假 设 检 验

目录 CONTENTS

01

原假设与备择假设

4 假 设 检 验

定义

在假设检验中，我们首先对总体参数做一个尝试性的假设，该尝试性的假设被称为原假设(null hypothesis)，记作H_0。然后，定义另一个与原假设的内容完全对立的假设，称之为备择假设(alternative hypothesis)，记作H_α。

假设检验的过程就是根据样本数据对这两个对立的假设进行检验。

在确定假设检验和备择假设检验时，关键的问题是思考搜集样本的目的是什么，我们想要得到怎样的结论。

5 假设建立

将研究中的假设作为备择假设

许多假设检验的应用都是试图收集证据来支持研究中的假设。在这种情形下，通常最好从备择假设开始，然后得到研究者希望支持的结论。

案例

制造一批新型燃油喷射系统，把它们安装在汽车上，计算这些汽车每加仑燃油行驶里程的样本均值，并据此在假设检验中判断是否可以得出结论：新型燃油喷射系统的平均效率超过24英里/加仑。令燃油效率的总体均值为μ，在研究中的假设$\mu > 24$是备择假设。

那么原假设与备择假设：

$$H_0: \mu \le 24$$

$$H_\alpha: \mu > 24$$

如果抽样结果表明得出拒绝H_0的结论，则可以做出$H_\alpha: \mu > 24$的推断。如果抽样结果不能拒绝H_0的结论。

在接纳一项事物之前，我们希望通过研究来判定是否有统计依据支持我们得出新方法确实更好的结论，在这种情形下，通常将研究中的假设表述为备择假设。

将受到挑战的假设作为原假设

从一种信念或假定开始，即从有关总体参数值的说法是真实的开始。然后，**我们将利用假设检验对这种假定提出怀疑**，并确定是否有统计证据支持得出假定不正确的结论。在这种情形下，原假设H_0表述了对总体参数值的信念或者假说；在备择假设H_α中，认为这种信念或者假说不正确。

案例

某瓶装饮料的制造商在标签上注明每瓶67.6液盎司。我们通常假设标签上的标注是正确的，但是可以对制造商的说法提出质疑。也就是我们可以挑战这种标签注明，认为它没有达到67.6液盎司，这种质疑表述作为我们的备择假设$\mu < 67.6$。于是，原假设和备择假设分别为：

$$H_0: \mu \ge 67.6 \qquad H_\alpha: \mu < 67.6$$

再从制造商本身出发，制造商不想让饮料分量不足，因为这会导致消费者投诉，但是也不想让产品超量，这会增加成本。企业的目标是使得饮料的容量恰好是67.6液盎司。但是生产操作无法保证100%正确，会发生饮料灌装不足或超量的情形。因此建立假设

$$H_0: \mu = 67.6 \qquad H_\alpha: \mu \ne 67.6$$

如果抽样的结果使得我们得到拒绝H_0的结论，那么说明灌装操作流程不合理，需要重新调整生产线！

6 假设形式小结

对于总体均值的假设检验，我们令μ_0代表假定值并且必须采用以下三种形式之一进行假设检验。

$$H_0: \mu \geq \mu_0 \quad H_0: \mu \leq \mu_0 \quad H_0: \mu = \mu_0$$
$$H_\alpha: \mu < \mu_0 \quad H_\alpha: \mu > \mu_0 \quad H_\alpha: \mu \neq \mu_0$$

前两种形式被称为单侧检验，第三种形式被称为双侧检验。注意：等号总是出现在原假设H_0中。
在选择H_0和H_α的适当形式时，记住将检验试图建立的结果设为备择假设！

7 第一类错误和第二类错误

由于假设检验是基于样本信息得到的，不可能得出的结论总是正确的，所以必须考虑发生误差的可能性

		总体情况	
		H_0是真的	H_α是真的
结论	接受H_0	结论正确	第二类错误
	拒绝H_0	第一类错误	结论正确

案例

考虑之前的汽车新型燃油喷射系统案例，建立了如下的假设

$$H_0: \mu \leq 24 \quad H_\alpha: \mu > 24$$

备择假设$H_\alpha: \mu > 24$，表明研究者致力于寻找样本证据支持结论，新型燃油喷射系统能使汽车燃油效率的总体均值超过24。

解　　读

表中第一行说明当得出接受H_0的结论时可能发生的情况。这时，如果H_0为真，则该结论正确，如果H_α为真，那么发生第二类错误，即当H_0为假时，我们却接受了H_0。表中第二行说明当得出拒绝H_0的结论时可能发生的情况。这时，如果H_α为真，则该结论正确，如果H_0为真，那么发生第一类错误，即当H_0为真时，我们却拒绝了H_0

