

什么是假设检验的两种错误？

关注问题

写回答

邀请回答

好问题 2

1 条评论

分享

15 个回答

默认排序



Minitab Users Group

已认证账号

+ 关注

20 人赞同了该回答

任何假设检验都不是 100% 确定的。由于检验基于概率，因此始终会存在得出不正确结论的几率。在执行假设检验^Q时，可能会犯两种类型的错误：类型 I 和类型 II。这两种错误的风险逆相关，并且由显著性水平和检验功效来确定。因此，在定义错误所带来的风险之前，应当确定哪种错误对于您所处的情形造成的后果更严重。

类型 I 错误

如果原假设^Q为真，但您否定它，则会犯类型 I 错误。犯类型 I 错误的概率为 α （即您为假设检验设置的显著性水平）。 α 为 0.05 表明，当您否定原假设时，您愿意接受 5% 的犯错概率。为了降低此风险，必须使用较低的 α 值。但是，使用的 **alpha 值**^Q 越小，在差值确实存在时检测到实际差值的可能性也越小。

类型 II 错误

如果原假设为假，但您无法否定它，则会犯类型 II 错误。犯类型 II 错误的概率为 β ， β 依赖检验功效。可以通过确保检验具有足够大的功效来降低犯类型 II 错误所带来的风险。方法是确保样本数量足够大，以便在差值确实存在时检测到实际差值。

在原假设为假时否定原假设的概率等于 $1-\beta$ 。此值是检验的功效。

基于样本的决策	有关总体的事实	
	H_0 为真	H_0 为假
无法否定 H_0	正确的决策（概率 = $1 - \alpha$ ）	类型 II 错误 - 当 H_0 为假时无法否定它（概率 = β ）
否定 H_0	类型 I 错误 - 当 H_0 为真时否定它（概率 = α ）	正确的决策（概率 = $1 - \beta$ ）

类型 I 和类型 II 错误的示例

为了理解类型 I 错误和类型 II 错误之间的相互关系，并确定对于您所处的情形哪个错误产生的后果更严重，请参考下面的示例。

某医疗研究员想要比较两种药物的有效性。原假设和备择假设如下：

- 原假设 (H_0): $\mu_1 = \mu_2$
两种药物等效。
- 备择假设 (H_1): $\mu_1 \neq \mu_2$
两种药物不等效

如果研究员否定原假设并在两种药物实际上没有差异时得出二者不同的结论，则会出现类型 I 错误。如果两种药物具有相同的有效性，则研究员无法将此错误视为太严重，因为患者无论服用哪种药物，仍受益于相同水平的有效性。但是，如果出现类型 II 错误，研究员将无法在应当否定原假设时否定原假设。也就是说，研究员会在两种药物不同时得出它们相同的结论。如果向公众出售有效性低的药物（而不是有效性高的药物），此错误可能会危及生命。

在执行假设检验时，请考虑犯类型 I 错误和类型 II 错误的风险。如果与一种类型的错误相比，犯另一种错误所带来的后果更严重或者犯错成本更高，请选择将反映这些后果的相对严重性的显著性水平和检验功效。

发布于 2019-08-09 09:54



下载知乎客户端

与世界分享知识、经验和见解

相关问题

假设检验的两类错误 怎么知道原假设本身是正确的还是错误的？ 5 个回答

假设检验为什么选择防止第一类错误？ 1 个回答

假设检验中，为何要避免第一类错误，而不避免第二类错误？ 59 个回答

假设检验两类错误的求法？ 0 个回答

假设检验中的的两类错误和两种风险为什么不可能完全消除？ 2 个回答

帮助中心

知乎隐私保护指引 申请开通机构号 联系我们

举报中心

涉未成年举报 网络谣言举报 涉企虚假举报 更多

关于知乎

下载知乎 知乎招聘 知乎指南 知乎协议 更多

京 ICP 证 110745 号 · 京 ICP 备 13052560 号 - 1 ·
京公网安备 11010802020088 号 · 京网文
[2022]2674-081 号 · 药品医疗器械网络信息服务备
案（京）网药械信息备字（2022）第00334号 · 广
播电视节目制作经营许可证：（京）字第06591号 ·
服务热线：400-919-0001 · Investor Relations · ©
2023 知乎 北京智者天下科技有限公司版权所有 · 违
法和不良信息举报：010-82716601 · 举报邮箱：
jubao@zhihu.com



