



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



生物医学统计概论(BI148-2)

Fundamentals of Biomedical Statistics

授课：林关宁

2022 春季



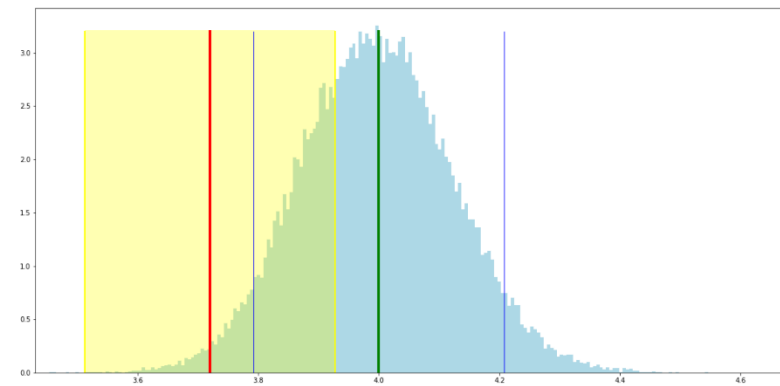
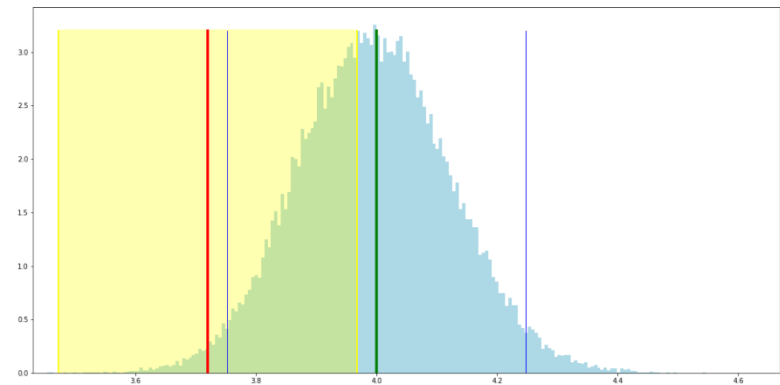
90%置信区间给出的值范围可能包括100个样本数据值中的90个。是否正确？

