假设检验

本章主要内容

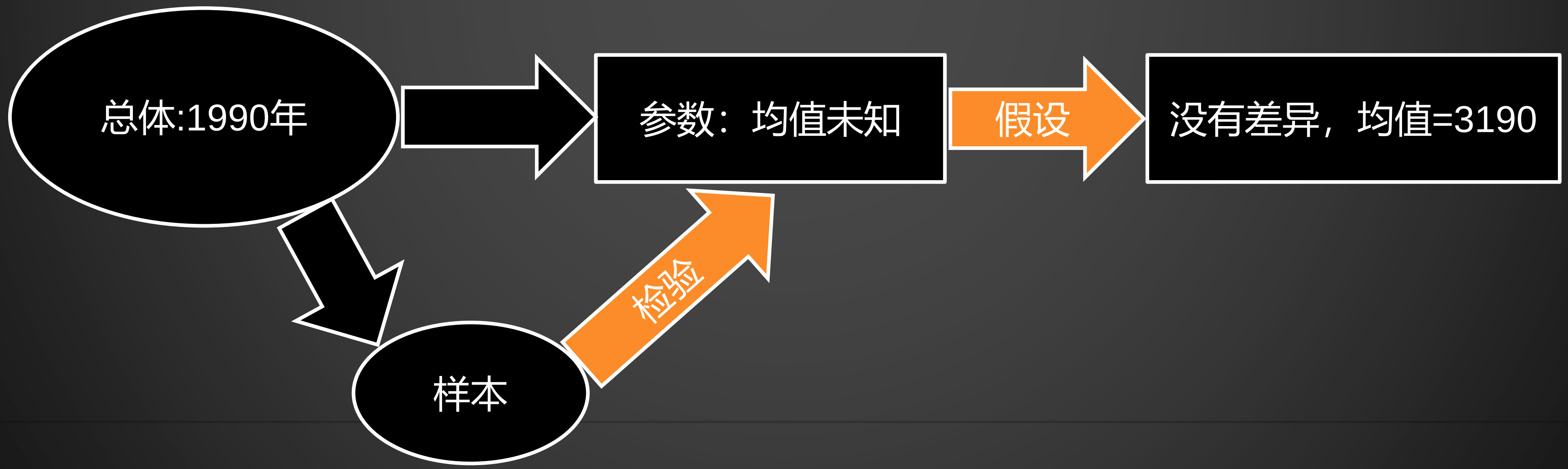
1. 假设检验的基本原理
2. 一个总体参数的检验
3. 两个总体参数的检验
4. 卡方检验

本章主要内容

1. 假设检验的基本原理
2. 一个总体参数的检验
3. 两个总体参数的检验
4. 卡方检验

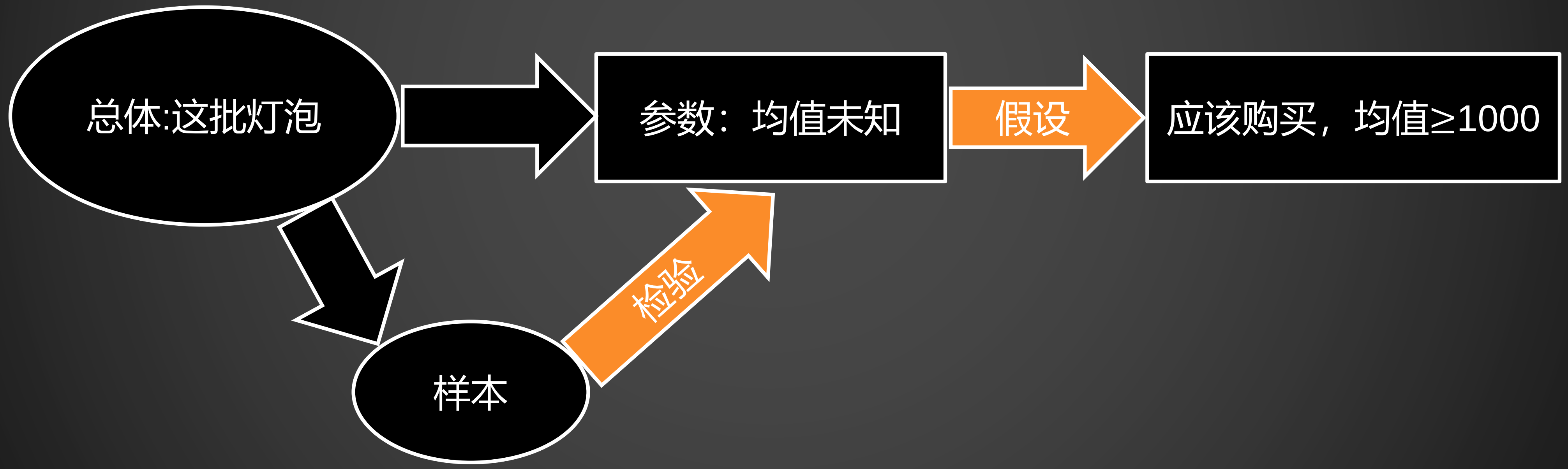
假设检验的基本原理

- 假设检验，hypothesis testing，是推断统计的一项重要内容，也就是说是一个根据样本去推断总体的过程。
- 例如，由统计资料得知，1989年某地新生儿平均体重为3190克，标准差为80克。现从1990年新生儿中随机抽取100个，测得其平均体重为3210克，问1990年的新生儿与1989年相比，体重有无显著差异?



假设检验的基本原理

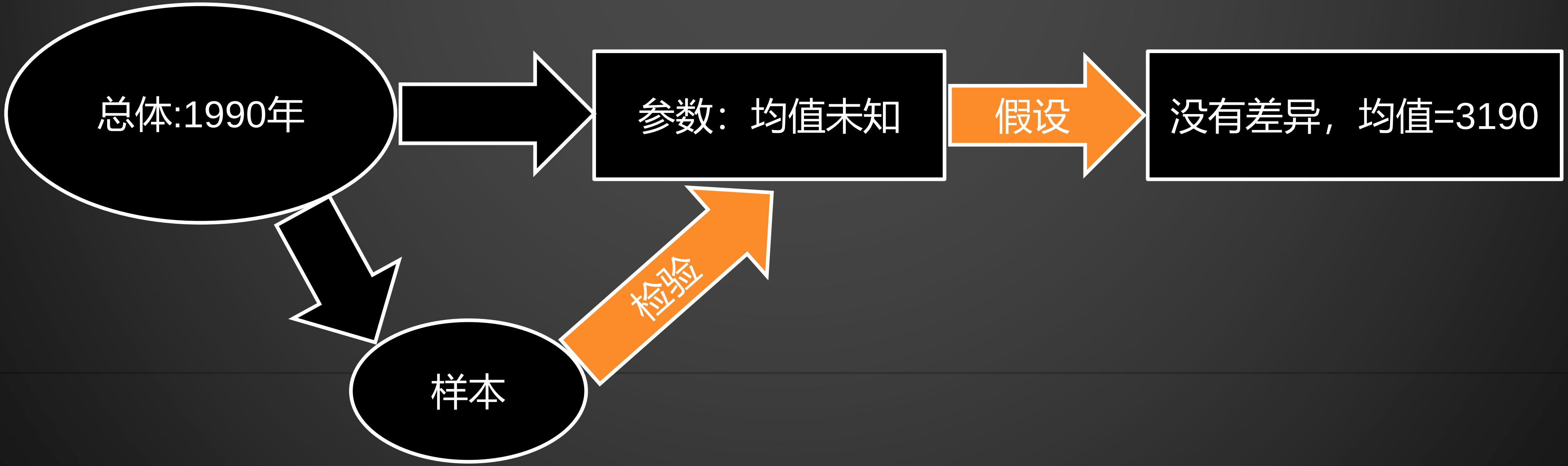
- 例如，某批发商欲从厂家购进一批灯泡,根据合同规定灯泡的使用寿命平均不能低于1000小时。已知灯泡燃烧寿命服从正态分布，标准差为200小时。在总体中随机抽取了100个灯泡，得知样本均值为960小时，问批发商是否应该购买这批灯泡?



