

Lecture 6 – Hypothesis testing 假设检验

- Outline for today
 - 知识回顾 L01-L05: statistics and uncertainty
 - 假设检验 Hypothesis testing
 - I类和II类错误 Type I and Type II errors
 - 单侧和双侧检验 One-sided and two-sided tests
 - 课堂练习 Practice Problems
 - 小结 Summary

Lecture 6 – Hypothesis testing 假设检验

- 学习目标

- 理解假设检验的统计意义及其步骤;
- 掌握假设检验的步骤并运用其来尝试解释生物学现象;
- 熟知假设检验中的错误及其应对方式;

- 这节课会非常多的**术语**

1. Recall from L01 – L05

- 统计学的目的是基于从总体中的样本所获得的信息
 - 通过随机抽样获取样本（这也是我们做统计推断的常用假设）
- 基于样本应用不同估计值对总体进行推断（参数估计）
 - 均值、标准差、中值、四分位距
- 同时提供推断（估计值）的精确性
 - 标准误、置信区间

1. Recall from L01 – L05

- 随机样本 a random sample
 - 子集：总体包含的每个个体有相同概率被抽中，且相互独立
- 抽样误差 sampling error
 - 因随机抽样造成的估计值与真实值（总体参数）之间的偶然差异（偶然差异by chance）
 - 准确性与精确性

1. Recall from L01 – L05

- 估计值及其计算
 - 均值mean、中值median
 - 方差variance、标准差standard deviation、四分位距IQR
- 估计值的准确性/不确定性的量化
 - 标准误 standard error
 - 置信区间 confidence interval (2 SE rule)



2. 假设检验 hypothesis testing

- 统计学：基于样本来对总体进行推断

- mean
- median
- SD
- proportion
- SE, CI

参数估计
parameter
estimation

为总体特征给出数值及范围
(values and bounds)

“新药的治疗效果有多大？”
“How large is the effect?”

假设检验
hypothesis
testing

H_0
 H_A

询问参数是否不同于特定的期望
(a specific “null” expectation H_0)

“新药有治疗效果吗？”
“Is there any effect at all?”

2.1 基于概率的假设检验

- 例子：脊髓灰质炎疫苗/小儿麻痹症疫苗
 - 试验：401 974名小学生被随机分为两组：
 - 实验组接种疫苗（预防）
 - 对照组注射生理盐水（安慰剂）
 - 不知道自己接种的是什么（编码）
- 1954年索尔克疫苗
 - polio vaccine by Jonas Salk
 - 史上规模最大的公共卫生试验
 - “它开启了现代疫苗评估的大门”



理查德·马尔瓦尼医生将索尔克的脊髓灰质炎疫苗注入了六岁的兰迪·克尔的左臂，美国历史上规模最大的公共卫生试验就此拉开帷幕。

<https://www.guokr.com/article/440644/>

2.1 基于概率的假设检验

- 例子：脊髓灰质炎疫苗/小儿麻痹症疫苗
 - 试验：401 974名小学生被随机分为两组：
 - 实验组接种疫苗（预防）
 - 对照组注射生理盐水（安慰剂）
 - 结果
 - 实验组 0.016% 学生患上麻痹性脊髓灰质炎
 - 对照组 0.057% 学生患上麻痹性脊髓灰质炎
 - 疫苗似乎将发病率降低了三分之二
 - 但两组组之间的差异非常小，每 10 000 人中的差别只有约 4 例。
 - 是疫苗起作用，还是纯属偶然？

2.1 基于概率的假设检验

- 假设检验

- 零假设 (the null hypothesis) 是疫苗没有作用，观察到的组间差异只是偶然发生的 (by chance)。

- 评估零假设则需要在疫苗没有作用的假设下，计算两组间的患病差异比观察到的差异更大的概率。
 - 结果发现这个概率很小（即观察到的差异相比而已较大）。

- 因此，否定零假设

- 尽管疫苗接种组和对照组的发病率并无太大差异（0.057% vs 0.016%），但由于索尔克疫苗试验的规模非常大（超过 40 万人参与），因此能够证明两者之间确实存在差异。
 - 疫苗有预防效果，使许多孩子免于患病，在随后几十年中的成功也证明了这一点。

2.2 提出并使用统计假设

- 假设检验

- 将实际数据与我们在特定零假设为真时所期望看到的结果进行比较。
- 如果数据与零假设成立时的预期数据相比过于异常，则拒绝零假设。

- 对单一总体中的比例进行检验

- 在本讲中我们将此此为例（最简单的示范）来说明假设检验的基本原理。
- 目标是用最少的计算来介绍主要概念。
- 本课以后的内容将介绍假设检验的许多具体方法。

2.2 提出并使用统计假设

- 统计假设与科学假设
 - 统计假设，包括零假设(the null hypothesis)和备择假设(the alternative hypothesis)都是关于总体特征的简单陈述。
 - 科学假设是关于自然现象(natural phenomena)的存在和可能原因的陈述。
 - 不要混淆统计假设与科学假设。
- 科学家设计实验研究和观察研究来检验科学假设的预测。
 - 将统计假设应用于所得数据有助于确定这些科学假设的哪些预测与数据一致，哪些不一致。



2.2 提出并使用统计假设

- 统计假设与科学假设
 - 统计假设，包括零假设(the null hypothesis)和备择假设(the alternative hypothesis)都是关于总体特征的简单陈述。
 - 科学假设是关于自然现象(natural phenomena)的存在和可能原因的陈述。
 - 不要混淆统计假设与科学假设。
- 例子？

2.3 零假设 A null hypothesis - H_0

- 零假设是关于总体参数值的一种具体说法 (a specific claim).
 - 它是为了论证的目的(the purposes of argument)而提出的，通常是一个可被质疑的观点。
 - 通常，零假设是指我们感兴趣的总体参数为零，即无影响、无偏好、无相关、或无差异 (i.e., no effect, no preference, no correlation, or no difference)。
 - 一般来说，零效假设是人们倾向于拒绝的一个陈述。