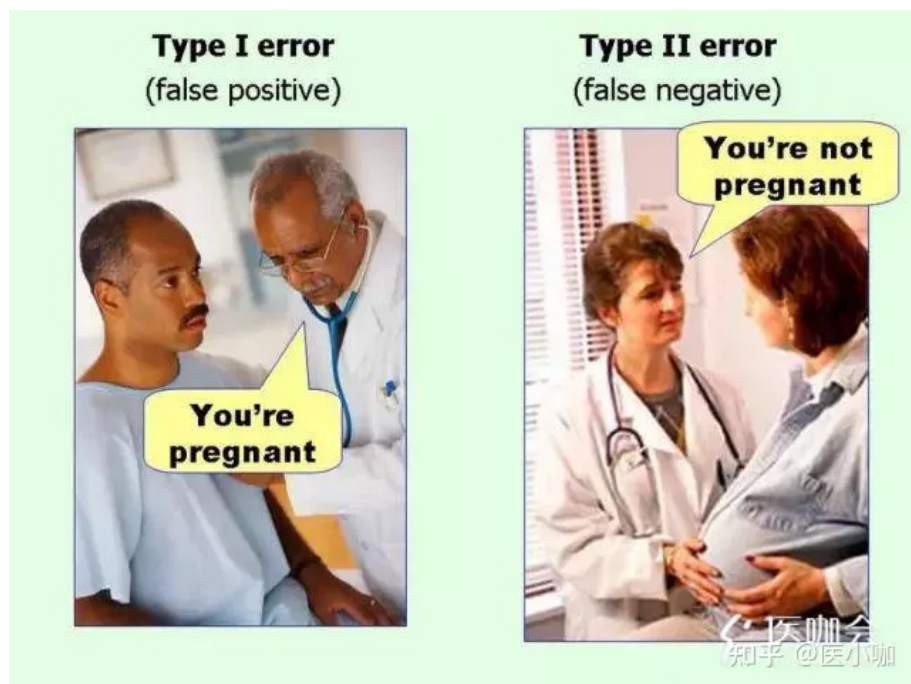
无障碍
服务

2 人赞同了该回答

Type I error (I类错误)：犯I类错误的概率，也就是当零假设^Q是真实的，却拒绝零假设的概率

Type II error (II类错误)：犯II类错误的概率，也就是当零假设是不真实的，却接受零假设的概率

我们用一个生动形象的例子来讲一下：零假设（H0）：病人没有怀孕。备择假设（H1）：病人怀孕。在下图中，我们可以看到，左侧的男性病人不可能怀孕（当然是指自然状态下），但医生却认为病人怀孕，这就是当零假设是真实的却拒绝了零假设，犯了I类错误；而右侧的女性病人，可以看到其怀孕，但是医生却认为病人没有怀孕，这就是零假设不成立却接受了零假设，犯了II类错误。



(图片来源: flowingdata.com)

一、研究实例

某研究团队拟研究多种生活方式的干预对于学龄前儿童^Q有氧运动能力和肥胖的影响，设计了一个整群随机对照试验^Q（cluster randomised controlled trial），一共纳入了40名学龄前儿童作为研究对象，将他们随机分为干预组和对照组。

对照组儿童仅接受常规的课程学习，包括每周一节45分钟的体育课。而干预组儿童在对照组的基础上，接受包括参加体育活动项目，增加营养知识课程等多方位的生活方式干预，干预共持续一年。研究的主要评价结局为有氧运动能力（20分钟往返跑^Q）和BMI指数。

结果显示，干预组儿童有氧运动能力高于对照组儿童(平均差异: 0.32, 95%CI: 0.07~0.57, $P=0.01<0.05$)，而两组儿童BMI的差异却无统计学^Q显著性(-0.07kg/m², -0.19~0.16, $P=0.31>0.05$)。

作者由此得出结论：多种生活方式干预可以提高学龄前儿童的有氧运动能力，但对BMI影响不大。

二、问题

上述研究于2011年发表在BMJ杂志上，研究结果很容易理解，那么问题来了，请大家来判断一下，以下三种说法，哪一项是正确的呢？

1. 如果实际上在该人群中，多种生活方式干预对于儿童的有氧运动能力没有影响，两组儿童的有氧运动能力并无差异，那么作者针对有氧运动能力进行假设检验^Q，得出的结论就会产生I类错误。
2. 如果实际上在该人群中，多种生活方式干预可以改善儿童的BMI指数，干预组儿童BMI指数低于对照组儿童，那么作者针对BMI进行假设检验，得出的结论就会产生II类错误。

