frac{s_1^2}{s_2^2} = \frac{51.5^2}{108.6^2} \approx 0.225$$

$$F_{\alpha/2}(4,\ 4) = 9.60$$

$$F_{1-\alpha/2}(4,\ 4) = \frac{1}{9.60}$$

因为 $F_{1-\alpha/2}(4,\ 4) < F < F_{\alpha/2}(4,\ 4)$ ，所以可认为甲、乙两种玉米的方差没有显著差异，即可认为 $\sigma_1 = \sigma_2$ 。

对于方差的检验并不严格，实际应用较少

再对均值作检验：

$$H_{20}: \mu_1 = \mu_2, \quad H_{21}: \mu_1 \neq \mu_2$$

前面已检验方差相等，故用 T 检验，自由度为5＋5－2＝8

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$\overline{x} = 998.0, \quad s_1^2 = 51.5^2; \qquad \overline{y} = 820.0, \quad s_2^2 = 108.6^2$$

$$|T| = \frac{|\, 998 - 820 \,|}{\sqrt{\frac{51.5^2 \times 4 + 108.6^2 \times 4}{8}} \sqrt{\frac{1}{5} + \frac{1}{5}}} \approx 3.31 > 2.306 = t_{0.025}(8)$$

所以拒绝原假设 H_0。即认为两种玉米的产量差异有统计意义。

8.4 非正态总体参数的假设检验——大样本法

总体X 服从某一分布(不是正态分布)，概率分布或概率密度中含有未知参数θ，则总体均值及方差都依赖于参数θ，即 $E(X)=\mu(\theta)$，$D(X)=\sigma^2(\theta)$.

从总体中抽取样本 $X_1, X_2, \cdots, X_n$，它们是相互独立，且与总体 X 服从相同分布。

在原假设 H_0：$\theta=\theta_0$ 成立的条件下，当样本容量充分大(一般要求$n\geq 50$) 时，统计量 $Z=\dfrac{\bar{X}-\mu(\theta_0)}{\sigma(\theta_0)/\sqrt{n}}$ 近似地服从标准正态分布$N(0, 1)$

8.4 非正态总体参数的假设检验－－大样本法

检验假设 H_0： $\theta=\theta_0$； H_1： $\theta\neq\theta_0$

对给定的显著水平α，确定 $z_{\alpha/2}$ 使

$$P\left\{\, |Z|\geq z_{\alpha/2} \right\}=P\left\{\, \left|\frac{\bar{X}-\mu(\theta_0)}{\sigma(\theta_0)/\sqrt{n}}\right|\geq z_{\alpha/2} \right\}\approx\alpha$$

若由样本观测值 $x_1,x_2,\cdots,x_n$ 算得Z的观测值的绝对值大于$z_{\alpha/2}$，

则拒绝原假设 H_0： $\theta=\theta_0$，接受备择假设 H_1： $\theta\neq\theta_0$ 。

否则，接受 H_0： $\theta=\theta_0$。

即在显著水平下α，原假设的H_0： $\theta=\theta_0$ 拒绝域为 $(z_{\alpha/2},\infty)$

或$(-\infty,-z_{\alpha/2})$。

例12 一枚硬币掷495次，结果为：出现正面220次；出现反面275次.
问：这枚硬币是否均匀（显著水平α=0.05下）

解：设 $X=\begin{cases}1, & \text{出现正面}\\ 0, & \text{出现反面}\end{cases}$，$X_i=\begin{cases}1, & \text{第}i\text{次出现正面}\\ 0, & \text{第}i\text{次出现反面}\end{cases}$ $i=1,\cdots,495$

则 $X_1, X_2, \cdots, X_{495}$ 是总体X的样本，X~B(1,*p*)，其中*p*为未知参数，且

$$P\{X=x\}=\begin{cases}1-p & x=0\\ p & x=1\end{cases} \qquad E(X)=p, \qquad D(X)=p(1-p)$$

检验这枚硬币的均匀性，即检验假设

$$H_0: \ p=p_0=0.5, \quad H_1: \ p\neq p_0$$

由于样本容量 n=495 足够大，因此在原假设H_0成立的条件下，统计量 $Z=\dfrac{\overline{X}-\mu(\theta_0)}{\sigma(\theta_0)/\sqrt{n}}=\dfrac{\overline{X}-p_0}{\sqrt{p_0(1-p_0)/n}}$ 近似地服从标准正态分布 $N(0,1)$ 。

抽样检查的结果表明，在495个样本观测值中恰有220个为1，

其余为0，所以样本均值为 $\overline{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\dfrac{220}{495}=0.4444$

而 $|z|=\dfrac{|0.4444-0.5|}{\sqrt{0.5\times 0.5/495}}=2.427$ $\quad z_{\alpha/2}=z_{0.025}=1.96$