案　　例

第一类错误是指H_0为真时却拒绝了H_0，既研究者认为新系统提高了燃油效率，但实际情况却是新系统并不比目前的系统好。

第二类错误是指H_0为假时却接受了H_0，既研究者认为新系统没有提高燃油效率，但实际情况却是新系统目前的系统好。

8 第一类错误和第二类错误

		总体情况	
		H_0是真的	H_α是真的
结论	接受H_0	结论正确	第二类错误
	拒绝H_0	第一类错误	结论正确

显著性水平
当原假设为真并且以等式形式出现时犯第一类错误的概率称为检验的显著性水平。

在案列中，显著性水平是当μ=24时拒绝H_0：$\mu \le 24$的概率

解　　读

原假设为H_0：$\mu \le 24$。假定原假设为真并且等号成立，即μ=24。当原假设为真并且以等式形式出现时，犯第一类错误的概率被称为检验的显著性水平。

显著性水平

α代表显著性水平，通过选择α控制犯第一类错误的概率。如果犯第一类错误的成本很高，则选择较小的α值，如果犯第一类错误的成本不高，则选择较大的α值。

一般将只控制第一类错误的假设检验称为显著性检验。

第二类错误

假设检验没有对第二类错误进行控制，因此如果决定接受H_0，我们无法确定该决策有多大的可信度。统计学家建议我们只使用两种结论：不能拒绝H_0或拒绝H_0。采用不能拒绝H_0，这种说法意味着我们对行动或判断持保留意见，避免了发生第二类错误的风险。

02

总体均值的检验

总体标准差已知情形

10

总体均值的检验:σ已知情形

总体均值的单侧检验(one-tailed test):

下侧检验	上侧检验
$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$
$H_\alpha: \mu < \mu_0$	$H_\alpha: \mu > \mu_0$

案例

美国联邦贸易委员会（FTC）定期设计统计调查，用以检验制造商的产品说明。列如，大号听装Hilltop咖啡的标签上标明装有3磅/听。FTC把大号听装咖啡标签上的信息理解为Hilltop的承诺：听装咖啡重量的总体均值为3磅/听。我们将说明，FTC如何通过下侧检验来验证Hilltop的承诺。

设定假设

接受的挑战为，Hilltop信守承诺：灌装总体均值大于等于3磅/听，这构成了原假设。反之则是备择假设。

显著性水平

假设选取样本的数据均值小于总体均值的估计，那么就要拒绝原假设。我们想要知道的是当样本均值比3磅少多少的时候，我们才能断言差异明显，并且甘愿犯第一类错误的风险控告Hilltop没有信守承诺。**这个问题中，一个关键因素就是决策者选取的显著性水平。**

检验统计量

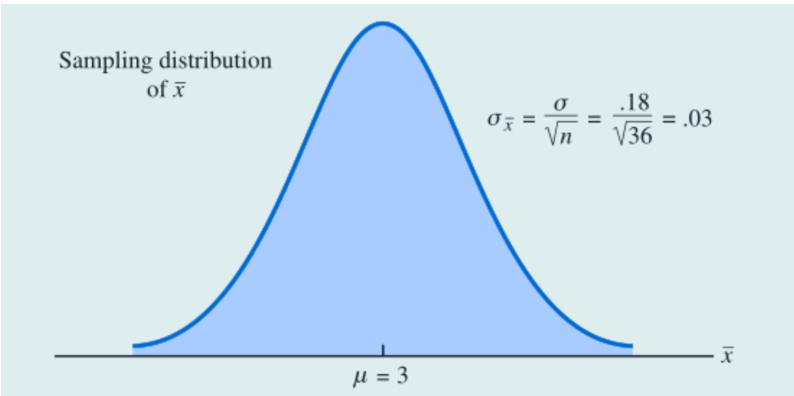
收集样本数据和设计实验来计算检验统计量

11 检验统计量

在Hilltop咖啡的研究中，假定总体标准差σ =0.18，另外如果总体服从正态分布或者抽样容量足够大，那么抽样样本也服从正态分布。因此样本均值$\bar{x}$的标准误差为$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 0.03$，因此

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - 3}{0.03}$$

的抽样分布是标准正态分布。z=-1表明$\bar{x}$的值位于比均值的假定值小1个标准差的位置。z=-3的下侧面积为0.0013，从而，得到z值小于均值3个或3个以上标准差的概率为0.0013。因此, $\bar{x}$的值比总体均值的假定值μ_0=3小3个或3个以上标准差的概率为0.0013.