假设检验的基本原理

【例】由统计资料得知，1989年某地新生儿平均体重为3190克，标准差为80克。现从1990年新生儿中随机抽取100个，测得其平均体重为3210克，问1990年的新生儿与1989年相比，体重有无显著差异。

分析：从调查结果看，1990年的新生儿的平均体重比1989年新生儿的平均体重增加了20g，但是这20g的差异可能是由于抽样的随机性造成的，也可能是1990年新生儿体重确实比1989年新生儿体重有所增加。



假设的表达式

- 为了用统计学的语言来表示上述信息，假设1990年新生儿的平均体重用μ来表示，1989年新生儿的平均体重用μ_0表示， $\mu_0 = 3190$。
- 假设1989年和1990年新生儿的体重没有显著差异，可以表示为$\mu = \mu_0$，称之为原假设（null hypothesis），表示为$H_0: \mu = 3190(g)$
- 如果原假设不成立，则需要在另一个假设中作出选择，另一个假设称为备择假设（alternative hypothesis），表示为$H_1: \mu \neq 3190(g)$

说明:

1. H_0表示原假设（null hypothesis），这里用0作为下标，所以有些文献中也称为“零假设”。
2. 由于假设检验是围绕着对原假设是否成立而展开的，所以有些文献也把备择假设称为替换假设，从英文名称也可以看出，alternative hypothesis，alternative就是可替换的意思。

确定检验统计量

已知新生儿的体重X服从正态分布，则$X \sim N(\mu, \sigma^2)$，根据前面所学的抽样分布的知识知道，样本的平均体重$\bar{X}$服从正态分布，均值为μ，方差为σ^2/n，即

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

假设原假设为真，即$\mu = 3190$，为了方便计算，将样本均值$\bar{X}$转化为标准正态分布，即

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

又已知标准差$\sigma = 80$，则样本均值$\bar{x} = 3210$在标准正态分布下的z值为

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3210 - 3190}{80/\sqrt{100}} = 2.5$$