- ☐ A 是
- ☒ B 否

提交

如何才能获得更窄区间的置信区间？

- ☒ A 增加置信水平，增加样本量
- ☐ B 增加置信水平，降低样本量
- ☒ C 降低置信水平，增加样本量
- ☐ D 降低置信水平，降低样本量



提交

一家跑步杂志想写个关于运动手表A和手表B的对比评论。随机抽取555名志愿者，让他们同时戴着两块手表跑10公里。结束后，参与者记录下每只手表所显示的距离。

请问如要计算两手表的距离均值差异的置信区要用哪种分布？

- ☐ A Z-distribution
- ☒ B Student's t-interval

提交

两个独立样本，总体方差都未知，且不等，请问要使用哪种区间分布？

- ☐ A Z-distribution
- ☐ B Student's t-interval
- ☒ C Welch's t-interval

提交

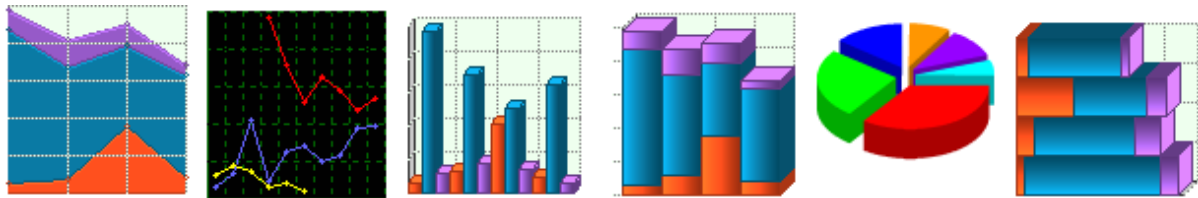
课程内容安排



上课日期	章节	教学内容	教学要点	作业	随堂测	学时
2.16	1	数据可视化, 描述性统计	1. 课程介绍 & 数据类型	作业1 (8%)	测试1 (8%)	2
2.23			2. 描述性统计 Descriptive Statistics & 数据常用可视化			2
3.2			3. 常用概率分布			2
3.9			4. 极限定理			2
3.16	2	推断性统计, 均值差异检验	5. 统计推断基础-1: 置信区间 Confidence Interval *	作业2 (12%)	测试2 (12%)	2
3.23			6. 统计推断基础-2: 假设检验 Hypothesis Test			2
3.30			7. 数值数据的均值比较-1: 单样本t-检验			2
4.6			8. 数值数据的均值比较-2: 独立双样本t-检验, 配对样本t-检验			2
4.13			9. 数值数据的均值比较-3: One-Way ANOVA			2
4.20			10. 数值数据的均值比较-4: Two-way ANOVA			2
4.27	3	比例差异检验	11. 类别数据的比例比较-1: 单样本比例推断 *	作业3 (4%)	测试3 (4%)	2
5.7 (调)			12. 类别数据的比例比较-2: 联立表的卡方检验			2
5.11	4	协方差, 相关分析, 回归分析	13. 相关分析 (Pearson r, Spearman rho, Kendal' s tau) *	作业4 (6%)	测试4 (6%)	2
5.18			14. 简单回归分析			2
5.25			15. 多元回归 Multiple Regression			2
6.1	5	Course Summary	16. 课程总结 *			2
			Total	30%	30%	32