2.3 零假设 A null hypothesis - H_0

- 例如：科学问题——药物 X 是否有正面效果？
 - H_0 : “X 药物 不会影响 Y 疾病患者的平均寿命”。
- 如果我们能够拒绝这一陈述，那么我们就获得了一些有用的信息：
 - 这些患者在服用 X 药物后，平均寿命会更长或更短。
- 因此，拒绝零假设将为预测药物 X 有正面效果的科学假设提供支持。
 - 如果服用 X 后平均寿命会更长的话
- 而不拒绝零假设则不会提供这种支持。

2.3 零假设 A null hypothesis - H_0

- H_0 : 特定的 (*specific*) 或基于理论和已有知识预期。
 - 确定了我们感兴趣的参数的一个特定值。
- 一项调查流网捕鱼 (drift-net fishing) 对海豚密度影响的研究
- H_0 : 在有流网捕鱼的区域和没有流网捕鱼的区域, 海豚密度相同。
(The density of dolphins is the same in areas with and without drift-net fishing.)
- H_0 : 一对父母眼睛为棕色, 他们各有一个蓝色眼睛的父母, 那么, 他们子女中棕色和蓝色眼睛的比例为 3:1。
- H_0 : 健康人的平均体温为37 摄氏度。

2.4 备择假设 An alternative hypothesis - H_A

- 备择假设包括零假设中所述值之外的所有其它可能的总体参数值。
 - 通常包括在生物学上比零假设中的值更有趣的可能性 (biologically more interesting)。
- 备择假设通常包括科学假设所预测的参数值。
 - 因此研究人员通常希望备择假设为真（但并非总是如此）。
- 与零假设相比，备择假设是非特异性的。
 - 除了零假设中特定值以外，总体特征参数的所有可能值或任何对比都包括在内。
 - 一样 vs. 不一样 → 或大或小

2.4 备择假设 An alternative hypothesis - H_A

- 例如:

- 一项调查流网捕鱼 (drift-net fishing) 对海豚密度影响的研究

- H_0 : 在有流网捕鱼的区域和没有流网捕鱼的区域, 海豚密度相同。

(The density of dolphins is the same in areas with and without drift-net fishing.)

- H_A : 在有流网捕鱼的区域和没有流网捕鱼的区域, 海豚密度不相同。

(The density of dolphins *differs* between areas with and without drift-net fishing.)

- H_0 : 一对父母眼睛为棕色, 他们各有一个蓝色眼睛的父母, 那么, 他们子女中棕色和蓝色眼睛的比例为 3:1。

- H_A : 一对父母眼睛为棕色, 他们各有一个蓝色眼睛的父母, 那么, 他们子女中棕色和蓝色眼睛的比例是 3:1以外的比例。

- H_0 : 健康人的平均体温为37 摄氏度。

- H_A : 健康人的平均体温不是37 摄氏度。

2.4 备择假设 An alternative hypothesis - H_A

- 零假设和备择假设的地位并不相同，零假设是用数据进行检验的唯一陈述。
 - 我们假设 H_0 为真，而假设检验是要看数据告诉我们什么信息。
 - 如果数据与零假设一致，意味着我们没有拒绝零假设(failed to reject it)。
 - 我们从不 "接受 "零假设 (we never “accept” the null hypothesis) 。
 - 如果数据与零假设不一致，那么拒绝零假设，并说数据支持备择假设。
- 拒绝 H_0 意味着我们排除了零假设的值。
 - 它还告诉我们，与零假设值相比，真实值可能的方向。
 - 但是，拒绝假设本身并不能揭示群体参数的大小或处理间差异的大小。
 - 我们使用估计值本身来提供大小。

2.5 假设检验: 示例——handedness in toads

- 假设检验的步骤:
 - 1. 提出假设 (state the H_0 & H_A) 。
 - 2. 利用数据计算检验统计量 (compute the test statistic with data)。
 - 3. 确定 P 值 (determine the P-value)。
 - 4. 得出适当的结论 (draw the appropriate conclusions) 。
- 示例: [European toads \(*Bufo bufo*\)](#) (Bisazza et al. 1996)
 - 人类主要是右撇子。
 - 其它动物是否也有惯用手的现象?
- 注意:
 - 这个示例针对的是比例——以说明基本过程, 而不涉及概率计算的细节。
 - 假设检验可以应用于多种统计量, 如均值、方差、均值差异、相关性等。
 - 在未来的学习中会讨论有关其它统计量计算的细节。



Hintau Aliaksei/Shutterstock.com



2.5 假设检验: 示例——handedness in toads

- 步骤一: 提出假设
 - 怎么提出 (或确定) 零假设 vs 备择假设?
- 实验体系 the experiment:
 - 样本: 从野外捕获18只蟾蜍
 - 假设这是一个随机样本 (a random sample)
 - 测试方法:
 - 研究人员给每只蟾蜍头上都缠着一个气球;
 - 记录下每只蟾蜍用哪只前肢去除掉气球 (remove the balloon)。



Hintau Aliaksei/Shutterstock.com

2.5 假设检验: 示例——handedness in toads

- 步骤一: 提出假设

- 科学问题:

- 在蟾蜍总体中右撇子(right-handed)和左撇子(left-handed)出现的频数相同, 还是像人类那样一种类型比另一种类型更常见?

- 统计假设:

- H_0 : 频数相同 (the sample frequency)

$$H_0: p_{\text{left}} = p_{\text{right}} = 0.05$$

- H_A : 频数不同 (not equally frequent)

$$H_A: p_{\text{left}} \neq p_{\text{right}}$$

- 双边检验 two-sided test

- $p_{\text{left}} > p_{\text{right}}$ or $p_{\text{left}} < p_{\text{right}}$

- 在获取数据前无法排除任何一种可能性 (方向)

2.5 假设检验: 示例——handedness in toads

- 步骤一: 提出假设

- H_0 : 频数相同 (the sample frequency)
- H_A : 频数不同 (not equally frequent)

$$H_0: p_{\text{left}} = p_{\text{right}} = 0.05$$

$$H_A: p_{\text{left}} \neq p_{\text{right}}$$

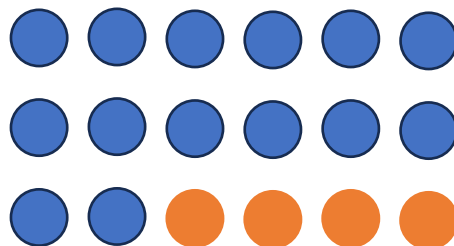
- 实验体系: 野外捕获的18只蟾蜍

- 记录用哪只前肢除掉缠在头上的气球

- 实验结果:

