介绍

A/ B测试

治疗组（无广告）的购买量比对照组（有广告）多43.4%

认为“展示一个广告会增加销售”的直觉是错误的

这个结果在统计上有意义还是偶然？

控制处理

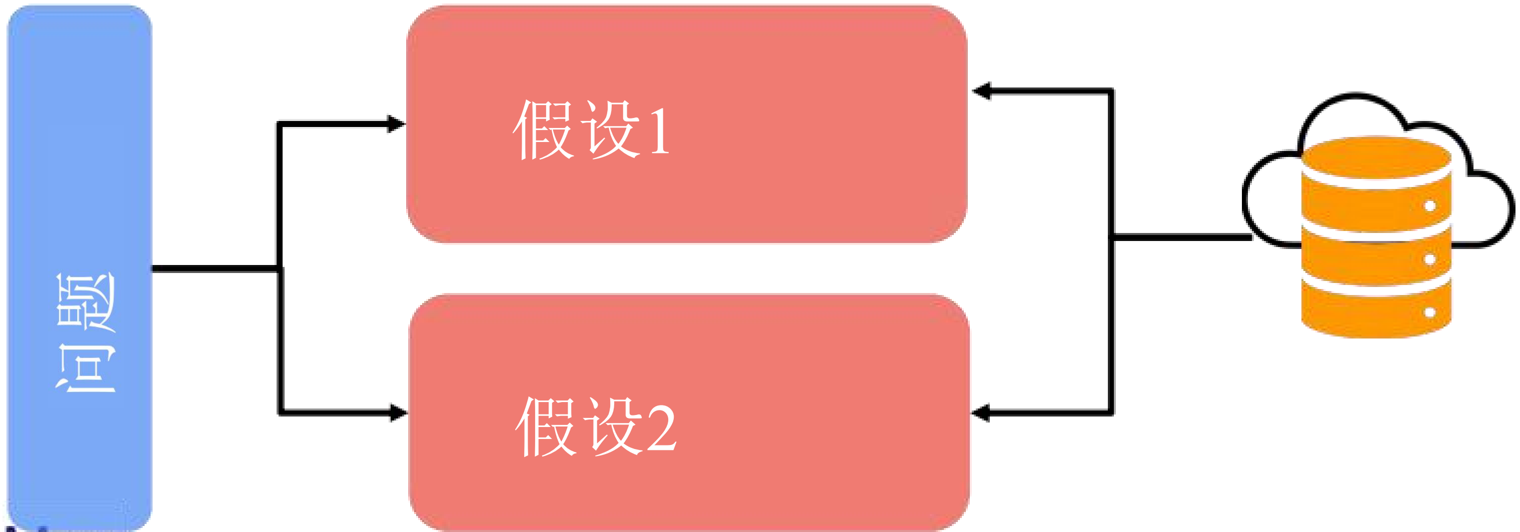
介绍

统计假设检验是一种统计推理的方法，用于决定手头的数据是否充分支持特定的假设。假设检验允许我们做出结论
关于总体参数的概率陈述。

-将一个问题转化为假设。

-收集数据来证明哪个假设很可能是正确的。

假设检验和置信区间允许使用样本数据来得出关于整个人群的结论。



介绍

样例

想象一下，你和你的女朋友正在一场关于世界上最受欢迎的冰淇淋口味是什么的辩论中。

Subarby假设：

- H !当前位置巧克力是最受欢迎的口味
  香草是最流行的口味

我们如何才能真正知道这个问题呢除非我们和世界上的每个人说话？我们如何知道我们的结论是可靠的？

使用假设检验，仅使用样本数据得出关于单个种群的结论

Queen Mary

设置组合试验

设置假设测试

建立零假设和替代假设的一些一般规则

问题

#

!空

的

- The#!在你收集任何数据之前都是真的。
- The#!通常是状态下没有两组相等的效应器。

•#!包含某种等量符号：=，≤，≥

的#！和#1是相互竞争的、不重叠的假设。

#：

备选方案

#1是我们想要证明的事实。

#1包含了空值的反对意见：=，>，<

例如：无辜的，直到被证明有罪

- H !无辜的
- H1：有罪

Queen Mary University of London

设置假设测试

练习：

想象一下，你创建了一个新的网页布局，你想知道这个新页面是否比现有的页面吸引了更多的流量。你如何建立零假设和替代假设？

Go to

www.menti.com

Enter the code

1939 7765

Or use QR code

Queen Mary University of London

错误类型

讨论：为什么这些假设和替代假设很重要？

在假设检验中，可能存在两种类型的错误：

类型I错误和类型II错误。

类型的错误通常是被认为是最糟糕的错误类型。

仅仅基于这些信息，你就能猜出每种类型的错误的真相/决策组合吗？

		真相	
		有罪	无辜的
决定	有罪		
	无辜的		

Go to

www.menti.com

Enter the code

1939 7765

Or use QR code

Queen Mary University of London

错误类型

Subarbl型错误

-错误地拒绝一个实际上是测试程序的结果的零假设。（决定替代方案（H 1）实际（H！）这是正确的。)。

-也被称为“假阳性”

-I型错误率表示为byα

-例子：一个无辜的罪犯被定罪

SubabII型错误

-未能拒绝一个与测试程序的结果相比实际上是错误的零假设。（决定null（H ！）（H1）是真的。

-也被称为“假阴性”

-第二型错误率表示为byβ

-例子：一个罪犯未被定罪

错误率

你应该设置你的零假设和替代假设，这样你最糟糕的错误是第一类错误。

类型I错误率是拒绝空值的概率

假设它是正确的。这个测试的设计目的是为了保持类型I的错误率低于一个预先指定的边界，称为显著性水平，通常用α表示，也称为阿尔法水平。通常，α=为0.05（5%）

应用，这意味着有5%的概率错误地拒绝真正的零假设是可以接受的（医学上的1%）。

二型错误率由theβ表示，与检验的幂有关，等于1−β

这两种类型的错误率是相互权衡的：对于任何给定的样本集，努力减少一种类型的错误率通常会导致增加另一种类型的错误率。

总结

错误类型表

表错误类型		零假设（H 0）为 真	零假设（H 0）为 假的
决定关于空假设（H 0）	拒绝失败	正确的推理（真负值）（概率= 1−α）	II错误（假阴性）（概率=β）
	拒绝	I类错误（假阳性）（概率= α）	正确的推理（真阳性）（概率= 1−β）

图像来源： https://en.wikipedia.org/wiki/Type Iand_type_II_errors#类型I错误

错误类型

案例研究：一个新的市场宣传活动

考虑到我们有兴趣测试一个新营销活动的平均收入是否比之前的营销
更好
运动

我们用以下方法建立了零假设和备选假设：

– H !:μn − μ$≤ 0
– H1 :μn −μ$ > 0

其中μ$是旧竞选的平均收入，andμ是新竞选的主要收入。

零假设是，新活动的平均收入低于或等于旧活动的平均收入。

另一种假设是，新活动的平均收入大于旧活动的平均收入。

第一类错误：决定新页面更好，但实际上旧页面更好。
第二类错误：决定旧页面更好，但实际上新页面更好。

错误类型

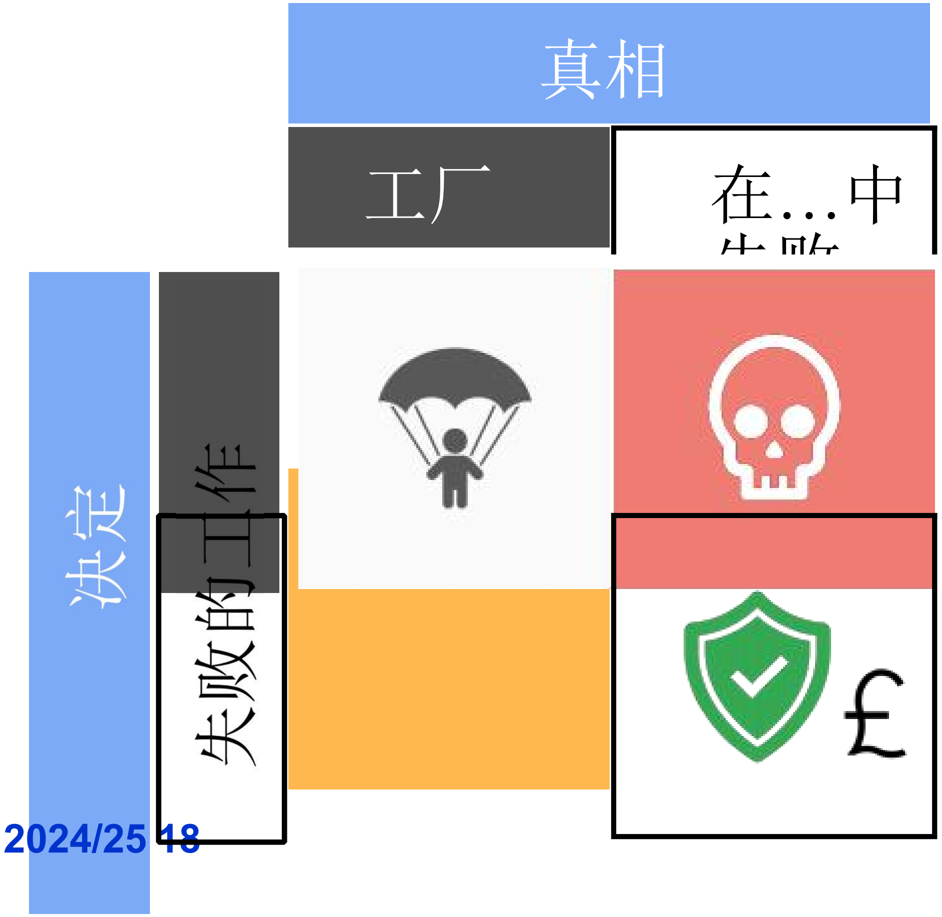
案例研究：天空业务（一个极端情况）

想象一下，你拥有一家跳伞公司。你必须检查降落伞，以确保它们能正常工作。
有两种潜在的结果，一是降落伞
工作或不工作。你检查每个降落伞来做决定。

如果你确定它不起作用，你就把它扔掉
。
降落伞真的不起作用，那太好了。
降落伞真的很有效，你损失了50英镑

如果你确定它有效，你就把它留给跳伞者使
用。
降落伞确实工作，跳伞者安全着陆
降落伞真的了，
可能出现的最严重的错误类型。-£50

讨论：α=1%足够吗？



假设检验的常见类型

测试填充参数

-例如，测试所有JP学生每周平均玩游戏的时间是否为10个小时，而调查数据显示JP学生玩游戏的时间（双尾测试）。

- H !:$\mu = 10$
- H1 :$\mu \neq 10$

-或者，测试所有JP级选手的每周平均游戏时间是否超过每周10个小时（单尾测试）。

- H !:$\mu \leq 10$
- H1 :$\mu > 10$

-一个样本t检验/z检验。

假设检验的常见类型

检验不同人群中参数的差异

-测试所有JP男学生每周平均玩游戏的平均时间是否与所有JP女学生相同。（双尾试验）

H o：μ男孩，=μ女孩

H 1：μ男孩，=μ女孩

-或者，测试是否所有男性JP每周的平均游戏时间学生比所有JP女学生都要长（单尾测试）。

H o：μ男孩≤μ女孩

H 1：μ男孩，>μ女孩

-双样本t检验/z检验

假设检验的常见类型

对同一个人进行某些治疗前后的差异（配对t检验）

-例如，测试罪犯在有女朋友后每周玩游戏的平均时间是否会减少。

- H !：每个男学生女朋友之前和之后的游戏时间没有显著差异（μd=0，其中μd表示差异的总体平均值）。

H1：在每个男性学生有女朋友之前和之后，游戏玩的时间有显著的差异（μd=0，其中μd表示差异的人口平均值）。

配对t检验

有很多不同的假设检验

假设检验总是在总体参数上，而不是在统计数据上

单样本测试

单样本测试

单样本检验是一种统计假设检验，用于

确定单个样本的平均值是否与已知或假设的总体平均值有显著不同。

-它通常是当你有大量的数据时使用的，并且你想评估这个样本是否提供了足够的证据来拒绝关于总体均值的零假设。

单样本检验的常用方法是单样本z检验和单样本t检验。

单样本z测试

例如：大学生的睡眠时间

假设我们对大学生进行了调查，收集了一些关于他们大学生活的数据。

-76名学生，22种信息类型

在这个例子中，我们将重点关注睡眠大学生得到的数量。

-有一名学生的睡眠时间数据缺失。

ID	Gender	Classification	Height	Shoe Size	Phone Time	# of Shoes	Birth order	Pets	Happy	Funny	College
1	male	senior	67.75	7	12	12	youngest	5	0.8	7	Natural Sciences
2	male	freshman	71	7.5	1.5	5	middle	4	0.75	8	Natural Sciences
3	female	freshman	64	6	25	15	oldest	8	0.9	6	Natural Sciences
4	female	freshman	63	6.5	30	30	middle	12	0.98	9	Education
5	male	senior	69	6.5	23	8	oldest	4	0.75	6	Natural Sciences
6	female	senior	64	8.5	13	25	oldest	1	0.95	5	Natural Sciences
7	female	freshman	62	8.5	23	12	oldest	2	0.95	7	Nursing
8	female	freshman	64	6	50	50	youngest	10	0.9	4	Liberal Arts
9	female	freshman	66	8	10	15	youngest	25	0.9	9	Natural Sciences
10	female	freshman	68	6.5	40	20	oldest	4	0.95		Nursing