3、如果增加样本量，则可以降低I类错误和II类错误的发生概率。

三、假设检验

首先，我们来聊一下假设检验。假设检验就是根据研究目的提出某种假设，然后利用收集的样本信息，去推断这一假设是否成立。

建立假设是进行假设检验的第一步，通常会先建立一个**原假设**，或者也叫零假设或**无效假设**^Q（null hypothesis），记为 H_0 ，例如某两个（或多个）总体参数相等，或总体参数之差为0。在本例中，原假设为干预组儿童和对照组儿童有氧运动能力相同，BMI均数相等。

与原假设对立的为**备择假设**，也称**对立假设**^Q（alternative hypothesis），记为 H_1 ，例如某两个（或多个）总体参数不相等，或总体参数之差不为0。在本例中，备择假设为干预组儿童和对照组儿童有氧运动能力不相同，BMI均数不相等。

通常备择假设包括大于或者小于两种情况，故一般为双侧检验^Q。若凭借专业知识有充分把握认为只存在大于或小于两者中的一种可能，则可采用**单侧检验**^Q。

四、P值

进行假设检验，就不得不提我们可爱又可恨的P值，它是用来判定假设检验结果的一个参数，是基于原假设 H_0 和抽样样本进行统计推断的一个工具。通俗点讲，就是在原假设 H_0 成立的条件下，所得到的与样本观察结果相同或更为极端结果的概率。

如果P值很小，说明原假设 H_0 的发生概率很小，可认为是**小概率事件**^Q，当P值小到一定程度时，我们就有理由拒绝原假设 H_0 的成立。但需要注意的是，P值的大小并不能代表所检验的差异的大小，也就是说P值越小，并不能说明差异越大。这一点很容易引起误解，因此我们在报告结果的时候，提倡使用“差异有统计学显著性”的描述，而非“有**显著性差异**^Q”。

那么，P值一般要小到什么程度才能被认为是小概率事件呢？此时我们就要设立一个检验水准，即 α ，它确定了小概率事件的标准。**通常设定 $\alpha=0.05$ 或 0.01** ，但 α 的取值并非一成不变，可以根据研究目的的不同给予不同的设置。

当 $P \leq \alpha$ 时，在设定 α 的检验水准下，可认为原假设 H_0 为小概率事件，因此拒绝 H_0 ，接受备择假设 H_1 ，差异有统计学显著性。

例如在上述研究实例中，两组儿童有氧运动能力差异性检验 $P=0.01 < 0.05$ ，在 $\alpha=0.05$ 的检验水准下，可认为干预组儿童有氧运动能力高于对照组儿童，说明多种生活方式干预可提高儿童的有氧运动能力。

当 $P > \alpha$ 时，在设定 α 的检验水准下，不能认为原假设 H_0 为小概率事件，因此不拒绝 H_0 ，差异无统计学显著性。例如在上述研究实例中，两组儿童BMI指数的差异性检验 $P=0.31 > 0.05$ ，在 $\alpha=0.05$ 的检验水准下，尚不能认为两组儿童的BMI指数不同，说明多种生活方式干预对于BMI无明显改善作用。

五、I类错误和II类错误

假设检验是基于抽样样本来进行结果推断的，而抽样样本只是总体的一小部分，从总体中抽取不同的样本，可能会得出不同的结果，因此我们通常希望抽样样本是一个能够很好地反映总体特征的具有代表性的样本。但由于**抽样误差**^Q的存在，在进行假设检验根据P值做出推断时具有一定的概率性，因此所得的结论就不一定完全正确，这就是我们常见的假设检验的陷阱：**I类错误和II类错误**。

I类错误，也称为假阳性错误，就是说实际上总体并无差异，原假设 H_0 是成立的，但是通过假设检验 $P \leq \alpha$ ，在设定 α 的检验水准下，拒绝了 H_0 ，认为有差异，出现了假阳性的现象。前面提到的检验水准 α ，就是预先设定允许犯I类错误概率的最大值，此时犯I类错误的概率即为 α 。

例如在上述研究实例中，如果实际在该人群中，干预措施对儿童的有氧运动能力没有影响，两组儿童的有氧运动能力并无差异，那么作者通过假设检验得出多种生活方式干预可提高儿童的有氧运动能力这一结论，就犯了I类错误，因此问题1的描述是对的。