➤ 14 right-handed

➤ 4 left-handed



Hintau Aliaksei/Shutterstock.com

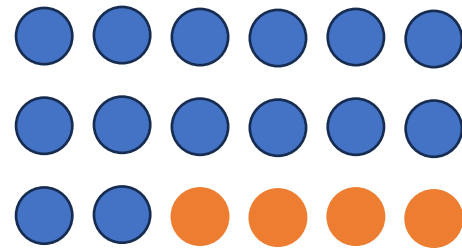


2.5 假设检验: 示例——handedness in toads

- 步骤二: 根据数据计算检验统计量 the test statistic
 - 检验统计量是根据数据计算出来的数值, 用以评估观察结果与零假设 H_0 条件下的预期结果的吻合程度。
- 18只蟾蜍中使用右前肢的偏好（右撇子）：个数或者比例。

➤ 14 right-handed

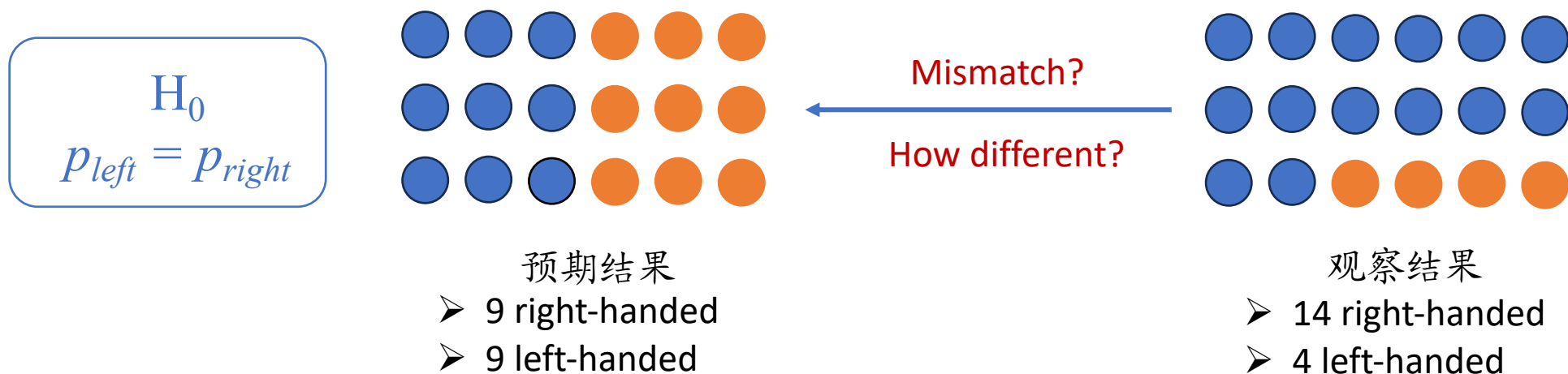
➤ 4 left-handed



- 这个结果是否说明蟾蜍有手性偏好呢？

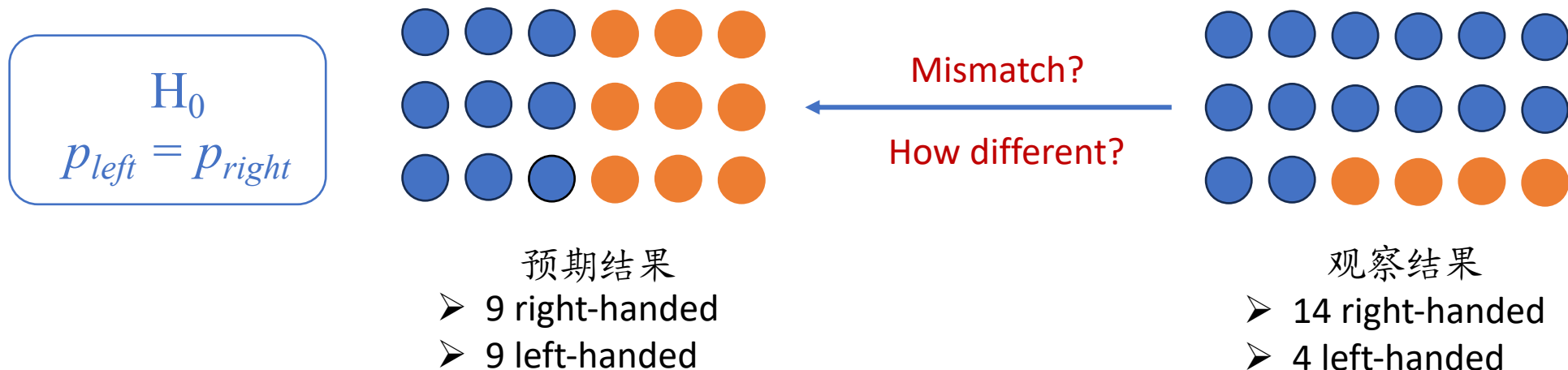
2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - 检验统计量用以评估观察结果与零假设 H_0 条件下的预期结果的吻合程度。
 - 由于抽样过程中偶然性的影响 (特别是在数据不多的情况下), 观察结果与 H_0 下的预期结果之间通常会存在差异。
 - 所以, 要判断数据是否符合 H_0 , 我们必须计算出当假设零假设为真时, 出现与观察结果一样或比观察结果更极端的情况的概率。