ID	Bfast Calories	Exercise	Stat Pre	Stat Post	Phone Type	Sleep	Social Media	Impact of SocNetworking	Political	Animal	Superhero
1	500	360	3		iPhone	7	180	worse	Democrat	Dog person	Batman
2	0	200	9		Android smartphone	7	20	better	Democrat	Dog person	Batman
3	200	30	7	5	Android smartphone	8	60	better	Republican	Dog person	Batman
4	200	180	6	7	iPhone	6	60	better	Republican	Both	Superman
5	0	180	4	7	iPhone	5.5	60	worse	Independent	Dog person	Superman
6	250	310	7	7	iPhone	6.5	90	no impact	Democrat	Dog person	Batman
7	200	60	7	8	Android smartphone	7	120	better	Republican	Both	Superman
8	200	0	5	7	iPhone	7	60	no impact	Independent	Both	Superman
9	0	0	8	6	iPhone	7	3	no impact	Democrat	Both	Batman
10	150	240	6	10	iPhone	6	180	better	Democrat	I don't like either	Batman

数据来源： http://sites.utexas.edu/sos/guided/inferential/numeric/claim/one-sample-t/

单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

第一步：建立一个假设。

假设我们有兴趣知道大学生的平均睡眠时间是否比成年人一样多。我们从a中就知道了
报告称，成年人的平均睡眠时间为7.25小时。我们可以设置假设检验如下。

$$H\ !:\mu = 7.25$$
$$H\ 1: \mu \neq 7.25$$

单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

第二步：收集一个样本，并找出该样本的汇总统计数据

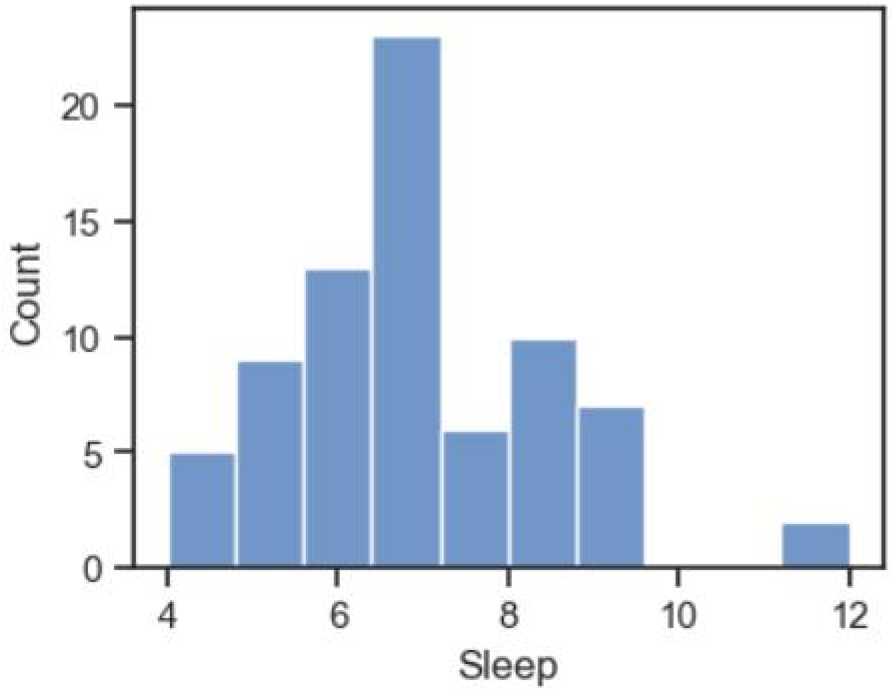
因为我们已经有了数据，让我们来探索样本数据。

-样本量=75

-样本平均值=6.873

-样本标准=1.538

-睡眠时间大致遵循正常分布



单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

步骤3：定义一个显著性水平

在我们开始执行我们刚刚建立的假设之前，我们需要定义一个显著的高度来表示我们可以指示的类型错误

接受在这种情况下，我们设置

$$\alpha = 0.05$$

单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

步骤4：生成原假设下感兴趣统计学的抽样分布

在一个假设检验中

-假设它为真的。

-当零假设为真时，估计一个样本是否具有类似于我们样本的平均值的可能性。

如果绝对不可能有一个平均值像我们的样本的平均值，我们会认为零可能不是真的（拒绝零支持替代）。

否则，我们认为零将是正确的（不能拒绝nu ll）。

为了做出这个决定，我们需要在零假设下生成样本均值的抽样分布

单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

步骤4a：生成空分布（z检验）

通常，我们认为样本si ze> 30很大。在这种情况下，我们可以使用S（样本标准差）作为总体标准deviationσ的估计值

$$SE \approx \frac{S}{\sqrt{n}}$$

根据CLT，在零假设下的均值的抽样分布（零分布）将近似为：

$$\mathrm{N}(\mu\ !, SE")$$

whereμ!是零假设下的平均值。

单样本z测试

例如：大学生睡眠时间的单样本双尾z检验

步骤4a：生成空分布（z检验）

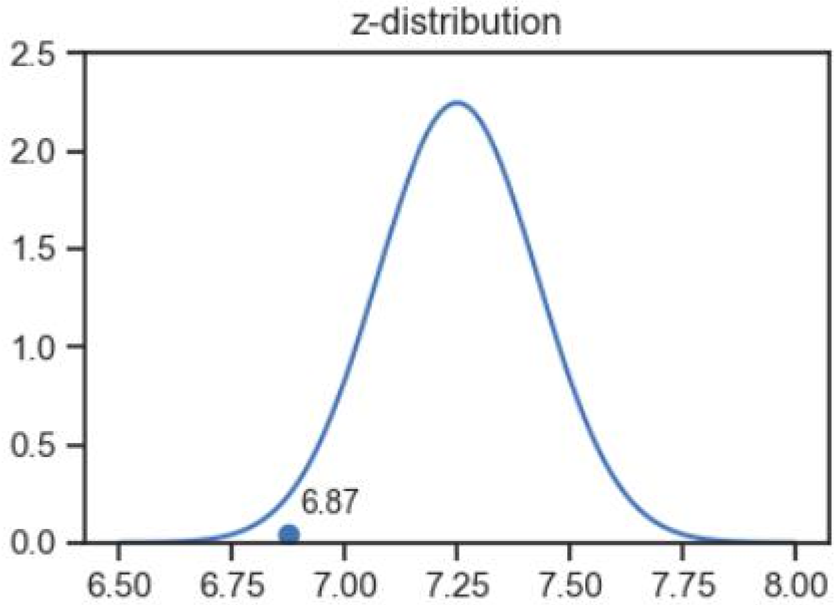
在这个例子中

$$SE \approx \frac{S}{\sqrt{n}} = \frac{1.538}{\sqrt{75}} = 0.1776$$

基于clt，第e次采样

零假设下的平均值的分布（零分布）将近似为：

$$N(7.25, 0.1776)$$



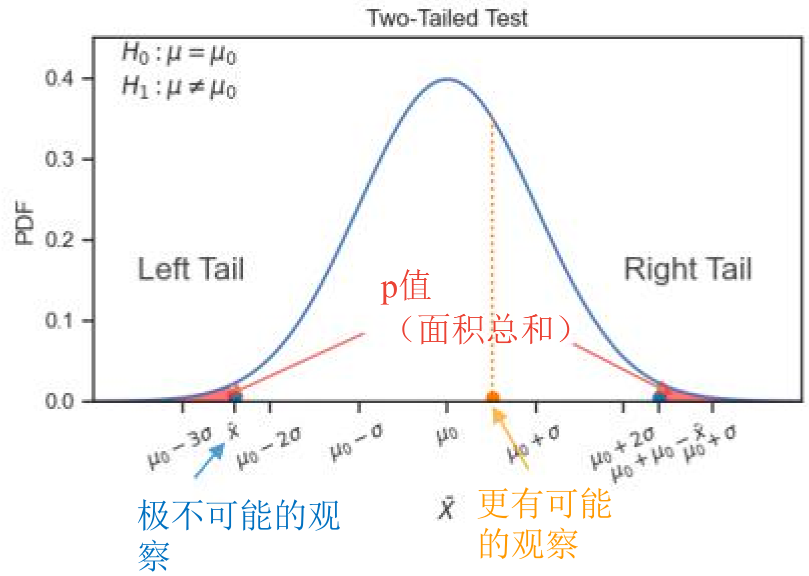
讨论：当空值为真时，我们如何确定一个随机样本的平均值为6.87的可能性？

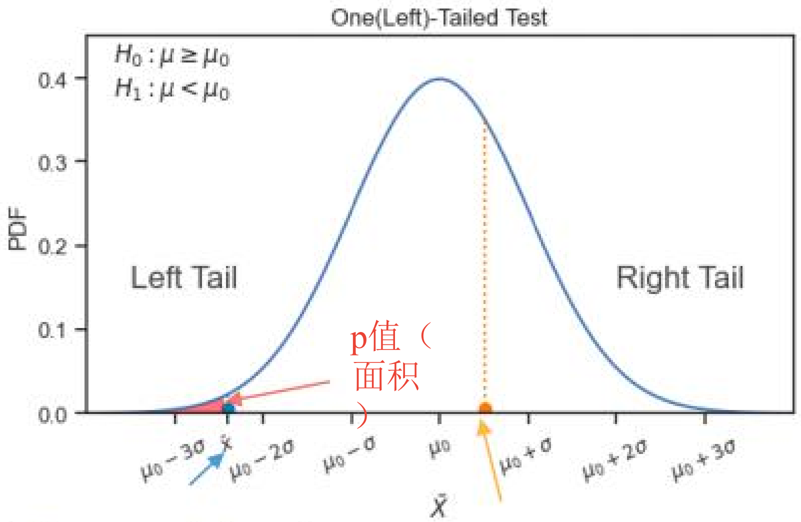
p值

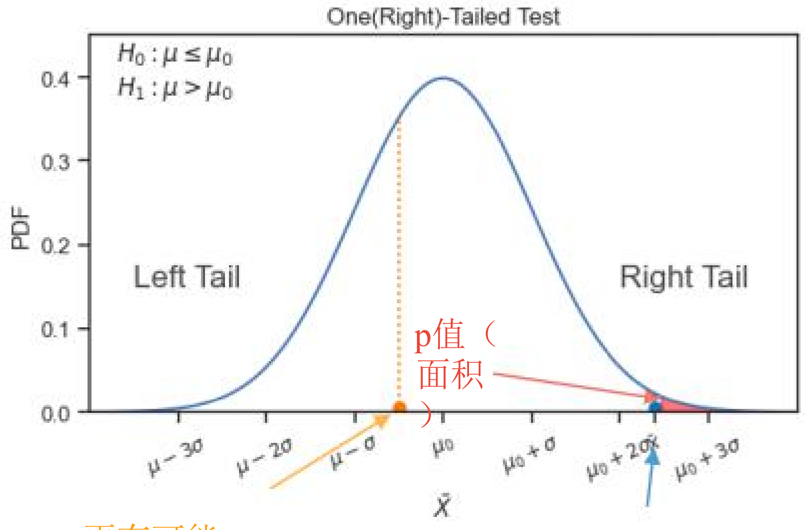
定义

这个概率，假设是H！检验统计量（例如，样本平均值）是真的吗 _ 会一个值当作极端或更多比实际观察到的更极端（即，更极端的支持替代假设）被称为检验的p值。

– “极端”的意思是“如果H0是真的，这远远不是我们的预期。”