所以得知 $|z|=2.427>1.96$，小概率事件发生了。

因此，在显著水平α=0.05下拒绝原假设。

练习：某建筑公司宣称其麾下建筑工地平均每天发生事故的次数不超过0.6起，现记录了该公司麾下建筑工地200天的安全生产情况如下，试检验该建筑公司的宣称是否成立（取$\alpha = 0/5$）

注：一天内发生事故的次数可认为服从泊松分布

一天发生的事故数	0	1	2	3	4	5	≥ 6	合计
天数	102	59	30	8	0	1	0	200

例13 (指数分布的假设检验)设我们要检验某种元件的平均寿命是否小于6000h ($\alpha = 0.05$)，假定原件寿命服从指数分布，现取5个原件，观测到如下5个失效时间(单位：h):

$$395, 4094, 119, 11572, 6133$$

样本数很少，无法应用中心极限定理

解：指数分布密度函数形式如下

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, x > 0 \\ 0, x \leq 0 \end{cases}$$

总体X服从指数分布也可以写成$X \sim \Gamma(1, \frac{1}{\theta})$

我们已经证明过$\frac{2n\bar{X}}{\theta} \sim \chi^2(2n)$（第六章作业）

于是可以用$\frac{2n\bar{X}}{\theta}$作为检验统计量并利用$\chi^2(2n)$的分位数建立检验的拒绝域

总体均值$E(X)=\theta$，而样本均值$\bar{x}=4462.6$

令$\theta_0=6000$，本题检验的假设为：

$$H_0\colon \theta \geq \theta_0, \qquad H_1\colon \theta < \theta_0$$

取检验统计量为$\frac{2n\bar{X}}{\theta_0}$，前面讨论过$\chi^2$检验，可知拒绝域为

$$\frac{2n\bar{X}}{\theta_0} < \chi^2_{1-\alpha}(2n) = \chi^2_{0.95}(10) = 3.94$$

再计算得到检验统计量的观测值为

$$\frac{2n\bar{X}}{\theta_0} = 7.4377 > 3.94$$

故接受原假设，可以认为这种元件的平均寿命不低于6000h.

上述检验方法在实际应用中可能会有什么问题？

考虑一下抽样的过程：拿n个元件同时开始使用，到其全部失效时试验才能停止；这n个元件中难免有几个寿命特长的，这样一来，就必须等待很长时间才能结束试验。

为免除这个不便，实际工作常采用截尾法：

（1） 定数截尾法

（2） 定时截尾法

（1） 定数截尾法

取n个元件做试验观察，定下一个自然数$r < n$，试验进行到有r个元件失效时为止，全部n个元件的工作时间加起来记为T，即

$$T = Y_1 + \cdots + Y_r + (n-r)Y_r$$

这里Y_1是最先失效的元件的寿命，Y_2为第二个失效的元件的寿命，以此类推，Y_r就是第r个失效的元件的寿命，试验也到此为止。余下还有$(n-r)$个元件，它们已工作的总时间为$(n-r)Y_r$。

不难理解，T越大，越使我们相信元件的平均寿命大；反之，如果T越小，我们越不太相信元件的寿命大。

对于如下检验：

$$H_0: \theta \geq \theta_0, \qquad H_1: \theta < \theta_0$$

当H_0成立时，θ比θ_0大，即一般而言，$\bar{X}$应该比θ_0大

可以证明：$\frac{2T}{\theta} \sim \chi^2(2r)$

仿照前面的推理，在给定的显著性水平α下，可得到如下拒绝域：

$$\frac{2T}{\theta_0} < \chi^2_{1-\alpha}(2r)$$

例：要检验某种元件平均寿命不小于5000小时，可以作如下假设检验：

$$H_0: \theta \geq \theta_0, \qquad H_1: \theta < \theta_0$$

其中$\theta_0 = 5000$，取15个元件做试验，预定到第5个失效时停止试验，于是$n = 15, r = 5$，若前5个失效元件的工作时间一次是800，1200，1500，2000，2200，则

$$T = 800 + 1200 + 1500 + 2000 + 2200 + 10 \times 2200 = 27500$$

取$\alpha = 0.05$，$\chi^2_{0.95}(10) = 3.94$

而$\frac{2T}{\theta_0} = 11 > 3.94$，故应接收原假设，认为这批元件的平均寿命不小于5000小时

（2） 定时截尾法

指定一个时刻T_0，抽取n个元件做试验，直到时刻T_0为止，把到这时为止全部n个元件的工作时间加起来记为T。显然，平均寿命越大，T越倾向于较大的值。

可以证明：$\frac{2T}{\theta} \sim \chi^2(2u+1)$，其中$u$是到$T_0$时失效的元件个数；

对于如下检验：

$$H_0: \theta \geq \theta_0, \qquad H_1: \theta < \theta_0$$

可以得到拒绝域为：$\frac{2T}{\theta_0} < \chi^2_{1-\alpha}(2u+1)$

例如，仍然取$\theta_0 = 5000$，$\alpha = 0.05$，取10个元件做试验，T_0定为1000；假设到T_0时，已有5个元件失效，时刻分别为100, 150, 230, 500, 580，则