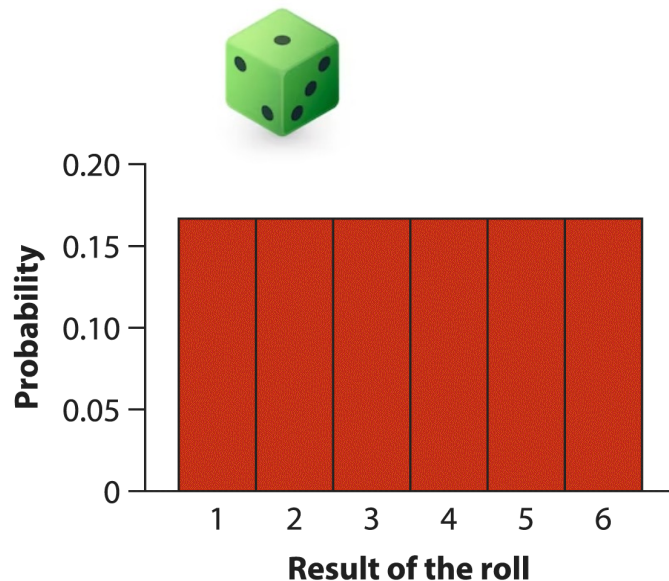
2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - 当假设零假设 H_0 为真时, 出现与观察结果一样或比观察结果更极端的情况的概率。
 - 要获得这个概率, 我们需要确定当 H_0 为真时的检验统计量的抽样分布 (the sampling distribution of the test statistic) 。
 - H_0 下检验统计量的可能值及其相关概率 → 零分布 (the null distribution)。
 - 可利用计算机的高性能来进行计算。

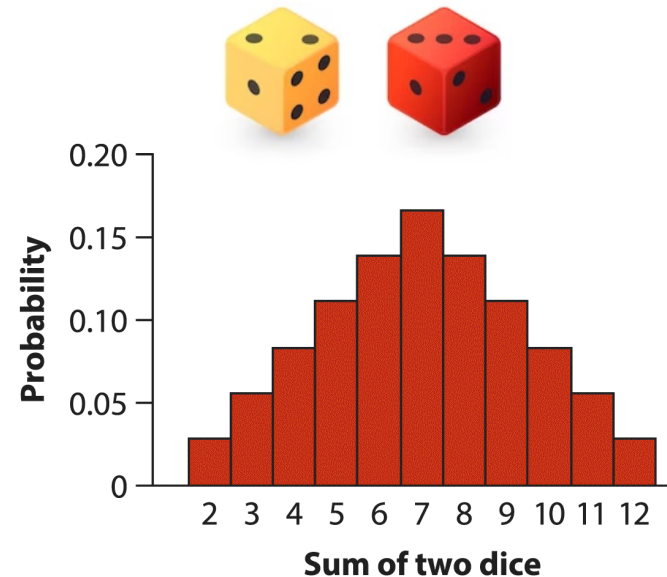


2.5 假设检验: 示例——handedness in toads

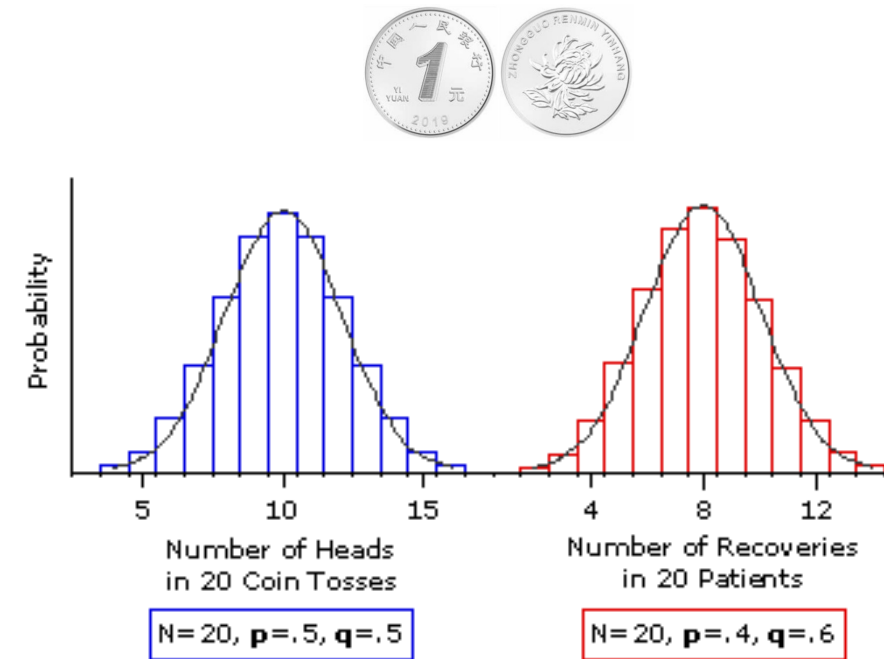
- 步骤三: 确定 P 值 (量化不确定性)
 - 检验统计量的抽样分布——零分布 the null distribution
 - 同等条件下大量重复试验的概率分布 (包含各种可能结果)



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

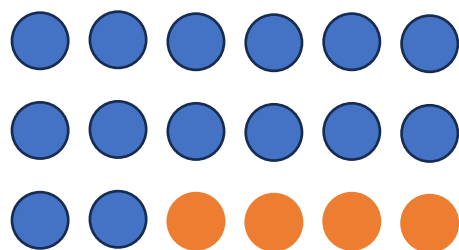


Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company



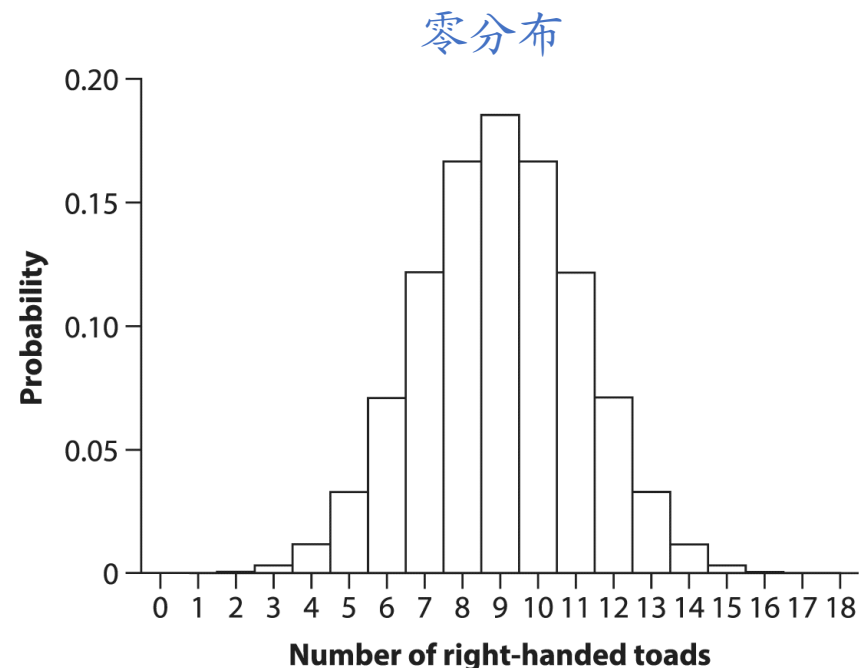
2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - 检验统计量的抽样分布——零分布 the null distribution
 - 18只蟾蜍中使用右前肢的偏好 (右撇子): 个数或者比例。
 - 类似于抛18枚硬币, 然后记录其中数字朝上 (两种结果之一) 的个数。
 - 18只蟾蜍中右撇子的个数 $\subset [0, 18]$