p值

计算

在生成null（null分布）下的采样分布后，

可以计算p值

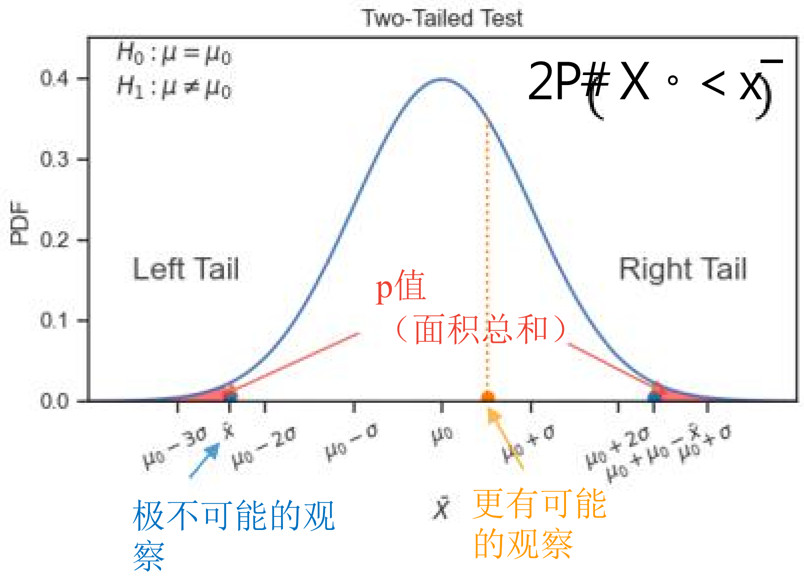
数字，哪里！是随机变量吗

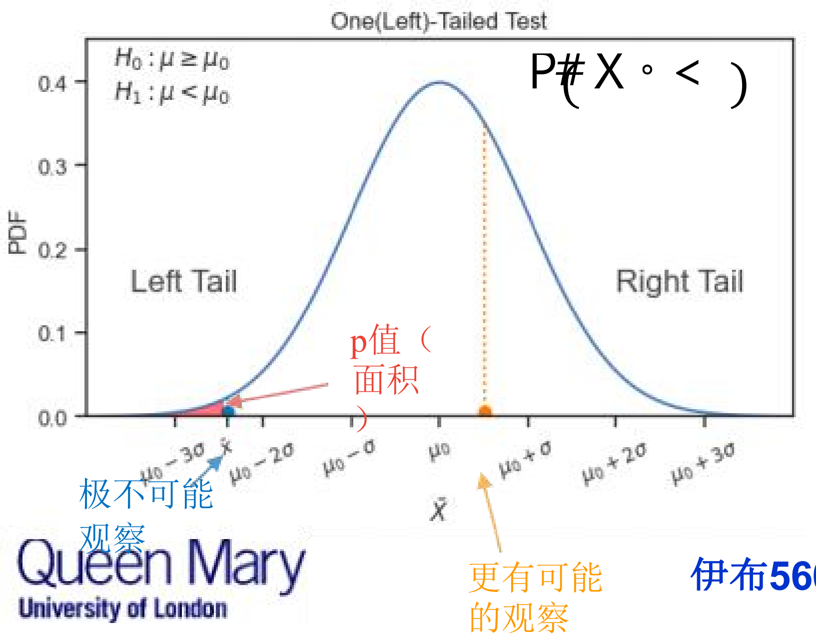
它遵循零分布，即

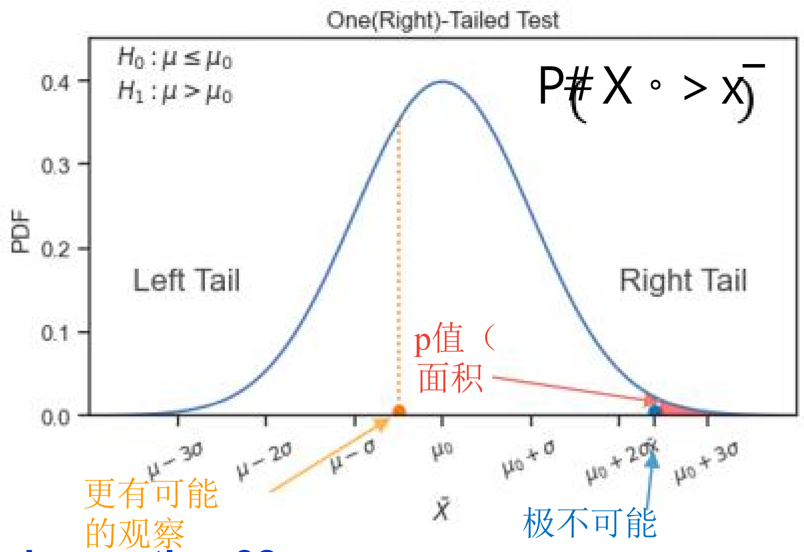
样本均值

-最大的价值，对#的大量支持。

-p值小，受#影响的系数系数强。







p值

一个较小的p值表明它不太可能从零值中观察到我们的统计量，也就是说，我们的统计量很可能来自于替代值，而不是零值。

当p值很大时，我们有证据表明我们的统计量很可能来自于零假设。因此，我们没有证据来拒绝无效。

做出决定/得出结论

- -如果p_值为≤α，我们拒绝为空，而支持替代方案
- -如果p_值为≥α，则我们无法拒绝空

单样本z测试（续）

例如：大学生睡眠时间的单样本双尾z检验

步骤5a：计算p值（z检验）

空分布N 7.25,0。“1776年”，在这个双尾测试中

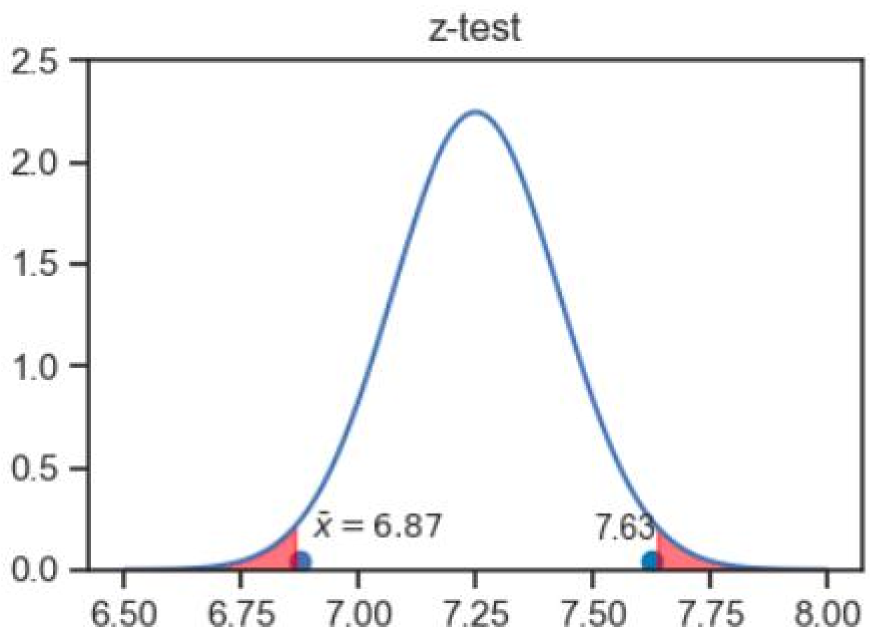
p-value=2P(8 X< 6.873).

沙痂z分数

$$z = \frac{\bar{x} - \mu_0}{SE} = \frac{6.873 - 7.25}{0.1776} = -2.12$$

通过参考z表，我们可以得到Pz<−2。12= 0.017，所以：

p−value= 2∗ 0.017= 0.034

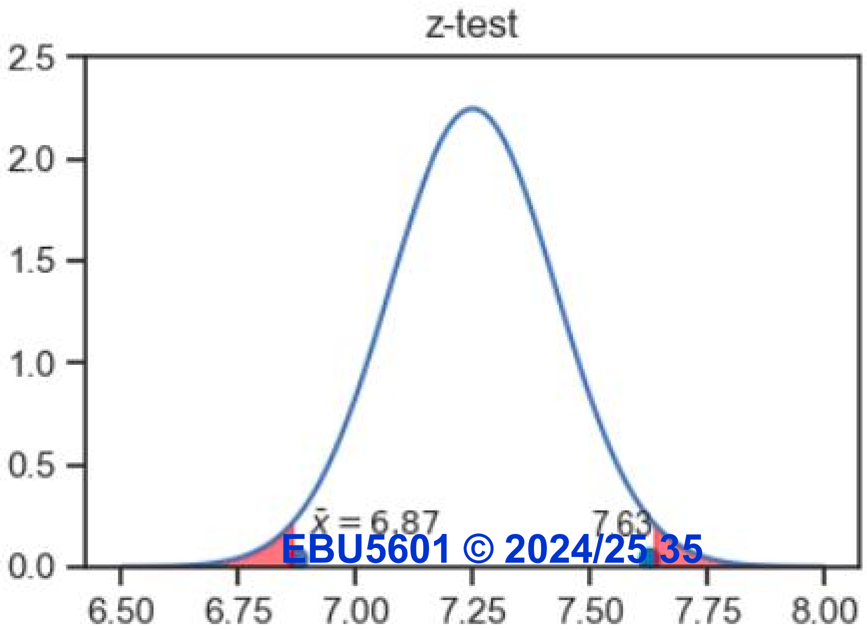


单样本z测试（续）

例如：大学生睡眠时间的单样本双尾z检验

步骤六：得出结论

因为p值=为0.034< e（=0.05），我们拒绝了无效的选择，即，有足够的证据支持大学生睡眠的小时数与7.25小时是不同的。



Queen Mary University of London

单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

第一步：建立一个假设。

假设我们现在想测试一下大学生的平均睡眠时间是否比成年人少。我们可以建立假设检验为
跟随

H !:μ≥ 7.25
H 1: μ< 7.25

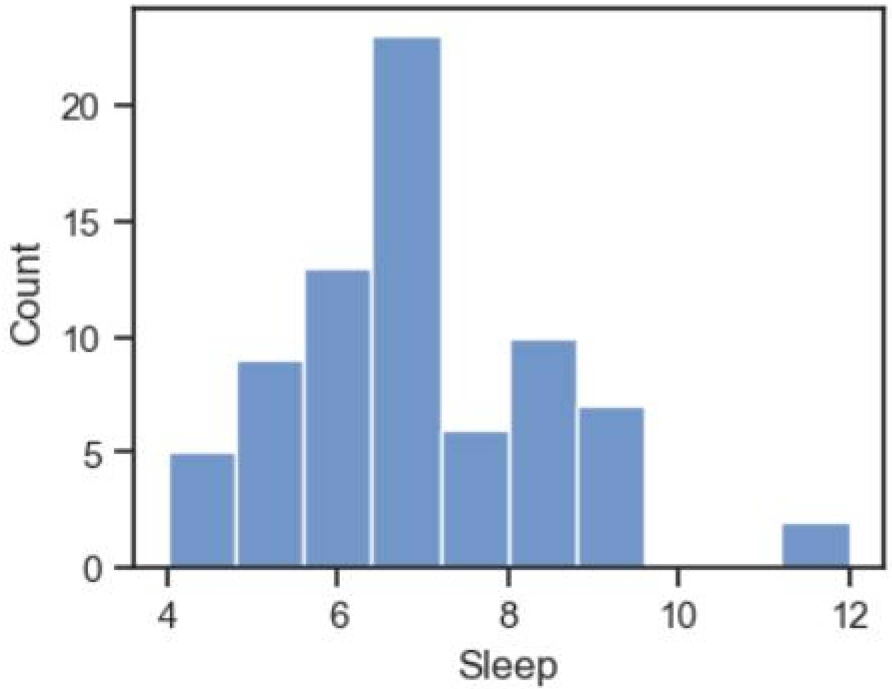
单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

第二步：收集一个样本，并找出该样本的汇总统计数据

我们已经有了这些数据和汇总的统计数据。

- -样本量=75
- -样本平均值=6.873
- -样本标准=1.538
- -睡眠时间大致遵循正常分布



单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

步骤3：定义一个显著性水平

我们保持显著性水平与最后的演示$\alpha = 0.05$相同

单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

步骤4a：生成空分布（z检验）

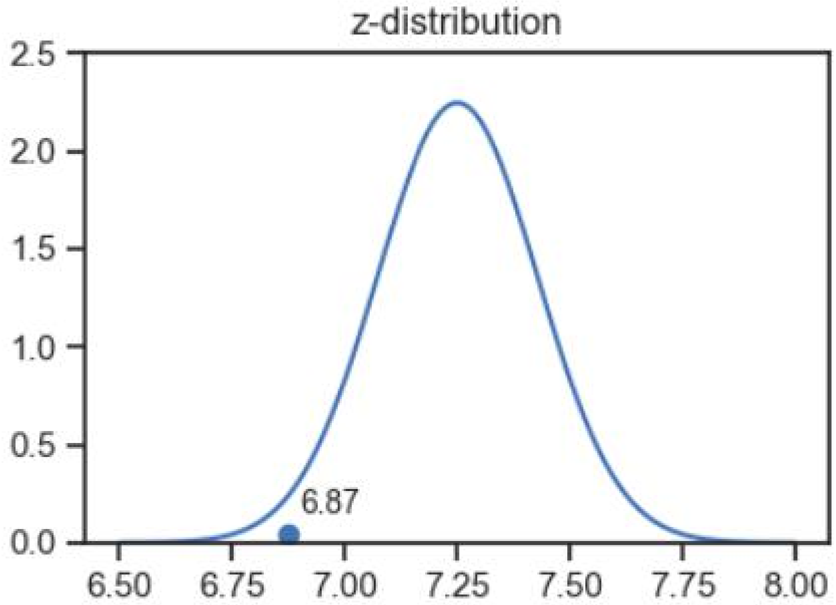
在这个例子中

$$SE \approx \frac{S}{\sqrt{n}} = \frac{1.538}{\sqrt{75}} = 0.1776$$

基于clt，第e次采样

零假设下的平均值的分布（零分布）将近似为：

N7.25, 0. 1776"



