

Lecture 11 – Experimental Design 试验设计

- Outline for today
 - Recall comparison of means 回顾均值的比较
 - Handling violations of assumptions 违背前提假设的处理
 - Nonparametric test 非参检验
 - Permutation test 置换检验
 - Experimental Design 试验设计
 - Summary
 - R Lab & Discussion

生物统计学

李 勤

生态与环境科学学院



1. 知识回顾

- 1.1 均值的比较: t 检验
 - 一个样本的均值 (与 H_0 中的值做比较)
 - 配对样本的均值;
 - 两个独立样本的均值;

1. 知识回顾

- 1.2 违背前提假设的处理
 - 数值的正态性（单样本/双样本t检验）
 - 忽视（基于检验的稳健性）
 - 数据转换（对数、反正弦、平方根转换）
 - 正态性检验（谨慎使用）
 - 两个样本方差相等（双样本t检验）
 - 近似的检验方法

2. 非参检验方法

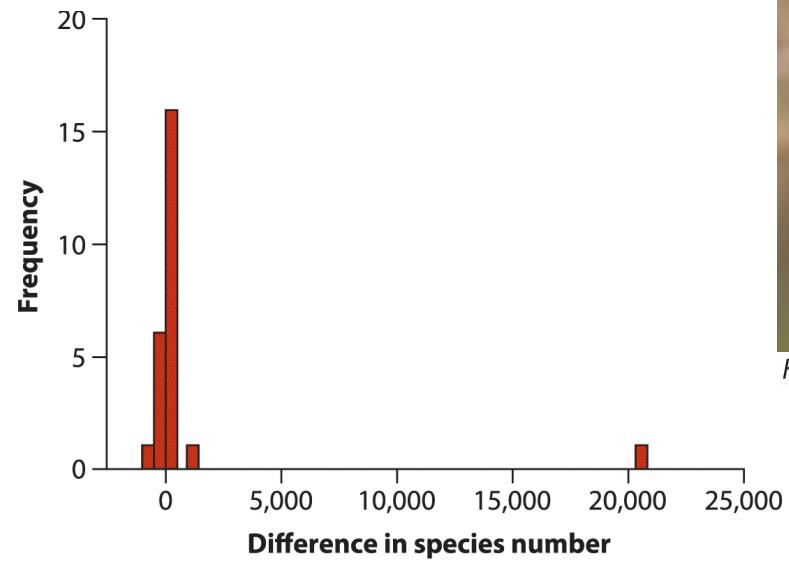
- 违背前提假设的处理
 - 1. 忽视违背
 - 2. 数据转换
 - 3. 非参方法
 - 符号检验
 - 符号秩检验
 - 秩和检验
 - 4. 置换检验

2.1 非参方法

- 3. 参数方法 vs. 非参方法
 - 参数方法 parametric test
 - 对抽样数据的概率分布进行了假设;
 - 到目前为止，我们学到的所有检验方法都是参数方法;
 - 非参方法 nonparametric test
 - 对数据的概率分布进行的假设较少;
 - 适用于数据的频数分布不是正态分布（例如存在异常值）的情况;
 - 替代单样本和配对t检验的非参方法
 - (1) 符号检验
 - (2) 符号秩检验
 - 替代双样本t检验的非参方法
 - (3) 秩和检验

2.2 非参方法 — 符号检验

- (1) 符号检验 Sign test
- EXAMPLE 13.4: Sexual conflict and the origin of new species
 - 交配多次 vs. 交配一次
 - 25个数据点: 昆虫类群配对组数
 - 数据: 把配对组间差异转化为 “+/-” (sign)



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

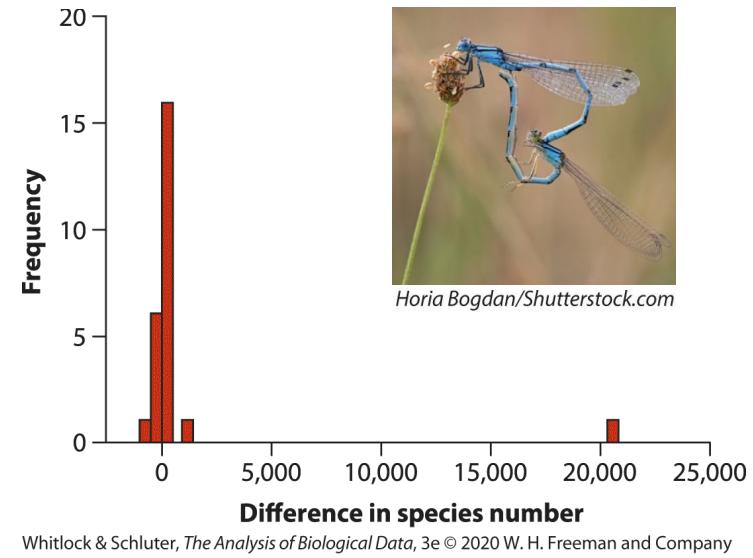


Horia Bogdan/Shutterstock.com

	taxonPair	nSpeciesMultipleMating	nSpeciesSingleMating	difference	sign
1	A	53	10	43	+
2	B	73	120	-47	-
3	C	228	74	154	+
4	D	353	289	64	+
5	E	157	30	127	+
6	F	300	4	296	+
7	G	34	18	16	+
8	H	3400	3500	-100	-
9	I	20	1000	-980	-
10	J	196	486	-290	-
11	K	1750	660	1090	+
12	L	55	63	-8	-
13	M	37	115	-78	-
14	N	100	30	70	+
15	O	21000	60	20940	+
16	P	37	40	-3	-
17	Q	7	5	2	+
18	R	15	7	8	+
19	S	18	6	12	+
20	T	240	13	227	+
21	U	15	14	1	+
22	V	77	16	61	+
23	W	15	14	1	+
24	X	85	6	79	+
25	Y	86	8	78	+

2.2 非参方法 — 符号检验

- (1) 符号检验 Sign test
- EXAMPLE 13.4: Sexual conflict and the origin of new species
 - 交配多次 vs. 交配一次 → 是否促进物种的演化?
 - 25个数据点: 昆虫类群配对组数
 - “+” 共计18组; “-” 共计7组;
 - 符号检验评估一个总体的中位数是否等于零假设中的值;
 - H_0 : 昆虫类群间物种数差异的中位数为零 (“+” 和 “-” 组数一致);
 - H_A : 昆虫类群间物种数差异的中位数不为零;

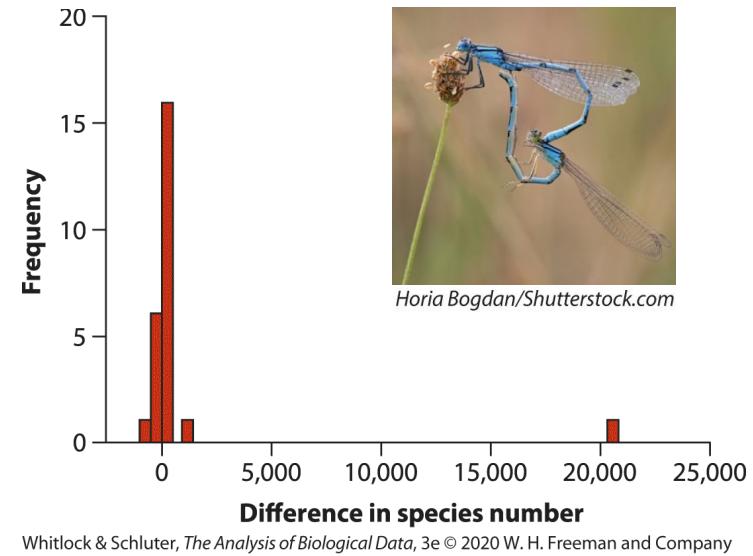


Horia Bogdan/Shutterstock.com

2.2 非参方法 — 符号检验

- (1) 符号检验 Sign test
- EXAMPLE 13.4: Sexual conflict and the origin of new species