II类错误，也称为假阴性错误，就是说实际上原假设 H_0 不成立，但是通过假设检验 $P > \alpha$ ，在设定 α 的检验水准下，不拒绝 H_0 ，得出了阴性的结论，此时犯II类错误的概率为 β 。例如在上述研究实例

中，如果实际在该人群中，干预措施对儿童的BMI有改善作用，那么作者通过假设检验得出干预后两组儿童的BMI差异无统计学显著性这一结论，就犯了II类错误，因此问题2的描述也是对的。

表1. 统计推断的I类错误和II类错误

实际情况	假设检验结果	
	拒绝 H_0 ，有差异	不拒绝 H_0 ，无差异
H_0 成立，无差异	I 类错误（假阳性， α ）	推断正确（ $1-\alpha$ ）
H_1 成立，有差异	推断正确（ $1-\beta$ ）	II 类错误（假阴性， β ）

注意：I类错误和II类错误只是一个统计学上的概念，在进行假设检验时无法确定其发生的实际概率。由于[两类错误](#)主要受样本量的影响，因此可以通过增大样本量的方法，使得我们的抽样样本尽可能的接近总体，具有更好的代表性，以达到降低两类错误发生概率的目的，因此问题3的描述也是对的。针对以上三个问题，你都判断对了么？最后，再次提醒大家谨防假设检验的陷阱，当[统计分析](#)出现阳性结果， $P<0.05$ 时，不要高兴的太早，认真思考一下是否有可能犯I类错误，当出现阴性结果， $P>0.05$ 时，也不要太灰心，想想是不是有可能II类错误在作怪，可以参考医咖会推送的有关处理阴性结果的系列文章（[【合集】试验主要结局为阴性，应该思考的12个问题！](#)），找找原因看看是否能有新的发现。

参考文献^Q：BMJ. 2014; 349:g4287 (PMID: 24994622)

发布于 2023-04-13 17:17

▲ 赞同 2 ▼

● 添加评论

🔗 分享

★ 收藏

♥ 喜欢

收起 ^

 **口语小卡片**
小程序 还在开发中...

+ 关注

9 人赞同了该回答

Ho: American girl的屁股 和 Chinese girl的屁股一样大

Ha: American girl的屁股 比 Chinese girl的屁股大

你找了100个American girl 和 100个Chinese girl，测量了她们的屁股，发现American girl的屁股平均比Chinese girl的屁股大10cm。

这时候有四种情况：

当American girl^Q的屁股和Chinese girl的屁股确实一样大（Ho正确）

- 你正确地没有拒绝原假设
- 你错误地拒绝了原假设 (Type 1 Error)

当American girl的屁股确实比Chinese girl的屁股大（Ha正确）

- 你正确的拒绝了原假设^Q
- 你错误的没有拒绝原假设 (Type 2 Error)

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	Type I error, α
	H_A true	Type 2 error, β	$1 - \beta$

- ▶ **Type I error** is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level).
- ▶ **Type 2 error** is failing to reject H_0 when you should have, and the probability of doing so is β .
- ▶ **Power** of a test is the probability of correctly rejecting H_0 , and the probability of doing so is $1 - \beta$

知乎 @纽约Johnny哥

Type 1 Error的严重性要高一点，因为你推翻了一个常识。

If Type I Error is dangerous or especially costly, choose a small significance level (e.g. 0.01).

Goal: we want to be very cautious about rejecting H_0 , so we demand very strong evidence favoring H_A before we would do so.

choosing α



If a Type 2 Error is relatively more dangerous or much more costly, choose a higher significance level (e.g. 0.10).

Goal: we want to be cautious about failing to reject H_0 when the null is actually false.

知乎 @纽约Johnny哥

更新：

The probability of type 1 error is α , the significance level.

假设检验^Q的原理就是当 α 足够小的时候，我们认为犯第一种错误（错误地拒绝了原假设）的概率足够小 --- 于是我们就可以 拒绝原假设。

编辑于 2019-04-13 19:10

▲ 赞同 9 ▼

● 收起评论

➦ 分享

★ 收藏

♥ 喜欢

收起 ^

写下你的评论...