检验统计量

在总体标准差已知的前提下，我们用标准正态分布随机变量 z 作为检验统计量来确定$\bar{x}$ 是否偏离假定值μ足够远，从而有理由拒绝原假设。

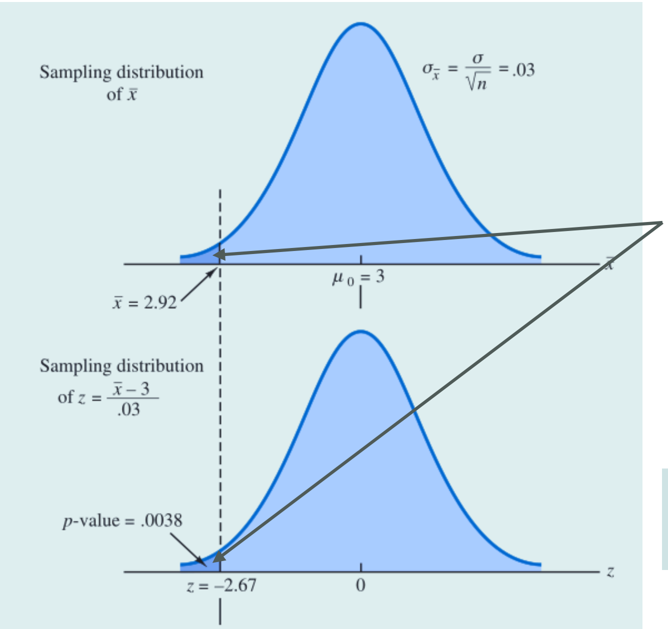
总体均值假设检验的统计量：σ已知情形

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

关键问题

下侧检验的一个关键问题，检验统计量z必须达到多小的时候，我们才能拒绝原假设。有两种方法：p-值法和临界值法。

12 p-值法



对于下侧检验，$p-$值是检验统计量小于或等于样本所给出的检验统计量的值的概率，因此必须得到标准正态分布曲线下在检验统计量的值左边部分的面积。

p-值拒绝法则

如果p-值≤ α，则拒绝H_0

$p-$值法

$p-$值是一个概率值，它度量样本所提供的证据对原假设的支持程度。$p-$值越小说明拒绝原假设的证据越多。

案 例 解 决

抽取的36听咖啡样本均值$\bar{x}=2.92$磅。计算检验统计量 $z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=-2.67$. 从而p-值为检验统计量小于或等于$-2.67$的概率。

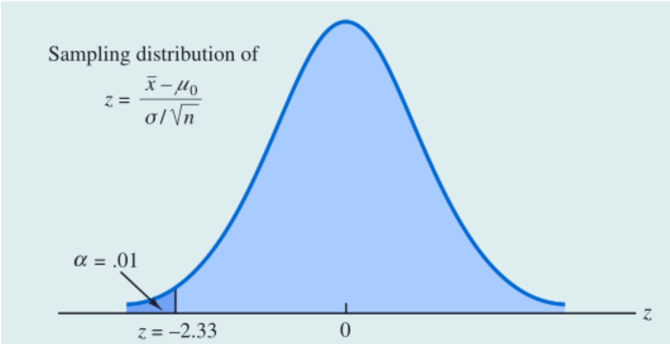
利用标准正态概率分布表，我们可知z= -2.67下侧的面积为0.0038. 那么这个值足以能够使得我们拒绝原假设么？答案是依赖于检验的显著水平。

研究人员选取0.01位显著性水平，那么由于0.0038小于0.01，故我们拒绝原假设，从而在0.01的显著性水平下我们发现有足够的统计证据拒绝原假设。

13

临 界 值 法

对于下侧检验，临界值是确定检验统计量的值是否小到足以拒绝原假设的一个基准。在检验统计量的抽样分布中，与下侧面积α相对应的值是检验统计量的临界值。换句话说，临界值是使得我们拒绝原假设的检验统计量的最大值。



下侧检验的拒绝法则：临界值

如果$z \le -z_\alpha$，则拒绝H_0

式中，$-z_\alpha$为临界值，即标准正态概率分布下侧的面积α时对应的z值。

案 例 解 决

我们发现z=-2.33的下侧面积等于0.01的值。根据抽样计算出的检验统计量值为 -2.67，由于小于 -2.23，得知p-值小于或等于0.01，所以我们拒绝H_0并且得出结论认为Hilltop咖啡的分量不足。，

14 总体均值的检验:σ已知情形

总体均值的双侧检验(two-tailed test):

$$H_0: \mu = \mu_0$$
$$H_\alpha: \mu \neq \mu_0$$

案例

Maxflight生产的高尔夫球必须达到平均驱动距离295码，才可以参与USGA的赛事。距离或大或小都是不被允许的，因此该公司需要检验其生产的高尔夫球平均驱动距离是否等于295码。

设定假设

$$H_0: \mu = 295$$
$$H_\alpha: \mu \neq 295$$

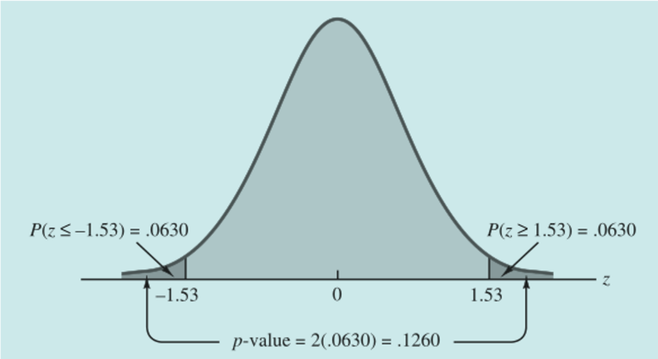
显著性水平

选取0.05作为显著性水平

检验统计量

总体标准差为12，样本容量为50，样本均值为297.6，那么检验统计量$z = \frac{x - \mu_0}{\frac{\sigma}{\sqrt{n}}} = 1.53$