- 交配多次 vs. 交配一次 → 是否促进物种的演化?
- 25个数据点: 昆虫类群配对组数
- “+” 共计18组; “-” 共计7组;
- 符号检验评估一个总体的中位数是否等于零假设中的值;
 - H_0 : 昆虫类群间物种数差异的中位数为零 (“+” 和 “-” 组数一致);
 - H_A : 昆虫类群间物种数差异的中位数不为零;
- 至此, 符号检验转化为二项检验 (the binomial test)
 - P 值 = $2 \Pr[X \leq 7] = 2 \sum_{i=0}^7 \binom{25}{i} 0.5^i 0.5^{25-i} = 2 \times 0.02164 = 0.043$
 - 结论: 因 P 值小于0.05, 拒绝 H_0 ;
 - 雌性多次交配的昆虫群体具有更多的物种, 与性冲突假说一致。



2.2 非参方法 — 符号检验



- (1) 符号检验 Sign test
- EXAMPLE 13.4: Sexual conflict and the origin of new species

- 符号检验评估一个总体的中位数是否等于零假设中的值;
 - H_0 : 昆虫类群间物种数差异的中位数为零（“+”和“-”组数一致）；
 - H_A : 昆虫类群间物种数差异的中位数不为零；
- 至此，符号检验转化为二项检验 (the binomial test)
 - R语言操作:

```
binom.test(7, n = 25, p = 0.5)
```

```
##  
## Exact binomial test  
##  
## data: 7 and 25  
## number of successes = 7, number of trials = 25, p-value = 0.04329  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.1207167 0.4938768  
## sample estimates:  
## probability of success  
## 0.28
```

2.2 非参方法 — 符号检验

- (1) 符号检验 Sign test
 - 基于 “ $+/-$ ” 的计数，符号检验转化为二项检验，来评估一个总体的中位数是否等于零假设中的值；
 - 数据：把配对组间差异转化为 “ $+/-$ ” (sign)，分别计数；
 - 位于零假设中位数上方的测量值被标记为 “+”，而位于下方的数被标记为 “-”；
 - 如果存在刚好等于零假设中的中位数值的数据，删掉并修正样本大小 n ；
 - 符号检验功效较低，需要较大的样本大小。
 - 当样本大小 $n \leq 5$ 时，则无论真实值与零假设值有多么不同，都无法从 $\alpha=0.05$ 的双侧符号检验中拒绝零假设；
 - 将数值数据转化为二元符号数据，丢失了大量信息（差异幅度）；但在其它统计方法不适用的情况下，依然是一个有用的统计工具。

2.3 非参方法 — 符号秩检验

- (2) 符号秩检验 The Wilcoxon signed-rank test
 - 是对符号检验的改进，用于评估总体的中位数是否等于零假设中常数；
 - 与符号检验不同，保留了关于幅度的信息；
 - 即每个数据点位于假设中位数上方或下方的距离有多远；
 - 但该检验假设在总体中测量值的分布围绕中位数对称，即不存在偏斜。
 - 这个假设几乎和单样本t检验的正态性假设一样具有限制性；
 - 由于这个限制，我们不详细解释其计算的细节；
 - 计算机上的大多数统计软件都可以执行该检验，但并未提示其局限性；

```
> library(MASS)          # load the MASS package
> head(immer)
  Loc Var     Y1     Y2
1  UF   M  81.0  80.7
2  UF   S 105.4  82.3
....
```

```
> wilcox.test(immer$Y1, immer$Y2, paired=TRUE)

Wilcoxon signed rank test with continuity correction

data: immer$Y1 and immer$Y2
V = 368.5, p-value = 0.005318
alternative hypothesis: true location shift is not equal to 0
```

2.4 非参方法 — 秩和检验

- (3) 秩和检验
 - the Wilcoxon rank-sum test
 - the Mann–Whitney U -test (曼–惠特尼 U 检验)
- 基于数据点从小到大的顺序排名（第 x 位），而非实际测量值；
 - 即可避免对测量值的概率分布做出假设；
- 检验两组数据的频率分布是否相同；
 - 如果两个分布的形状相同，那么比较其集中位置（中位数或均值）；
 - 从而替代双样本 t 检验（不需要那么多假设）；

2.4 非参方法 — 秩和检验

- (3) 秩和检验（曼-惠特尼U检验）
 - 基于数据排名来检验两组数据的频率分布是否相同（中位数或均值）；
 - 例子： [EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets](#) 蟋蟀的性食性
 - “雌性在饥饿状态下更容易交配吗？”
 - 样本： 11只饥饿的蟋蟀 vs. 13只正常喂食的蟋蟀
 - 数据： 与雄性蟋蟀交配的等待时间

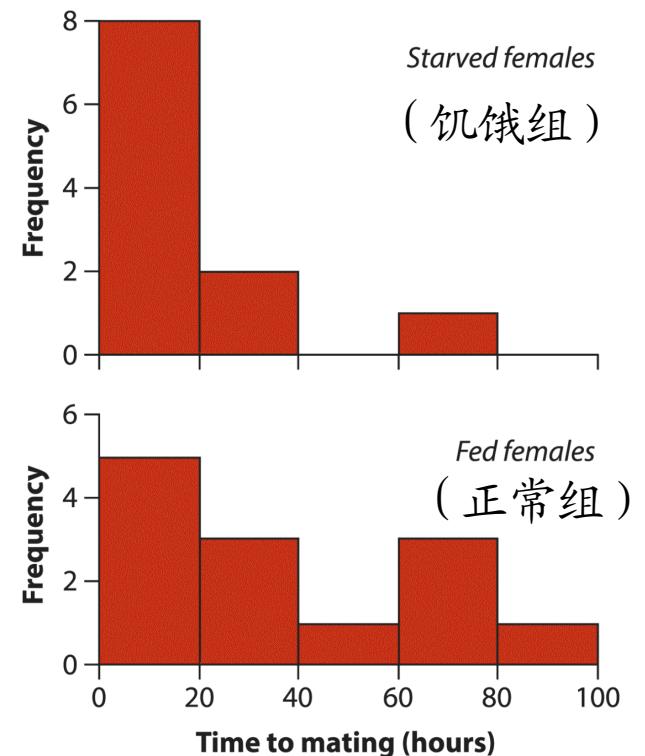


© David H. Funk

feedingStatus	timeToMating	feedingStatus	timeToMating
starved	1.9	fed	1.5
starved	2.1	fed	1.7
starved	3.8	fed	2.4
starved	9	fed	3.6
starved	9.6	fed	5.7
starved	13	fed	22.6
starved	14.7	fed	22.8
starved	17.9	fed	39
starved	21.7	fed	54.4
starved	29	fed	72.1
starved	72.3	fed	73.6
		fed	79.5
		fed	88.9

2.4 非参方法 — 秩和检验

- (3) 秩和检验（曼-惠特尼U检验）
 - 基于数据排名来检验两组数据的频率分布是否相同（中位数或均值）；
 - 例子： [EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets](#) 蟋蟀的性食性
 - “雌性在饥饿状态下更容易交配吗？”
 - 样本： 11只饥饿的蟋蟀 vs. 13只正常喂食的蟋蟀
 - 数据： 与雄性蟋蟀交配的等待时间
 - 数据呈偏态； 样本量较小；
 - 不能依赖中心极限定理假设均值抽样分布呈正态；
 - 假设检验
 - H_0 : 饥饿和被喂养的雌性蟋蟀的交配时间的分布相同；
 - H_A : 这两组之间的交配时间的分布不同。