*** 随堂测试**

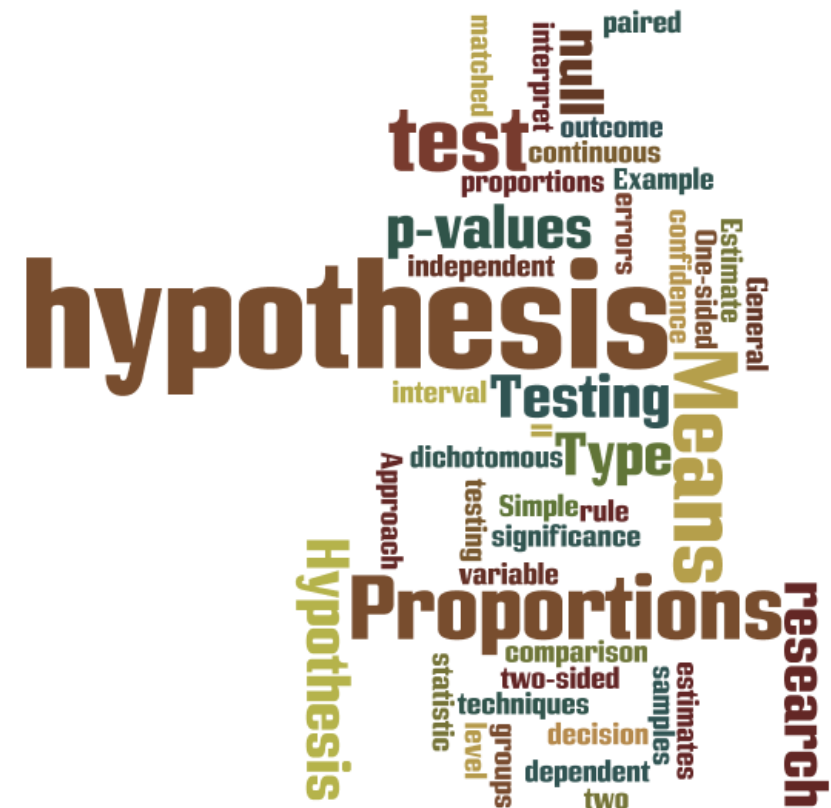




Week 6

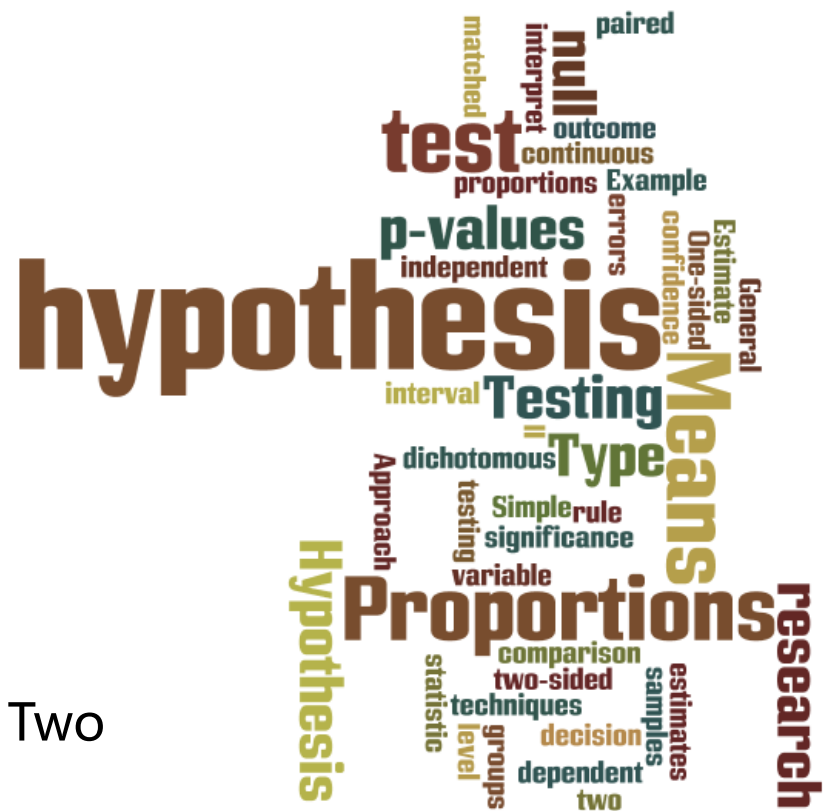
假设检验

back



单元2内容 (week 5-10)

- 总体、样本
 - 参数、统计量
- 置信区间
 - 区间估计、总体方差
- 均值差异
 - One-sample 均值估计
 - Two-sample 均值差异
- 零假设检验
 - H_0 vs H_a , Type I & II errors, P-value, One-sample vs Two sample test
- 方差分析
 - ANOVA test
- 非参数检验
- 样本量、效应量、功效 (Sample size, Effect size, Power)



Why do we use hypothesis testing?

- **Hypothesis testing** is an essential procedure in statistics. **假设检验**是统计学中一个必不可少的过程
- A **hypothesis test** evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data. 假设检验评估关于总体的两个相互排斥的陈述，来确定样本数据最能支持哪种陈述
- When **we** say that a finding is statistically significant, it's thanks to a **hypothesis test**. 因为我们有了**假设检验**这个工具，我们才能对一个发现/结果来描述它是否具有统计学意义



Why hypothesis testing?

科学中的许多重要决策，如是否应批准一种疾病的新疗法上市，**主要都是由数据驱动产生的。**

例如，一项针对新的降胆固醇药物的临床研究是否提供了强有力的证据证明对心脏病风险的患者有益？

- **可以从研究数据中计算出置信区间，以提供总体参数的合理值范围**
 - 例如总体胆固醇水平的**平均下降**。如果群体**平均下降效应**大到足以具有临床重要性，则认为药物对患者群体具有有益效果。
- **但同时还需要评估药物有效的证据强度；换句话说，观察到的效果是否比所预期的仅凭机会变化的要大？**
 - 假设检验是一种计算在常见的假设（或**零假设**）下**得到特定观察值的概率**的方法。通过假设数据来自零假设指定的分布，可以计算观察到与样本所代表的值一样极端的值的可能性。如果这种极端观察的机会很小，则有足够的证据可以拒绝零假设而支持替代假设（或**备择假设**）。



Statistical Hypotheses

A **statistical hypothesis**: a claim about the value of a parameter, population characteristic (could be a combination of parameters), or about the form of an entire probability distribution.

统计假设：关于参数值、总体特征（可能是参数的组合）或整个概率分布形式的表述

统计假设的一些例子：

- $H: \mu = 75$ cents, where μ is the true population average of daily per-student candy+soda expenses in US high schools $\mu = 75$ 美分是美国高中生每日糖果+苏打水的总体平均支出
- $H: p < 0.10$, where p is the population proportion of defective helmets for a given manufacturer $p < 0.10$ 是制造商的缺陷头盔的总体比例
- If μ_1 and μ_2 denote the true average breaking strengths of two different types of twine, one hypothesis might be the assertion that $\mu_1 - \mu_2 = 0$, and another is the statement $\mu_1 - \mu_2 > 5$
如果 μ_1 和 μ_2 表示两种不同类型麻线的真实平均断裂强度，一个假设可能是 $(\mu_1 - \mu_2 = 0)$ ，另一个假设是陈述 $(\mu_1 - \mu_2 > 5)$



Null vs Alternative Hypotheses 零假设 vs 备择假设

In any hypothesis-testing problem, there are always two competing hypotheses under consideration:

1. The status quo (null) hypothesis
2. The research (alternative) hypothesis

现状假设 (零假设、原假设)

研究假设 (备择假设)

一个在一直以来都被认为是常见的分布

由某种干预或条件变化引起的新分布

For example,

$\mu = .75$ versus $\mu \neq .75$

$p \geq .10$ versus $p < .10$

The objective of **hypothesis testing** is to decide, based on sample information, *if the alternative hypotheses is actually supported by the data.*

假设检验的目的是根据样本信息确定替代假设是否得到数据的支持

We usually do new research to challenge the existing (accepted) beliefs.