单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

步骤5a：计算p值（z检验）

空分布N 7.25,0。“1776年”，在这个单尾试验中

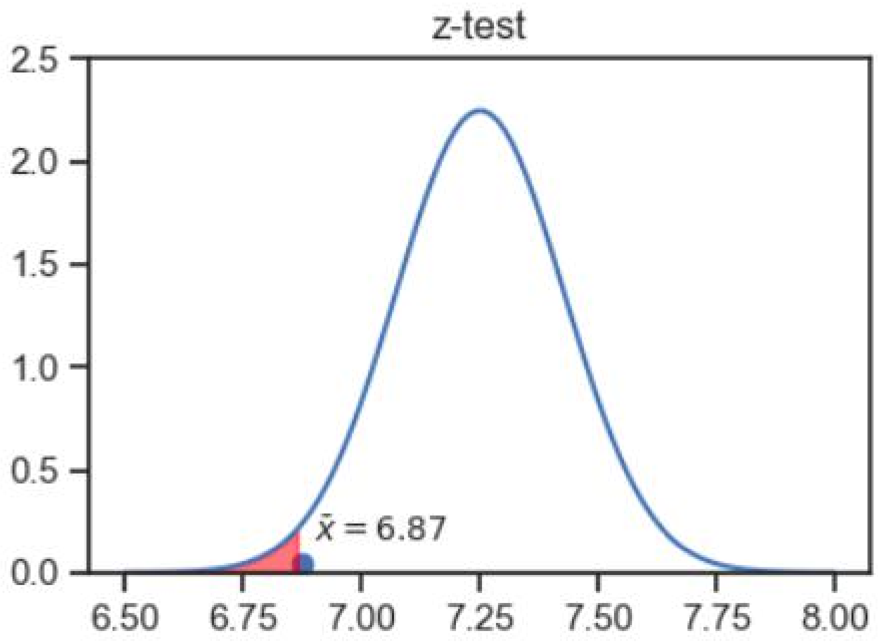
p-value=P(8 X< 6.873).

沙痂z分数

$$z = \frac{\bar{x} - \mu_0}{SE} = \frac{6.873 - 7.25}{0.1776} = -2.12$$

通过参考z表，我们可以得到Pz<−2。12= 0.017，所以：

p− value= 0.017



单样本z测试（续）

例如：大学生睡眠时间的单样本单尾z检验

步骤六：得出结论

因为p值=0.017<e（=0.05），我们拒绝了空值，也就是说，有足够的证据支持该主张

大学生的睡眠时间不到7.25小时。

概述：学生的T型分布

钟形，波峰较短，尾巴较粗。设计时要考虑到与小样本量相关的更大的不确定性。

t-分布描述了距离的可变性

在样本均值和总体均值之间

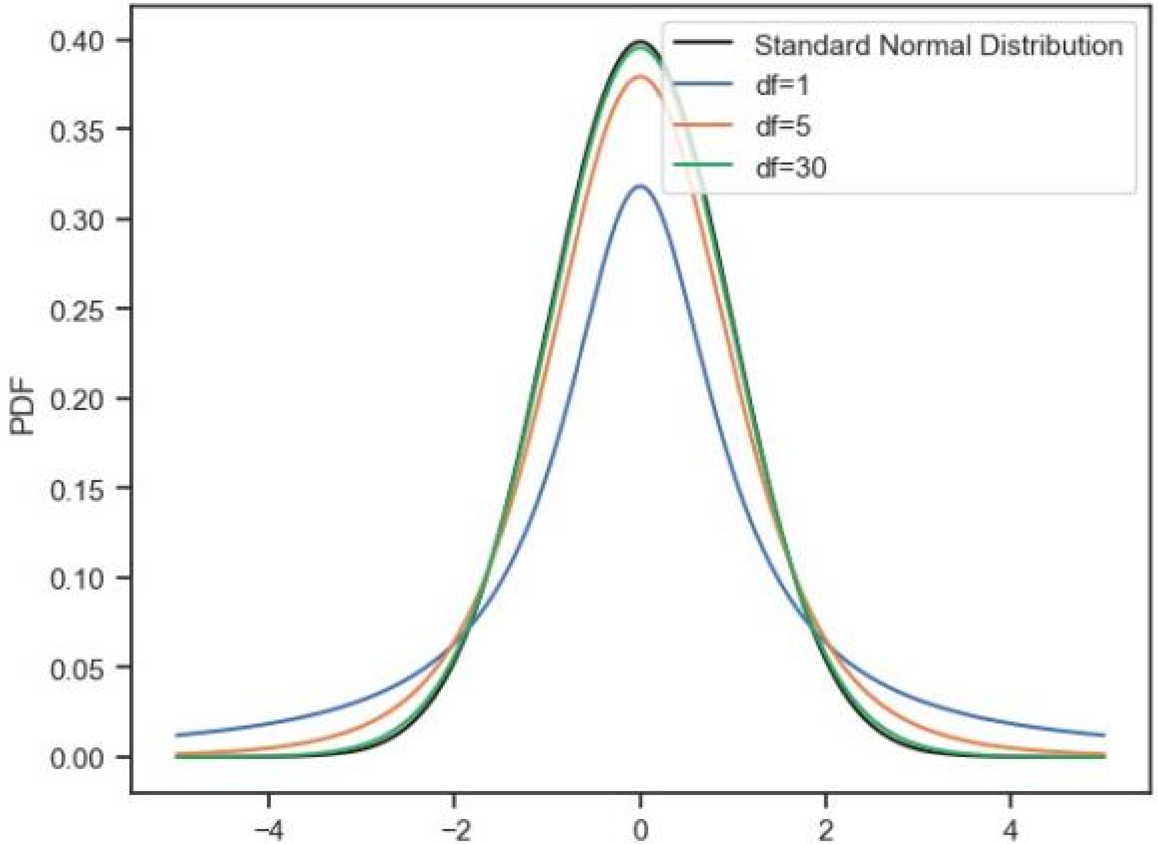
种群标准差未知，同质体近似服从正态分布。

这个分布只有一个参数，自由度，基于（但不等于）样本大小。

概述：学生的T型分布

自由度较小的t分布有厚尾和较低的峰。在大约30个自由度处，t-分布非常接近于标准正态分布分布（z分布）。

在开发中广泛使用
样本分布
统计量当你需要评估平均值时，你应该使用t分布
不了解人口
标准差
当你有一个小（n < 30）样本时，使用它就特别重要。



概述：学生的T型分布

从CLT中，我们知道该样本均值的抽样分布遵循正态分布

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

然而，大多数时候，人口标准deviationσ是未知的。如果样本量足够大，我们考虑

样本标准差s是一个很好的估计ofσ

$$Z \approx \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1), if\ n > 30$$

如果样本量较小，则不遵循标准正态分布，而是遵循具有特定自由度v的t-分布。(n < 30), $\frac{X-\mu}{s/\sqrt{n}}$

τ 2 ( )