2.4 非参方法 — 秩和检验

- (3) 秩和检验：蟋蟀的性食性

- H_0 : 饥饿和被喂养的雌性蟋蟀的交配时间的分布相同；
- H_A : 这两组之间的交配时间的分布不同。
- Mann-Whitney U 检验: 计算秩和及 U 统计量
 - 饥饿组/group 1:
 - $R_1 = 3+4+7+9+10+11+12+13+14+17+21 = 121$
 - $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 11 \times 13 + 11 \times 12 / 2 - 121 = 88$
 - $U_2 = n_1 n_2 - U_1 = 11 \times 13 - 88 = 55$
 - (通过 R_2 也会得出同样结果)
 - U 统计量: U_1 和 U_2 中较大的一个 $\rightarrow U_1$
 - 与关键值/临界值做比较: $U_{0.05(2),11,13} = 106$
 - P值 > 0.05 ; 因此不拒绝零假设; 即没有足够的证据表明饥饿和被喂养的雌性蟋蟀的交配时间分布不同。

	feedingStatus	group	timeToMating	rank
1	fed	2	1.5	1
2	fed	2	1.7	2
3	starved	1	1.9	3
4	starved	1	2.1	4
5	fed	2	2.4	5
6	fed	2	3.6	6
7	starved	1	3.8	7
8	fed	2	5.7	8
9	starved	1	9.0	9
10	starved	1	9.6	10
11	starved	1	13.0	11
12	starved	1	14.7	12
13	starved	1	17.9	13
14	starved	1	21.7	14
15	fed	2	22.6	15
16	fed	2	22.8	16
17	starved	1	29.0	17
18	fed	2	39.0	18
19	fed	2	54.4	19
20	fed	2	72.1	20
21	starved	1	72.3	21
22	fed	2	73.6	22
23	fed	2	79.5	23
24	fed	2	88.9	24

2.4 非参方法 — 秩和检验

- (3) 秩和检验：蟋蟀的性食性

- H_0 : 饥饿和被喂养的雌性蟋蟀的交配时间的分布相同；
- H_A : 这两组之间的交配时间的分布不同。
- 计算秩和及 U 统计量，再与关键值做比较得出结论。

- R 语言操作：

Wilcoxon rank-sum test

This test, available in R, is equivalent to the Mann-Whitney U -test.

```
wilcox.test(timeToMating ~ feedingStatus, data = cannibalism)
```

```
##  
## Wilcoxon rank sum test  
##  
## data: timeToMating by feedingStatus  
## W = 55, p-value = 0.3607  
## alternative hypothesis: true location shift is not equal to 0
```

2.4 非参方法 — 秩和检验

- (3) 秩和检验的其它注意事项
 - 当多个相同数值出现时的排序:
 - 取顺序的平均数
 - 如: 两个相同的数值, 占据4-5位, 那么更正后的“秩”为: $(4 + 5)/2 = 4.5$
 - 大样本和正态近似
 - 对于中等和较大的样本大小, 曼-惠特尼U检验中的U统计量进行转化后, 其抽样分布在零假设成立时可近似为标准正态分布:
 - $Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/3}}$

2.5 非参方法的一些前提假设和应用限制

- 非参方法的前提假设：随机样本
 - 虽然非参数检验不依赖于正态分布，但样本均为所代表总体的随机样本。
- 一些限制：
 - Wilcoxon符号秩检验假定总体中测量值的分布是对称的；
 - Mann-Whitney U 检验的 H_0 中认为两组数据的分布是相同的；
 - 因此，拒绝 H_0 意味着这些分布不相同；
 - 但拒绝 H_0 并不一定意味着这两个分布具有不同的位置（均值或中位数）；
 - 如果要得出位置不同的结论，需要额外假设这两组的分布具有相同的形状；但 U 检验对不等方差或不同偏斜引起的分布之间差异非常敏感（即该方法不够稳健）。

2.5 非参方法 — 小结

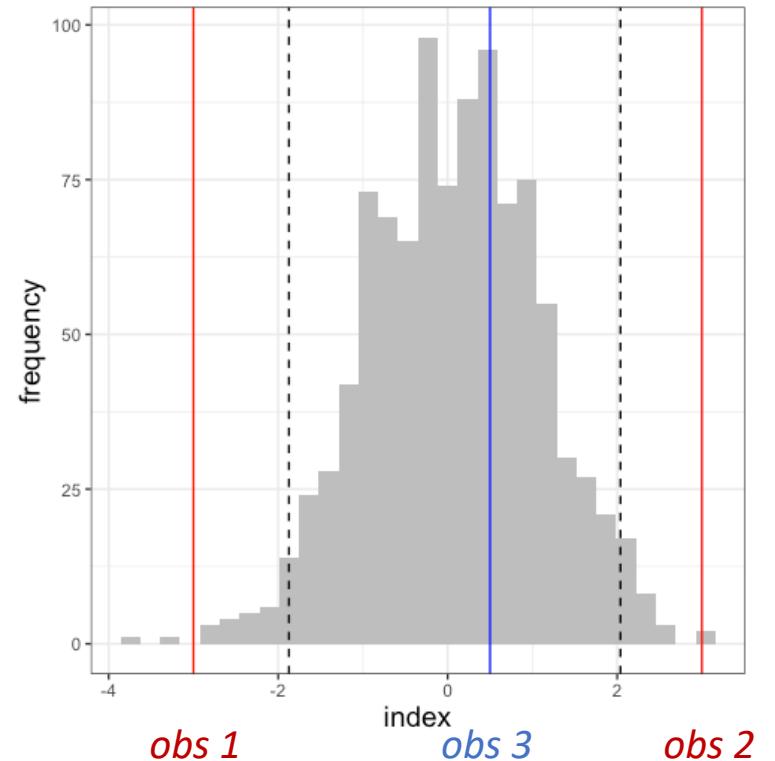
- 关于I类错误：
 - 满足特定检验的前提假设时，无论是参数检验还是非参数检验，出现I类错误（即拒绝true H_0 ）的概率都为 α ；
 - 当不符合参数检验的前提假设时，例如测量值的分布严重偏态，那么I类错误率就会大于设定的 α 值；
 - 因此，在前提假设被严重违反时不使用参数检验的主要原因就是过高的I类错误率；
 - 非参数检验只要满足限制性较小的前提假设，其I类错误率等于 α ；

2.5 非参方法 — 小结

- 关于II类错误：
 - 非参检验通常比参数检验的功效（power）要低；
 - 多种非参方法使用的是排序数据，因此造成的信息丢失会导致非参数检验的功效小于相应的参数检验；
 - 功效较小意味着非参数检验拒绝错误的零假设的概率较低，因此II类错误率高于参数检验；
 - 大多数生物学家倾向于在符合前提假设的情况下使用参数检验，只有在数据转换不符合参数检验假设的情况下才使用非参数检验方法；
 - 当参数检验的假设条件遭到严重破坏时，非参数检验的功效下降就不重要了；因为在这种情况下根本无法使用参数检验。

3. 置换检验 Permutation Test

- 违背前提假设的处理
 - 1. 忽视违背
 - 2. 数据转换
 - 3. 非参方法
 - 4. 置换检验
 - a computer-intensive method
 - 替代双样本 t 检验和 Mann-Whitney U 检验（其它方法参考Ch19）



3.1 置换检验

- “置换”是指重新排列（re-arrangement）
- 也称之为随机化检验（randomization test）
- 方法：
 - 通过反复随机地重新排列数据中两个变量中一个变量的值，多次计算检验统计量，生成两个变量之间关联的零分布；
 - 如果根据原始数据计算出的检验统计量的观测值，用来与零分布相比；如果有异常，我们就拒绝变量之间没有关联的零假设。

3.1 置换检验

- 如果进行置换操作？（随机化过程）

- 例子： EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets 蟋蟀的性食性

- 样本：11只饥饿的蟋蟀 vs. 13只正常喂食的蟋蟀

- 数据：与雄性蟋蟀交配的等待时间

- 2个变量：

- 蟋蟀的进食状态（类型变量）

- 交配等待时间（数值变量）

- 问题：

- “雌性在饥饿状态下更容易交配吗？”

- 进食状态与交配等待时间是否有相关性？

- 不同进食状态下交配等待时间是否一样？

feedingStatus	timeToMating	feedingStatus	timeToMating
starved	1.9	fed	1.5
starved	2.1	fed	1.7
starved	3.8	fed	2.4
starved	9	fed	3.6
starved	9.6	fed	5.7
starved	13	fed	22.6
starved	14.7	fed	22.8
starved	17.9	fed	39
starved	21.7	fed	54.4
starved	29	fed	72.1
starved	72.3	fed	73.6
		fed	79.5
		fed	88.9

mean: 17.73

mean: 35.98
23

3.1 置换检验

- 例子: EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets 蟋蟀的性食性
- 问题: 不同进食状态下交配等待时间是否一样?