$$T = 100 + 150 + 230 + 500 + 580 + 5 \times 1000 = 6560$$

此处$u = 5$，故临界值为$\chi^2_{0.95}(11) = 4.575$

而$\frac{2T}{\theta_0} = 2.624 < 4.575$，故应该拒绝原假设，认为元件寿命不到5000小时。

二项分布参数p的检验

设某事件的概率为p，p未知，做n次独立试验，每次观察该事件是否发生，记X为该事件发生的次数，则$X \sim B(n, p)$。要根据X去检验以下假设：

（1） $H_0: p \le p_0$，$H_1: p > p_0$

（2） $H_0: p \ge p_0$，$H_1: p < p_0$

（3） $H_0: p = p_0$，$H_1: p \ne p_0$

以第一个假设检验问题为例，从直观上看，我们检验的规则是

φ：当$X \le K$时接受H_0，反之就拒绝H_0

φ： 当$X \leq K$时接受H_0，反之就拒绝H_0

给定显著性水平α，上述检验就相当于要找到整数K，使得

$$\sum_{i=0}^{K} C_n^i {p_0}^i (1 - p_0)^{n-i} = 1 - \alpha$$

很难找到一个K刚好满足上式，但是可以找到K_0，满足

$$\sum_{i=0}^{K_0} C_n^i {p_0}^i (1 - p_0)^{n-i} < 1 - \alpha < \sum_{i=0}^{K_0+1} C_n^i {p_0}^i (1 - p_0)^{n-i}$$

这时，我们只好取K_0或者$K_0 + 1$

当n充分大时， $C_n^{K_0+1} {p_0}^{K_0+1} (1 - p_0)^{n-K_0-1}$一般比较小，不影响实际应用场景

同理，我们可以得到问题（2）（3）的解决方法：

（2） $H_0: p \geq p_0$，$H_1: p < p_0$

φ：当$X \geq K$时接受H_0，反之就拒绝H_0。

其中，K由$\sum_{i=0}^{K-1} C_n^i {p_0}^i (1-p_0\ )^{n-i} = \alpha$确定

（3） $H_0: p = p_0$，$H_1: p \neq p_0$

φ：当$K_1 \leq X \leq K_2$时接受H_0，反之就拒绝H_0。

其中K_1，K_2分别由下面两个等式确定：

$$\sum_{i=0}^{K_1-1} C_n^i {p_0}^i (1-p_0\ )^{n-i} = \frac{\alpha}{2}$$

$$\sum_{K_2+1}^{n} C_n^i {p_0}^i (1-p_0\ )^{n-i} = \frac{\alpha}{2}$$

例：一工厂向商店供货，商店要求废品率不超过$p = 0.05$，经双方同意，制定抽样方案：每批（假定批量很大）抽$n = 24$件，检查其中废品个数X，当$X \leq K$时，商店接受该批产品，否则就拒收。双方约定检验水平为$\alpha = 0.05$（这意味着废品率为0.05的批次有95%的概率可通过检查，若$p < 0.05$，通过的可能性更高）。问如何决定K。

解：$H_0: p \leq p_0 = 0.05,\ H_1: p > p_0 = 0.05$

令$I = \sum_{i=0}^{K} C_n^i {p_0}^i (1 - p_0)^{n-i} = 1 - \alpha$

计算可得：$K = 2$时，$I = 0.884; K = 3$时，$I = 0.970$

若取$K=2$，对商店有利；取$K=3$，对工厂有利

照数字看，$K=3$时，I与0.95接近，取$K=3$更合理，但如果商店不同意，一定要坚持0.95这个数，可以按照如下方式处理：

a) 当$X\le 2$时，接受这批产品

b) 当$X\ge 4$时，拒收这批产品

c) 当$X=3$时，随机决定是否接受，接受的概率是

$$r=\frac{0.95-0.884}{0.97-0.884}=0.767$$

这样的检验，称为**随机化检验**，其中包含了一个随机机制去决定原假设是否被接受。在涉及离散型分布的检验中，如果要坚持约定的显著性水平，往往需要随机化检验。

8.5 分布律的假设检验

背景：关于总体参数的假设检验是以总体分布形式已知为前提的。

但实际问题，有时不能预先知道总体分布的形式。

解决方法：用假设检验的方法，根据样本的观察值判断总体是否具有某种分布

这类对总体分布形式的检验问题就是分布律的假设检验，

称为分布拟合检验。

8.5 分布律的假设检验

利用总体X的样本观测值可以求出其统计分布，若选用某一理论分布去拟合，则无论理论分布如何选取，一般来说，它与统计分布之间总存在某些差异。

问题：

这些差异仅仅是试验次数有限导致的随机性的差异，还是所选择的理论分布与统计分布之间存在的实质性的差异呢?