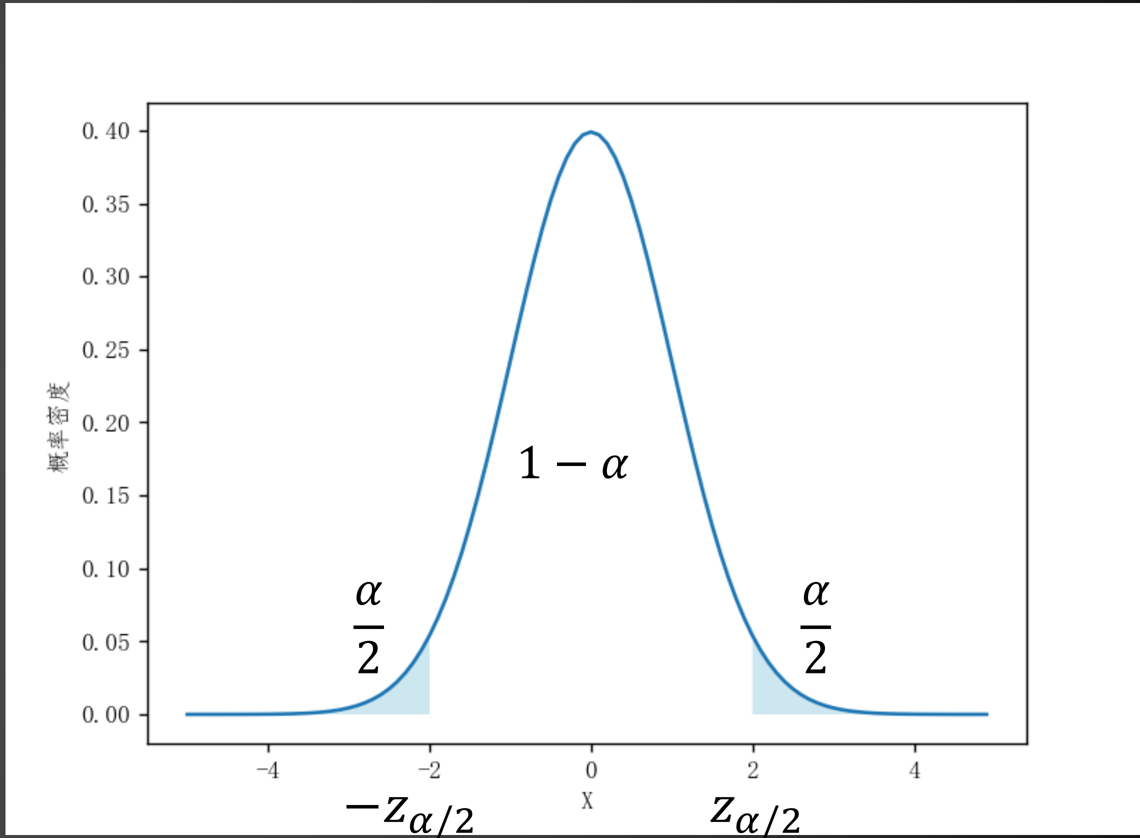
进行统计决策

规定一个很小的概率$\alpha = 0.05$，如右图所示，左右两侧各占$\alpha/2$，而$z_{\alpha/2} = 1.96$。

从图中可以看出$z = 2.5$落在阴影部分内，属于小概率事件，不太可能发生。

说明原假设为真不成立，即$\mu = 3190$不成立，即$\mu \neq 3190$，也就是说1990年新生儿平均体重与1989年有显著差异。

至此，假设检验完成。



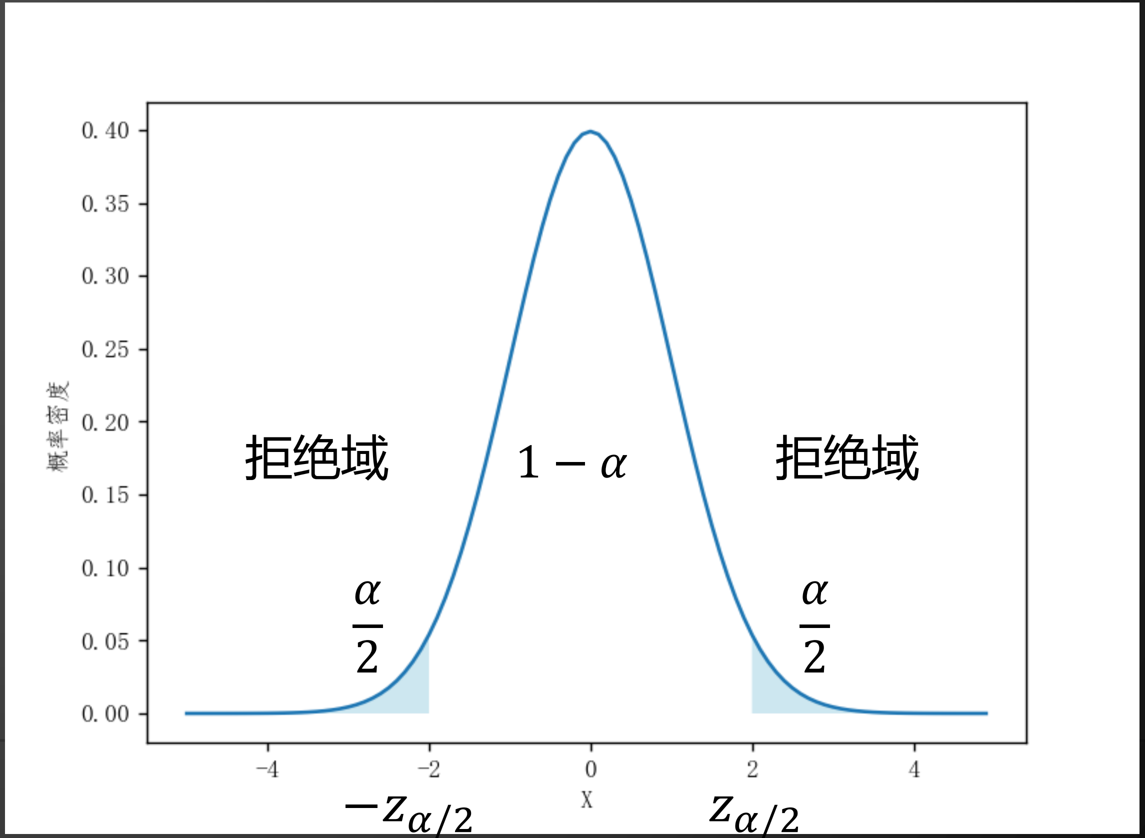
假设检验的基本原理

前面的结论说明，如果$z > z_{\alpha/2}$，则拒绝原假设，图中$z_{\alpha/2}$右侧区域称为拒绝域。

如果根据样本数据计算出来的z值小于-1.96，即$z < -z_{\alpha/2}$，也要拒绝原假设，图中$-z_{\alpha/2}$左侧区域也称为拒绝域。

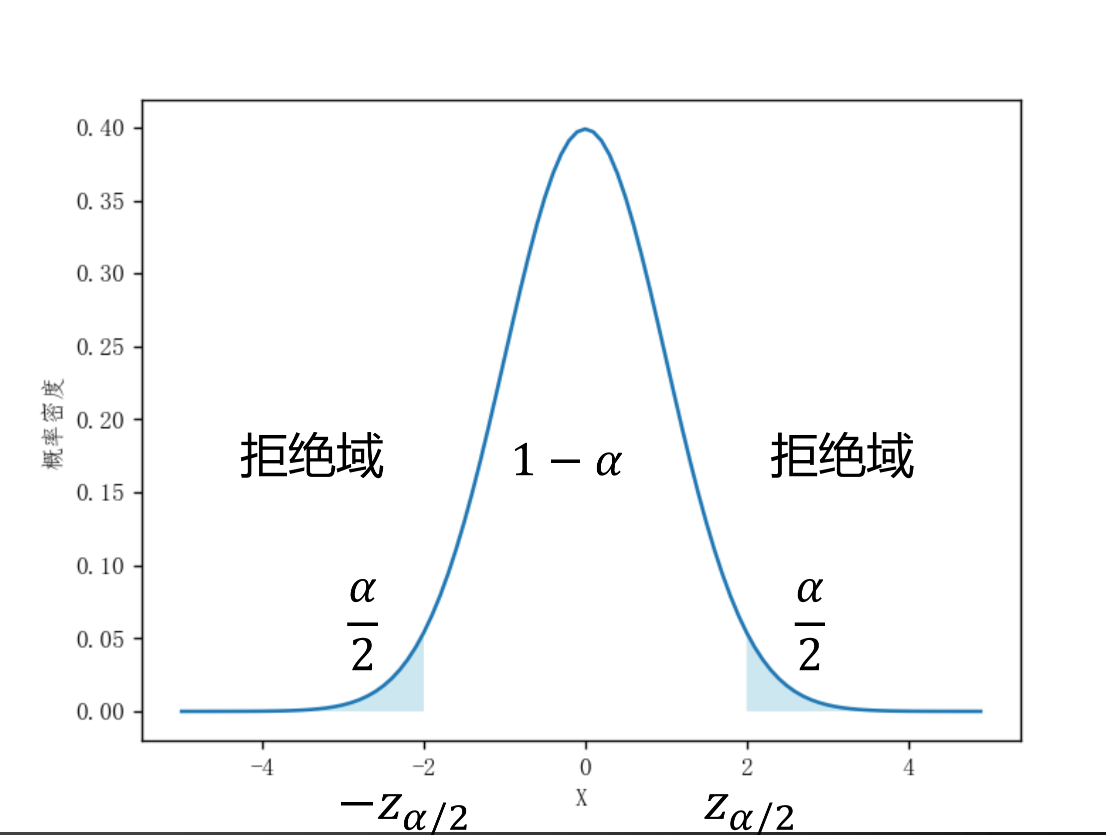
在前面的讲解中，涉及以下几个概念：

- α称为显著性水平，或者风险性水平。
- 给定α，对应的$z_{\alpha/2}$称为临界值。
- 统计量$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$称为检验统计量。



两类错误

- 根据前面的讲解，我们是依据小概率原理来做出拒绝原假设的决策。
- 如果小概率事件发生了，即虽然根据样本计算出来的z值落在了两侧的拒绝域内，但原假设确实为真，而我们却拒绝了原假设，意味着我们的决策犯错了，犯错的概率等于α，这个错误也叫弃真错误。
- 当然，也有可能原假设为假，我们却做出了接受了原假设的决策，这个错误叫取伪错误。
- 所以，在假设检验中，错误不可避免，而且有两类错误：弃真错误和取伪错误。



两类错误

- 根据前面的讲解，将原假设的实际情况（真、假）与对原假设的决策（不拒绝、拒绝）归纳为一张表。

原假设	没有拒绝	拒绝
真	$1-\alpha$	α
为假	β	$1-\beta$

- 第一类错误（弃真错误）

第一类错误是指原假设H_0为真，却被拒绝了，用α表示第一类错误，也称α错误。

- 第二类错误（取伪错误）

第二类错误是指原假设H_0为假，却没有拒绝原假设，用β表示第二类错误，也称β错误。

- 在前面的讲解中，我们依据α来做出决策，即在假设检验中，我们更关心第一类错误。

假设检验的流程

以前面提到的新生儿体重为例：

由统计资料得知，1989年某地新生儿平均体重为3190克，标准差为80克。现从1990年新生儿中随机抽取100个，测得其平均体重为3210克，问1990年的新生儿与1989年相比，体重有无显著差异。

1、提出假设：原假设和备择假设

$H_0: \mu = 3190(g)$

$H_1: \mu \neq 3190(g)$

2、确定检验统计量

$n = 100$，属于大样本，且总体标准差σ已知，所以采用z统计量，有

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{(3210 - 3190)}{80/\sqrt{100}} = 2.5$$

3、根据显著性水平进行统计决策

假设显著性水平$\alpha = 0.05$，则$z_{\alpha/2} = 1.96$，因为$z > z_{\alpha/2}$，所以$z = 2.5$落入拒绝域，所以拒绝原假设。也就是说，我们认为与1989年相比，1990年新生儿的体重有显著差异。

