

以一个研究为例，细说I类错误和II类错误！

原创 龚志忠 医咖会 2022-07-15 19:00 发表于山西

- Type I error (I类错误)：犯I类错误的概率，也就是当零假设是真实的，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

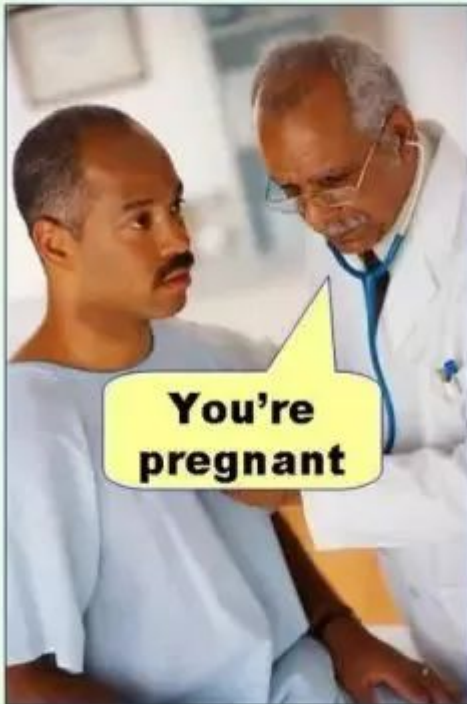
我们用一个生动形象的例子来讲一下：

零假设 (H_0)：病人没有怀孕。

备择假设 (H_1)：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。

Type I error (false positive)



Type II error (false negative)



(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（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）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的差异却无统计学显著性(-0.07kg/m^2 ， $-0.19\sim 0.16$ ， $P=0.31>0.05$)。

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，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，通常我们会先建立一个**原假设**，或者也叫零假设或无效假设（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（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。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。

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，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”。

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：

BMJ. 2014; 349:g4287 (PMID: 24994622)

医咖会
MEDICO GROUP

面向医生个人、医院/单位、企业提供各种科研服务

医咖会

www.medecogroup.com

中国科研医生平台

知识学习 / 社区交流 / 科研服务 →



遇到科研问题，得马上解决？
赶快扫码，找个大咖来咨询！

点击左下角“[阅读原文](#)”，免费学习70多篇SPSS教程、30多篇R教程，以及更多免费科研教程！

[阅读原文](#)

喜欢此内容的人还喜欢

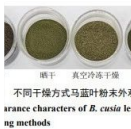
EHJ综述：什么是孟德尔随机化研究？(IF=39.3)

医咖会



基于UHPLC-Q-Orbitrap HRMS和HPLC-DAD法研究不同干燥方式对马蓝叶化学成分的影响

中草药杂志社



睡觉都比坐着好？30分钟睡眠代替静坐能改善心血管健康

医咖会