为解决这个问题，数理统计中有几种不同的拟合检验法，在此仅讨论最常用的皮尔逊χ^2准则.

8.5 分布律的假设检验

设总体 X 的分布未知，从总体中抽取一个容量为 n 的样本 $X_1, X_2, \cdots, X_n$，观察值 $x_1, x_2, \cdots, x_n$，检验总体分布是否等于某确定的分布$F_0(x)$时的 χ^2 检验分下面四个步骤进行。

（1）分布假设

$$H_0: F(x) = F_0(x); \qquad H_1: F(x) \neq F_0(x).$$

要求当H_0为真时，$F(x)=F_0(x)$的形式及参数都是已知的。但实际上参数值往往是未知的。这时，需要先用参数估计法（如矩估计法，极大似然估计法）来求出参数的估计。

（2）由样本构造相应的统计量

在实数轴上选取 k-1 个分点 $t_1, t_2, \cdots, t_{k-1}$, 将数轴分成 k 个互不相交的区间 $S_i = (t_{i-1}, t_i]$, $i = 1, 2, \cdots, k$ ，其中 $t_0 = -\infty$, $t_k = +\infty$ ，H_0为真时，记p_i为总体 X 落在 $S_i = (t_{i-1}, t_i]$ 内的概率:

$$p_1 = P(X \le t_1) = F_0(t_1)$$

$$p_2 = P(t_1 < X \le t_2) = F_0(t_2) - F_0(t_1)$$

$$\vdots$$

$$p_i = P(t_{i-1} < X \le t_i) = F_0(t_i) - F_0(t_{i-1})$$

$$\vdots$$

$$p_k = P(t_{k-1} < X \le +\infty) = 1 - F_0(t_{k-1})$$

（2）由样本构造相应的统计量

记m_i为 n 个样本值中落入的 $S_i=(t_{i-1},t_i]$个数，即组频数。

显然有 $\sum_{i=1}^{k} m_i = n$ 。

以上的目的是把X分布区间划分成多个小区间，而后判断一下样本落入此区间的频率与此区间概率的比，当H_0为真时，频率应该趋于概率。

由频率的稳定性可知，在H_0为真的条件下，$\left|\frac{m_i}{n}-p_i\right|$ 的值很小。

作统计量 $\chi^2=\sum_{i=1}^{k}\frac{(m_i-np_i)^2}{np_i}$，称为$\chi^2$统计量。

皮尔逊证明了：则当$n\to\infty$时，不论总体属于什么分布，统计量χ^2的分布趋于自由度为 $k-r-1$ 的 χ^2 分布：

$$\chi^2=\sum_{i=1}^{k}\frac{(m_i-np_i)^2}{np_i}\sim\chi^2(k-r-1)$$

区间	频数	频率	概率
$(t_0,t_1]$	m_1	f_1	p_1
$(t_1,t_2]$	m_2	f_2	p_2
⋮	⋮	⋮	⋮
$(t_{k-1},t_k]$	m_k	f_k	p_k
总计	n	1	1

其中

k 为所分区间的个数

r 为被估计参数的个数

（3）对于给定的显著水平α，确定自由度为 $(k\text{-}r\text{-}1)$ 的 χ_{α}^{2} 值，使

$$P(\chi^2 \geq \chi_{\alpha}^{2}) = \alpha$$

（4）由样本观察值计算出χ^2的值。

若由样本观测值计算得差异度χ^2的值大于 χ_{α}^{2} ，则拒绝原假设H_0，否则接受H_0.

Remark 1: 利用皮尔逊 χ^2 准则检验总体分布的假设时，要求样本容量 n 及观测值落在各个区间内的频数 m_i 都相当大，一般要求

$np_i \geq 5$ (教材第200页)

$$n \geq 50, \qquad m_i \geq 5, \ (i=1,2,\ldots l)$$

若某些区间内的频数太小，则应把相邻的若干个区间（行或列）合并，使合并后的区间内的频数足够大，但注意必须相应减少差异度χ^2的自由度. (之所以如此要求：大样本才可能成立大数定律；m_i足够大时，m_i/n才有意义)

Remark 2: 当假设的理论分布中含有未知参数时，一般应利用极大似然估计法求这些参数的估计值

例13 纺织厂检查某种布匹每 100 m^2 内的疵点数，下表为检验 75 块布 (每块100 m^2) 所得的疵点数样本记录：

疵点数x_i	0	1	2	3	4	5	≥6
频数m_i	20	30	15	7	2	1	0

问：每块布的疵点数X是否服从泊松分布 $\pi(\lambda)$ $(\lambda>0)$ (取显著水平$\alpha=0.05$)?