我们通常就是做新的研究来挑战现有的 (公认的) 知识



Burden of proof 举证责任

The **burden of proof** is placed on those who believe in the alternative claim. 举证责任落在那些相信 H_A 备择假设的人身上

In testing statistical hypotheses, the problem will be formulated so that one of the claims is initially favored. 在检验统计假设时，问题的表述将先肯定其中一个假设

This initially favored claim (H_0) will not be rejected in favor of the alternative claim (H_a or H_1) unless the sample evidence contradicts it and provides strong support for the alternative assertion. 除非样本证据与之相矛盾，并为备择假设提供有力支持，否则不会拒绝最初支持的零假设 (H_0) 而被备择假设 (H_a 或 H_1) 取代

If the sample does not strongly contradict H_0 , we will continue to believe in the plausibility of the null hypothesis. 如果样本与 H_0 没有强烈矛盾，我们将继续相信无效假设的合理性

The two possible conclusions: 1) *reject H_0* **拒绝零假设 H_0**
2) *fail to reject H_0* . **无法拒绝零假设 H_0**



No proof ... only evidence

We can never prove that a hypothesis is true or not true. 我们永远无法证明一个假设是真的还是假的

We can only conclude that it is or is not *supported by the data*. 我们只能得出数据是否支持这一假设的结论

A **test of hypotheses** is a method for *using sample data to decide whether the null hypothesis should be rejected* in favor of the alternative. 假设检验是一种使用样本数据来决定是否应该拒绝零假设，以支持备择假设的方法

Thus we might test the null hypothesis $H_0: \mu = .75$ against the alternative $H_a: \mu \neq .75$. Only if sample data strongly suggests that μ is something other than 0.75 should the null hypothesis be rejected. 因此，我们可以检验零假设 $H_0: \mu = .75$ vs. 备择假设: $H_a: \mu \neq .75$ 。只有在样本数据强烈证明 μ 不等于0.75的情况下，我们才可以拒绝零假设

In the absence of such evidence, H_0 should not be rejected, since it is still considered plausible. 在缺乏此类证据的情况下，我们不应拒绝 H_0 ，因为它仍被认为是合理的



通常，调查/科研人员先怀疑零假设 H_0 是不真实的，并进行假设检验，以评估针对零假设 H_0 的证据强度。拒绝或不拒绝零假设背后的逻辑类似于许多法律制度中的无罪推定原则。

innocent until proven guilty

例如在美国，被告在被证明有罪之前只能被假定为无罪；只有毫无疑问被告并非无辜的条件下，才能做出有罪判决。

所以在正式的假设检验方法中，**通常是先不会拒绝零假设** (H_0)，除非与之相矛盾的证据非常有力，以至于唯一合理的结论就是去拒绝 H_0 并支持 H_A 。

为什么要如此执着于无效假设？

Why be so committed to the null hypothesis?

- sometimes we do not want to accept a particular assertion unless (or until) data can show strong support 有时我们不想接受一个特定的假设，除非（或直到）数据能显示出强有力的支持
- reluctance (cost, time) to change 不愿改变（成本、时间）

Example: Suppose a company is considering putting a new type of coating on bearings that it produces. 例如：假设一家公司正在考虑在其生产的轴承上涂抹一种新型涂层

The true average wear life with the current coating is known to be 1000 hours. With μ denoting the true average life for the new coating, the company would not want to make any (costly) changes unless evidence strongly suggested that μ exceeds 1000. 已知当前涂层的实际平均磨损寿命为 $\mu = 1000$ 小时。除非有有力证据表明 μ 会超过1000，否则该公司不会进行任何（昂贵的）更改。



Hypotheses and Test Procedures

An appropriate problem formulation would involve testing

$H_0: \mu = 1000$ against $H_a: \mu > 1000$.

The conclusion that a change is justified is identified with H_a , and it would take conclusive evidence to justify rejecting H_0 and switching to the new coating.

H_a 的结论是可以更换新涂层，但需要充足的证据来拒绝零假设，去接受新结论

Scientific research often involves trying to decide whether a current theory should be replaced, or “elaborated upon.”

科学研究通常都是涉及到试图决定是否应该替换现有理论或是更新现在理论



What are basic of hypothesis testing?