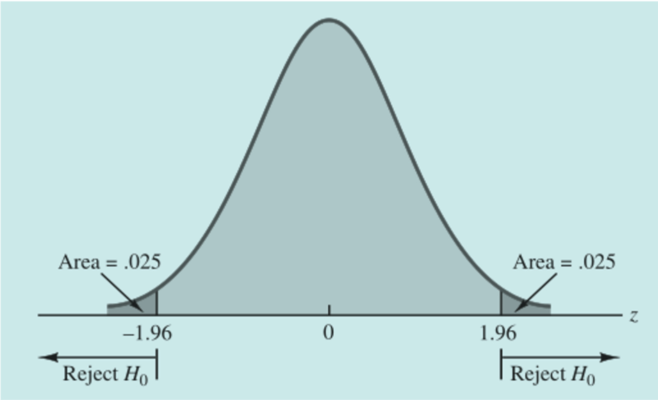
15 p-值法与临界值法



对于双侧检验而言, $z \le -1.53$和$z \ge 1.53$都满足要求。因此$p-$值为$P(z \le -1.53)+P(z \ge 1.53)$.

标准正态分布的两侧是对称的，因此$P(z \ge 1.53)$ = 1- $P(z \le 1.53)$=0.063.

p-值= 0.063*2=0.1260



双侧检验$p-$值法

1. 计算出检验统计量
2. 如果检验统计量的值位于上侧(z>0)，计算z大于或等于检验统计量值的概率(上侧面积)。如果检验统计量的值位于下侧(z<0)，计算z小于或等于检验统计量值的概率(下侧面积)。
3. 将上一步算出的概率乘以2得到p-值

临 界 值 法

检验的临界值位于标准正态分布的上侧尾端和下侧尾端。取显著性水平0.05，每侧尾端临界值所对应的面积为0.05/2=0.025。查标准正态概率表可知，检验统计量的临界值$-z_{0.025}=-1.96$和$z_{0.025}=1.96$。从而利用临界值法，双侧检验的拒绝法则是：

如果$z \le -1.96$ 或者$z \ge 1.96$，则拒绝H_0

因为Maxflight研究中的检验统计量z=1.53，在0.05显著性水平下统计证据不允许我们拒绝原假设

总结

总体均值的检验:σ已知情形

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

17 区间估计与假设检验的关系

在σ已知的情形下，总体均值的100(1-α)%置信区间估计为

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

总体均值的双侧检验(two-tailed test):

$$H_0: \mu = \mu_0$$
$$H_\alpha: \mu \neq \mu_0$$

用置信区间的方法进行如下形式的假设检验

$$H_0: \mu = \mu_0$$
$$H_\alpha: \mu \neq \mu_0$$

1. 从总体中抽取一个简单随机样本，并利用样本均值$\bar{x}$建立总体均值μ的置信区间
$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
2. 如果置信区间包含假设值μ_0，则不能拒绝H_0；否则，拒绝H_0

我们也可以应用到单侧检验，建立单侧置信区间，但是这种应用比较少。

常用的p-值指导

p值越小，则拒绝H_0的证据越多，从而支持H_α的证据越多。

p-值小于0.01

强有力的证据断定H_α为真

p-值介于0.01~0.05

有力的证据断定H_α为真

p-值介于0.05~0.10

弱的证据断定H_α为真

p-值大于0.10

没有足够的证据断定H_α为真

课堂练习

CNN和AcMedia提供了一个专门的电视频道，向那些在超市收银台前等待结账的顾客播放新闻、短讯和广告。假定顾客在超市收银台前等待结账时间的总体均值为8分钟，并以此为依据决定电视节目的长度。由实际等待时间组成一个样本，并利用样本进行检验，从而判断实际等待时间的均值与假设之间是否存在差异。

1. 提出这一应用的假设
2. 由120名购物者组成一个样本，等待时间的均值为8.4分钟。假定总体标准差为3.2分钟，求p-值
3. 在显著性水平0.05下，你的结论是什么
4. 计算总体均值的95%置信区间。它支持你的结论吗？

解答

1. $H_0: \mu = 8 \quad H_\alpha: \mu \neq 8$
2. 检验统计量$z = \frac{8.4-8}{3.2/\sqrt{120}}$=1.3693，对应的曲线上侧面积为0.08545，那么p-value = 2 *0.08545 = 0.171
3. p-value=0.171 > 0.05，因此不能拒绝原假设，也就是说实际等待时间的均值与假设之间不存在差异
4. 置信区间 [8.382, 8.418] 包含了8.4，因此不能拒绝原假设