- H_0 : 两组雌蟋蟀的等待交配平均时间相同;
- H_A : 两组雌蟋蟀的等待交配平均时间不同;
- 检验统计量: $\bar{Y}_1 - \bar{Y}_2 = 17.73 - 35.98 = -18.26$
- 步骤:
 - (1) 将两组所有数据 (24个) 合并成一个数据池;
从中随机选择一个值并分配给处理1中的第1个个体;
重复以上步骤, 保持两个处理组的样本量大小, 直到
用完所有数据;
 - (2) 计算新构成的数据集的检验统计量;
 - (3) 重复以上步骤 (至少1000次), 可计算每次随
机分配数据后的统计检验量 (1000个 $\bar{Y}_1 - \bar{Y}_2$);

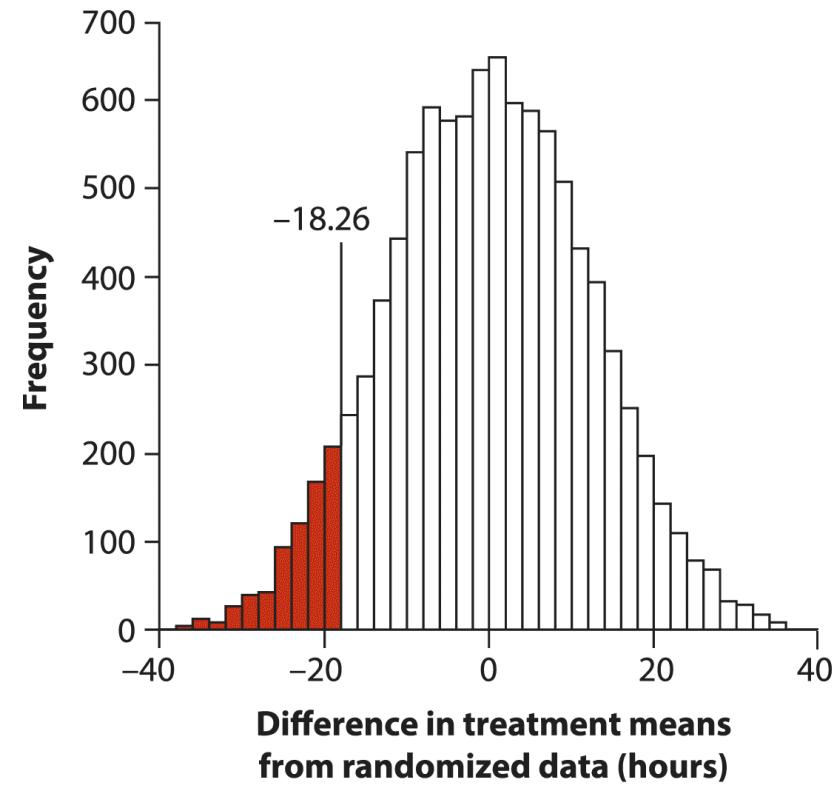
		One Permutation	
	feedingStatus	permSample	
	feedingStatus	permSample	
1	starved	1.7	12 fed
2	starved	17.9	13 fed
3	starved	22.6	14 fed
4	starved	73.6	15 fed
5	starved	72.1	16 fed
6	starved	14.7	17 fed
7	starved	13	18 fed
8	starved	5.7	19 fed
9	starved	88.9	20 fed
10	starved	21.7	21 fed
11	starved	29	22 fed
			23 fed
			24 fed

$$\bar{Y}_1 = 32.81$$

$$\bar{Y}_2 = 23.22$$

3.1 置换检验

- 例子: EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets 蟋蟀的性食性
- 步骤:
 - (1) 随机分配数据到两个处理组;
 - (2) 计算新构成的数据集的检验统计量;
 - (3) 重复以上步骤计算出10000个统计检验量; 由此可以做出统计检验量的零分布 (null distribution);
 - (4) 确定P值, 即把原始数据的检验统计量观测值 (observed value) 和其零分布做比较:
 - 观察值所在的位置, 及与它一致或更极端的概率;
 - 10000个数据中有712个这样的检验统计量;
 - $P = 2 \times \Pr[-18.26] = 2 \times (712/10000) = 0.142$
 - (5) 结论: 不拒绝 H_0 .
 - 即不同进食状态的蟋蟀的交配等待时间无显著差异。



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

3.1 置换检验

- 例子: EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets 蟋蟀的性食性
- 步骤:
 - (1) 随机分配数据到两个处理组;
 - (2) 计算新构成的数据集的检验统计量;
 - (3) 重复以上步骤计算出1000个统计检验量; 由此可以做出统计检验量的零分布 (null distribution);
 - (4) 确定P值, 即把原始数据的检验统计量观测值 (observed value) 和其零分布做比较:
 - 观察值所在的位置, 及与它一致或更极端的概率;
 - 10000个数据中有712个这样的检验统计量;
 - $P = 2 \times \Pr[-18.26] = 2 \times (712/10000) = 0.142$
 - (5) 结论: 不拒绝 H_0 .

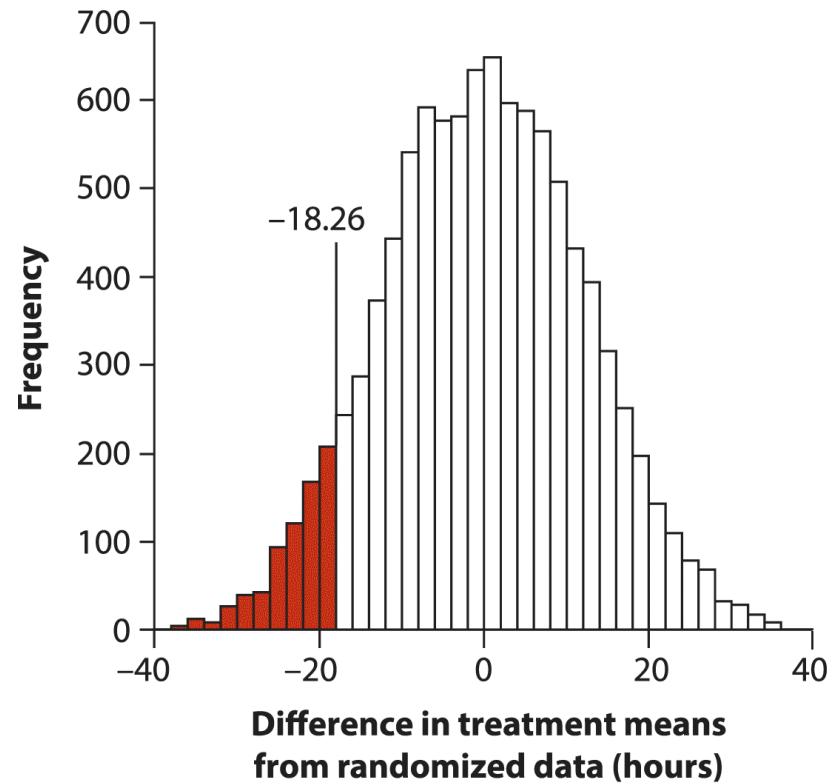
进行随机化操作的关键

```
permResult <- vector() # initializes
for(i in 1:nPerm){
  if(i%>100) print()
  # step 1: permute the times to mating
  cannibalism$permSample <- sample(
    cannibalism$timeToMating, replace = FALSE)
  # step 2: calculate difference between means
  permSampleMean <- as.data.frame(summarize(
    group_by(cannibalism, feedingStatus),
    permMean = mean(permSample, na.rm = TRUE)))
  permResult[i] <- permSampleMean$permMean[1] -
    permSampleMean$permMean[2]
}
```



3.1 置换检验

- 例子: EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets 蟋蟀的性食性
- 步骤:
 - (1) 随机分配数据到两个处理组;
 - (2) 计算新构成的数据集的检验统计量;
 - (3) 重复以上步骤计算出1000个统计检验量; 由此可以做出统计检验量的零分布 (null distribution);
 - (4) 确定P值, 即把原始数据的检验统计量观测值 (observed value) 和其零分布做比较:
 - 观察值所在的位置, 及与它一致或更极端的概率;
 - 10000个数据中有712个这样的检验统计量;
 - $P = 2 \times \Pr[- < -18.26] = 2 \times (712/10000) = 0.142$
 - (5) 结论: 不拒绝 H_0 。
 - 如果重复以上步骤, 会得出相同的P值吗?