• I 类错误 & II 类错误 (Type I/II error)

- **第一类错误 (I 类错误, 假阳性)** 也称为 α 错误, 是指当**零假设 (H_0) 正确时, 却拒绝了 H_0** , 所犯的这个错误。这意味着研究者的结论**不正确**, 即**观察到了实际上并不存在的效应**。
- **第二类错误 (II 类错误, 假阴性)** 也称为 β 错误, 是指当**零假设 (H_0) 错误时, 反而去接受了 H_0** , 把它当真了, 所犯的这个错误。这意味着研究者的结论**不正确**, 即**没有观察到本就存在的效应**。
- $\alpha + \beta$ **不一定** 等于 1

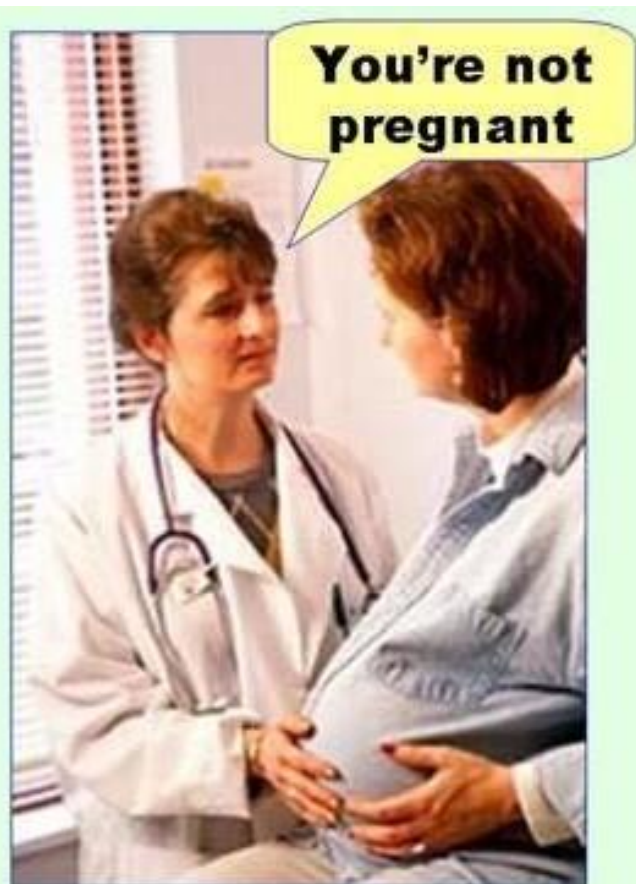
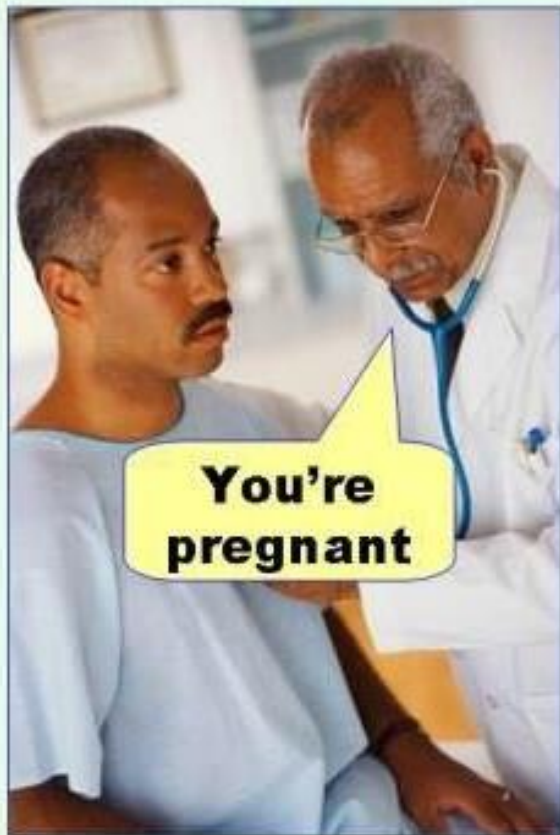
		真实情况	
		H_0 真实的	H_0 错误的
根据研究判断结论	拒绝 H_0	I类错误 α (假阳)	正确判断
	接受 H_0	正确判断	II类错误 β (假阴)

陪审团审判案子		真实情况	
		无罪	有罪
裁决	有罪	错误	正确
	无罪	正确	错误

以下各是
哪类错误?

零假设 (H_0) :
病人没有怀孕

备择假设 (H_1) :
病人怀孕



提交

A

第I类错误 (False positive)

C

第I类错误 (False positive)

B

第II类错误 (False negative)

D

第II类错误 (False negative)

显著性水平 α (Level of significance)

The significance level α can be thought of as the value that quantifies how rare or unlikely an event must be in order to represent sufficient evidence against H_0 .

Typically, α is chosen to be 0.05, 0.01, or some other small value.

In the context of decision errors, α is the probability of making a Type I error.