解：检验假设 H_0：$X\sim\pi(\lambda)$；H_1：X不服从泊松分布$\pi(\lambda)$

先求出原假设H_0成立时λ的极大似然估计值

$$\hat{\lambda}=\bar{x}=\frac{1}{75}\left(0\times20+1\times30+2\times15+3\times7+4\times2+5\times1\right)=1.25$$

再按X~ π (1.25)计算概率

$$P_i = P\{i \le X < i+1\} = P\{X = i\} = \frac{1.25^i}{i!} e^{-1.25}, \quad i = 0, \cdots, 4$$

$$P_5 = P\{5 \le X < +\infty\} = \sum_{i=5}^{+\infty} \frac{1.25^i}{i!} e^{-1.25} = 1 - \sum_{i=0}^{4} \frac{1.25^i}{i!} e^{-1.25}$$

然后计算出 np_i 及 $(m_i - np_i)^2 / np_i$，$i = 0, \cdots, 5.$

列表如下

<table>
<tr><th>x_i</th><th>频数m_i</th><th>p_i</th><th>np_i</th><th>$(m_i - np_i)^2 / np_i$</th></tr>
<tr><td>0</td><td>20</td><td>0.287</td><td>21.5</td><td>0.10</td></tr>
<tr><td>1</td><td>30</td><td>0.358</td><td>26.9</td><td>0.36</td></tr>
<tr><td>2</td><td>15</td><td>0.224</td><td>16.8</td><td>0.19</td></tr>
<tr><td>3</td><td rowspan="3">$\left.\begin{matrix}7\\2\\1\end{matrix}\right\}10$</td><td rowspan="3">$\left.\begin{matrix}0.094\\0.029\\0.008\end{matrix}\right\}0.131$</td><td rowspan="3">$\left.\begin{matrix}7.1\\2.1\\0.6\end{matrix}\right\}9.9$</td><td rowspan="3">0.01</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td></tr>
<tr><td>≥6</td><td>0</td><td></td><td></td><td></td></tr>
</table>

得差异度χ^2=0.66，合并后区间数 $l = 4$， $\pi(\lambda)$未知参数个数r =1

χ^2自由度 $l = k - r - 1 = 2$，查表得 $\chi^2_{0.05}(2) = 5.99$

因 $\chi^2 = 0.66 < \chi^2_{0.05}(2) = 5.99$

$X \sim \pi(1.25)$?

接受原假设H_0，即认为 $X \sim \pi(\lambda)$。

例14　测量100个某种机械零件的质量(单位：g)，统计如下表

零件质量区间	频数	零件质量区间	频数
236.5~239.5	1	251.5~254.5	22
239.5~242.5	5	254.5~257.5	11
242.5~245.5	9	257.5~260.5	6
245.5~248.5	19	260.5~263.5	1
248.5~251.5	24	263.5~266.5	2

问：这种机械零件的质量是否服从正态分布 (取显著水平$\alpha=0.05$)?

解：依题意，检验假设

H_0：$X\sim N(\mu, \sigma^2)$，H_1：X不服从正态分布$N(\mu, \sigma^2)$.

解： 已知 n=100，把各个区间的中点值取作x_i，分别计算参数μ，σ^2的极大似然估计值：

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 250.6 \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 26.82$$

$\hat{\sigma} = 5.18$，计算 $X \sim N(250.6,\ 5.18^2)$ 时，X 落在各个区间的概率

$$p_i = P\{\alpha_{i-1} < X < \alpha_i\} = \Phi\left(\frac{\alpha_i - 250.6}{5.18}\right) - \Phi\left(\frac{\alpha_{i-1} - 250.6}{5.18}\right)$$

$$i = 1,2,\cdots,10$$

其中Φ(X) 为标准正态分布的分布函数，随机变量的取值为区间(-∞，+∞) 应把第一个区间扩大为(-∞，239.5)，把最后一个区间扩大为(263.5，+∞) 把检验所需数据整理并列表如下：

零件质量区间	m_i	p_i	np_i	$(m_i - np_i)^2 / np_i$
(-∞，239.5)	1	0.0166	1.66	
239.5~242.5	5	0.0428	4.28	
242.5~245.5	9	0.1041	10.41	
245.5~248.5	19	0.1774	17.74	
248.5~251.5	24	0.2266	22.66	
251.5~254.5	22	0.2059	20.59	
254.5~257.5	11	0.1348	13.48	
257.5~260.5	6	0.0637	6.37	
260.5~263.5	1	0.0117	1.17	
(263.5，+∞)	2	0.0064	0.64	
总计	100	1	100	

零件质量区间	m_i	p_i	np_i	$(m_i-np_i)^2/np_i$
(-∞，239.5)	6	0.0594	5.94	0.001
239.5~242.5				
242.5~245.5	9	0.1041	10.41	0.191
245.5~248.5	19	0.1774	17.74	0.089
248.5~251.5	24	0.2266	22.66	0.079
251.5~254.5	22	0.2059	20.59	0.097
254.5~257.5	11	0.1348	13.48	0.456
257.5~260.5	9	0.0918	9.18	0.004
260.5~263.5				
(263.5，+∞)				
总计	100	1	100	0.917