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

3.2 置换检验的前提假设

- 置换检验的假设很少，可以应用于多种情况。
 - 随机样本；
 - 变量的分布在不同总体间具有相同的形状；
 - 当样本量较大时，置换检验对违反这一前提假设具有很强的稳健性，比秩和检验（曼-惠特尼 U 检验）更强。
 - 当样本量较小时，置换检验的功效（即拒绝假零假设的能力）比参数检验低，但比秩和检验更强。
 - 当样本量较大时，上述两种方法与参数检验的功效相似。

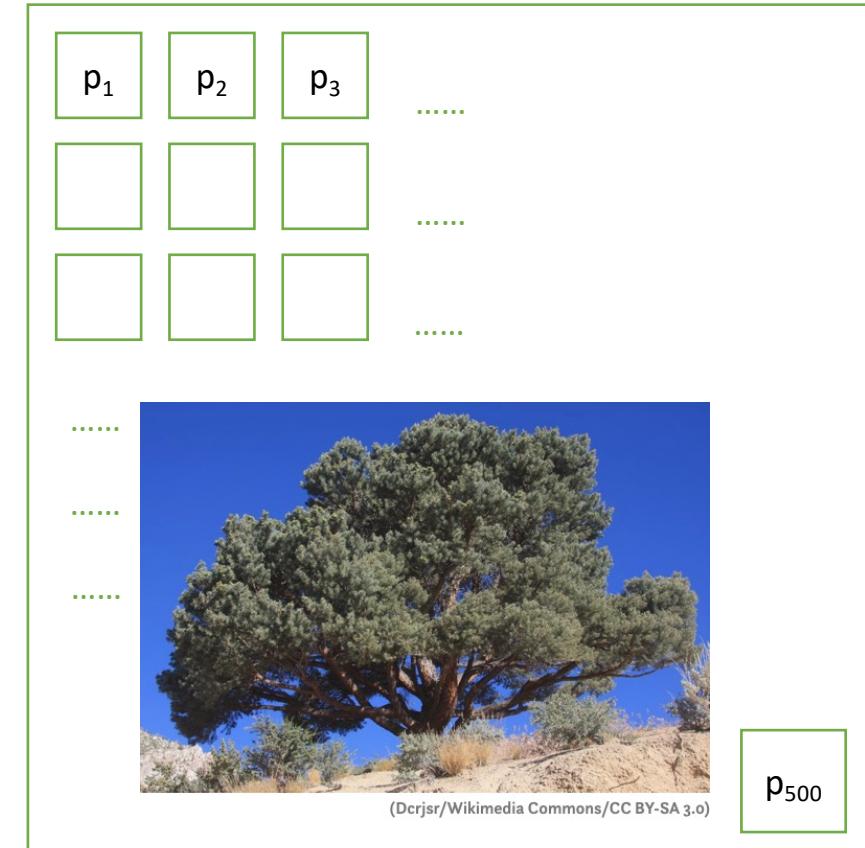
4. 两组数据比较时的统计学注意事项

- 正确的抽样单元
- 间接比较的谬误
- 正确解释置信区间的重叠



4.1 正确的抽样单元 sampling unit

- 抽样单元间的独立性 (the independent units of replication)
 - 正确判断重复的抽样单元才能保证结论的准确性;
- 例子
 - 1. 某种松树的年龄: 500个样方
 - 年龄的测量: 树的年轮;
 - 每个样方内的松树之间不独立;
 - 样方间的数据才相互独立;
 - Q: 抽样单元是什么?



4.1 正确的抽样单元

- 抽样单元间的独立性 (the independent units of replication)
 - 正确判断重复的抽样单元才能保证结论的准确性；

- 例子

- 2. 生物入侵的影响
 - 12条溪流，有或无溪红点鲑（入侵种）；
 - 检验：溪红点鲑是否影响奇努克鲑的生存？



Encyclopaedia Britannica/Universal Images Group/
Getty Images

brook trout
溪红点鲑



Chinook salmon
奇努克鲑

troutTreatment	nReleased	nSurvivors	proportionSurvived	存活比例
present	820	166	0.202	
absent	467	180	0.385	
present	960	136	0.142	
present	700	153	0.219	
absent	959	178	0.186	
present	545	103	0.189	
absent	1029	326	0.317	
present	769	173	0.225	
absent	27	7	0.259	
absent	998	120	0.12	
absent	936	135	0.144	
present	1001	188	0.188	

4.1 正确的抽样单元

- 抽样单元间的独立性 (the independent units of replication)
 - 正确判断重复的抽样单元才能保证结论的准确性；

- 例子

- 2. 生物入侵的影响
 - 12条溪流，有或无溪红点鲑（入侵种）；
 - 检验：溪红点鲑是否影响奇努克鲑的生存？
 - The χ^2 contingency test — $R \times C$ 列联表？

	Trout absent	Trout present
Survived	946	919
Not survived	3470	3876

troutTreatment	nReleased	nSurvivors	proportionSurvived
present	820	166	0.202
absent	467	180	0.385
present	960	136	0.142
present	700	153	0.219
absent	959	178	0.186
present	545	103	0.189
absent	1029	326	0.317
present	769	173	0.225
absent	27	7	0.259
absent	998	120	0.12
absent	936	135	0.144
present	1001	188	0.188

4.1 正确的抽样单元

- 抽样单元间的独立性 (the independent units of replication)