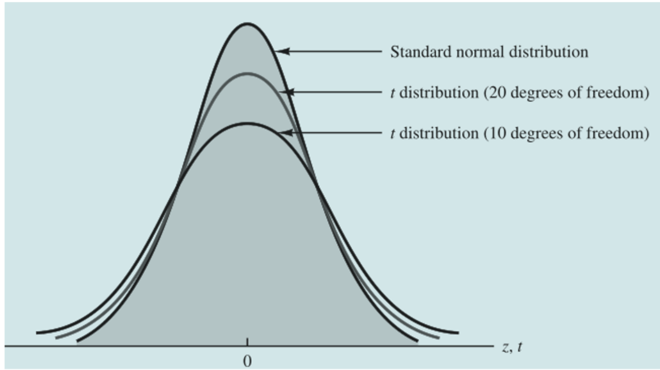
03

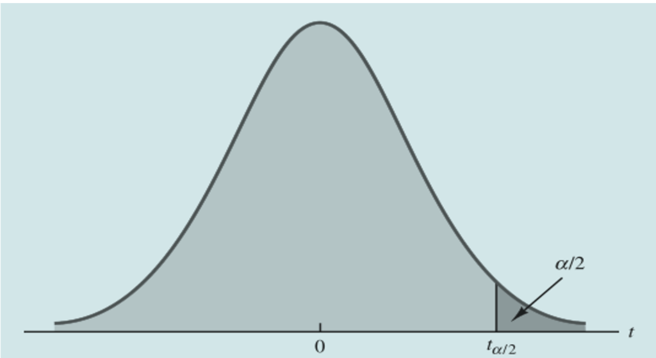
总体均值的检验

总体标准差未知情形

21

t 分布回顾





给t加上下标表明其在t分布上侧的面积

σ未知

在建立总体均值的区间估计时，我们通常并没有关于总体标准差一个号的估计。在这种情形下，我们必须利用同一样本估计μ和σ两个参数。

t 分布

定义：在概率论和统计学中，t-分布（t-distribution）用于根据小样本来估计呈正态分布且方差未知的总体的均值。如果总体方差已知（例如在样本数量足够多时），则应该用正态分布来估计总体均值。

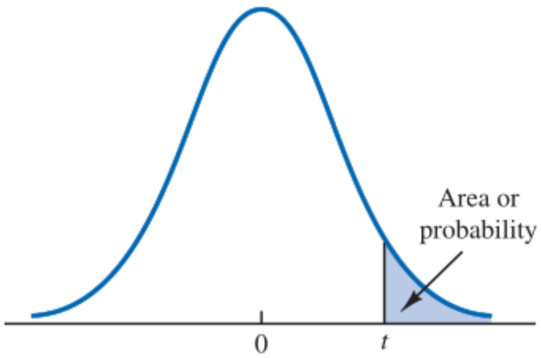
当利用s估计σ时，边际误差和总体均值的区间估计都以t分布的概率分布为依据进行的。虽然t分布的数学推导是以假设总体服从正态分布为依据的，但是许多研究表明在总体分布偏态的情形下，t分布效果也相当不错。

t 分布依赖于自由度的参数。当自由度为1,2,3,...时，有且仅有唯一的t分布与之相对应。随着自由度的增大，t分布与标准正态分布之间的差别变得越来越越小

t分布的均值为0

22

t分布表回顾



Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	.848	1.296	1.671	2.000	2.390	2.660
61	.848	1.296	1.670	2.000	2.389	2.659
62	.847	1.295	1.670	1.999	2.388	2.657
63	.847	1.295	1.669	1.998	2.387	2.656
64	.847	1.295	1.669	1.998	2.386	2.655
65	.847	1.295	1.669	1.997	2.385	2.654
66	.847	1.295	1.668	1.997	2.384	2.652
67	.847	1.294	1.668	1.996	2.383	2.651
68	.847	1.294	1.668	1.995	2.382	2.650
69	.847	1.294	1.667	1.995	2.382	2.649
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	.846	1.291	1.662	1.987	2.368	2.632
91	.846	1.291	1.662	1.986	2.368	2.631
92	.846	1.291	1.662	1.986	2.368	2.630
93	.846	1.291	1.661	1.986	2.367	2.630
94	.845	1.291	1.661	1.986	2.367	2.629
95	.845	1.291	1.661	1.985	2.366	2.629
96	.845	1.290	1.661	1.985	2.366	2.628
97	.845	1.290	1.661	1.985	2.365	2.627
98	.845	1.290	1.661	1.984	2.365	2.627
99	.845	1.290	1.660	1.984	2.364	2.626
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