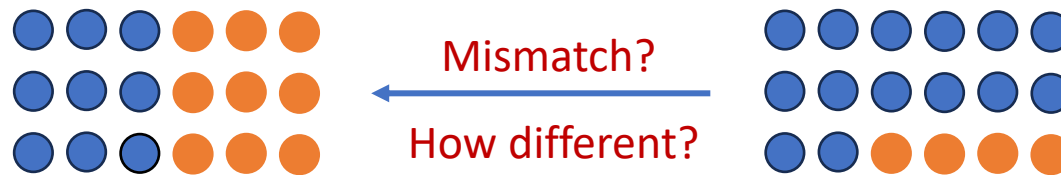
观察结果

- 14 right-handed
- 4 left-handed

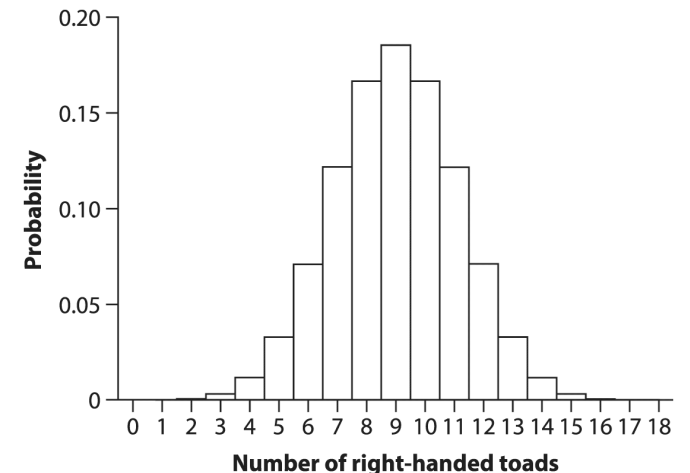


2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - $14/18 \neq 9/18$? 但这种不匹配是否大到足以排除偶然性?



- 要判断数据是否符合 H_0 , 我们必须计算出当假设零假设为真时, 出现与观察结果一样或比观察结果更极端的情况的概率? (the probability of all results *as unusual as or more unusual than* that exhibited by the data)



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

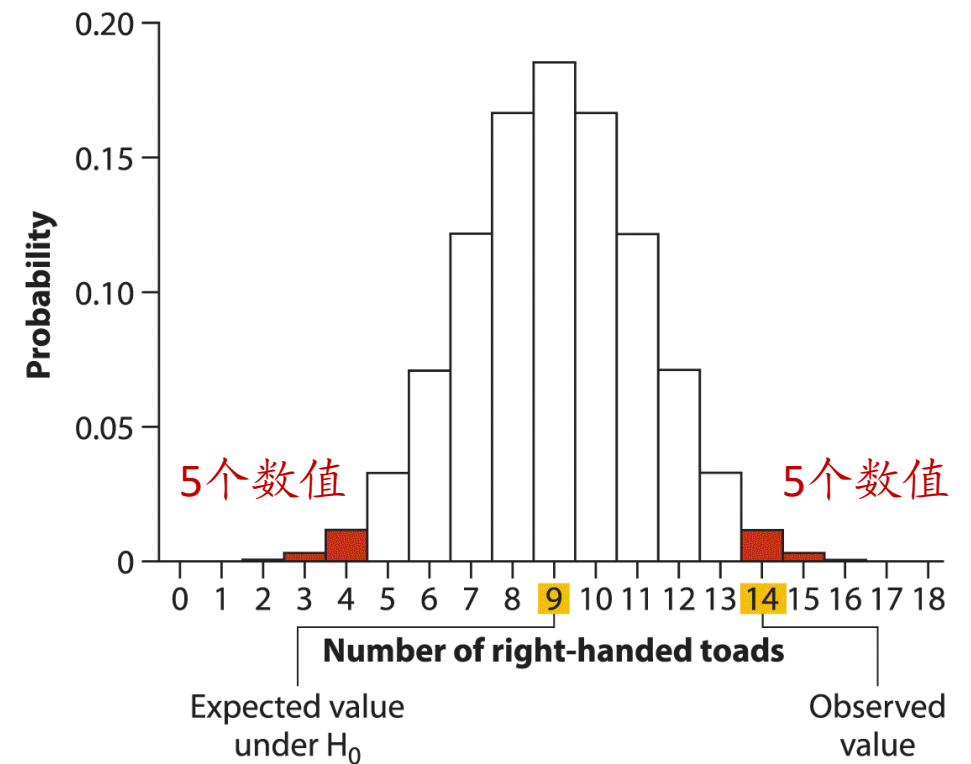
- $\Pr[\text{大于等于14只蟾蜍是右撇子}] = ?$

2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - P 值是在零假设 H_0 为真的情况下获得观察数据 (或更极端结果) 的概率。
 - 如果 P 值很小, 那么意味着数据与 H_0 的预期结果不一致, 我们需要拒绝 H_0 ;
 - 否则, 我们不拒绝 H_0 。
 - 一般来说, P 值越小, 反对 H_0 的证据就越强。
 - P 值不是零假设为真的概率
 - “假设” (hypothesis) 不是随机试验的结果, 因此没有概率。
 - P 值指的是在零假设下进行数据抽样时发生特定事件的概率。