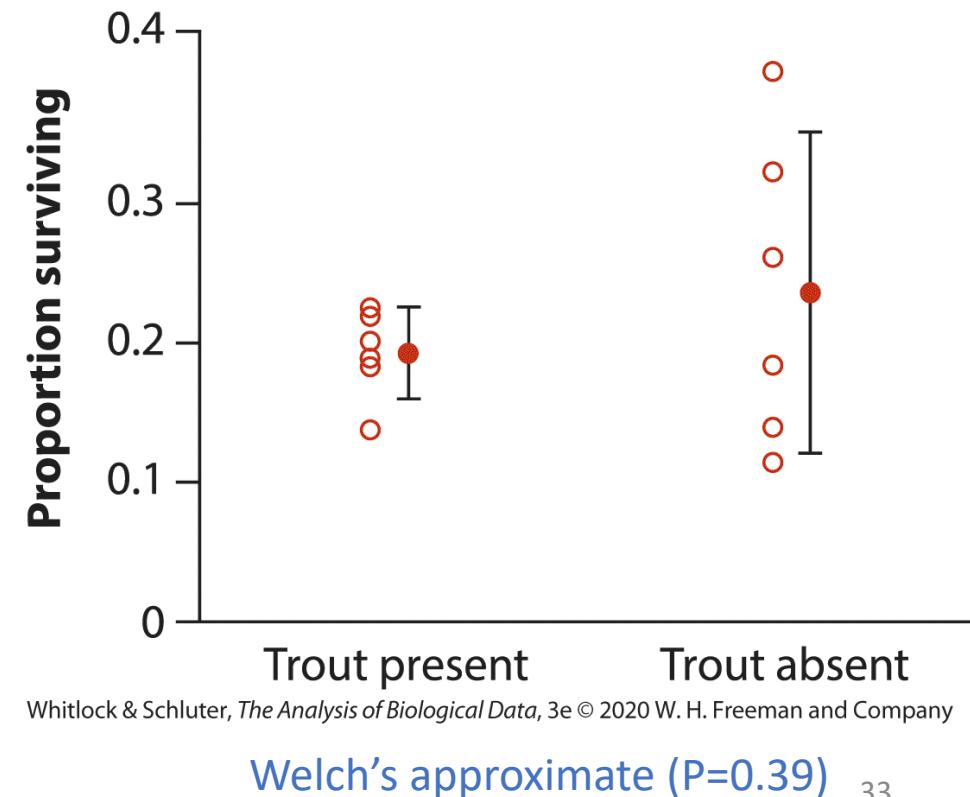
- 正确判断重复的抽样单元才能保证结论的准确性；

- 例子

- 2. 生物入侵的影响

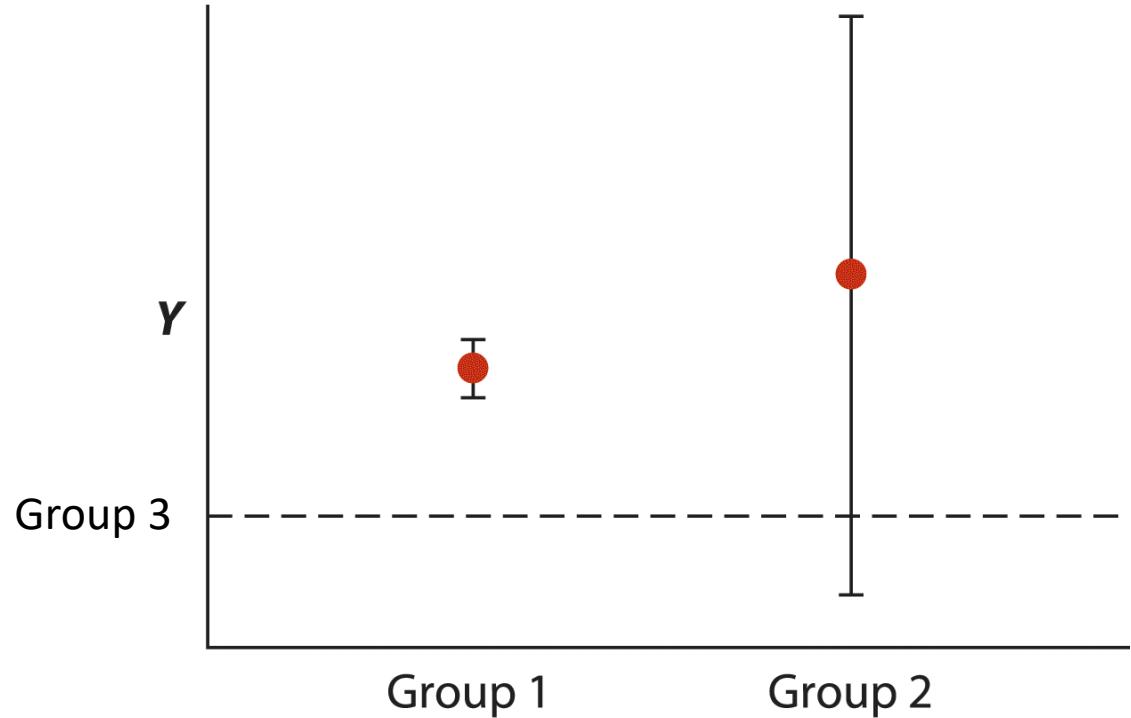
- 12条溪流，有或无溪红点鲑（入侵种）；
 - 检验：溪红点鲑是否影响奇努克鲑的生存？
 - 同一条溪流中的奇努克鲑情况更相似；

	Sample mean	Standard deviation	Sample size
Trout present	0.194	0.0297	6
Trout absent	0.235	0.1036	6



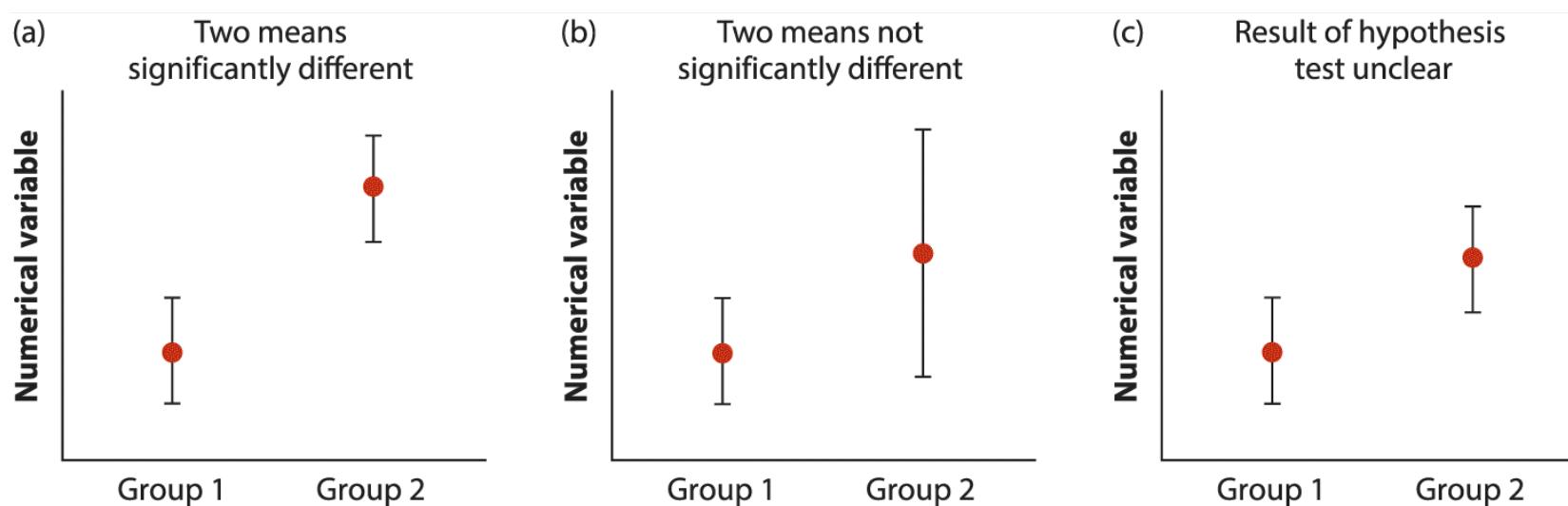
4.2 间接比较的谬误

- 直接的比较: Group 1 vs. Group 2
- 错误的是间接比较:
 - (the fallacy of indirect comparison)
 - Group 1 vs. Group 3: 有显著不同
 - Group 2 vs. Group 3: 无显著不同
 - → Group 1 vs. Group 2: 有显著不同



4.3 置信区间重叠的正确解释

- 置信区间的重叠情况 Group 1 vs. Group 2 (肉眼判断)
 - (a) CI不重叠，均值都没有被包括在另一组的CI中：显著不同
 - (b) CI重叠，且其中一组的均值被包括在另一组的CI中：无显著不同
 - (c) CI重叠，但两组均值没有被包括在另一组的CI中：不确定
 - 需要具体的两样本t-test来确定统计检验结果。



5. 试验设计 Experimental Designs

- 试验研究 (Experimental study) 的目的: 因果推断
 - *vs.* 观察研究 (Observational study) —— 相关性推断
- 试验设计的黄金原则: 临床试验 (clinical trial)
 - 针对人体进行的两个或两个以上处理的试验;
- 非人体试验 (place)
 - 实验室试验 laboratory experiments
 - 野外试验 field experiments