3 条评论

默认

最新



wish
恶心

2022-04-01

● 回复

♥ 喜欢



景行止

理解了，谢谢大佬

2020-12-26

● 回复

♥ 1



徐浩洋

一般都写 H_0 吧

2019-04-10

● 回复

♥ 喜欢

Baidu

广告 ▼

i.hao61.net/d.js?cid=40308



绀青丝

+ 关注

53 人赞同了该回答

1.拒真和取伪

2.总体随机抽取样本，根据样本的数据去选择判断总体。

例如:

你高中参加过那么多次考试包括高考。总体说明了你这个人的能力水平。但是最终决定的是高考（实际操作时，样本肯定是随机抽取的，这里举例假设随机抽取到了高考作为样本），而高考本身就带了不确定性。

某高校分数很高90吧，你不巧就是高考考了85，差到低于它的分数（但高中总体是好的93，足够匹配这个高校），可是对方看的就是高考，不好意思，你被淘汰。那么它犯了拒真错误。

另一种情况，你很幸运的超常发挥（95），超常到高于录取分。（而高中整体来看你80并没有这个能力被这个学校录取）可是它录取了你，那么它就犯了取伪错误。

以上条件不变，高校惜才，要降低拒真或者说降低拒真的概率，它就要降低它的分数线直到能录取“本来优秀的”你，或者说录取到其他像你这样发挥失常的人，但是这样的话，就容易录取到发挥超常的人，也就是取伪的概率升高。

高校很严格想剔除超常发挥的人，要降低犯取伪错误的可能，就要提高分数线，提高到那个超常发挥的人进不来，但是这样的话，拒绝掉有实力的人的概率就加大。

在检验假设时都是要拒真概率要低于显著性水平 α （通常取0.05）

看学校怎么选择，宁愿哪个高些哪个低些就会相应的调整分数线。

无论怎么定选择标准，合格水平线，犯**两类错误**的可能性都存在，只是可能性高或低而已。

[仅仅是一个例子，我也不知道老师们怎么定线]

所以两个错误除了你增加**样本容量**以外可以都减少犯错概率以外，只要保证犯其中一个**错误**的可能小了犯另一个错误几率就会大（教材引进**势函数**证明）



写下你的评论...

7 条评论

默认

最新



时间

妙妙妙

2021-12-05

回复

喜欢



白白

妙啊

2021-06-15

回复

喜欢



嘉佳

讲的不错，谢谢！

2020-09-22

回复

喜欢



逸宸

比那个陪审团例子靠谱

2020-06-17

回复

1



Burning

答主，请问假设检验都可以写出势函数啊

2020-04-16

回复

喜欢



阿招别吃了

牛！谢谢！

2019-12-14

回复

喜欢



玉生烟

说得挺好的

2019-09-26

回复

喜欢



Jason

CFA 特许金融分析师资格证持证人

+ 关注

19 人赞同了该回答

举个例子，假设有一个银行的工作人员，在判断是不是要放贷给某个客户。如果客户的信用好，就贷款给他；如果客户的信用不好，就不贷款给他。

Alternative Hypothesis (HA) 备择假设：这个客户信用不好

Null Hypothesis (H0) 原假设：这个客户信用好

如果我判定这个客户的信用不好 (Reject H0, 或者说统计学上这个客户显著地不同于“significantly different from”信用好的客户的特征)，而实际这个客户的信用是好的 (实际: H0 is True)，那么我就犯了第一类错误 (Type I Error)。如果我犯了这个错误，这个客户的贷款业务我就没做，我的损失是这笔业务的利息收入我挣不到了。

犯第一类错误的概率，也叫做显著性水平 significance level (α)。

如果我判定这个客户的信用好 (Do not reject H0, 或者说统计学上这个客户没有显著地不同于“not significantly different from”信用好的客户的特征)，而实际这个客户的信用是不好的 (实际: H0 is False)，那么我就犯了第二类错误 (Type II Error)。如果我犯了这个错误，我就把钱贷给了这个客户，那么最后我的本金都收不回来了。

不犯第二类错误的概率，也就是“1-犯第二类错误的概率”，叫做检验能力 power of the test。

通常，我们希望 power of the test 越高越好，因为我们不希望把钱贷给信用差的客户，以至于我们损失本金。

The power of a test

Decision	True	Condition
Do not Reject	H_0 is true ✓	H_0 is falso Type II Error
Reject	Type I Error (α)	$1 - P(\text{Type II Error})$ power of the test

知乎 @Jason

编辑于 2019-05-16 11:48

赞同 19



收起评论

分享

收藏

喜欢

写下你的评论...

2 条评论

默认

最新



王二



2022-02-28

回复

喜欢



Jackie

解释得很通俗易懂!

2021-05-13

回复

喜欢

写回答