由此得差异度$\chi^2=0.917$

因为合并后区间数为$k=7$，需要估计的参数个数 $r=2$，即χ^2

自由度 $l=k-r-1=4$

查表得 $\chi^2_{0.05}(4)=9.49$

因为 $\chi^2=0.917<\chi^2_{0.05}(2)=9.49$ ，所以接受原假设H_0

即在显著水平a=0.05下，可以认为这种机械零件的质量服从正态分布

例15：自1965年1月1日至1971年2月9日共2231天中，全世界记录到里氏震级4级以上地震共162次，统计如下：

相继两次地震间隔天数	0~4	5~9	10~14	15~19	20~24	25~29	30~34	35~39	≥ 40
出现的频数	50	31	26	17	10	8	6	6	8

试检验两次地震间隔天数X服从指数分布(显著水平α＝0.05)

解：根据题意，需要检验假设　H_0：X的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & x\leq 0\end{cases}$$

此处H_0中的参数θ未给出，先由最大似然估计法求的θ的估计值　$\hat{\theta}=\bar{x}=\dfrac{2231}{162}=13.77$

在H_0为真的情况下，X的分布函数的估计为

$$\hat{F}(x)=\begin{cases}1-e^{-x/13.77} & x>0\\ 0 & x\le 0\end{cases}$$

在H_0下，X可能取值的全体是[0，∞)。参照所给出的统计数据，将区间[0，∞)分为k＝9个互不重叠的小区间

$$A_1=[0,\ 4.5]$$

$$A_2=(4.5,\ 9.5]$$

$$A_3=(9.5,\ 14.5]$$

$$\cdots\cdots\cdots$$

$$A_8=(34.5,\ 39.5]$$

$$A_9=(39.5,\ \infty)$$

获得概率 $p_i = P(A_i) = P\{a_i < X \le a_{i+1}\} = \hat{F}(a_{i+1}) - \hat{F}(a_i)$

A_i	m_i	$\hat{p}_i$	$n\hat{p}_i$	$m_i^2 / n\hat{p}_i$
A_1: $0 \le x \le 4.5$	50	0.2788	45.1656	55.3519
A_2: $4.5 < x \le 9.5$	31	0.2196	35.5752	27.0132
A_3: $9.5 < x \le 14.5$	26	0.1527	24.7374	27.3270
A_4: $14.5 < x \le 19.5$	17	0.1062	17.2044	16.7980
A_5: $19.5 < x \le 24.5$	20	0.0739	11.9718	8.3530
A_6: $24.5 < x \le 29.5$	8	0.0514	8.3268	7.6860
A_7: $29.5 < x \le 34.5$	6	0.0358	5.7996	6.2073
A_8: $34.5 < x \le 39.5$	6	0.0248	4.0176	14.8269
A_9: $39.5 < x < \infty$	8	0.0568	9.2016	Σ=163.5633

对于显著水平α＝0.05，自由度为$k - r - 1$的分位点为

$$\chi^2_{0.05}(k-r-1)=\chi^2_{0.05}(8-1-1)$$

$$=12.592$$

现在

$$\chi^2=163.5633-162=1.5633$$

$$\chi^2_{0.05}(6)=12.592>\chi^2=1.5633$$

故在水平α＝0.05下，接受H_0，认为 X 服从指数分布

8.6 假设检验问题的p值法

以上讨论的假设检验方法称为临界值法

- 临界值法在应用中可能会带来一些麻烦，若主张不同的显著性水平，会得到矛盾的结论，不便于处理
- 不能直观的知道拒绝和接受的可靠程度

显著性水平α	拒绝域	对应的结论($Z=-2.25$)
0.1	$Z \le -1.282$	拒绝H_0
0.05	$Z \le -1.645$	拒绝H_0
0.025	$Z \le -1.96$	拒绝H_0
0.01	$Z \le -2.326$	接受H_0
0.005	$Z \le -2.576$	接受H_0

8.6 假设检验问题的p值法

现代计算机统计软件系统中，一般都给出检验问题的p值

假设检验问题的p值（probability value）是原假设成立时，检验统计量取比观查到的结果更为极端的数值的概率。

[定义] 在一个假设检验问题中，利用样本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。

例16. 设总体$X \sim N(\mu, \sigma^2)$，μ未知， $\sigma^2 = 100$，现有52个样本，其均值为$\bar{x} = 62.75$。现在做如下假设检验