5.1 试验设计的原则

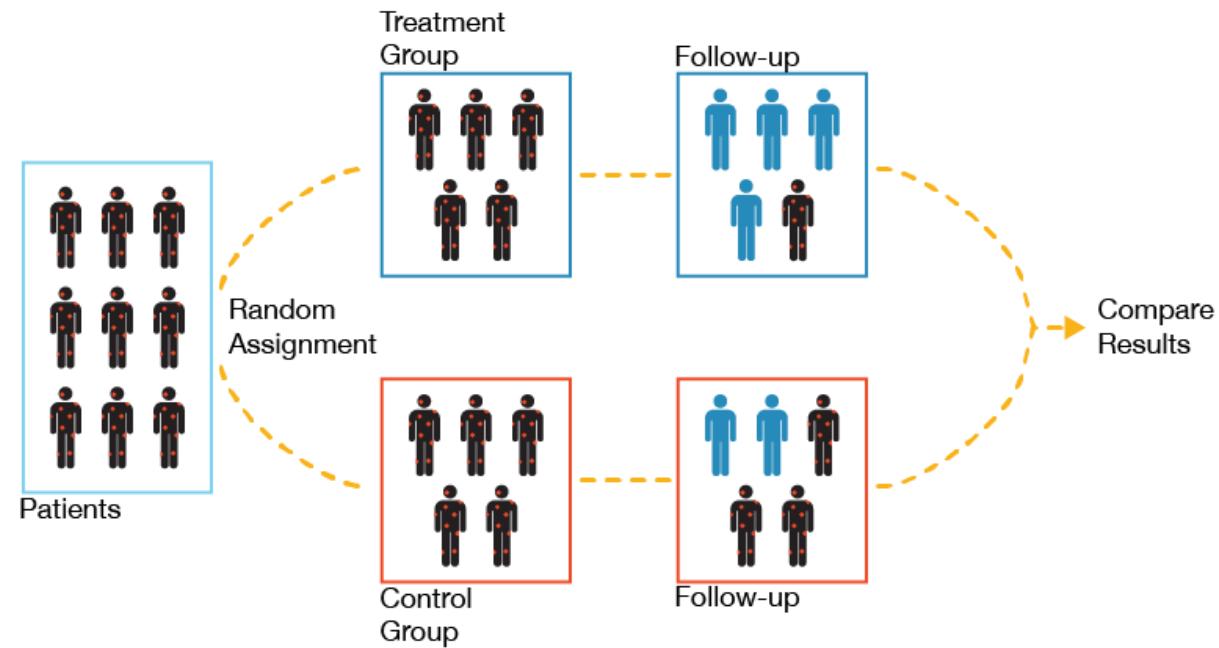
- 两项重要原则
 - 减少偏差: to reduce **bias** in estimating and testing treatment effects
 - 减少抽样误差: to reduce the effects of **sampling error**
- 试验单元 (experimental unit)
 - 等同于抽样单元;
 - 相互独立的一个或一组个体（接受相同或不同的处理）;
- 一个好的试验
 - A good experiment randomly assigns treatments to multiple, independent experimental units, and there will be multiple, independent experimental units per treatment.

5.2 试验设计的原则1 — 减少偏差

- 目的：测量值能准确反应试验中（感兴趣的）处理的效应；
 - 排除confounding variables对判断试验处理效应影响；
 - E.g., 巧克力和诺贝尔奖
- 1. 控制组 (a simultaneous control group)
- 2. 随机化 (randomization)
- 3. 单盲或双盲 (blinding)

5.2 试验设计的原则1 — 减少偏差

- 1. 控制组 (a simultaneous control group)
 - 同时设置感兴趣的处理组和对照组;
 - 原因: 安慰剂效应、测试效应、时间差异等;
 - 做法:
 - 临床试验:
 - 安慰剂、生理盐水等;
 - 实验室试验:
 - 类似的处理方式;
 - 野外试验:
 - 理论上需要类似的处理方式;



5.2 试验设计的原则1 — 减少偏差

• 2. 随机化 (randomization)

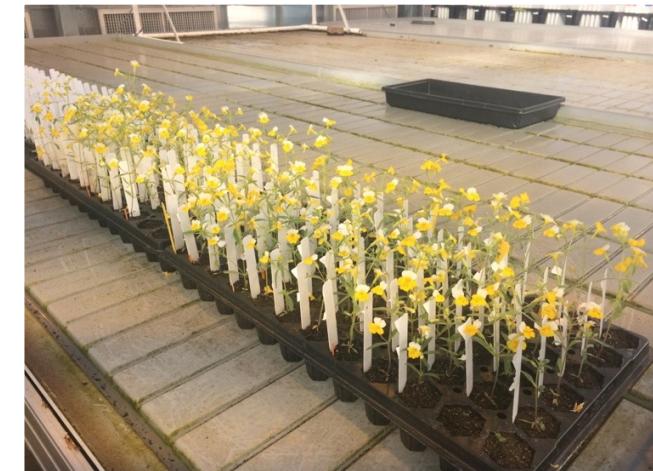
- 不同处理应随机地对样本个体进行实施；
- 目的：并非完全去除confounding factor的影响，而是去除这类因子与处理之间的相关性；
 - 让这种相关性更加均匀地分散在不同处理组中，从而避免偏差；

• 方法

- 采用随机数生成器来将不同处理分配给不同的试验单元；
- a computer random-number generator

Experimental unit	1	2	3	4	5	6	7	8
Random number	11	18	87	55	76	70	90	4
Treatment	A	A	B	A	B	B	B	A

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



© Qin Li @ Greenhouse

5.2 试验设计的原则1 — 减少偏差

• 3. 单盲或双盲 (blinding)

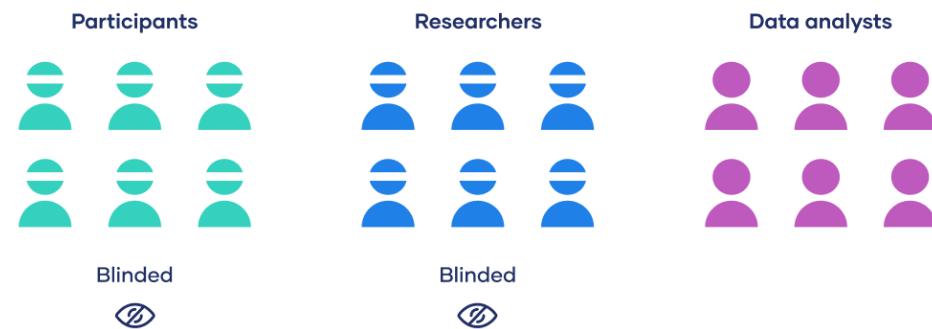
- 目的：去除主观性可能导致的偏差；
 - 如主观判断、过多关注等造成的夸张效应
 - 相比双盲实验的效应：27% ↑
- (1) 单盲试验
 - 试验单元（样本个体）不应知晓何种处理；
- (2) 双盲试验
 - 试验单元及试验操作者不应知晓何种处理；
 - E.g., 编码标签；