- Type I error refers to incorrectly rejecting the null hypothesis.
- ❖ 显著性水平 α 可以被认为是，作为拒绝 H_0 假设的概率，量化**事件的罕见程度或不太可能程度**的值
- ❖ 通常， α 为0.05、0.01或其他一些小值
- ❖ 在决策错误的情况下， α 就是发生I-型错误的概率



统计量

“检验统计量” 测量观察数据与零假设为真时预期数据之间的差异

The test statistic measures the discrepancy between the observed data and what would be expected if the null hypothesis were true.

- Specifically, how many standard deviations is the observed sample value from the population value under the null hypothesis? 具体说，在零假设下，观察到的样本值与总体值之间有多少标准差？

When testing hypotheses about a mean, the test statistic is

当检验关于均值的假设时，检验统计量是

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}},$$

where the test statistic T follows a t distribution with $n - 1$ degrees of freedom.

其中检验统计量 T 服从具有 $n - 1$ 个自由度的 t 分布

在统计学中，自由度 (degree of freedom df) 指的是计算某一统计量时，取值不受限制的变量个数通常 $df = n - k$ ，其中 n 为样本数量， k 为被限制的条件数或变量个数，或计算某一统计量时用到其它独立统计量的个数，自由度通常用于抽样分布中。



p-value p值

p值 (P value) 就是**当零假设为真时**，出现偏离零假设值的观测值以及比观测值更极端的值的概率，说白了P值是个概率值，是用来判定假设检验结果的一个参数。p值是根据实际统计量计算出的显著性水平。

p值是Fisher先提出来的“显著性检验”理论体系中的概念，假设检验之所以可行，其理论背景是**小概率理论**，小概率事件在一次试验中几乎是不可能发生的，但是它一旦发生，我们就有理由拒绝原假设，p值越小，拒绝原假设的理由越充分；反之，小概率事件没有发生，则认为原假设是合理的。

正常步骤：计算与检验统计相关的p值，并将其与显著性水平 α 进行比较

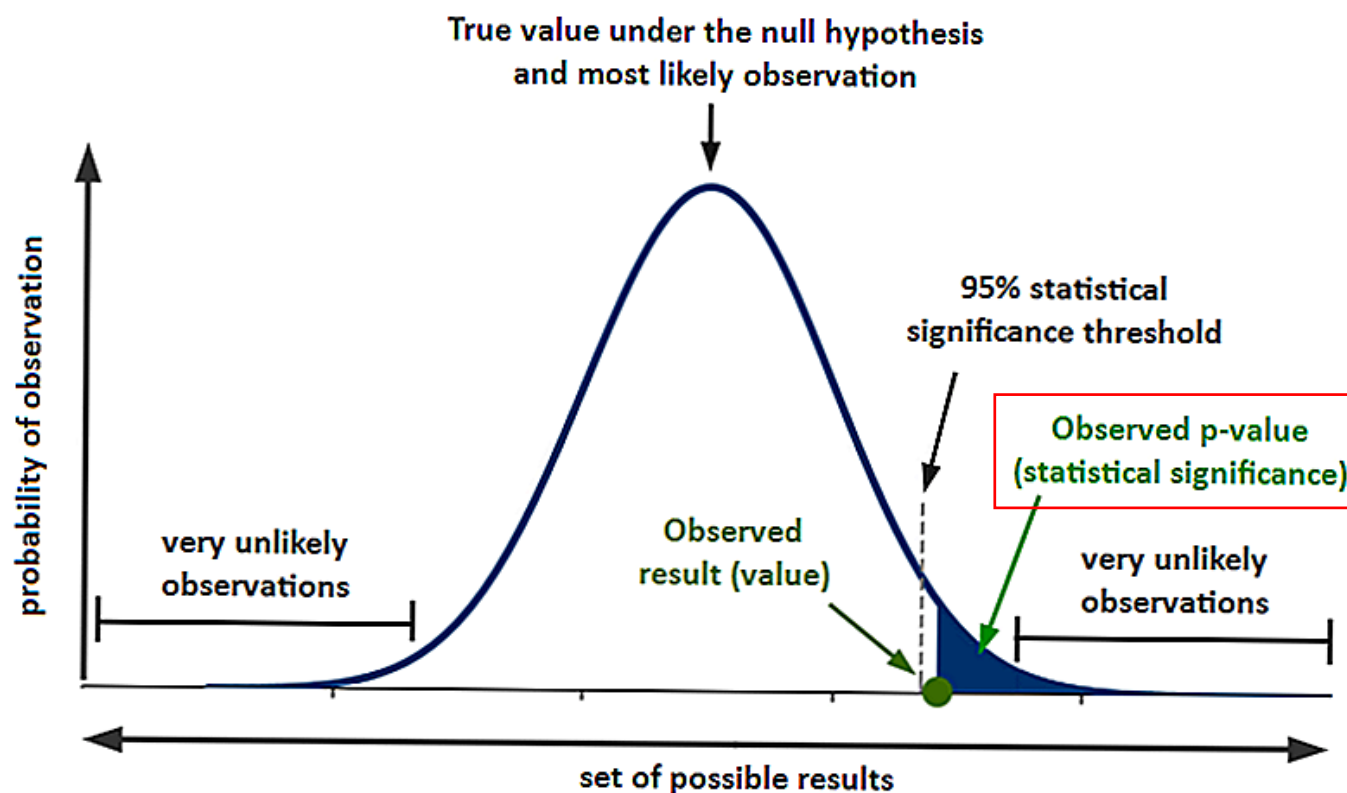
最常见是比较 p 值与显著性水平 α (alpha)。 α 是当 H_0 为真时否定 H_0 的概率