$$H_0: \mu \leq \mu_0 = 60，H_1: \mu > \mu_0$$

采用Z检验法，检验统计量为$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

带入数据，得Z的观测值为$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 1.983$

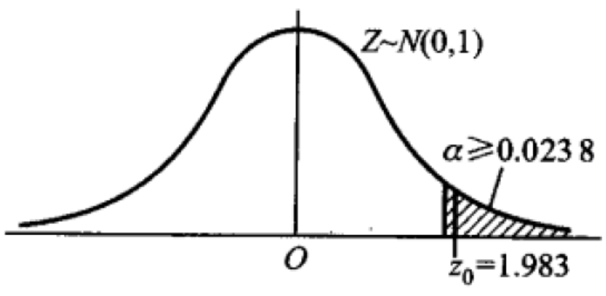
概率

p值 = $P\{Z \geq z_0\} = 1 - \Phi(1.983) = 0.0238$

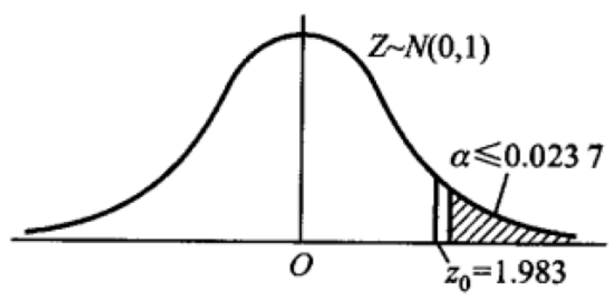
$$p\text{值} = P\{Z \geq z_0\} = 1 - \Phi(1.983) = 0.0238$$

1) 若显著性水平$\alpha \geq p = 0.0238$，则对应的临界值$z_\alpha \leq 1.983$，则表示观测值落在拒绝域内，因而拒绝H_0；

2) 若显著性水平$\alpha < p = 0.0238$，则对应的临界值$z_\alpha > 1.983$，则表示观测值不落在拒绝域内，因而接受H_0。



(1)



(2)

因此，p值 $= P\{Z \geq z_0\} = 0.0238$是原假设$H_0$可被拒绝的最小显著性水平

在上例中，如果σ^2未知，则采用t检验，对应的检验统计量为

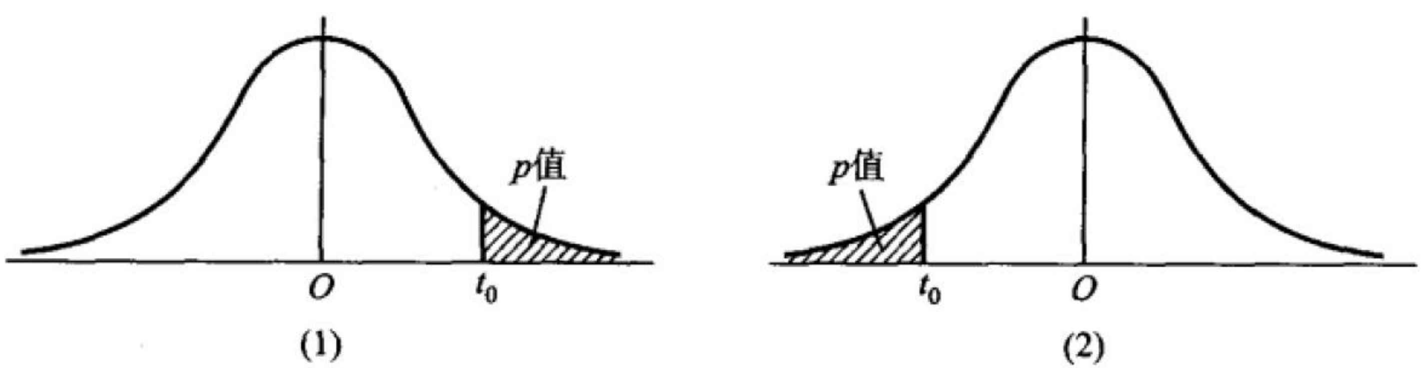
$$T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$$