where $T = \frac{\bar{X}-\mu}{\tau} \sim t_{\nu}$

自由度

在统计学中，自由度的数量是在最终计算中可以自由变化的统计量的值的数量。

一般来说，一个参数估计的自由度等于进入估计减去在参数本身估计的中间步骤中使用的参数数。

样例

如果变量需要从sizeN的随机样本中估计，那么自由度等于N减去作为中间步骤估计的参数数（样本平均值），因此等于N−1。

对于单样本t检验，需要花费一个自由度估计平均值，和剩余的n -1度距离估计变异性。

单样本t测试

例如：大学生睡眠时间的单样本单尾t检验

第一步：建立一个假设。

假设我们现在想测试一下大学生的平均睡眠时间是否比成年人少。我们可以建立假设检验为

跟随

H !:μ≥ 7.25

H 1: μ< 7.25

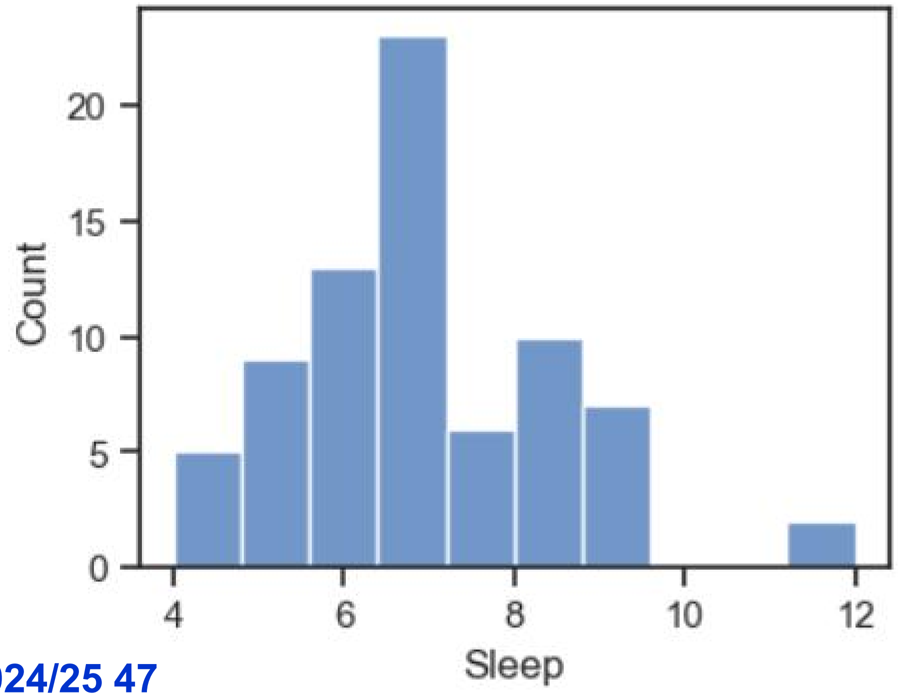
单样本t测试

例如：大学生睡眠时间的单样本单尾t检验

第二步：收集一个样本，并找出该样本的汇总统计数据

我们已经有了数据和汇总统计数据。

- -样本量=75
- -样本平均值=6.873
- -样本标准=1.538
- -自由度(v)=n−1=74
- -睡眠时间大致遵循正态分布



单样本t测试

例如：大学生睡眠时间的单样本单尾t检验

步骤3：定义一个显著性水平

我们保持显著性水平与最后的演示$\alpha = 0.05$相同

单样本t测试

例如：大学生睡眠时间的单样本单尾t检验

步骤4b：生成空分布（t检验）

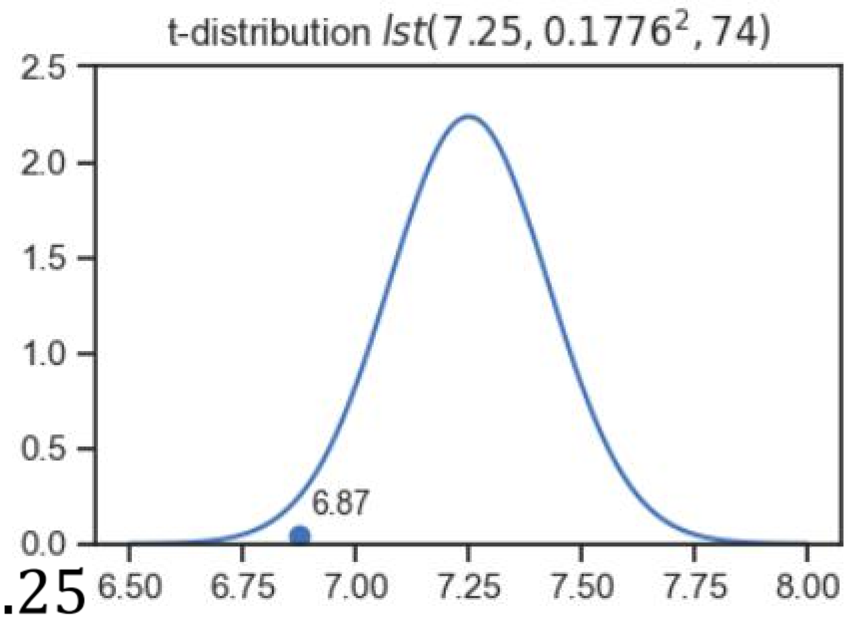
在这个例子中

$$SE = \frac{s}{\sqrt{n}} = \frac{1.538}{\sqrt{75}} = 0.1776$$

根据t分布的定义，在原假设下的抽样分布

分布)遵循一个位置规模的t-distributionlst(μ=7.25，τ2 = 0.1776 2 , v= 74) ,i.e., ~t 74

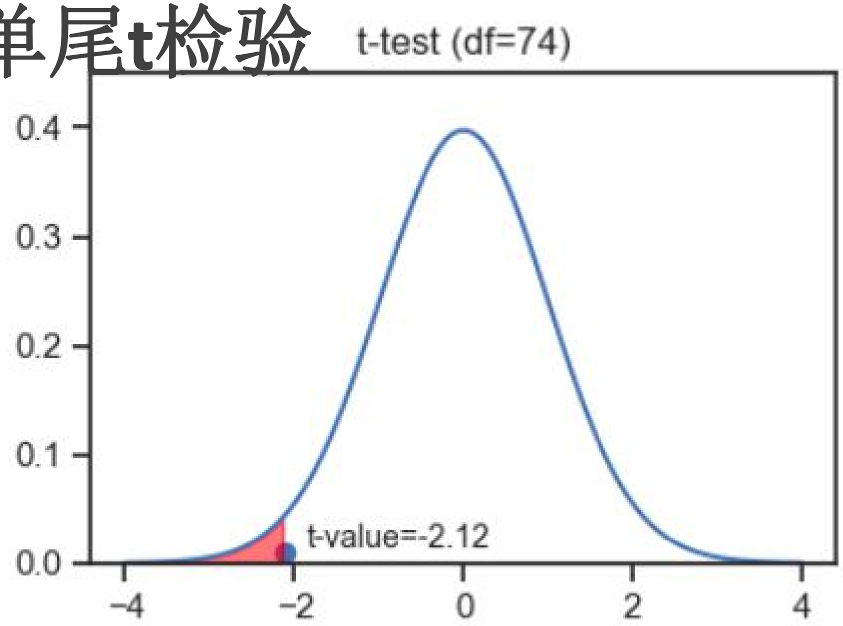
$$T = \frac{\bar{X}_0 - 7.25}{0.1776}$$



其中，t74是具有自由度的（标准化的）t分布的74。

单样本t检验

例如：大学生睡眠时间的单样本单尾t检验



步骤5b：计算p值（t检验）

在这个左尾检验中，我们可以计算p值为

$$P(\bar{X}_0 < \bar{x})$$

$$= P\left(\frac{\bar{X}_0 - 7.25}{0.1776} < \frac{6.873 - 7.25}{0.1776}\right)$$

沙痂t值

$$t = \frac{\bar{x} - \mu_0}{SE} = \frac{6.873 - 7.25}{0.1776} = -2.12$$

通过在SciPy中使用stats.t.cdf（），我们可以得到

p−值=PT<−2。12= 0.01856

```
t_value = (sample_mean - 7.25) / std_err
stats.t.cdf(t_value, df=74)
```

```
0.01863897743847032
```

单样本t检验

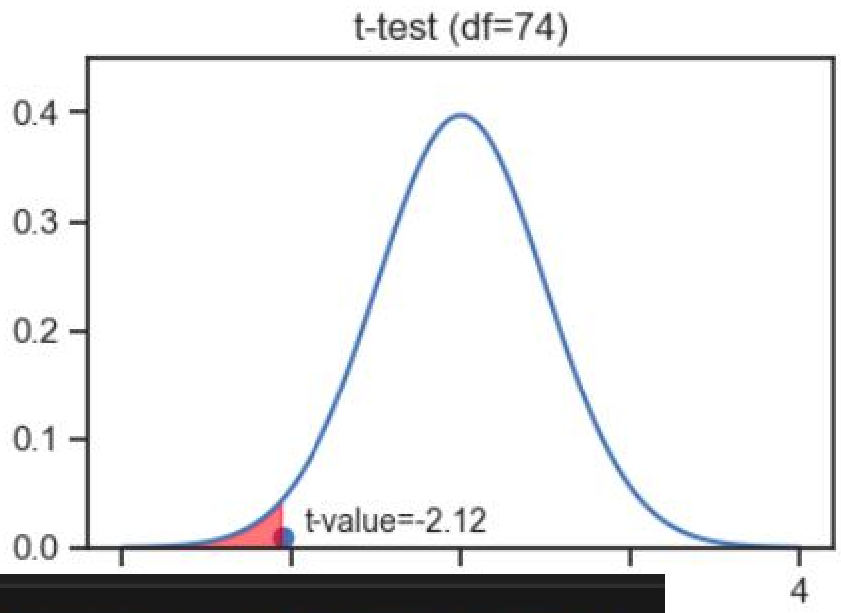
例如：大学生睡眠时间的单样本单尾t检验

步骤4b + 5b：使用SciPy进行单样本t检验

由usingscipy.stats.ttest_1 samp（）inSciPy，我们得到

$$p- \text{value}= 0.01856$$

在返回的数据中，统计量是t值，求值是p值，df是自由度。




```
stats.ttest_1samp(stu_survey_sleep_df['Sleep'], popmean=7.25, alternative='less')
```

```
TtestResult(statistic=-2.120932064022875, pvalue=0.018638977438470327, df=74)
```

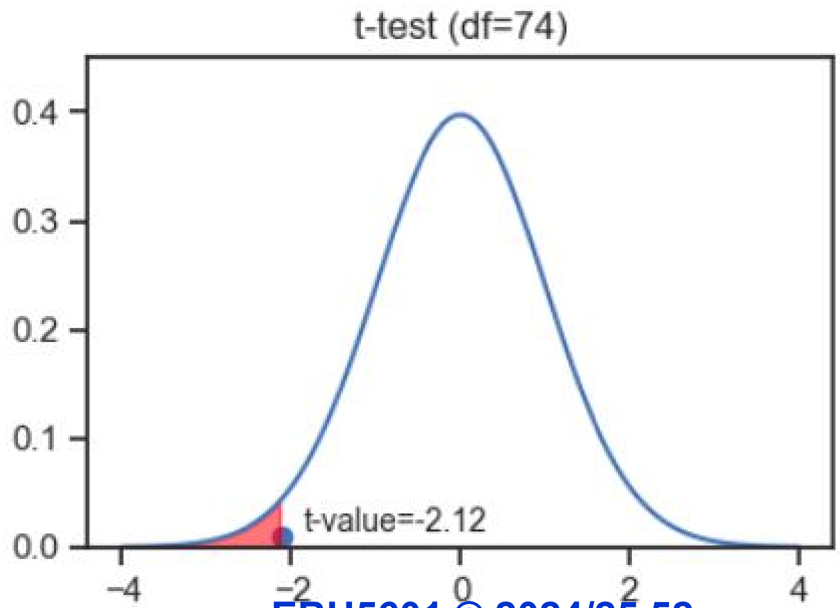
单样本t检验

例如：大学生睡眠时间的单样本单尾t检验

步骤六：得出结论

因为p值=0.01856<e（=0.0 5），我们拒绝了无效的替代方案，即，有足够的证据支持该主张

大学生的睡眠时间不到7.25小时。



Queen Mary
University of London

单样本t检验

例如：大学生睡眠时间的单样本单尾t检验

步骤4b - 6：使用t表进行单样本t检验

计算值ast=x−μ 0/SE。

-在本例中，t值为-2.12

基于onαand df，我们可以在t表中查找临界值tα，v（单尾检验）或tα/z，v（双尾检验），它实际上是当刚刚满足显著性水平时的t值。

-在本例中，我们可以找到临界值为1.664

Table B t distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390

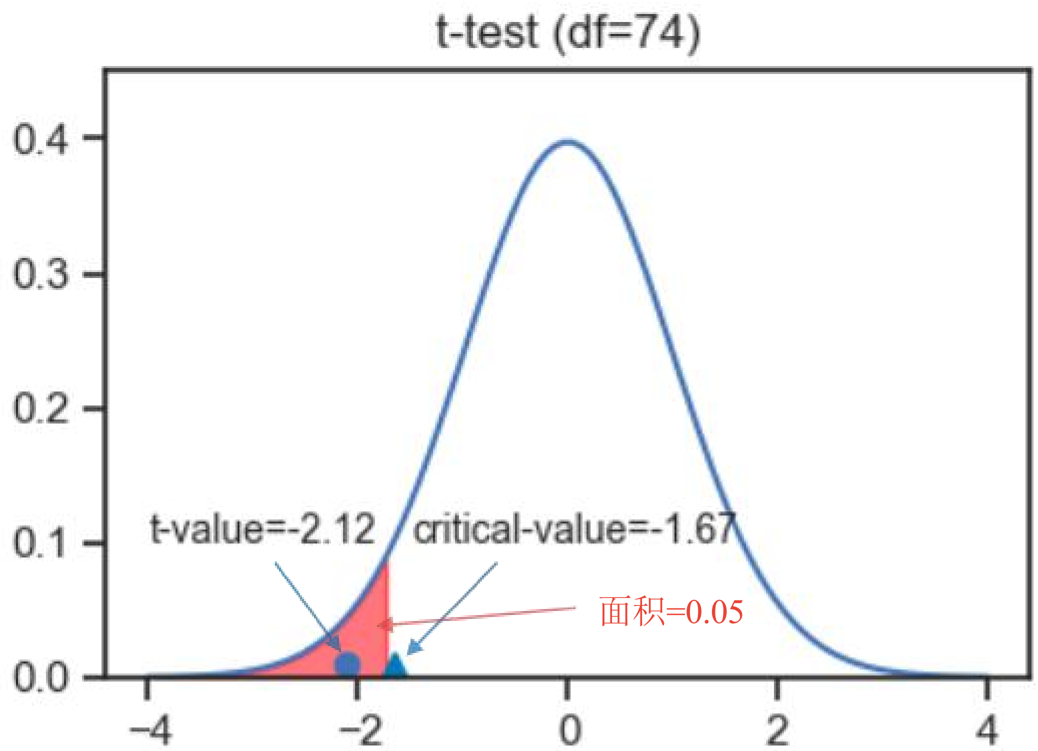
单样本t检验

例如：大学生睡眠时间的单样本单尾t检验

步骤4b - 6：使用t表进行单样本t检验

t表中的关键值
  实际上是t值，当意义水平刚刚达到

比较t值与临界值，如果|t值为|>临界值，拒绝nul l。
  否则，将无法拒绝该空值。
  -在这个例子中，由于|t值|=2.12>1.664，我们拒绝了替代方案的零偏好



基于引导的标准错误

例如：大学生睡眠时间的单样本单尾t检验

步骤4：生成空分布

Subabz测试/t测试：

基于的估计生成零分布

标准误差s/ n与一些假设，例如，

样本数据的分布应近似为正态分布。

沙痂靴子

使用引导来生成感兴趣的统计量的抽样分布

从自助采样分布中获得SE。

在步骤4中继续生成零分布，并继续从步骤4开始的假设检验。

Queen Mary University of London

基于引导的标准错误

演示：大学Stu的睡眠时间减少单样本单尾t检验

步骤4：生成空分布

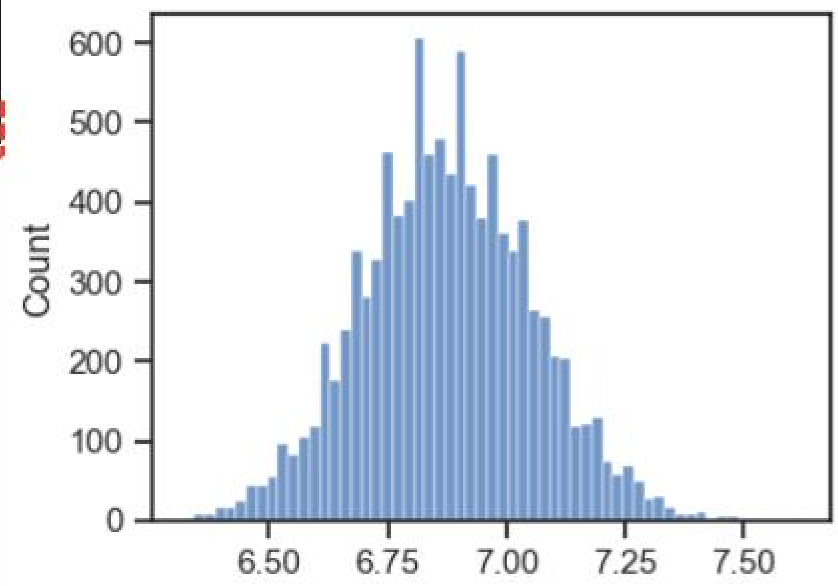
```python
# Bootstrap the sample of sleep hours
boot_means = np.array([stu_survey_sleep_df.sample(frac=1, replace=True)['Sleep'].mean() for i in range(10000)])
std_err_boot = boot_means.std()
print("standard error of sampling distribution from bootstrapping:", std_err_boot)

fig, ax = plt.subplots(figsize=(4, 3))
sns.histplot(data=boot_means, ax=ax)

plt.show()
```