利用P值进行决策

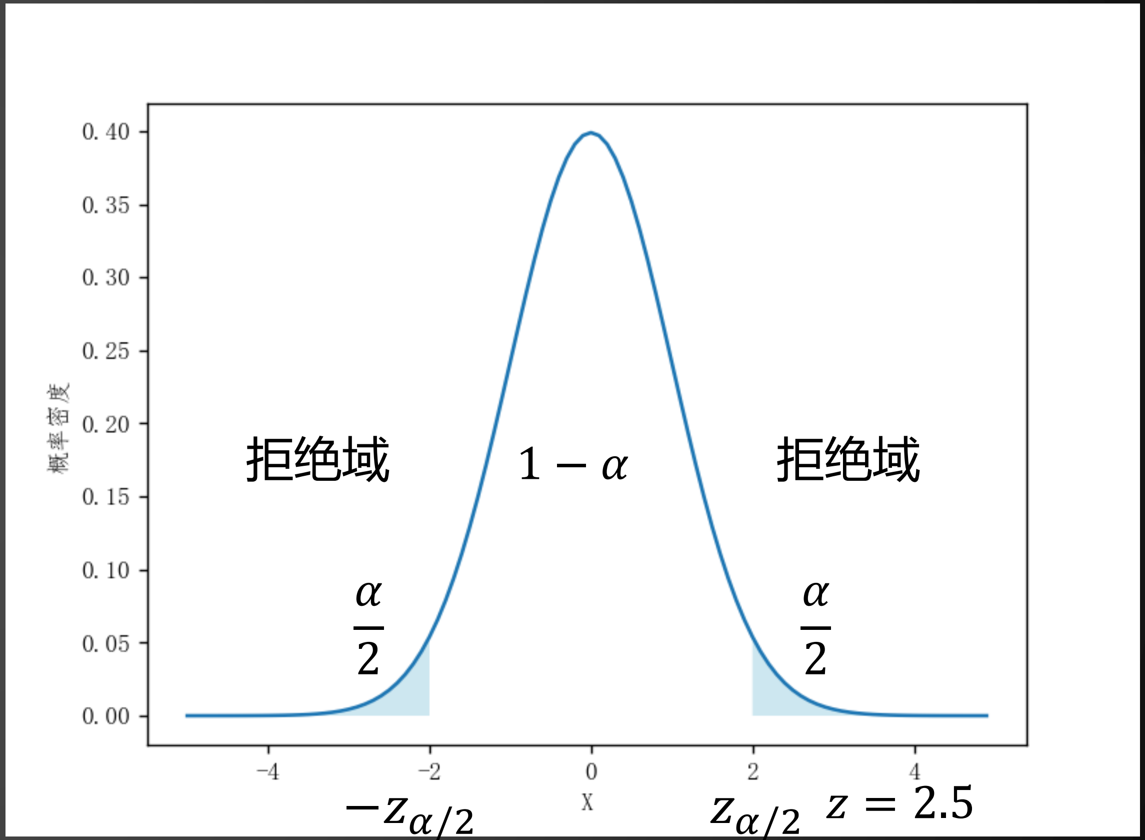
在前面的例子中，我们是利用z值与临界值$z_{\alpha/2}$比较做出决策。其实还可以依据z值对应的概率来做出决策。

$$z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{(3210-3190)}{80/\sqrt{100}}=2.5$$

根据$z=2.5$，可以计算出对应的概率为：0.00621

由于右侧拒绝域对应的概率为$\alpha/2=0.025$，一般将上述概率乘以2，与α比较，$0.00621\times 2=0.01242$，远小于α，所以拒绝原假设。

这里的概率0.01242就叫作P值，P值就是原假设为真时所得到的样本观察结果或更极端结果出现的概率。



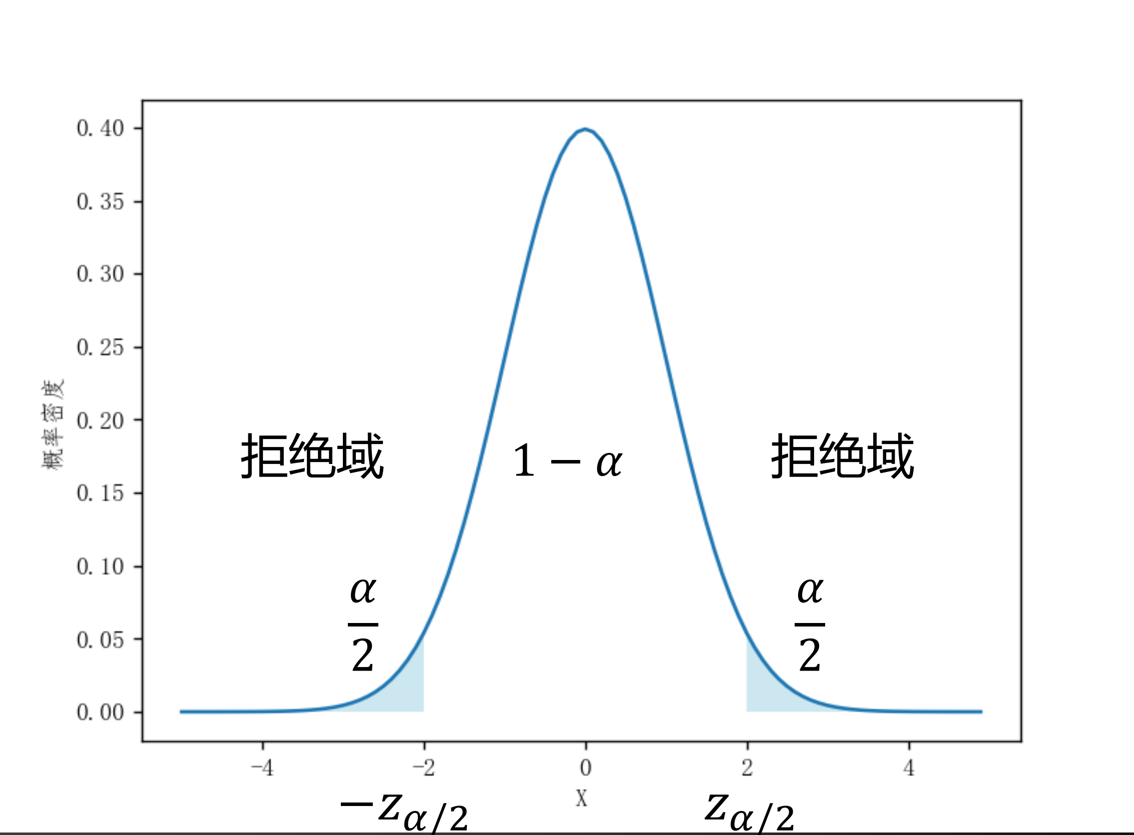
双侧检验

前面提到的新生儿体重问题，示意图如下，有两个拒绝域，分别位于两侧，每个拒绝域的面积为$\alpha/2$，这种属于双侧检验，原假设的形式为$H_0: \mu = \mu_0$。

对于双侧检验，如果$z > z_{\alpha/2}$或者$z < -z_{\alpha/2}$，

则拒绝原假设，可以简化为：

- 如果$|z| > |z_{\alpha/2}|$，则拒绝原假设；
- 如果$|z| < |z_{\alpha/2}|$，则接受原假设。



单侧检验

还有一些假设问题，带有方向性，例如，灯泡的使用寿命、轮胎行驶的里程数，所考察的数值越大越好，这种属于左单侧检验；另外一些，如废品率、生产成本，所考察的数值越小越好，这种属于右单侧检验。