- $H_0: \mu \le \mu_0$，$H_1: \mu > \mu_0$

  p值 $=P\{T \ge t_0\}=t_0$右侧尾部面积

- $H_0: \mu \ge \mu_0$，$H_1: \mu < \mu_0$

  p值 $=P\{T \le t_0\}=t_0$左侧尾部面积



- $H_0: \mu = \mu_0$，$H_1: \mu \neq \mu_0$
    - 当$t_0 > 0$时，

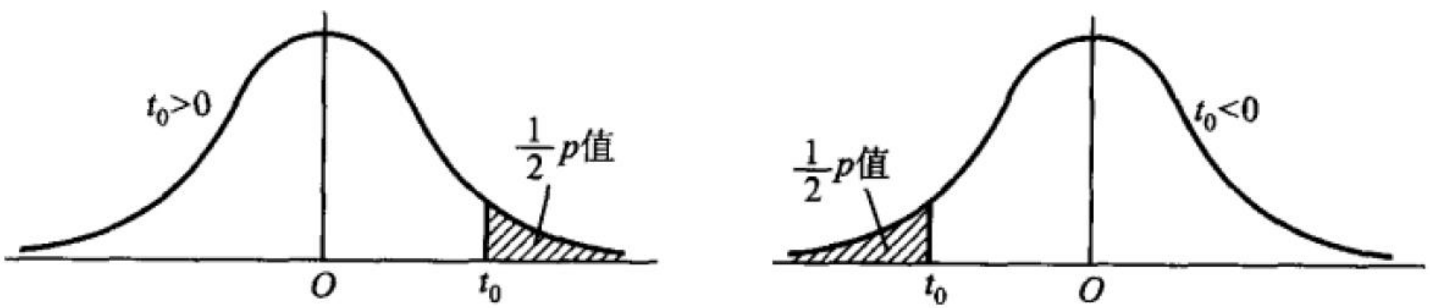
    p值$= P\{|T| \geq t_0\} = P\{T \geq t_0 \cup T \leq -t_0\}$

    $= 2 \times t_0$右侧尾部面积

    - 当$t_0 < 0$时，

    p值$= P\{|T| \geq -t_0\} = P\{T \geq -t_0 \cup T \leq t_0\}$

    $= 2 \times t_0$左侧尾部面积



(1) (2)

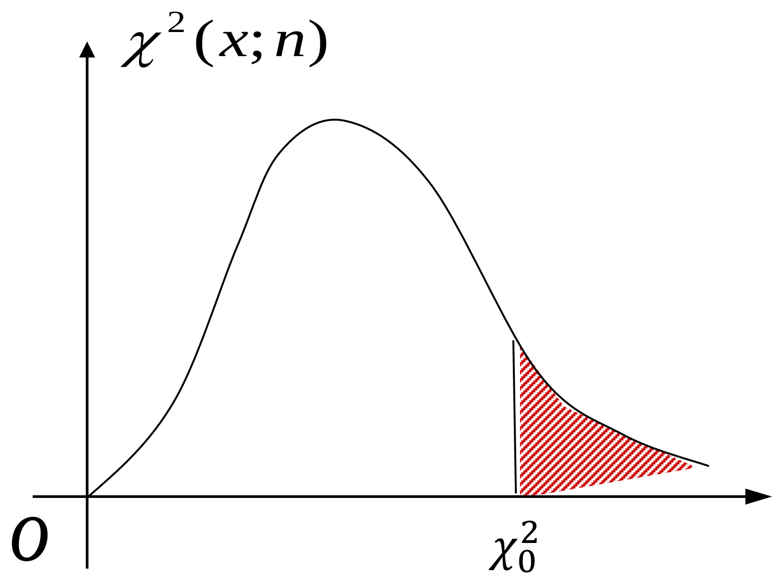
对于不对称的抽样分布，做如下检验

$$H_0: \sigma = \sigma_0, \qquad H_1: \sigma \neq \sigma_0,$$

检验统计量$\frac{(n-1)S^2}{{\sigma_0}^2} \sim \chi^2(n-1)$

可根据样本算得统计量的观测值为χ_0^2

p值 $= 2 \times \min\{P\{\chi^2 \le \chi_0^2\}, P\{\chi^2 \ge \chi_0^2\}\}$



Summary 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(蕴含假设，但又不能还有其他未知参数），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.

（一）关于单个正态总体N(μ，σ²) 参数的假设检验

(1) σ^2已知，关于μ的检验－－ Z 检验

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

(2) σ^2未知，关于μ的检验－－ t 检验

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

(3) 关于σ^2的检验－－ χ^2 检验

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

（二）关于两个正态总体参数的假设检验

(1) 已知σ_1、σ_2，检验 H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

$$Z=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

(2) 未知σ_1，σ_2，但是知道$\sigma_1=\sigma_2$

检验H_0：$\mu_1-\mu_2=\delta$，H_1：$\mu_1-\mu_2\neq\delta$

$$T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

(3)未知μ_1，μ_2，σ_1，σ_2，检验：H_0：$\sigma_1^2\leq\sigma_2^2$ H_1：$\sigma_1^2>\sigma_2^2$

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

（三）关于非正态总体的假设检验

总体X 服从某一分布(不是正态)，概率分布含有未知参数

θ， E(X)＝μ(θ)，D(X)＝σ^2(θ).

检验假设 H_0：θ＝θ_0 H_1：θ≠θ_0

样本容量充分大(一般要求n≥50) 时，统计量 $Z=\dfrac{\overline{X}-\mu(\theta_0)}{\sigma(\theta_0)/\sqrt{n}}$

近似地服从标准正态分布N(0，1)，类似Z检验进行。

（四）分布律的假设检验

检验假设

$$H_0: F(x) = F_0(x); \qquad H_1: F(x) \neq F_0(x).$$

皮尔逊χ^2准则.

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} \sim \chi^2(k-r-1)$$

谢谢！