✓ 0.7s

```
standard error of sampling distribution from bootstrapping: 0.178584734496292
```



标准误差（0.1786）非常大

接近我们的估计（0.1776）

假设检验的常见步骤

Queen Mary
University of London

假设检验的常见步骤

第一步：建立假设

-单样本、双样本、配对样本、多样本等。

-单尾，或双尾。

步骤2：采集样本。

-收集一个随机样本，并找出该样本的汇总统计数据，如样本量(n)、样本平均值（x）、样本标准差（s）

-检查必要的假设

步骤3：定义一个显著性水平。

-定义一个显著的高度来表示你可以接受的最大类型的I错误（例如，α=0.05）

假设检验的常见步骤

步骤4：生成中的统计量的抽样分布

基于收集的样本的零假设（零分布）下的兴趣

-z检验：抽样分布服从正态分布

当我们知道总体标准偏差(σ)时（很少出现情况）

- $SE_z=\sigma/\sqrt{n}$

当样本量（n）>为30时

- $SE_z\approx S/\sqrt{n}$

假设检验的常见步骤

步骤4：生成中的统计量的抽样分布

基于收集的样本的零假设（零分布）下的兴趣

-t检验：抽样分布遵循t分布

- $SE_z = S/\sqrt{n}$

特别是当样本量(n)较小时（<=30）

结果与大样本量的z检验的结果相似

Subarge更一般的thanz测试

假设检验的常见步骤

步骤4：生成中的统计量的抽样分布

基于收集的样本的零假设（零分布）下的兴趣

-z/t检验的假设

这些数据是连续的。

样本数据是从一个人群中随机抽样的。

样本数据近似为正态分布。

-SE可以通过引导样本获得，因为这违反了正常的身份假设。

假设检验的常见步骤

步骤5：确定p值

-z测试：

- z-order: $z = \frac{\bar{x} - \mu_0}{SE_z}$

Subarbp值：

  – Two-tailed: $P(|Z| > |z|)$
  – Left-tailed: $P(Z < z)$
  – Right-tailed: $P(Z > z)$

-t检验：

- t-value: $t = \frac{\bar{x} - \mu_0}{SE_t}$

p值：而不是p值，我们可以得到

  – Two-tailed: $P(|T| > |t|)$
  – Left-tailed: $P(T < t)$
  – Right-tailed: $P(T > t)$

critical value by referring to a t-table with the significance level α

- $t_{\alpha,\nu}$(one-tailed test)
- or $t_{\alpha/2,\nu}$ (two-tailed test)

假设检验的常见步骤

步骤六：得出结论

-如果p−值为≤α，我们拒绝支持替代方案

-如果p−值为≥α，我们无法重新为空

-对于t检验，如果我们选择使用临界值

如果|t值|>临界值t-，/或t-/1，/，我们拒绝null以支持替代方案

否则，我们无法拒绝null。

双样本试验

介绍

双样本检验是对随机样本数据进行的检验，每个样本都从不同的给定样本中独立获得

人口测试的目的是确定这两个人群之间的差异是否是统计上的重要的

假设设置

双样本**t**检验的常用设置有：

双尾试验

- H !： μ 1=μ2或H ! :μ1 − μ2 = 0

H1： μ1 =μ2或H1： μ1−μ2 = 0

单尾试验

- H !： μ1≤μ2或H ! :μ1 − μ2 ≤ 0

H1： μ1 >μ2或H1： μ1−μ2 > 0

或

H !： μ1≥μ2或H ! ： μ1−μ2≥0 H1： μ 1<μ2或H1 ： μ1−μ2 < 0

双样本t检验

双样本检验是对随机样本数据进行的检验，每个样本都从不同的给定样本中独立获得

人口测试的目的是确定这两个人群之间的差异是否是统计学上的重要的

双样本t检验可以看作是两个群体之间的平均差异的单样本t检验（μ diff=μ1−μ 2）。在原假设下，阈差的抽样分布的均值为0。

$(\bar{X}_{diff} = \bar{X}_1 - \bar{X}_2)$

双样本t检验

t值：

t值可计算为：

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE_{diff}} = \frac{(\bar{x}_1 - \bar{x}_2)}{SE_{diff}}$$

式中，x1是大小为n1、标准差为s1的总体1的样本的平均值，x2是大小为n2、标准差为s2的总体2的样本的平均值。SE是两个种群（8X18X2−8 X 2）的标准误差

双样本t检验

t值：

如果两个种群的方差相等

$$SE_{diff} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

其中，合并方差：

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

其中，n1和n2为样本量，s1和s2为两个样本的标准差，以及样本的度
自由

$$df = n1 + n\,2 - 2$$

双样本t检验

t值：

如果两个种群有一个等方差 $\sqrt{s^2 \quad s^2}$

自由度的近似自由度

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{(n_1 - 1)}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{(n_2 - 1)}\left(\frac{s_2^2}{n_2}\right)^2}$$

双样本t测试

例如：男女学生之间睡眠时间的差异

与上一个例子相同的数据集。

-75名学生的有效睡眠时间和性别的数据

在这个例子中，我们将重点关注男女学生睡眠时间的差异。

ID	Gender	Classification	Height	Shoe Size	Phone Time	# of Shoes	Birth order	Pets	Happy	Funny	College	Bfast Calories	Exercise	Stat Pre	Stat Post	Phone Type	Sleep	Social Media	Impact of SocNetworking	Political	Animal	Superhero
1	male	senior	67.75	7	12	12	youngest	5	0.8	7	Natural Sciences	500	360	3		iPhone	7	180	worse	Democrat	Dog person	Batman
2	male	freshman	71	7.5	1.5	5	middle	4	0.75	8	Natural Sciences	0	200	9		Android smartphone	7	20	better	Democrat	Dog person	Batman
3	female	freshman	64	6	25	15	oldest	8	0.9	6	Natural Sciences	200	30	7	5	Android smartphone	8	60	better	Republican	Dog person	Batman
4	female	freshman	63	6.5	30	30	middle	12	0.98	9	Education	200	180	6	7	iPhone	6	60	better	Republican	Both	Superman
5	male	senior	69	6.5	23	8	oldest	4	0.75	6	Natural Sciences	0	180	4	7	iPhone	5.5	60	worse	Independent	Dog person	Superman
6	female	senior	64	8.5	13	25	oldest	1	0.95	5	Natural Sciences	250	310	7	7	iPhone	6.5	90	no impact	Democrat	Dog person	Batman
7	female	freshman	62	8.5	23	12	oldest	2	0.95	7	Nursing	200	60	7	8	Android smartphone	7	120	better	Republican	Both	Superman
8	female	freshman	64	6	50	50	youngest	10	0.9	4	Liberal Arts	200	0	5	7	iPhone	7	60	no impact	Independent	Both	Superman
9	[illegible]	[illegible]	66	8	10	15	youngest	25	0.9	9	Natural Sciences	0	0	[illegible]	6	iPhone	7	3	no impact	Democrat	Both	Batman
10	female	freshman	68	6.5	40	20	oldest	4	0.95	[illegible]	Nursing	150	240	6	10	iPhone	6	180	better	Democrat	I don't like either	Batman

数据来源： http://sites.utexas.edu/sos/guided/inferential/numeric/claim/one-sample-t/

双样本t测试

例：男女学生睡眠之间的差异双样本双尾t检验

第一步：建立一个假设。

假设我们有兴趣知道男大学生每天的平均睡眠时间是否与
女学生每天的平均睡眠时间。Wecan建立了假设检验如下。

$$H\ !:\mu f\ -\ \mu\ m=0$$
$$H\ 1:\mu f\ -\mu\ m\neq 0$$

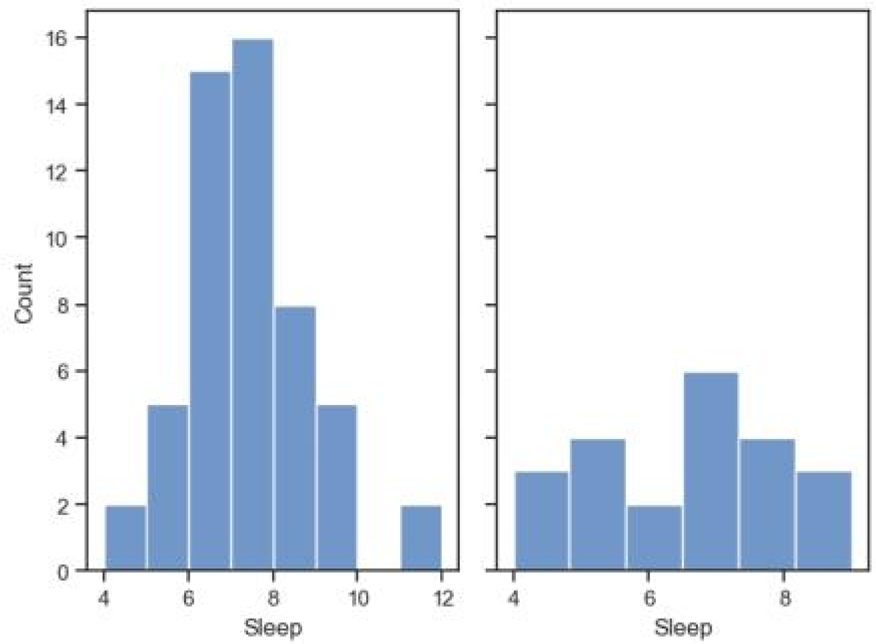
双样本t测试

例如：男女学生睡眠之间的差异双样本双尾t检验

第二步：收集一个样本，并找出该样本的汇总统计数据

我们已经有了这些数据和汇总的统计数据。

- -样本尺寸为nf =53，n m= 22
- – Samplemean xf =7.019, x m= 6.523
- -样品std sf = 1.535，s m= 1.523
- -sf≈sm，假设双种群分布近似为相同的方差，所以v=nf+nm−2= 73
- -这两个样本的分布呈粗糙的钟形



双样本t测试

例如：男女学生睡眠之间的差异双样本双尾t检验

步骤3：定义一个显著性水平

我们保持显著性水平与最后的演示α= 0.05相同

双样本t测试

例如：男女学生睡眠之间的差异双样本双尾t检验

步骤4c：生成空值分布（t检验）

由于sf≈sm，假设两个种群分布具有近似相同的方差

合并方差

$$s_p^2 = \frac{(n_f - 1)s_f^2 + (n_m - 1)s_m^2}{n_f + n_m - 2} = \frac{(53-1)1.535^2 + (22-1)1.523^2}{53+22-2}$$

$$= 2.346$$

合并标准差

$$s_p = \sqrt{s_p^2} = 1.532$$

零分布的标准误差

$$SE_{diff} = s_p\sqrt{\frac{1}{n_f} + \frac{1}{n_m}} = 1.532\sqrt{\frac{1}{53} + \frac{1}{22}} = 0.389$$

$$\boxed{\therefore T = \frac{\bar{X}_{diff} - 0}{0.389} \sim t_{73}}$$

双样本t检验

例如：男女学生睡眠之间的差异双样本双尾t检验

步骤5c：计算p值（t检验）

睡眠小时差的零分布 $T = \frac{\bar{X}_{diff} - 0}{\quad} \sim t$

沙痂t值

$$t = \frac{(\bar{x}_f - \bar{x}_m) - (\mu_1 - \mu_2)}{SE_{diff}} = \frac{(7.019 - 6.523) - 0}{0.389} = 1.275$$

通过在SciPy中使用stats.t.cdf（）withdf=73，我们可以

得到p−值=2p T > 1.275= 0.206

```
t_value = (sample_mean_female - sample_mean_male) / se_p
p_value = 2*(1 - stats.t.cdf(t_value, df))
print("t-value = {}, p-value = {}".format(t_value, p_value))
```

✓ 0.0s

```
t-value = 1.2773441806166717, p-value = 0.2055284062199625
```

双样本t检验

例如：男女学生睡眠之间的差异双样本双尾t检验

步骤4c + 5c：使用SciPy进行双样本t检验

我们可以在scipy.stats.ttest_ind（）中进行双样本

scipy.stats.ttest_ind（）检验

如果假设两个种群之间的方差相等，则将相等的_var设为真。

```python
stats.ttest_ind(stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'female']['Sleep'],
                stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'male']['Sleep'],
                equal_var=True,
                alternative='two-sided')
```

✓ 0.0s

```
TtestResult(statistic=1.2773441806166719, pvalue=0.20552840621996235, df=73.0)
```

在返回的数据中，统计量是t值，p值是p值，df是自由度。

双样本t检验

例如：男女学生睡眠之间的差异双样本双尾t检验

步骤4c + 5c：使用SciPy进行双样本t检验

我们也可以在不假设方差的情况下进行t检验，将_var设为False。

```
stats.ttest_ind(stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'female']['Sleep'],
                stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'male']['Sleep'],
                equal_var=False,
                alternative='two-sided')
```