2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - P 值是在零假设 H_0 为真的情况下获得观察数据 (或更极端结果) 的概率。
- 计算 P 值
 - 在计算 P 值时, 需要考虑到与观测数据一样不寻常或甚至更极端的结果。
 - 若 H_0 为真, 它们会出现的概率很低
 - 零分布的左右两个尾部的各5个数字
 - 包括 ≥ 14 & ≤ 4
 - 4只右撇子 = 14只左撇子



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

2.5 假设检验: 示例——handedness in toads

- 步骤三: 确定 P 值 (量化不确定性)
 - 假设零假设 H_0 为真的前提下, 计算P 值
(assuming that the null hypothesis is true)
 - $\Pr[\geq 14 \text{ right-handed toads}] = \Pr[14] + \Pr[15] + \Pr[16] + \Pr[17] + \Pr[18]$
 $= 0.0155$
 - $\Pr[\geq 14 \text{ left-handed toads}] = \Pr[4] + \Pr[3] + \Pr[2] + \Pr[1] + \Pr[0]$
 $= 0.0155$
 - $\text{P-value} = \Pr[\geq 14 \text{ right-handed toads}] + \Pr[\geq 14 \text{ left-handed toads}]$
 $= 2 * 0.0155 = \mathbf{0.031}$

Number of right-handed toads	Probability
0	0.000004
1	0.00007
2	0.0006
3	0.0031
4	0.0117
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	0.0117
15	0.0031
16	0.0006
17	0.00007
18	0.000004
Total	1



2.5 假设检验: 示例——handedness in toads

- 步骤四: 给出恰当的结论 Draw the appropriate conclusion
 - 如果 P 值很 "小", 则拒绝零假设; 否则, 不拒绝 H_0 。
 - P 值 = 0.031
 - P 值多小算足够小?
 - 按照大多数生物研究领域的惯例, P 值的临界值是 0.05。
 - 如果 $P \leq 0.05$, 我们就拒绝零假设; 如果 $P > 0.05$, 我们就不拒绝它。
 - 这个 P 的判定临界值 (即 $P = 0.05$) 称为显著性水平, 用 α 表示。
- 那么, 蟾蜍有手性偏好吗?



2.5 假设检验: 示例——handedness in toads

- 步骤四: 给出恰当的结论 Draw the appropriate conclusion
 - 如果 P 值很 "小", 则拒绝零假设; 否则, 不拒绝 H_0 。
 - P 值 = 0.031
 - P 值多小才算足够小?
 - 按照大多数生物研究领域的惯例, P 值的临界值是 0.05。
 - 如果 $P \leq 0.05$, 我们就拒绝零假设; 如果 $P > 0.05$, 我们就不拒绝它。
 - 这个 P 的判定临界值 (即 $P = 0.05$) 称为显著性水平, 用 α 表示。
- 那么, 蟾蜍有手性偏好吗?
 - 我们拒绝在蟾蜍总体中左撇子和右撇子出现频率相同的零假设。
 - 我们基于已有数据得出的结论是: 大多数蟾蜍是右撇子。

$$H_0: p_{\text{left}} = p_{\text{right}}$$

$$H_A: p_{\text{left}} < p_{\text{right}}$$



2.5 假设检验: 示例——handedness in toads

- 步骤五: 报告结果 Reporting the results
 - 在研究论文或实验报告中撰写统计检验结果时, 一般要包含以下信息:
 - 检验统计量的值 the value of the test statistic
 - 样本量 the sample size
 - P 值 the P-value
 - 此外, 最好能提供相关统计参数的置信区间或标准误差。
 - P 值表示反对零假设的证据的权重 (P 越小表示证据越有力)
 - 置信区间则衡量效应的大小 (the size of the effect based on the test statistic)
 - 例如:
 - 18 只蟾蜍中有 14 只是右撇子 (在这种情况下, 既给出了检验统计量, 也给出了样本量)
 - P-value = 0.031; $0.54 < \text{Proportion} < 0.91$ (95% CI);

3. 假设检验中的错误

- I 类错误和 II 类错误 Type I and Type II errors

- I 类错误: 拒绝为真的零假设 (rejecting a true null hypothesis)

- 显著性水平 α (the significance level) 设定了犯一类错误的概率。

- 生物学研究中通常将 $\alpha = 0.05$ 视为可接受的错误率 (an acceptable error rate)。

- 降低 α (如 $\alpha=0.01$) 会使得 $P\text{-value} > \alpha$, 这样会增加拒绝为真的 H_0 的难度 (好处: 没有拒绝 H_0), 同时也会增加拒绝为假的 H_0 的难度 (犯 II 类错误)。

		Actual Situation	
		H_0 True	H_0 False
Decision	Fail to Reject H_0	✓	Type II Error (β)
	Reject H_0	Type I Error (α)	✓

3. 假设检验中的错误

- I 类错误和 II 类错误 Type I and Type II errors
 - I 类错误: 拒绝为真的零假设 (rejecting a true null hypothesis)
 - II 类错误: 没有拒绝为假的零假设 (failing to reject a false null)
 - 如果一项研究犯II类错误的概率较低, 我们则说它功效 (power) 较高。
 - 功效 (power): 拒绝为假的零假设的概率。
 - 在其它条件相同的情况下, 一项研究的功效越大, 则是越好的研究。

		Actual Situation	
		H_0 True	H_0 False
Decision	Fail to Reject H_0	✓	Type II Error (β)
	Reject H_0	Type I Error (α)	✓

Power = 1 - β

3. 假设检验中的错误

- I 类错误和 II 类错误 Type I and Type II errors

- 功效 power

- 功效很难量化，因为拒绝零假设的概率取决于事实与零假设的差异程度。
 - 检测小的差异比检测大的差异更困难——因为我们永远不知道真实值是多少，所以我们通常无法有把握地预测一项研究的真实功效有多大。
 - 如果样本量越大、或者如果事实与零假设的差异越大、或者如果总体中个体的变异越低，那么一项研究的功效就越大。
 - 在后面学习实验设计时，我们将讨论如何计算功效以及如何设计研究以优化功效。