下面看一个左单侧检验问题。

【例】某批发商欲从厂家购进一批灯泡，根据合同规定灯泡的使用寿命平均不能低于1000小时。已知灯泡燃烧寿命服从正态分布，标准差为200小时。在总体中随机抽取了100个灯泡，得知样本均值为960小时，问批发商是否应该购买这批灯泡？

分析：

要检验的问题是批发商是否应该购买这批灯泡，即这批灯泡的平均使用寿命是否不低于1000小时，设这批灯泡的平均使用寿命为μ，则原假设为：$\mu \geq 1000$，表示批发商应该购买这批灯泡，相应的备择假设为：$\mu < 1000$，表示批发商不应该购买这批灯泡。

单侧检验

1、提出假设：原假设和备择假设

$H_0: \mu \geq 1000$

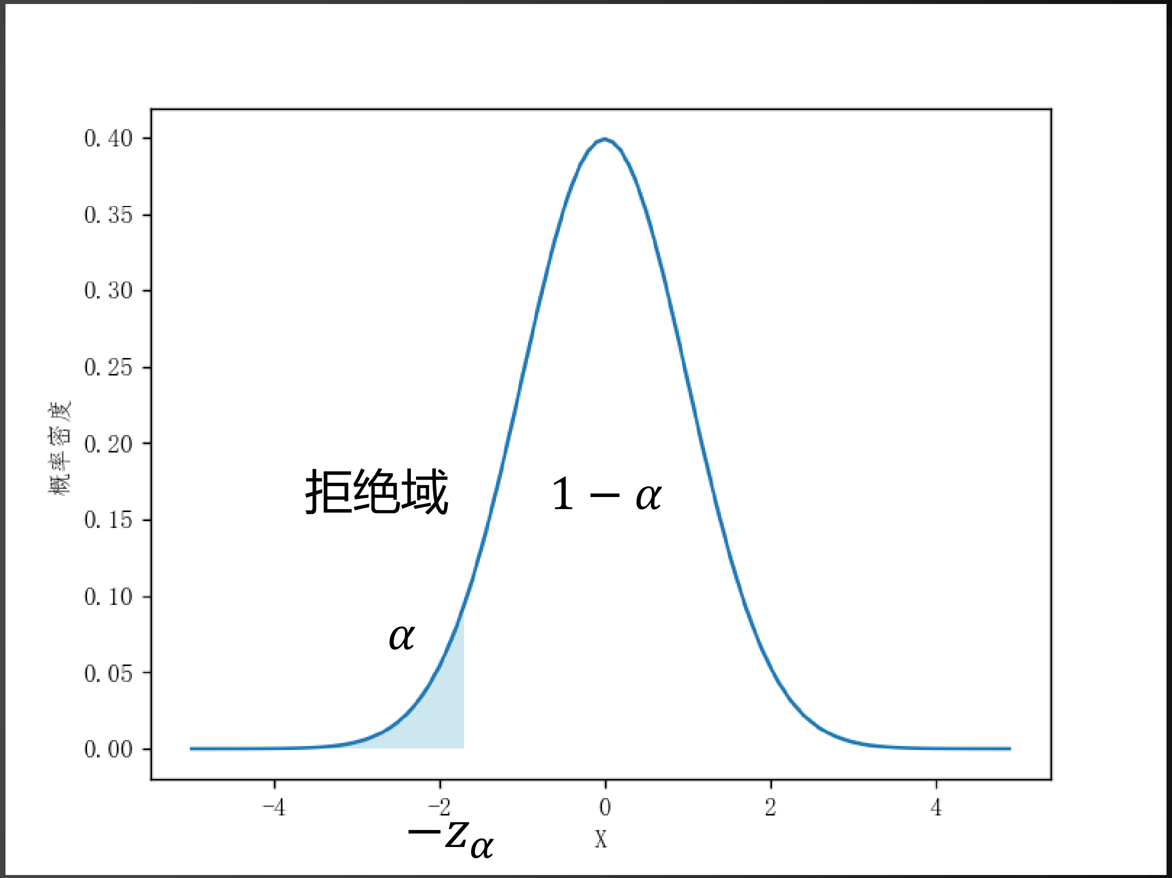
$H_1: \mu < 1000$

2、确定检验统计量

$n = 100$，属于大样本，且总体σ已知，所以采用z统计量，有

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{960 - 1000}{200/\sqrt{100}} = -2$$

3、根据显著性水平进行统计决策



假设显著性水平$\alpha = 0.05$，则$-z_\alpha = -1.65$，因为$z < -z_\alpha$（或者$|z| > |z_\alpha|$，与双侧检验统一），所以判断$z = -2$落入拒绝域，所以拒绝原假设，也就是说，批发商不该购买这批灯泡。

总结：双侧检验和单侧检验

双侧检验

1、原假设形式：

$H_0: \mu = \mu_0$。

2、临界值：

规定显著性水平α，临界值为$z_{\alpha/2}$

3、决策准则：

- 如果$|z| > |z_{\alpha/2}|$，则拒绝原假设；
- 如果$|z| < |z_{\alpha/2}|$，则接受原假设。

单侧检验

1、原假设形式：

左单侧检验，$H_0: \mu > \mu_0$

右单侧检验，$H_0: \mu < \mu_0$

2、临界值：

规定显著性水平α，临界值为z_{α}

3、决策准则：

- 如果$|z| > |z_{\alpha}|$，则拒绝原假设；
- 如果$|z| < |z_{\alpha}|$，则接受原假设。

假设检验问题的分类

对于总体来说，关心的参数有：均值、比例、方差，按照总体的个数把假设检验问题分为：

一个总体参数的检验	两个总体参数的检验
• 总体均值的检验 • 总体比例的检验 • 总体方差的检验	• 两个总体均值之差的检验 • 两个总体比例之差的检验 • 两个总体方差比的检验

本章主要内容

1. 假设检验的基本原理
2. 一个总体参数的检验
3. 两个总体参数的检验
4. 卡方检验

一个总体参数的检验问题

一个总体参数的假设检验有三种类型。

- 总体均值的检验
- 总体比例的检验
- 总体方差的检验

一个总体参数的检验问题

一个总体参数的假设检验有三种类型。

- 总体均值的检验
- 总体比例的检验
- 总体方差的检验

总体均值的检验

【例】某机床厂加工一种零件，根据经验知道, 该厂加工的零件的椭圆度渐近服从正态分布, 其总体均值为0.081 mm, 总体标准差为0.025mm, 今另换一种新机床进行加工, 取200个零件进行检验, 得到椭圆度均值为0.076mm, 问新机床加工零件的椭圆度总体均值与以前有无明显差别。

提出假设:

$H_0: \mu = 0.081mm$ 没有显著差别

$H_1: \mu \neq 0.081mm$ 有显著差别

因为$n > 30$，属于大样本问题，所以选用z统计量。

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0.076 - 0.081}{0.025/\sqrt{200}} = -2.83$$

假设风险性水平$\alpha = 0.05$，则临界值$z_{\alpha/2} = \pm 1.96$,因为$|z| > |z_{\alpha/2}|$，所以决绝原假设，即认为新老机床加工零件椭圆度的均值有显著差异。

一个总体参数的检验问题

一个总体参数的假设检验有三种类型。

- 总体均值的检验
- 总体比例的检验
- 总体方差的检验

总体比例的检验

在总体比例的检验中，通常采用z检验量，因为关于比例的问题一般采用大样本量。

当n充分大时（$n \geq 30$），p近似服从正态分布，均值为π，方差为$\frac{\pi(1-\pi)}{n}$，即

$$p \sim N(\pi, \frac{\pi(1-\pi)}{n})$$

对于比例问题的检验，z统计量的计算公式为：

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

【例】一项统计结果声称某市老年人口(年龄在65岁以上)所占的比例为14.7%, 该市老年人口研究会为了检验该项统计是否可靠, 随机抽取了400名居民, 发现其中有57人年龄在65岁以上。调查结果是否支持该市老年人口比例为14.7%的看法($\alpha = 0.05$)?