✓ 0.0s

```
TtestResult(statistic=1.2813008328404065, pvalue=0.2075436466844493, df=39.583166134257965)
```

在返回的数据中，统计量是t值，p值是p值，df是自由度。

双样本t检验

例如：男女学生睡眠之间的差异单样本单尾t检验

步骤六：得出结论

因为p值=0.206>e（=0.05），我们没有拒绝null，即没有足够的证据支持平均

男女学生每天的睡眠时间不同，女生与男生每天的平均睡眠时间差异无统计学意义。

双样本t检验

练习：男女学生睡眠时间的差异双样本双尾t检验

步骤4c - 6：使用t表进行双样本t检验

Go to

www.menti.com

Enter the code

1939 7765

Or use QR code

Table B t distribution critical values

df	Tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390

通过引导程序进行的双样本z测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤4：生成空分布

```python
boot_female_sleep_means, boot_male_sleep_means, boot_diff_means = [], [], []

for i in range(10000):
    # Boostrapping the original sample to generate a new sample
    boot_sample = stu_survey_sleep_df.sample(frac=1, replace=True)

    # Calculate the mean sleep hours for female and male students respectively.
    female_sleep_mean = boot_sample[boot_sample['Gender'] == 'female']['Sleep'].mean()
    male_sleep_mean = boot_sample[boot_sample['Gender'] == 'male']['Sleep'].mean()

    # Insert current mean to the list of means to generate the sampling distributions of the mean sleep hours of female and male studen
    # as well as the sampling distribution of the mean difference between female and male students.
    boot_female_sleep_means.append(female_sleep_mean)
    boot_male_sleep_means.append(male_sleep_mean)
    boot_diff_means.append(female_sleep_mean - male_sleep_mean)

print("standard error of the sampling distribution of mean sleep hours of female from bootstrapping: ", np.std(boot_female_sleep_means)
print("standard error of the sampling distribution of mean sleep hours of male from bootstrapping:", np.std(boot_male_sleep_means))
print("standard error of the sampling distribution of mean difference from bootstrapping: ", np.std(boot_diff_means))
```

✓ 2.3s Python

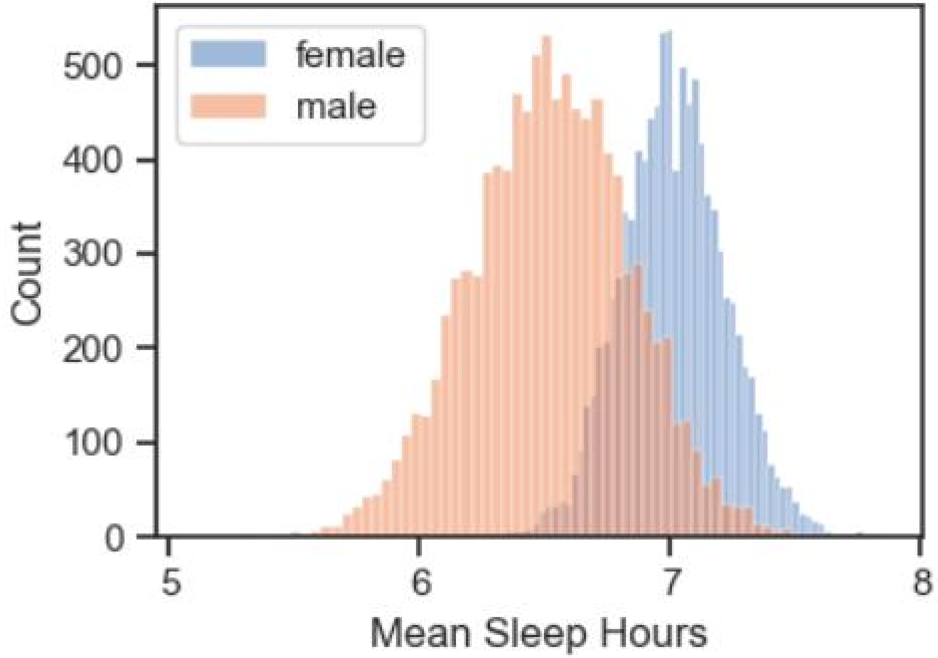
```
standard error of the sampling distribution of mean sleep hours of female from bootstrapping:  0.20957152939602225
standard error of the sampling distribution of mean sleep hours of male from bootstrapping: 0.32059062692473955
standard error of the sampling distribution of mean difference from bootstrapping:  0.38286697294521543
```

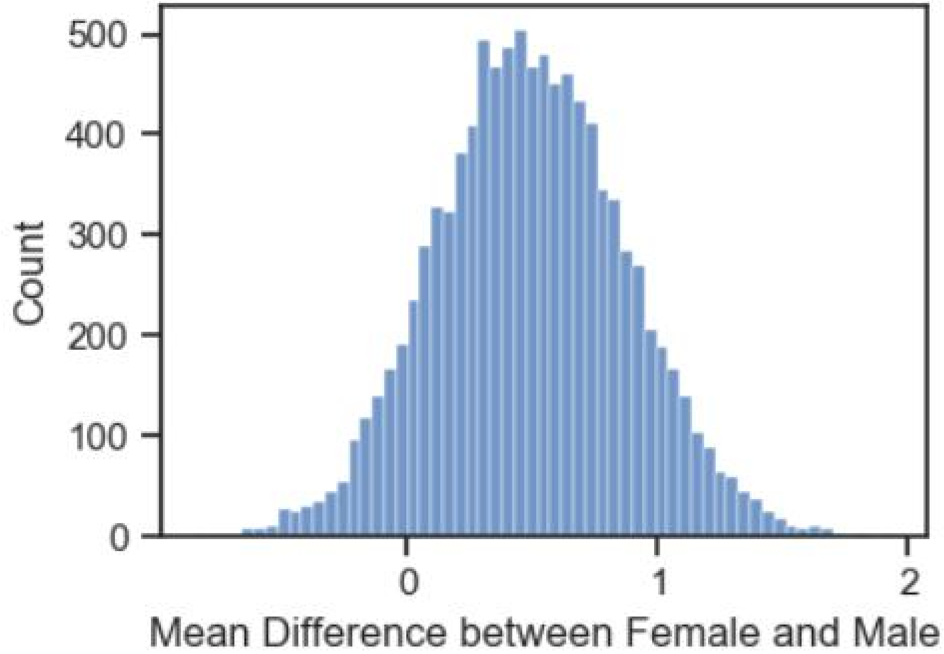
标准误差（0.383）非常接近我们的估计（0.389）

通过引导程序进行的双样本z测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤4：生成空分布





我们可以生成平均差分N0,0的零分布。383"

通过引导程序进行的双样本z测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤5：计算p值（z检验）

沙痂z分数

通过looki=1.295p−值=2pZ>1.295=2∗ （1−0.9032）= 0。

$$z = \frac{(\bar{x}_f - \bar{x}_m) - (\mu_1 - \mu_2)}{SE_{diff}} = \frac{(7.019 - 6.523) - 0}{0.383}$$

1936ng up the z-table, we can get p-value:

我们可以用scipy得到类似的结果。数据的统计数据。范数cdf () .

步骤六：得出结论

由于值-值>α=0.05，我们不能拒绝null。

排列测试

介绍

Subabz/t测试

-参数检验，其中零分布是从理论概率分布中得到的。

-依赖于种群应该遵循合理分布的假设

痂周排列试验

-也称为随机化检验或再随机化检验，是用于评估是否观察到的一个参数统计检验
组间的差异或在数据集中观察到的变量之间的关系具有统计学意义
。

-不依赖于关于底层人口分布的特定假设。当传统参数检验的假设，如z/t检验不满足时，或者当您想要进行假设检验而不进行分布假设时，它特别有用。

-涉及两个或两个以上的sa mples

介绍

零分布的排列测试

-零假设是所有的样本都来自相同的分布。

-在原假设下，检验统计量的分布是通过计算可能条件下的所有可能值的检验统计量得到的
观察到数据的重新排列。

-例如，当我们想要检验两个种群之间的平均差异时，排列检验只是在假设两组在测量变量方面没有差异的情况下，生成平均差异的分布。

-由此，我们使用观察到的统计数据来看看这个统计数据在多大程度上是特殊的，也就是说，如果治疗标签在治疗后只是随机的，那么观察到这个值的可能性（或更大）

排列测试程序说明

-孔变试验的总体步骤与az/t试验相同（第1步至第6步）。差异在于第4步（生成零分布）和第5步（计算p值）。

Queen Mary University of London

排列程序

第4步（生成空分布）和第5步（计算值）中的排列过程如下：

步骤4d：生成null分布

- 1.将来自不同组/样本的结果合并到不同的数据集。
- 2.打乱合并后的数据，然后随机抽取（不替换）与a组相同大小的重新样本。
- 3.从剩余的数据中，随机抽取（没有替换）是同组B组的样本。
- 4.对C、D组也做同样的事情。您现在已经收集了一组反映原始样本大小的重样本。
- 5.计算排列样本的兴趣统计量；这构成了一个置换迭代。
- 6.重复前面的步骤，生成检验统计量的排列分布（空分布）。