通常，p 值越小，否定 H_0 的样本证据越强大。更具体地说，p 值是**导致否定 H_0 的最小 α 值**。对于任何大于 p 值的 α 值，将无法否定 H_0 ，而对于任何 p 值的 α 值，可否定 H_0

- 显著性水平是原假设为真时拒绝零假设的概率，**也即上述所说的小概率的界限**，常取值0.05，0.01。
- 在显著性水平 α 下，P值规则为： $P \leq \alpha$ ，则拒绝 H_0 ；如果 $P > \alpha$ ，则不拒绝原假设。
- 我们通常把 $1-\alpha$ 称为**置信水平**，即对推断结果的把握度、可靠性



Probability & Statistical Significance Explained



正确使用

1. p值并不暗示任何影响、差异或关系的强度或大小。同时添加相关系数或平均值可以帮助读者更好地理解你的研究结果。
2. 为主要结果写上精确的p值，以维护科学的严谨性。如果确切的p值小于.001，你可以写 " $p < .001$ "。
3. 由于p值不能等于0，请将 " $p = .000$ " 替换为 " $p < .001$ "，因为后者被视为标准做法。
4. 确保在表述中使用 " $p < .05$ " 而不是 " $p < 0.05$ "，因为在数值不能大于1的情况下，大多数专家不赞成在小数点前加零。

Level of significance and types of tests

Level of significance (α)

1. **显著性水平 (Level of significance):** 指我们接受或拒绝零假设的显著性程度。接受或拒绝一个假设是不可能100%准确的, 因此我们选择一个通常为5%的显著性水平

The probability of rejecting a null hypothesis that is true; the probability of making this error.