23

总体均值的检验:σ未知情形

总体均值假设检验的检验统计量：σ未知情形。
服从t分布

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

案例

某杂志对大西洋通道的机场服务进行评分，最低分为0分，最高分为10分。评分的总体均值超过7的机场将被认为提供了优质服务。选取了希思罗机场60名客人组成的一个样本，样本均值为7.25分，样本标准差为1.052分。样本数据能否表明希思罗机场提供了优质的服务？

```
from scipy.stats import t
import numpy as np
p_value = 1 - t.cdf(t_value,df)
```

建立假设

$$H_0: \mu \leq 7$$
$$H_\alpha: \mu > 7$$

显著性水平设定为0.05

检验统计量

$z = \frac{x-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{7.25-7}{1.052/\sqrt{60}} = 1.84$, 自由度为60-1=59

p-value = 0.0354，小于显著性水平0.05.因此我们可以拒绝原假设。

临界值法

对应自由度59，上侧面积为0.05的临界值$t_{0.05} =$ 1.671。由于t=1.84>1.671，所以拒绝原假设。

24 总体均值的检验:σ未知情形

总体均值假设检验的检验统计量：σ未知情形。
服从t分布

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

案例

Holiday Toys生产产品并通过1000多家的零售商分销其产品。在即将到来的冬季制定生产规模计划时，企业必须在知道零售层面的实际需求量前确定每种产品的生产数量。对本年度的一款最新玩具，企业负责人预计平均每家零售商的需求量为40个。在根据这一估计做出最后的生产决策之前，Holiday Toys决定对25个零售商组成的样本进行调查。样本均值为37.4，标准差为11.79。

```
from scipy.stats import t
import numpy as np
p_value = 1-t.cdf(t_value,df)
```

建立假设

$$H_0: \mu = 40$$
$$H_\alpha: \mu \neq 40$$

显著性水平设定为0.05

检验统计量

$z = \frac{x - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$, 自由度为25-1=24

由于这是一个双侧检验，因此p-值是t分布曲线$t \le -1.10$部分面积的2倍。计算得出p-value = 0.1411*2=0.2822。在显著性水平0.05的情形下，我们不能拒绝原假设

临界值法

双侧检验的临界值为$t_{-0.025} = -2.604$和$t_{0.025} = 2.604$。检验统计量的拒绝法则为:

如果$t \le -2.604$或者$t \ge 2.604$，则拒绝H_0

检验统计量t = -1.10，所以不能拒绝原假设

25

总 结

总体均值的检验:σ未知情形

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0\colon \mu \geq \mu_0$ $H_a\colon \mu < \mu_0$	$H_0\colon \mu \leq \mu_0$ $H_a\colon \mu > \mu_0$	$H_0\colon \mu = \mu_0$ $H_a\colon \mu \neq \mu_0$
Test Statistic	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

04

第 二 类 错 误

27 假设检验与决策

显著性检验

显著性检验控制了发生第一类错误的概率，但没有控制发生第二类错误的概率。因此得出的结论是“不能拒绝H_0”，而不是“接受H_0”。**因为后者会使我们在H_0为假时接受H_0，从而承担发生第二类错误的风险**

第二类错误

如果假设检验的目的是当H_0为真时，做出一种决策；当H_α为真时做出另一种决策，这时统计学家建议控制第二类错误的概率。当同时控制第一类错误和第二类错误的可能性时，假设检验的结论要么是“接受 H_0”或者拒绝“H_0”

案例

对供应商的一批电池，某名质量控制管理人员必须决定是接受这批货物，还是因为其质量差而将货物退还。假定设计规格中要求这电池的平均使用寿命至少为120小时，为了评估这批货物的质量，我们选取36节电池组成样本进行检验。

对总体均值建立如下形式的假设检验：

$$H_0 : \mu \geq 120$$

$$H_\alpha : \mu < 120$$

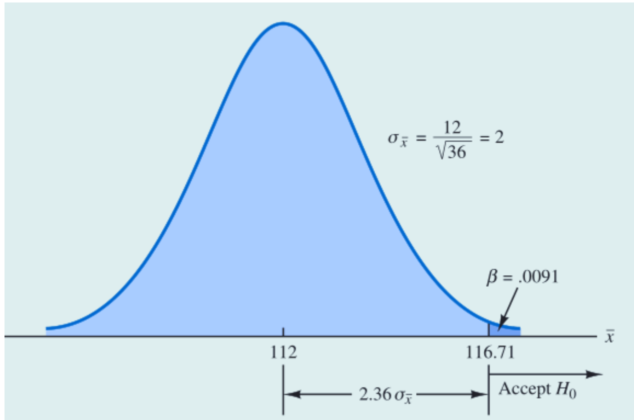
如果拒绝H_0，我们可以做出将货物退还的结论。但是如果不能拒绝H_0，决策者仍需要确定采取某种措施，因此我们需要控制第二类错误发生的概率，从而帮助决策者做决断！

28

计算第二类错误的概率

假定假设检验的显著性水平为0.05，在σ=12小时已知的情形下，检验统计量为

$$z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{\bar{x}-120}{12/\sqrt{36}}$$