排列程序

Step 5d：计算值

-将观察到的检验统计量与从排列中获得的检验统计量的分布进行比较（步骤4）。

-计算与观察到的检验统计量一样的极值或更极端的排列检验统计量的比例。这个比例是p值。

步骤六：得出结论

-与其他检验相同，如果p-值小于预定的显著性提升(α)，你将拒绝替代假设的零假设信息。

-否则，您无法拒绝原假设。

排列测试

样例

假设我们从两个样本中随机抽取两个样本

人口红色的4个值来自于一个分布，蓝色的5个值来自于另一个分布。我们想测试一下这两个种群的表达方法是否不同。

-零假设是两组样本都来自相同的分布。H !：μ 1=μ2或H！ :μ1 − μ2 = 0

-另一种假设是第一个分布的大小比第二个分布的大。H1 ：μ 1=μ2或H1 ：μ1−μ2 = 0

{58, 59, 64,55}: $\bar{x}1 = 61.5$

{51, 41, 51,71, 42}: $\bar{x}2 = 51.2$

排列测试

样例

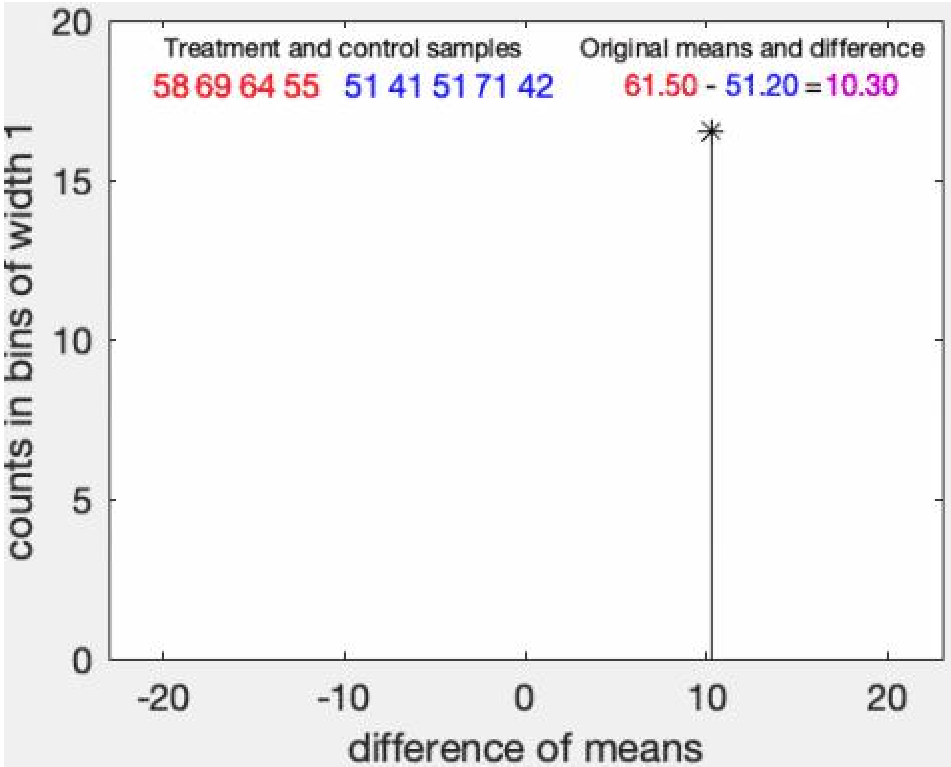
排列测试以随机的4和5计算价值

-有126种不同的表达方法

4评估分为一组，5个评估分为另一组（9选择4或9选择5）。

-假设的p值被估计为排列的比例差值大于或大于均值的差值原始样本。

-在这个例子中，我们没有做到拒绝p=5%水平。



排列测试

演示：男女学生睡眠时间的差异双样本双尾测试

第一步：建立一个假设。

假设我们有兴趣知道男大学生每天的平均睡眠时间是否与女学生每天的平均睡眠时间。Wecan建立了假设检验如下。

$$H\ !:\mu f\ -\ \mu\ m=0$$

$$H\ 1:\mu f\ -\mu\ m\neq 0$$

Queen Mary University of London

排列测试

演示：男女学生睡眠时间的差异双样本双尾测试

第二步：收集一个样本，并找出该样本的汇总统计数据

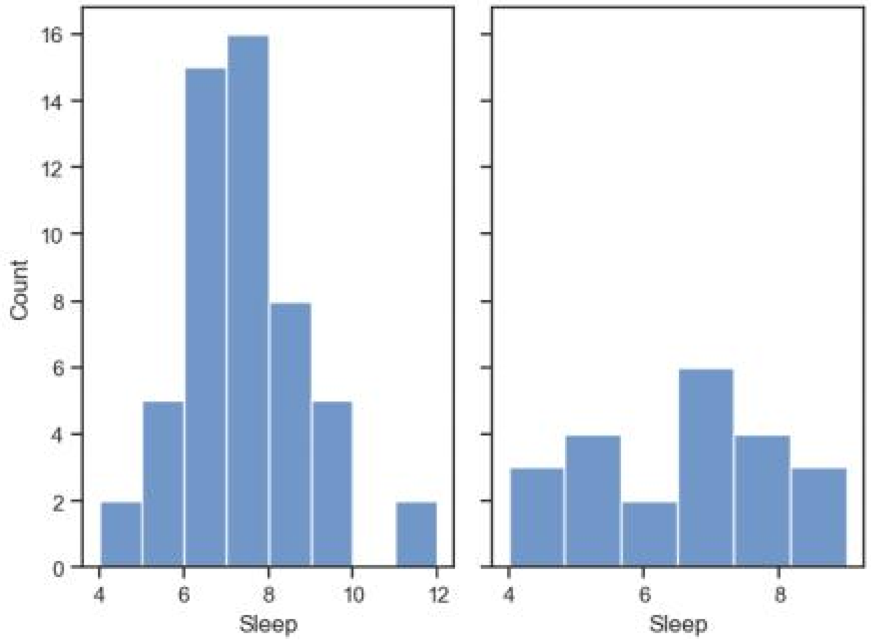
我们已经有了数据和汇总统计数据。

- -样本尺寸为nf =53，n m= 22
- – Samplemean xf =7.019, x m= 6.523
- -样品std sf = 1.535，s m= 1.523
- -样本平均数的差异

$$\bar{x}f - \bar{x}m = 7.019 - 6.523 = 0.496$$

- -排列检验不包含对两个变量的分布的任何假设

抽样资料



排列测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤3：定义一个显著性水平

我们保持显著性水平与最后的演示$\alpha = 0.05$相同

排列测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤4：生成空分布

```python
# Define the method to calculate the test statistic of interest
def diff_statistic(x, y):
    x_mean = np.mean(x)
    y_mean = np.mean(y)
    mean_diff = x_mean - y_mean
    return mean_diff

# Obtain the two original samples
female_sleep = stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'female']['Sleep']
male_sleep = stu_survey_sleep_df[stu_survey_sleep_df['Gender'] == 'male']['Sleep']

print(female_sleep.shape, male_sleep.shape)
```

✓ 0.0s Python

```
(53,) (22,)
```

统计量：两个原始样本之间的平均差值，null_distributionis为抽样

零值下的分布

由排列法产生的假设

```python
# Perform a permutation test for the two groups
pm_res = stats.permutation_test((female_sleep, male_sleep), diff_statistic, permutation_type='independent', alternative='two-sided')
pm_res
```

✓ 0.0s Python

```
PermutationTestResult(statistic=0.4961406518010296, pvalue=0.2188, null_distribution=array([-0.01843911,  0.1745283 , -0.17924528, ..., -0.661
       -0.01843911,  0.68910806]))
```

Queen Mary University of London

排列测试

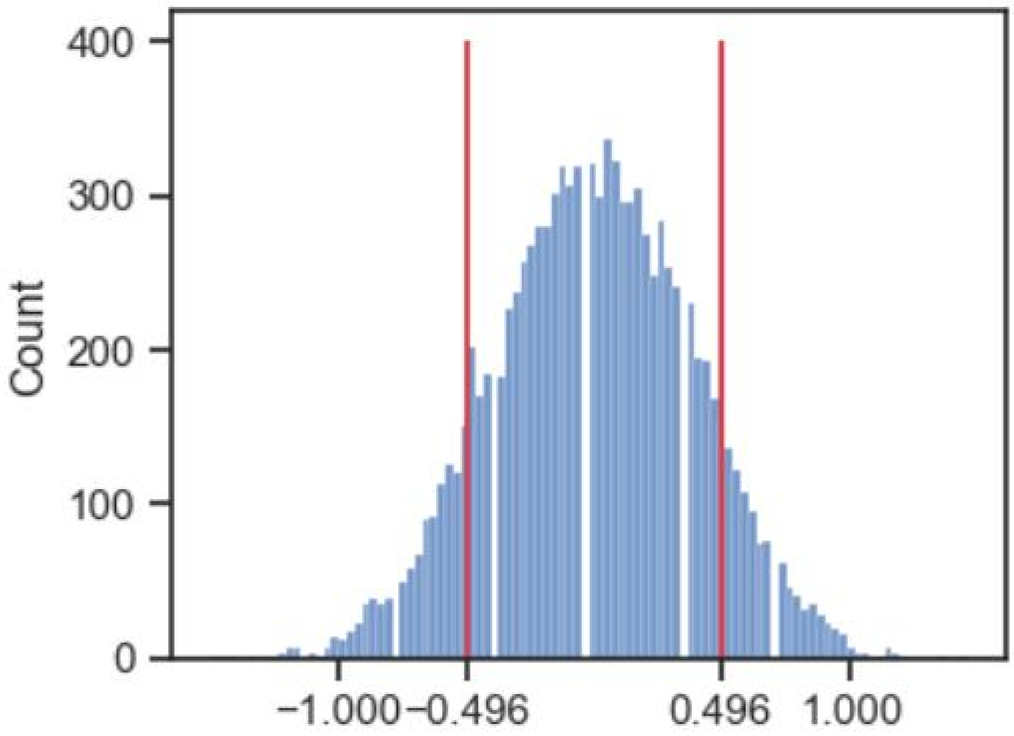
演示：男女学生睡眠时间的差异双样本双尾测试

步骤4：生成空分布

```
PermutationTestResult(statistic=0.4961406518010296, pvalue=0.2188, null_distribution=array([-0.01843911,  0.1745283 , -0.17924528, ..., -0.661
       -0.01843911,  0.68910806]))
```

可视化的

零分布



排列测试

演示：男女学生睡眠时间的差异双样本双尾测试

步骤5：计算p值（排列检验）

读取返回的结果中的求值

```
PermutationTestResult(statistic=0.4961406518010296, pvalue=0.2188, null_distribution=array([-0.01843911,  0.1745283 , -0.17924528, ..., -0.661
       -0.01843911,  0.68910806]))
```

根据这个定义。

```
1  (pm_res.null_distribution >= pm_res.statistic).mean() + (pm_res.null_distribution <= -pm_res.statistic).mean()
✓ 0.0s
0.2208220822082208
```

步骤六：得出结论

由于值-值>α=0.05，我们不能拒绝null。

其他需要考虑的事情

不考

大样本量的影响

由于样本量大，假设检验甚至会导致

最小的发现具有统计学意义。然而，这些发现可能实际上根本没有什么意义。

例如：大学生的睡眠时间

让我们来看看第一个单样本的例子，它测试大学生的平均睡眠时间是否等于或

不同于7.25小时。

H !:μ= 7.25

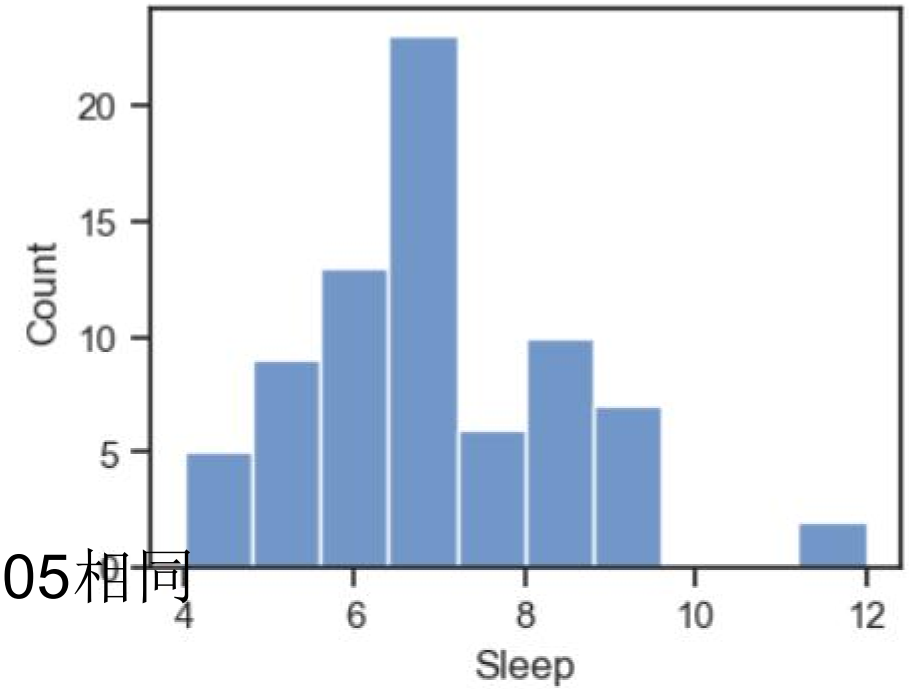
H 1: μ≠ 7.25

大样本量的影响

例如：大学生睡眠时间的单样本双尾t检验

我们已经有了数据和汇总统计数据。

- -样本量=75
- -样本平均值=6.873
- -样本标准=1.538
- -自由度(v)=n−1=74
- -睡眠时间大致遵循正态分布



我们保持显著性水平与最后的演示α= 0.05相同

大样本量的影响

例如：大学生睡眠时间的单样本双尾t检验

如果weuse单样本t检验

$$SE = \frac{s}{\sqrt{n}} = \frac{1.538}{\sqrt{75}} = 0.1776$$

沙痂t值

$$t = \frac{\bar{x} - \mu_0}{SE} = \frac{6.873 - 7.25}{0.1776} = -2.12$$

通过在SciPy中使用stats.t.cdf（），我们可以得到

$$p-\text{值}=2pT<-2.12=2*0.01856=0.0\,3712$$

因为-值=0.03712<α（=0.5），我们拒绝零支持替代，即有足够的证据支持该主张

大学生睡眠的小时数少于

7.25小时。

Queen Mary
University of London

大样本量的影响

练习：大学Stu组的睡眠时间凹陷单样本双尾t检验

现在假设样本数据的大小只有25。再次进行双尾t检验并得出结论。

-样本量=25

-样本平均值=6.873

-样本标准=1.538

-睡眠时间大致遵循正态分布

的significanceα = 0.05

Go to

www.menti.com

Enter the code

1939 7765

Or use QR code

统计与。实际意义

使用置信区间和假设检验，我们可以在决策中提供统计学意义。

-然而，在做决策时，考虑到实际意义也同样重要。

-实际意义考虑了您的假设检验置信区间的结果中可能不直接考虑的其他因素。

-诸如空间、时间或金钱等约束在商业决策中很重要，然而，它们可能不能在统计检验中直接得到解释。

统计与。实际意义

讨论：统计vs。实际意义

如果一种减肥药，他在1年内平均个人减肥0.2公斤，以下哪一个是关于这个结果的真实陈述？

这绝对没有统计学意义
这绝对不重要
如果样本量小，这将是非常可行的
如果我们有一个大的样本量，这将具有统计学意义

Go to
www.menti.com
Enter the code
1939 7765

Or use QR code

Queen Mary
University of London

回顾一下

假设测试

回顾一下

设置更新假设测试

- -零假设
- -替代假设

Subarb类型的错误

- -TypeI错误
  - I型错误率：α（符号缺陷等级）
- -TypeII错误
  - 第一类错误率：β

假设检验的常见类型

- -测试总体参数（单样本测试）
- -检验不同人群中参数的差异（双样本检验）
- -对同一个体进行治疗前后的差异（配对t检验）

回顾一下

假设检验的要点

- -假设检验的常见步骤
- -p值
- -t值分布和临界值

Subarb单样本测试

- -单样本z/t检验
- -通过引导程序进行的单样本测试

Subarban双样本测试

- -双样本z/t检验
- -通过自举程序进行的双样本测试
- -排列测试

其他需要考虑的事情

- -大样本量的影响

统计与统计。实际意义

问题

在

QM+chao.shu@qmul.ac.uk上使用学生论坛