解：

假设检验的表达式：

$H_0: \pi = 14.7\%$

$H_1: \pi \neq 14.7\%$

$p = \frac{57}{400} = 0.14 = 14\%$

$z = \frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.14-0.147}{\sqrt{\frac{0.147\times(1-0.147)}{400}}} = -0.254$

由于这是一个双侧检验，所以当$\alpha = 0.05$时，$|z_{\alpha/2}| = 1.96$

由于$|z| < |z_{\alpha/2}|$，所以原假设位于接受域，接受原假设，可以认为调查结果支持了该市老年人口所占比例为14.7%的看法。

一个总体参数的检验问题

一个总体参数的假设检验有三种类型。

- 总体均值的检验
- 总体比例的检验
- 总体方差的检验

回顾：样本方差的分布

样本方差的分布比较复杂，这里只讨论当总体分布为正态分布时，样本方差的分布。

当$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的样本，则样本方差S^2的分布为：

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

即服从自由度为$n-1$的卡方分布。

总体方差的检验

【例】某厂商生产出一种新的饮料瓶机器，按设计要求，该机器装一瓶1000ml的饮料误差上下不超过1ml。如果达到设计要求，表明机器的稳定性能非常好。现从该机器装完的产品中随机抽取25瓶,分别进行测定(用样本观测值分别减1000ml)，得到如表8-2所示的结果。

0.3	-0.4	-0.7	1.4	-0.6
-0.3	-1.5	0.6	-0.9	1.3
-1.3	0.7	1	-0.5	0
-0.6	0.7	-1.5	-0.2	-1.9
-0.5	1	-0.2	-0.6	1.1

试以$\alpha = 0.05$的显著性水平检验该机器的性能是否达到设计要求。

分析：该机器装一瓶1000ml的饮料误差上下不超过1ml，描述的是波动性，即方差不超过1ml，如果机器的性能达到设计要求，说明方差不超过1ml，原假设表示该机器的性能达到设计要求，即$\sigma^2 \leq 1$，这里是一个右单侧检验问题，因为我们希望方差越小越好。

解:

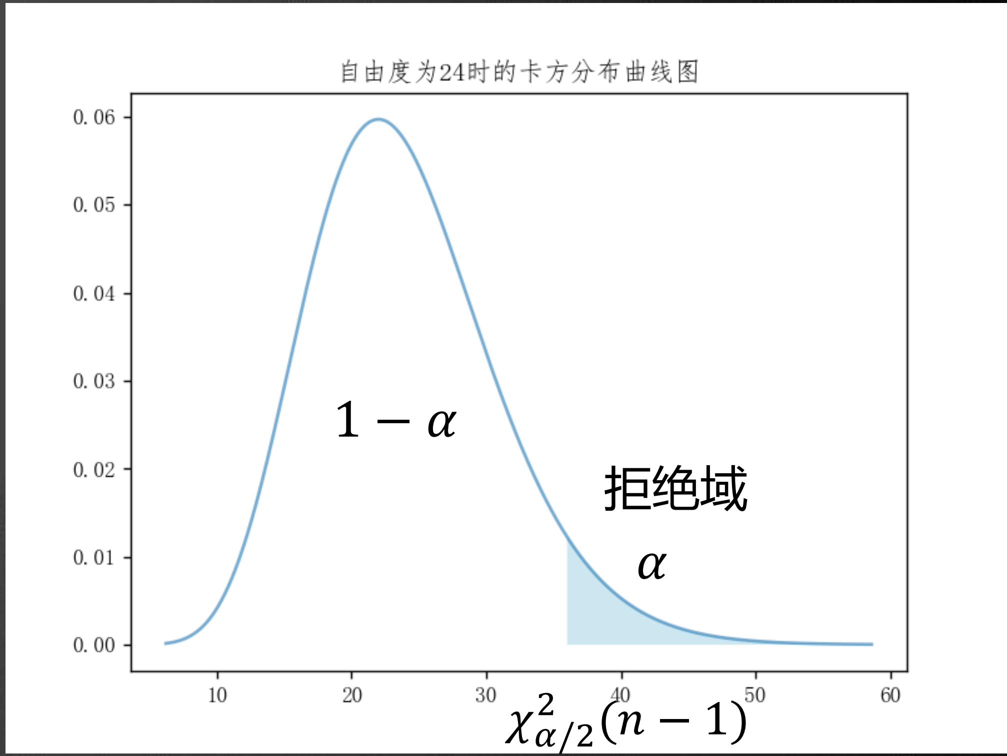
这是一个单侧检验问题，写出假设:

$H_0: \sigma^2 \le 1$

$H_1: \sigma^2 > 1$

根据样本数据计算出$s^2 = 0.866$，所以

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)\times 0.866}{1} = 20.8$$



由Excel公式可以计算出，$\chi^2_{0.05}(24) = 36.415$，故不能拒绝原假设$H_0$，可以认为该机器的性能达到设计要求。

本章主要内容

1. 假设检验的基本原理
2. 一个总体参数的检验
3. 两个总体参数的检验
4. 卡方检验

两个总体参数的检验

有时候，我们可能需要比较两个总体的参数，例如，两种方法生产出来的产品平均抗拉强度是否有显著差别，或者两个人群对于某种食品的偏好是否有显著差异等。

两个总体参数的检验有以下三类问题：

- 两个总体均值之差的检验
- 两个总体比例之差的检验
- 两个总体方差比的检验

两个总体参数的检验

有时候，我们可能需要比较两个总体的参数，例如，两种方法生产出来的产品平均抗拉强度是否有显著差别，或者两个人群对于某种食品的偏好是否有显著差异等。

两个总体参数的检验有以下三类问题：

- 两个总体均值之差的检验
- 两个总体比例之差的检验
- 两个总体方差比的检验

两个总体均值之差的检验

两个总体的方差已知

两个独立样本计算出的$\bar{x}_1 - \bar{x}_2$的抽样分布服从正态分布，标准差为：

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

此时，检验统计量z的计算公式为：

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

两个总体均值之差的检验：例题

【例】有两种方法可用于制造某种以抗拉强度为重要特征的产品。根据以往的资料得知，第一种方法生产出产品抗拉强度的标准差为8千克，第二种方法的标准差为10千克。从两种方法生产的产品中各抽一个随机样本，样本量分别为$n_1 = 32, n_2 = 40$，测得$\bar{x}_1 = 50$千克，$\bar{x}_2 = 44$千克。问这两种方法生产出来的产品平均抗拉强度是否有显著差别$(\alpha = 0.05)$?