拒绝规则

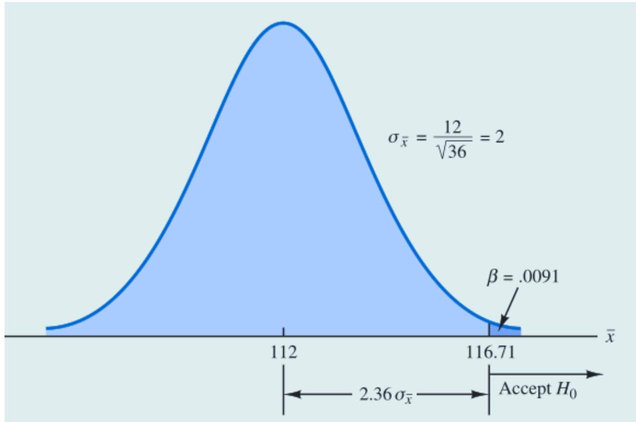
根据临界值法以及$z_{0.05}$=1.645，下侧检验的拒绝法则为：如果$z\le-1.645$，则拒绝原假设。

当$z=\frac{x-120}{12/\sqrt{36}}\le-1.645$ ，则可以拒绝H_0。求解$\bar{x}$ 可知$\bar{x}\le116.71$时，拒绝H_0

第二类错误

当电池的真值小于120小时而我们却做出接受H_0：$\mu\ge120$的决定时，我们就犯了第二类错误。因此我们必须选择一个小于120小时的μ值来计算发生第二类错误的概率。假定我们选取了μ=112小时，但我们却接受了H_0：$\mu\ge120$，因此发生第二类错误的概率是多少呢？它应该是当μ=112时样本均值大于116.71的概率。

29 计算第二类错误的概率



Value of μ	$z = \frac{116.71 - \mu}{12/\sqrt{36}}$	Probability of a Type II Error (β)	Power $(1-\beta)$
112	2.36	.0091	.9909
114	1.36	.0869	.9131
115	.86	.1949	.8051
116.71	.00	.5000	.5000
117	−.15	.5596	.4404
118	−.65	.7422	.2578
119.999	−1.645	.9500	.0500

计算概率

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{116.71 - 112}{12/\sqrt{36}} = 2.36$$

由标准正态概率分布表可知，当z=2.36时其上侧面积为1-0.9909=0.0091. 用β代表第二类错误发生的概率。我们可以得出结论，如果总体均值为112小时，则发生第二类错误的概率为0.0091.

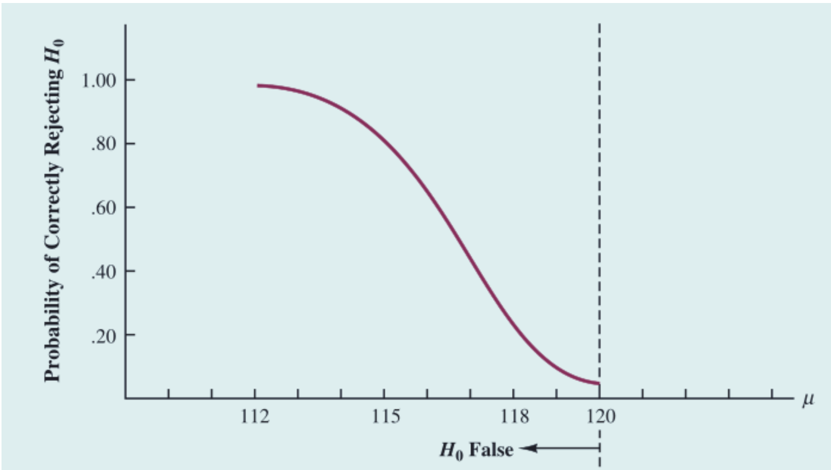
其它总体均值

从左边表中可以看出，随着μ逐渐增加到120，发生第二类错误的概率也随之增大，达到上限0.95。反之，μ越小，概率越小。当总体均值的真值在原假设的值μ=120附近时，发生第二类错误的概率很大。

30 计算第二类错误的概率

当H_0为假时，得出拒绝H_0的正确结论的概率被称作检验的**功效**(power)。对于给定的μ值，功效为$1-\beta$，即得出拒绝原假设正确结论的概率等于1减去发生第二类错误的概率。
对任意一个μ值，功效曲线的高度代表了当H_0为假时得出拒绝H_0这个正确结论的概率