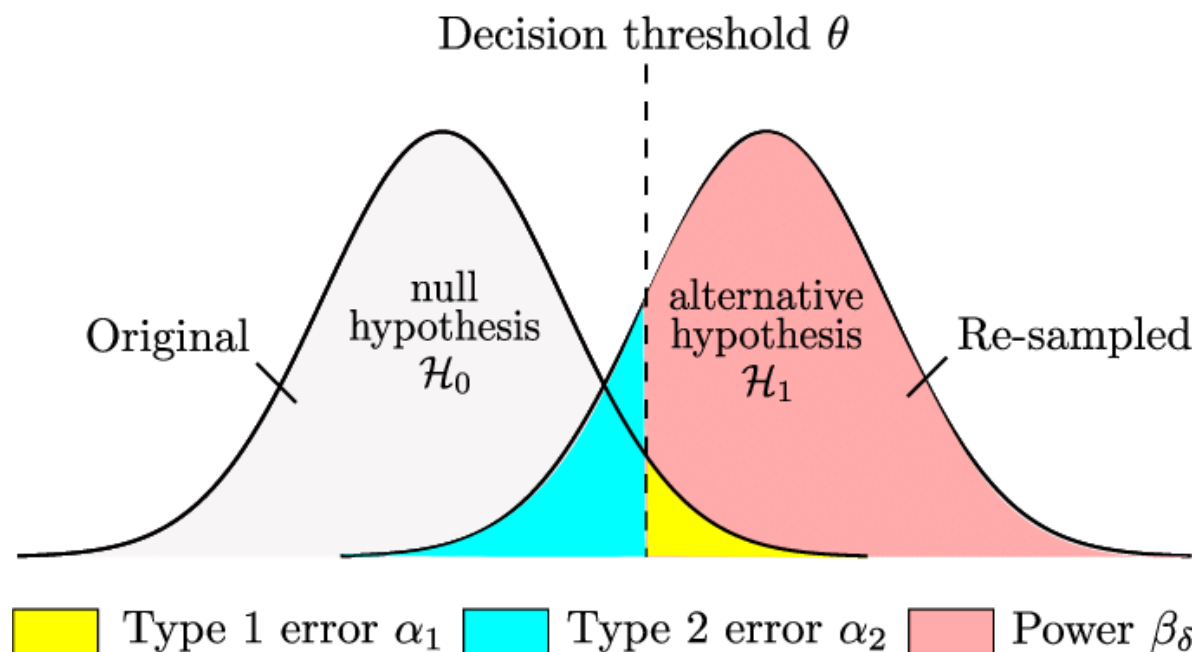
3. 假设检验中的错误

- I 类错误和 II 类错误 Type I and Type II errors

- 功效 power

- 功效很难量化，因为拒绝零假设的概率取决于事实与零假设的差异程度。

Decision	Actual Situation	
	H_0 True	H_0 False
Fail to Reject H_0	✓	Type II Error (β)
Reject H_0	Type I Error (α)	✓



3. 假设检验中的错误

- I 类错误和 II 类错误 Type I and Type II errors
 - I 类错误: 拒绝为真的零假设 (rejecting a true null hypothesis)
 - II 类错误: 没有拒绝为假的零假设 (failing to reject a false null)

		Actual Situation	
		H_0 True	H_0 False
Decision	Fail to Reject H_0	✓	Type II Error (β)
	Reject H_0	Type I Error (α)	✓



4. 假设检验中的单侧或双侧检验

- 单侧检验 a one-sided (or one-tailed) test
 - 备择假设中的参数值与零假设中的参数值的差异存在方向性。
 - 更大？更小？（而不是没有方向性的“不一样”）
 - 当数据偏离的方向符合 H_A 中描述的方向， H_0 被拒绝。
 - 单侧检验中的备择假设及其方向性必须在获得数据前设定。
 - 而不是依据数据结果。
 - 数据本身不能用来预测偏离的方向。

4. 假设检验中的单侧或双侧检验

- 例子：蟾蜍的手性偏好

- 单侧检验 a one-sided (or one-tailed) test

- H_0 : 频数相同 (the sample frequency)
- H_A : 右撇子更多 (more right-handed toads)

- $p_{\text{left}} < p_{\text{right}}$

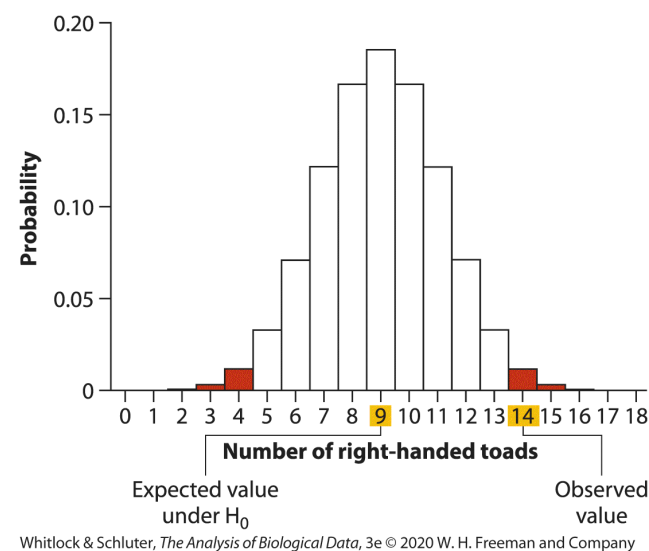
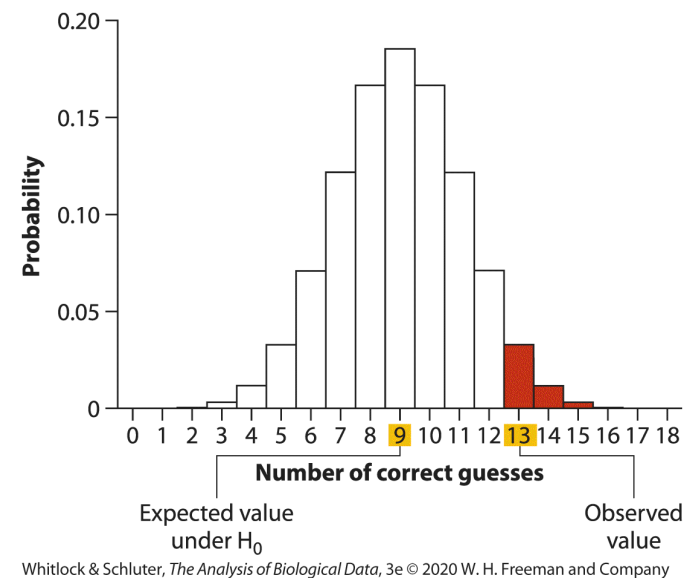
- 在获取数据前设定了一个方向