分析：要比较的这两种方法生产出来的产品平均抗拉强度是否有显著差别，说白了，就是这两种方法生产出来的产品平均抗拉强度之差是否等于零的问题，如果等于零，就表示没有显著差别，否则，表示有显著差别。

用μ_1表示第一种方法生产的产品的平均抗拉强度，用μ_2表示第二种方法生产的产品的平均抗拉强度，$\mu_1 - \mu_2$表示这两种方法生产出来的产品平均抗拉强度之差

原假设：$\mu_1 - \mu_2 = 0$，表示这两种方法生产出来的产品平均抗拉强度没有显著差别

解：这是一个双侧检验问题，写出假设：

$H_0: \mu_1 - \mu_2 = 0$ 没有显著差别

$H_1: \mu_1 - \mu_2 \neq 0$ 有显著差别

由于总体的方差已知，所以选用z统计量，有

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

将题目中已知条件代入，有

$$z = \frac{50 - 44 - 0}{\sqrt{\frac{64}{32} + \frac{100}{40}}} = 2.83$$

当$\alpha = 0.05$时，$z_{\alpha/2} = 1.96$，落入拒绝域，所以拒绝H_0，即两种方法生产出来的产品其抗拉强度有显著差别。

两个总体参数的检验

有时候，我们可能需要比较两个总体的参数，例如，两种方法生产出来的产品平均抗拉强度是否有显著差别，或者两个人群对于某种食品的偏好是否有显著差异等。

常见问题：

- 两个总体均值之差的检验
- 两个总体比例之差的检验
- 两个总体方差比的检验

回顾：两个样本比例之差的分布

样本1：从具有参数π_1的二项总体中抽取的包含n_1个观测值的样本

样本2：从具有参数π_2的二项总体中抽取的包含n_2个观测值的样本

则两个样本比例差为：

$$p_1 - p_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

当n_1和n_2很大时，该比例之差近似服从正态分布，均值和方差分别为：

$$E(p_1 - p_2) = \pi_1 - \pi_2$$

$$D(p_1 - p_2) = \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}$$

两个总体比例之差的检验

两个总体比例之差不为零

即检验$\pi_1 - \pi_2 = d_0 (d_i \neq 0)$，在这种情况下，两个样本比例之差$p_1 - p_2$近似服从以$\pi_1 - \pi_2$为数学期望，$\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$为方差的正态分布，因此选择z作为检验统计量：

$$
\begin{aligned}
z &= \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \\
&= \frac{(p_1 - p_2) - d_0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}
\end{aligned}
$$

两个总体比例之差的检验

【例】有一项研究报告说青少年经常上网聊天，男生的比例至少超过女生10个百分比，即 $\pi_1 - \pi_2 \geq 10\%$（π_1为男生比例，π_2为女生比例）。现对150个男生和150个女生进行上网聊天的频度调查，其中经常聊天的男生有68人，经常聊天的女生有54人，问调查结果是否支持研究报告的结论($\alpha = 0.05$)?

分析：

如果调查结果支持研究报告的结论，则说明男生的比例至少超过女生10个百分比，即$\pi_1 - \pi_2 \geq 10\%$，可以把这个式子作为原假设，表示调查结果支持研究报告的结论。

解:

$H_0: \pi_1 - \pi_2 \geq 10\%$

$H_1: \pi_1 - \pi_2 < 10\%$

根据题目已知，$n_1 = n_2 = 150, p_1 = \frac{68}{150} = 0.45, p_2 = \frac{54}{150} = 0.36, d_0 = 10\%$

根据上面的公式，得

$$z = \frac{(p_1 - p_2) - d_0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$
$$= \frac{0.45 - 0.36 - 0.1}{\sqrt{\frac{0.45\times(1-0.45)}{150} + \frac{0.36\times(1-0.36)}{150}}} = -0.177$$

这是一个左单侧检验，$z_\alpha = 1.645$，由于$|z| < z_{\alpha/2}$，则原假设落入非拒绝域，故无法推翻原假设，调查结果支持研究报告的结论。

检验中的匹配样本

匹配样本这个概念在前面的课程参数估计中提到过，可以参考前面课程的讲解。

【例】一个以减肥为主要目标的健美俱乐部声称，参加它的训练班至少可以使肥胖者平均体重减轻8.5千克以上。为了验证该声称是否可信，调查人员随机抽取了10名参加者，得到他们的体重记录如下表所示。

训练前后的体重记录，单位：千克

训练前	**94.5**	**101**	**110**	**103.5**	**97**	**88.5**	**96.5**	**101**	**104**	**116.5**
训练后	85	89.5	101.5	96	86	80.5	87	93.5	93	102

在$\alpha = 0.05$的显著性水平下，问调查结果是否支持该俱乐部的声称?

分析：对于匹配样本问题来说，可以先求这两个样本之间的差值$d_i(i = 1,2,\cdots,10)$，然后将这些差值看成一个样本，现在要检验的是这个样本的均值是否大于等于8.5，也就是说将这个问题转化为一个样本均值的检验问题。

上一页例题解答

解：

首先，将每个样本的差值计算出来，右表所示。

μ表示这两个样本的差值的均值，$\mu = \mu_1 - \mu_2$

写出假设表达式

$H_0: \mu \leq 8.5$　　平均减重没有超过8.5千克

$H_1: \mu > 8.5$　　平均减重超过8.5千克

由于是小样本，且总体标准差未知，所以采用t统计量，有

$$t = \frac{\bar{d} - \mu_0}{s/\sqrt{n}}$$

训练前	训练后	差值
94.5	85	9.5
101	89.5	11.5
110	101.5	8.5
103.5	96	7.5
97	86	11
88.5	80.5	8
96.5	87	9.5
101	93.5	7.5
104	93	11
116.5	102	14.5
合计	—	98.5

上一页例题解答

根据样本数据，差值的均值和标准差分别为：

$$\bar{d} = \frac{\sum d_i}{n} = \frac{98.5}{10} = 9.85$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 2.199$$

根据t值的计算公式，有

$$t = \frac{\bar{d} - \mu_0}{s/\sqrt{n}} = \frac{9.85 - 8.5}{2.199/\sqrt{10}} = 1.94$$

这是一个右单侧检验，计算$t_\alpha(n-1)$，自由度为$n-1=9$，$\alpha = 0.05$，$t_\alpha(9) = 1.833$

因为$|t| > |t_\alpha(n-1)|$，位于拒绝域，所以拒绝原假设，即认为该俱乐部的声称是可信的。

假设检验与参数估计的联系

相同点：

- 参数估计和假设检验都是推断统计的组成部分，都是利用样本对总体进行某种推断。

不同点：

- 参数估计讨论的是用样本统计量去估计总体参数的方法，总体参数是未知的。
- 而在假设检验中，则是先对总体参数提出一个假设，然后利用样本信息去检验这个假设是否成立。

假设检验与置信区间的关系

假设检验和置信区间之间有着明显的联系，考虑双侧检验问题。

原假设：$H_0: \mu = \mu_0$，样本统计量：$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