功效曲线



01 确定原假设与备择假设

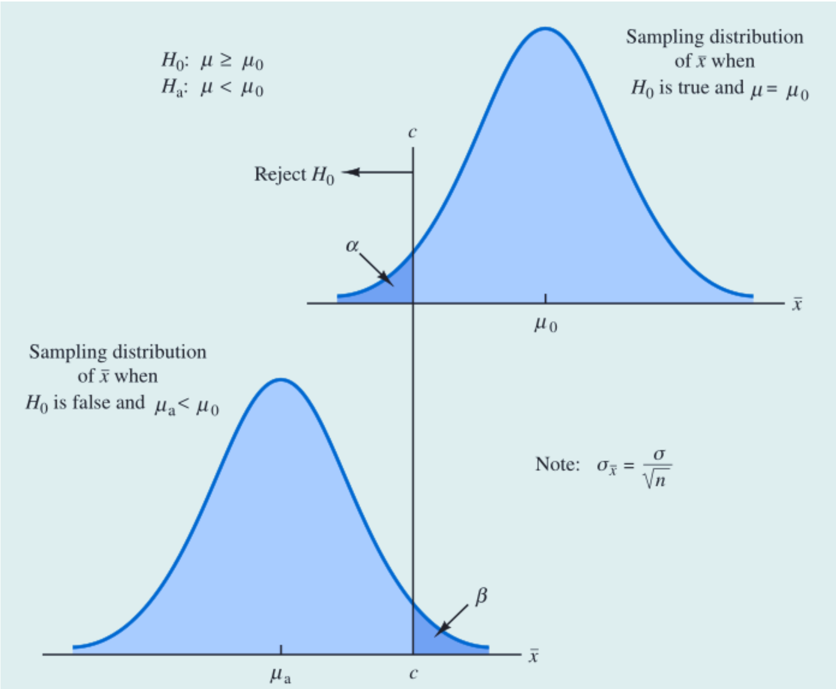
02 在显著性水平α下，根据临界值法确定临界值并建立检验的拒绝原则

03 利用2中所得的拒绝法则，求解与检验统计量的临界值相对应的样本均值的取值

04 利用3中的结果，得到接受H_0时所对应的样本均值的值，这些值构成了检验的接受域

05 对于满足备择假设的μ值，利用样本均值的抽样分布和步骤4中的接受域，计算样本均值落在接受域的概率。这一概率值既定为在选定μ值处发生第二类错误的概率

31 对总体均值进行假设检验时样本容量的确定



第一类错误

当原假设为真。对于下侧检验，检验统计量的临界值记作$-z_\alpha$，图中上半部分中垂线c与$\bar{x}$相对应。如果当$\bar{x} \leq c$时拒绝H_0，则发生第一类错误的概率为α。以z_α表示标准正态分布上侧面积为α时所对应的的z值，可利用如下公式计算c

$$c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

第二类错误

当备择假设H_α为真。阴影区域的面积β恰好是当$\bar{x} > c$却接受了原假设时决策者发生第二类错误的概率。以z_β表示标准正态分布上侧面积为β时所对应的z值。

$$c = \mu_\alpha + z_\beta \frac{\sigma}{\sqrt{n}}$$

对总体均值进行假设检验时样本容量的确定

由于我们想要选取一个c值是的当拒绝H_0接受H_α时，发生第一类错误的概率等于选取的值α，发生第二类错误的概率等于选取的值β。因此上述两式所给出的c值是相等。于是，如下方程一定成立

$$\mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_\alpha + z_\beta \frac{\sigma}{\sqrt{n}}$$

求解得出$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_\alpha)^2}$

总体均值单侧假设检验的样本容量

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_\alpha)^2}$$

z_α代表标准正态分布的上侧面积为α时对应的z值；z_β代表标准正态分布的上侧面积为β时对应的z值；σ为总体的标准差；μ_0为假设检验中样本均值的值；μ_α为第二类错误中所采用的总体均值的值。

如果是双侧检验，使用$z_{\alpha/2}$代替z_α

对总体均值进行假设检验时样本容量的确定

案例

回顾货物接收的案例。设计规格要求电池使用寿命的均值至少为120小时。如果H_0：$\mu \geq 120$被拒绝，则该批货物被拒收。我们假定质量控制管理人员要求第一类错误和第二类错误的可接受概率达到如下要求：

- 关于第一类错误：如果货物中电池寿命的均值为120，那么甘愿冒0.05的风险概率拒绝该货物。
- 关于第二类错误：如果货物中电池寿命的均值比规则要求少5小时，那么甘愿冒β=0.10的风险概率接受这批货物。

利用正态概率表分布表：

$$z_{0.05}=1.645，z_{0.10}=1.28$$

因此在接受货物的例子中，我们建议使用的样本容量为

$$n = \frac{(1.645+1.28)^2 12^2}{(120-115)^2} 49.3$$

我们可以观察到α，β，样本容量n之间的如下三种关系：

1. 当三者中有二者已知时，即可计算得到第三者
2. 对于给定的显著性水平α，增大样本容量将会减少β
3. 对于给定的样本容量，减小α将会使β增大，相反增大α将会使β减小

当未对第二类错误的概率加以控制的时候，我们应该牢记第三条，它说明不能毫无必要的选择太小的显著性水平α。对于给定的样本容量，选择较小的显著性水平意味着将发生第二类错误的风险增大！