- 双侧检验 a two-sided (or two-tailed) test

- H_0 : 频数相同 (the sample frequency)
- H_A : 频数不同 (not equally frequent)

- $p_{\text{left}} > p_{\text{right}}$ or $p_{\text{left}} < p_{\text{right}}$

- 在获取数据前无法排除任何一种可能性/方向





5. 课堂练习 Practice Problems

- Q1: 设想你正在使用随机数据样本来检验一个零假设。请回答以下说法正确与否：与抽样误差小的估计值相比，抽样误差大的参数估计值会导致检验的I类错误率更高。
 - 正确
 - 错误



5. 课堂练习

- 假设一个统计检验的零假设为真。请说出下列说法正确与否：
 - Q2: 如果你通过检验拒绝了这个 H_0 , 你就犯了第二类错误。
 - 正确
 - 错误
 - Q3: 如果你未能通过检验拒绝这个 H_0 , 你就犯了 I 类错误。
 - 正确
 - 错误



5. 课堂练习

- Q4: 把显著性水平 α 的值从 0.05 降到 0.01, 对以下各项有怎样的影响?
 - a. 犯第一类错误的概率?
 - b. 犯第二类错误的概率?
 - c. 检验的功效?
 - d. 样本大小?

(降低? 增大? 无影响?)



5. 课堂练习

- Q5: 假设你进行了一项研究，以确定儿子是否像母亲。在 18 次独立试验中，你分别向一名参与者展示了一名男孩的照片和两名成年女性的照片，其中一名是男孩的母亲，另一名是随机选择的女性。你要求被试者猜测哪位女性是男孩的母亲。如果儿子和母亲不相似，那么被试猜对的概率是 $1/2$ 。如果儿子真的很像母亲，那么被试猜对的概率大于 $1/2$ 。回答下列问题，完成假设检验的四个步骤。
 - a. 说明适当的零假设和备择假设。
 - b. 这是单侧检验还是双侧检验？如何判断？
 - c. 你会用什么作为本研究的检验统计量？
 - d. 假设你发现 18 位参与者中有 7 位猜对了，11 位猜错了。计算这一结果的 P 值。零分布与蟾蜍手性偏好相同（每种结果的概率）。
 - e. 你的测试结论是什么？
 - f. 完成假设检验后，下一步应该怎么做来解释你的结论并确定总体参数 P 最合理的取值范围？



5. 课堂练习

- Q6: 随着大脑神经元活动的增加，流向大脑的血流量也会增加，以满足对氧气的需求。Sheth 等人（2004 年）测量了大鼠大脑体感皮层的血液动态，以确定血量是否随着神经元活动的增加而线性增加，或者这种关系是否可能是非线性的（即神经元活动量越大，流向大脑的血量增加越快）。如果两个变量之间是线性关系，则比率为 1；如果数值与 1 不同，则表示是非线性关系。
- 他们的实验结果显示，随着神经元活动的增加，脑部组织中血红蛋白总量的增加率为 1.17（意味血流量增大）。该比率的 95% 置信区间为 $0.74 \leq \text{比率} \leq 1.59$ 。并且，研究人员以变量间关系为线性关系的零假设（ H_0 : 比率=1）进行了假设检验。
- 他们的检验结果会拒绝零假设吗？请解释。

6. 小结 – 假设检验

- 假设检验是指将实际数据与特定零假设为真的期望结果进行比较的统计过程。如果数据与零假设成立时的预期数据相比过于异常，则拒绝零假设。
- 假设检验的步骤：
 - 1. 提出假设 (state the H_0 & H_A) 。
 - 2. 利用数据计算检验统计量 (compute the test statistic with data)。
 - 3. 确定 P 值 (determine the P-value)。
 - 4. 得出适当的结论 (draw the appropriate conclusions) 。
- 零假设 H_0 是关于总体参数值的一种具体说法（假设为真）。一般来说，零效假设是人们倾向于拒绝的一个陈述。
- 备择假设 H_A 包括零假设中所述值之外的所有其它可能的总体参数值。通常包括在生物学上比零假设中的值更有趣的可能性。

6. 小结 – 假设检验

- 检验统计量是根据事实数据计算的一个数字，用于评估该数据与零假设的预期结果的匹配程度。
 - 零分布是在假定零假设为真的前提下，检验统计量结果的抽样分布。
- P 值是在零假设 H_0 为真时获得观察数据（或更极端结果）的概率。
 - 如果 P 值很小，那么意味着数据与 H_0 的预期结果不一致，我们需要拒绝 H_0 ；
 - 否则，我们不拒绝 H_0 。
 - 所以，P 值表示反对零假设的证据的权重（P 越小表示证据越有力）
 - P 值不是 H_0 为真的概率。
- 在双侧检验中，备择假设 H_A 包括零假设 H_0 所指定参数值两侧的参数值。
 - 大多数假设检验都是双侧检验。
 - 在单侧检验中，备择假设只包括零假设所述参数值一侧的参数值（单侧检验限于参数值在另一侧情况极为罕见，e.g., 子女与父母的相似性）。

6. 小结 – 假设检验

- I 类错误是拒绝为真的零假设的错误，II 类错误是没有拒绝为假的零假设。
 - 显著性水平 α 设定了犯一类错误的概率。
 - 生物学研究中通常将 $\alpha = 0.05$ 视为可接受的错误率。
 - 功效是拒绝为假的零假设的概率；通常增大样本量可以提高功效。
- 不能拒绝零假设通常不能断定假设是真还是假。
 - 因此建议在给出假设检验结果时，应尽可能附上相关参数的置信区间。
 - 置信区间则衡量效应的大小。
- 阅读资料：
 - Whitlock & Schluter - Ch6
 - 生物统计学-李春喜-第四章第一节