2. 这通常用 α 表示, 通常是0.05或5%, 这意味着输出值应该是有95%的信心在每个样本中都能有类似的结果

Common significance levels

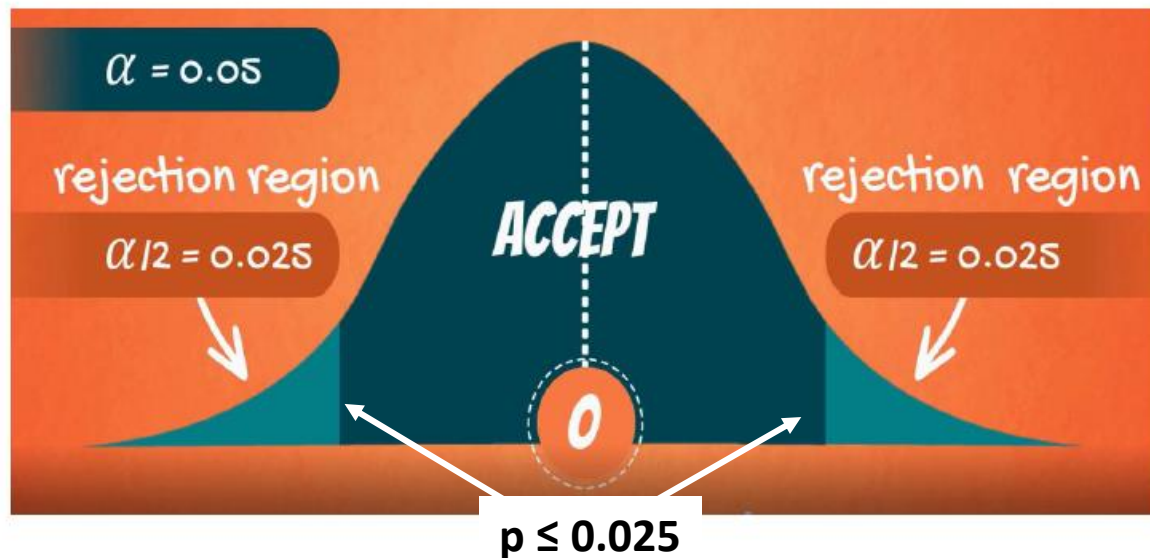
0.10

0.05

0.01

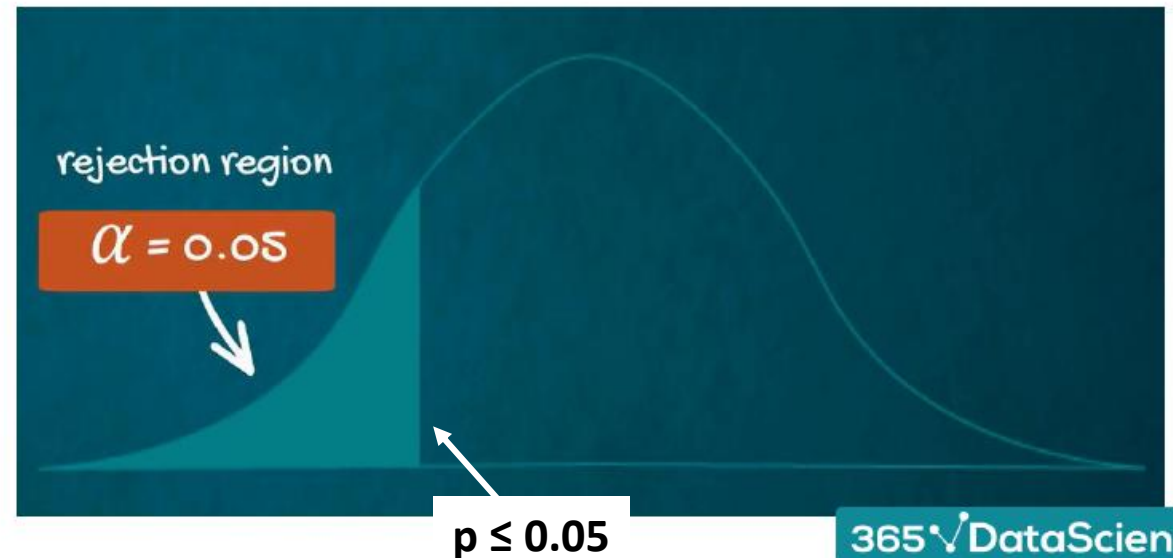
3. **双边检验 (Two-tailed test):** 双尾检验是一种统计检验, 分布的**临界区域**是双面的, 检验一个样本是否大于或小于一定范围的值。如果被测样本落在**任一临界区域**, 则接受替代假设, 而不是零假设

Used when the null contains an equality (=) or an inequality sign (\neq)



4. **单边检验 (One-tailed test):** 对一个统计假设的检验, 其中拒绝区域只在**抽样分布的一侧**, 称为单尾检验

Used when the null doesn't contain equality or inequality sign ($<$, $>$, \leq , \geq)



研究实例

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

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 randomized 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/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有氧运动能力，但对BMI影响不大



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。



先设好假设检验

首先，我们来看一下假设检验。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。



P-值

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。



上述研究于2014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，**以下说法是否正确？**

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**。

- A

是
- B

否

		真实情况	
		H ₀ 真实的	H ₀ 错误的
根据研究判断结论	拒绝H ₀	I类错误 α (假阳)	正确判断
	接受H ₀	正确判断	II类错误 β (假阴)

提交

上述研究于2014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，**以下说法是否正确？**

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**。

- ☒ A 是
- ☐ B 否

		真实情况	
		H ₀ 真实的	H ₀ 错误的
根据研究判断结论	拒绝H ₀	I类错误 α (假阳)	正确判断
	接受H ₀	正确判断	II类错误 β (假阴)

提交

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，**以下说法是否正确？**

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。

☒ A 是

☐ B 否

提交

问题

上述研究于2014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**。

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**

		真实情况	
		H ₀ 真实的	H ₀ 错误的
根据研究判断结论	拒绝H ₀	I类错误 α (假阳)	正确判断
	接受H ₀	正确判断	II类错误 β (假阴)



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**。

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**

		真实情况	
		H ₀ 真实的	H ₀ 错误的
根据研究判断结论	拒绝H ₀	I类错误 α (假阳)	正确判断
	接受H ₀	正确判断	II类错误 β (假阴)



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。

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**



提示

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，找找原因看看是否有新的发现。

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 ）



谢谢，下周见！



让开，
我要去学习了