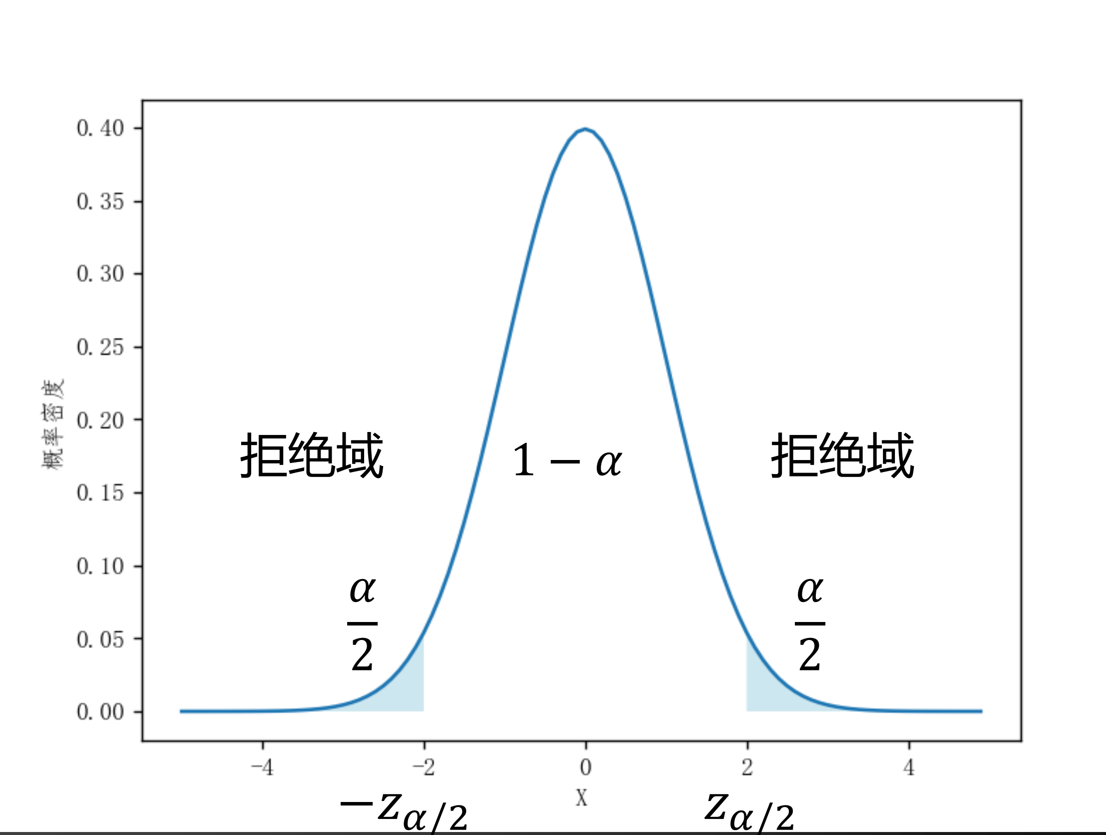
如果$|z| > |z_{\alpha/2}|$，则拒绝原假设，反之，则接受。

也可以通过置信区间进行决策：

置信区间：$(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

如果$\mu_0 \in (\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$，则接受原假设，

反之，则接受原假设。



假设检验与置信区间的关系

由统计资料得知，1989年某地新生儿平均体重为3190克，标准差为80克。现从1990年新生儿中随机抽取100个，测得其平均体重为3210克，问1990年的新生儿与1989年相比，体重有无显著差异。

以前面提到的新生儿体重为例：

1、提出假设：原假设和备择假设

$H_0: \mu = 3190(g)$

$H_1: \mu \neq 3190(g)$

2、确定检验统计量

$n = 100$，属于大样本，且总体标准差σ已知，所以采用z统计量，假设显著性水平$\alpha = 0.05$，则$z_{\alpha/2} = 1.96$，置信区间为：

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (3194,3226)$$

3、利用置信区间进行统计决策

因为$3190 \notin (3194,3226)$，所以拒绝原假设。也就是说，我们认为与1989年相比，1990年新生儿的体重有显著差异。

本章主要内容

1. 假设检验的基本原理
2. 一个总体参数的检验
3. 两个总体参数的检验
4. 卡方检验

卡方检验

卡方检验就是检验两个变量之间有没有关系。

例如，

- 卡方检验可以检验男性或者女性对线上买生鲜食品有没有区别；
- 想知道性别跟对猫和狗的偏爱程度是否有显著关系；
- 不同城市级别的消费者对买SUV车有没有什么区别；

卡方检验实例

我们要观察性别和在线上买不买生鲜食品有没有关系，现实生活中，女性通常去菜市场买菜的比较多，那么在线上是不是也这样。

	男	女	总计
线上不买生鲜	527	72	599
线上买生鲜	206	102	308
总计	733	174	907

来源：知乎，文文酱的数据课堂

1、提出假设

原假设：性别和在线上买不买生鲜食品没有关系

备择假设：性别和在线上买不买生鲜食品有关系

2、计算出各类别的频数，即实际值（观测值），上述表格已给出。

卡方检验实例

3、计算出比例，再根据比例计算出理论值

计算出线上不买生鲜和线上买生鲜的比例

- 线上不买生鲜：66%
- 线上买生鲜：34%

- 根据比例，得到理论值

	男	女	总计
线上不买生鲜	$733 \times 66\% = 484$	$174 \times 66\% = 115$	599
线上买生鲜	$733 \times 34\% = 249$	$174 \times 34\% = 59$	308
总计	733	174	907

卡方检验实例

4、计算出卡方值，和临界值比较，给出结论

- 接着，计算出卡方值

$$\chi^2 = \sum \frac{(A-T)^2}{T}$$

其中，A为实际值，T为理论值，卡方值衡量的是实际值与理论值的差异程度

- 根据卡方检验的计算公式，有

$$\chi^2 = \frac{(527-484)^2}{484} + \frac{(72-115)^2}{115} + \frac{(206-249)^2}{249} + \frac{(102-59)^2}{59} = 58.4$$

- 自由度：（行数-1）×（列数-1）

$$(2-1)\times(2-1) = 1$$

- 临界值：95%置信水平为3.84，结论：性别和在线上买不买生鲜食品有显著关系。

卡方检验总结

脚下留心：

- 卡方检验只适用于分类数据，而不是数值数据
- 对于样本的数量有一定的要求，例如100+、200+

卡方检验的步骤

1. 提出假设
2. 计算出各类别的频数，即实际值（观测值）
3. 计算出比例，再根据比例计算出理论值
4. 计算出卡方值，和临界值比较，给出结论

卡方检验实例

例如，想知道性别跟对猫和狗的偏爱程度是否有显著关系？

	猫	狗
男	207	282
女	231	242

1、提出假设

- 原假设：性别对于猫或狗的偏爱没有影响
- 备择假设：性别对于猫或狗的偏爱有影响

卡方检验实例

2、计算出各类别的频数，即实际值（观测值）

	猫	狗	总计
男	207	282	489
女	231	242	473
总计	438	524	962

3、计算出比例，根据比例计算出理论值

- 偏爱猫的比例：46%
- 偏爱狗的比例：54%

卡方检验实例

3、计算出比例，根据比例计算出理论值

偏爱猫的比例：46%，偏爱狗的比例：54%

根据比例计算出理论值

	猫	狗	总计
男	$489 \times 46\% = 225$	$489 \times 54\% = 264$	489
女	$473 \times 46\% = 218$	$473 \times 54\% = 255$	473
总计	438	524	962

卡方检验实例

4、计算出卡方值，和临界值比较，给出结论

根据公示计算出卡方值:

$$\chi^2 = \frac{(207-225)^2}{225} + \frac{(282-264)^2}{264} + \frac{(231-218)^2}{218} + \frac{(242-255)^2}{255} = 4.11$$

临界值：95%置信水平为3.84

结论：性别对于猫或狗的偏爱有影响