Single-blind study



Double-blind study



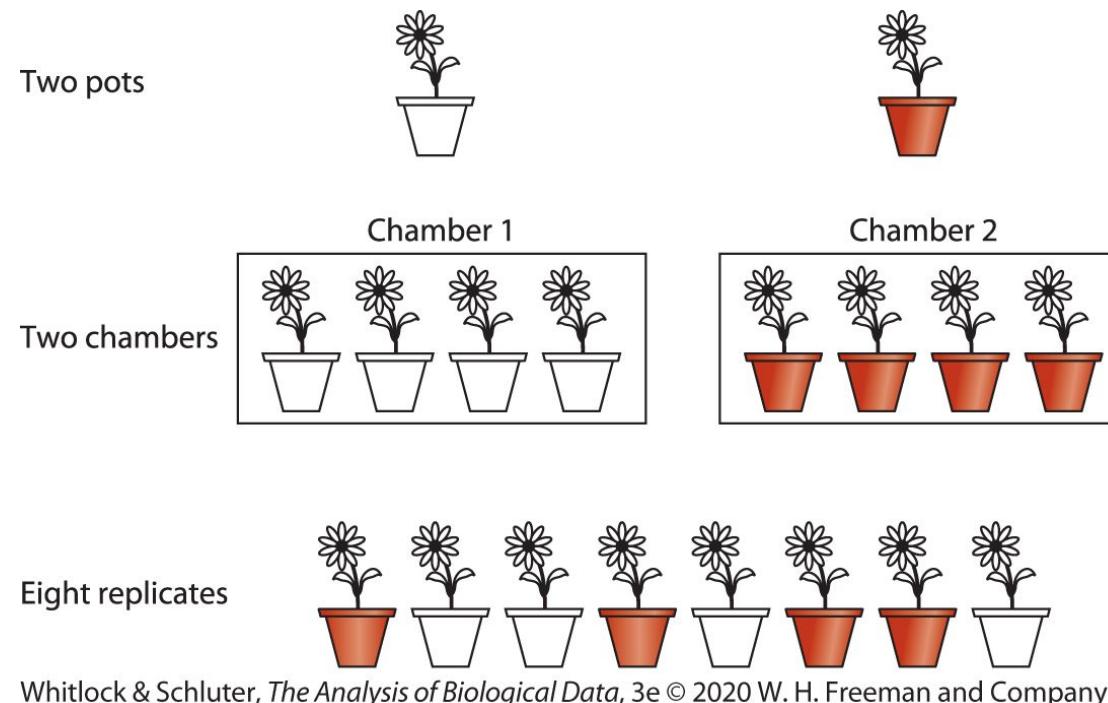
5.2 试验设计的原则2 — 减少抽样误差

- 目的：维系稳定/不变的试验条件；
 - 环境的一致性：温湿度等
 - 试验单元的一致性：同一性别、年龄、基因型等；
 - 让结论具有普适性；
- 1. 重复单元 (replication)
- 2. 权衡 (balance)
- 3. 设置组别 (blocking)

5.2 试验设计的原则2 — 减少抽样误差

- 1. 重复单元 (replication)

- 多个重复试验单元构成所需样本；
- 排除单元个体间差异对处理效应的影响；
- 原则上，大样本具有更小的标准误以及可获得假设检验的准确结论；



5.2 试验设计的原则2 — 减少抽样误差

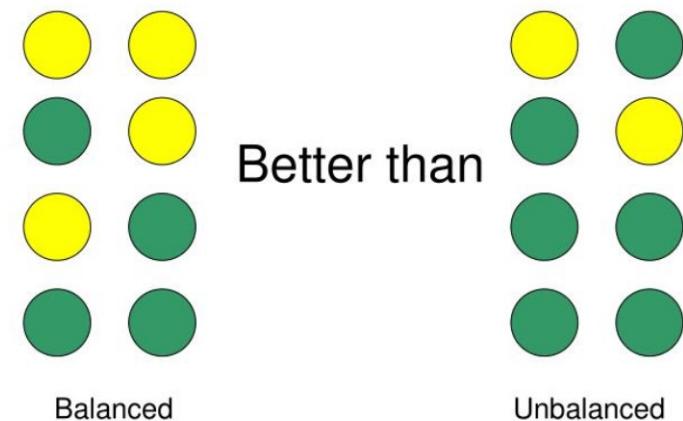
- 2. 权衡 (balance)

- 不同处理组具有相同的样本量;
 - balanced = the same sample size in each treatment
 - 当 $n_1 = n_2$ 时, SE 最小;

- 均值差异的标准误: $SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$

- 合并方差 (pooled sample variance): $s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$

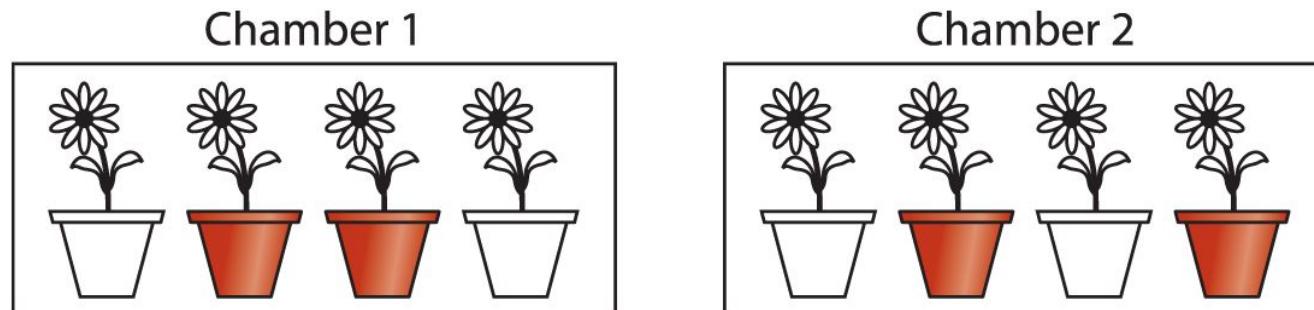
- 尤其在样本总量固定的情况下;



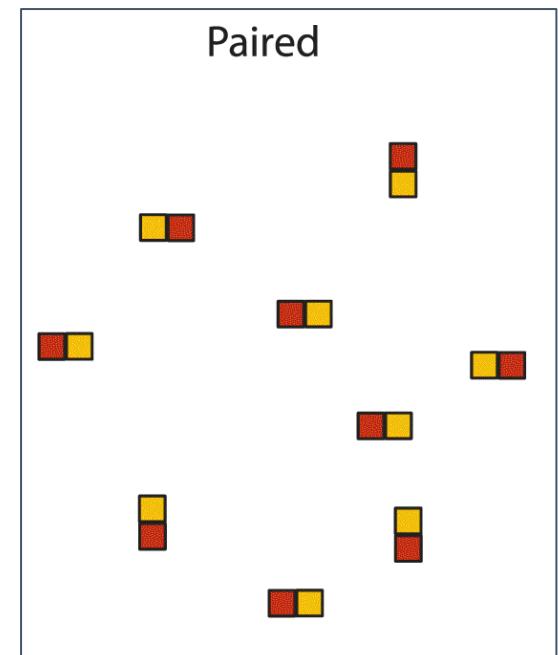
5.2 试验设计的原则2 — 减少抽样误差

- 3. 设置组别 (blocking)

- 排除其它存在变异的因素的影响;
 - E.g., 空间位置、就诊医院、同窝后代、同天的实验等;
- 组别内相对均质，且设置了各处理的试验单元;
- 处理间的差异也时在组别内进行评估;
 - 配对样本即组别设置;



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



5.3 样本大小的评估

- Next Lecture

6. 小结

- 当不满足一些检验方法的前提假设时，可以采取以下方法：
 - 适当的忽视、数据转换、非参方法、置换检验；
- 参数方法和非参方法
 - 对变量分布做出假设的被称为参数方法；不对变量分布做出假设的方法被称为非参数方法；
 - 非参数检验通常比参数检验功效低；
- 置换检验是一种零模型：
 - 检验统计量的观测值与计算机随机化数据生成的零分布进行比较；

6. 小结

- 试验设计的原则是减少偏差和抽样误差；
- 减少偏差的方法包括：
 - 1. 进行试验组的同时设置控制组；
 - 2. 将处理随机地分配给试验单元；
 - 3. 进行单盲或双盲试验来避免主观性效应；
- 减少抽样误差的方法包括：
 - 1. 大量重复的试验单元；
 - 2. 权衡不同处理组的样本量大小；
 - 3. 当存在其它变异条件下设置组别；

7. 讨论



- Ch14-1: Identify which goal of experimental design (i.e., reducing bias or limiting sampling error) is aided by the following procedures:
 - Using a genetically uniform animal stock to test treatment effects
 - Using a completely randomized design
 - Grouping related experimental units together
 - Taking the response measurements while unaware of the treatments assigned to experimental units
 - Using a computer to randomly assign treatments to experimental units within each block