

# Lecture 12 – 违背前提假设的处理

- 内容大纲
  - 回顾试验/实验设计
  - 违背前提假设的处理 Handling violations of assumptions
- 总结 Summary
- R Lab & Discussion

生态统计

李 勤

生态与环境科学学院

# 1. 选择检验方法

- 进行假设检验前的基本判断
- 假设检验的方法选择
- 违背前提假设时的处理方法
  - Handling violations of assumptions

## 1.1 假设检验前的判断及方法选择

- 选择合适的假设检验方法前的四个审问：
  - 检验涉及一个变量，还是检验两个或更多变量之间的关系？
  - 变量是分类的还是数值的？
  - 数据是否以配对形式出现？
  - 检验的假设是什么，数据是否符合这些假设？
- 参考：INTERLEAF 7 (Page 958)

# 1.1 假设检验前的判断及方法选择

表 1. 一个变量的检验方法 (一组数据)

Data type 数据类型	Goal	目标	Test	检验方法
Categorical 类型变量 (分类变量)	Use frequency data to test whether a population proportion equals a null hypothesized value	使用频数数据检测总体中的比例是否等于零假设中的值	Binomial test (7) $\chi^2$ Goodness-of-fit test with two categories (use if sample size is too large for the binomial test) (8)	二项检验 (7) $\chi^2$ 拟合度检验, 应用于变量分为两类 (如果样本量过大, 无法进行二项检验, 则使用该检验) (8)
	Use frequency data to test the fit of a specific population model	使用频数数据检测特定总体模型的拟合程度	$\chi^2$ Goodness-of-fit test (8)	$\chi^2$ 拟合度检验 (8)
Numerical 数值变量	Test whether the mean equals a null hypothesized value when data are approximately normal (possibly only after a transformation) (13)	当数据近似正态分布 (或经过转换后符合) 时, 检测平均值是否等于零假设中的值 (13)	One-sample $t$ -test (11)	单样本 $t$ 检验 (11)
	Test whether the median equals a null hypothesized value when data are not normal (even after transformation)	当数据不符合正态分布 (即使经过转换), 检测中位数是否等于零假设中的值	Sign test (13)	符号检验 (13)
	Use frequency data to test the fit of a discrete probability distribution	使用频率数据测试离散概率分布的拟合程度	$\chi^2$ Goodness-of-fit test (8)	$\chi^2$ 拟合度检验 (8)
	Use data to test the fit of the normal distribution	检测数据是否符合正态分布	Shapiro-Wilk test (13)	Shapiro-Wilk 检验 (13)

# 1.1 假设检验前的判断及方法选择

表 2. 两个变量相关性的检验方法

		Type of explanatory variable 解释变量			
		Categorical	类型变量	Numerical	数值变量
Type of response variable 响应变量	Categorical 类型变量	Contingency analysis (9)	独立性检验 (9)	Logistic regression (17)	逻辑斯蒂回归 (17)
	Numerical 数值变量	<i>t</i> -tests, ANOVA, Mann-Whitney <i>U</i> -test, etc. [See Table 3 for more details.]	<i>t</i> 检验 方差分析 <i>U</i> 检验等 [更多细节见表 3]	Linear and nonlinear regression (17) Linear correlation (16) Spearman's rank correlation (when data are not bivariate normal) (16)	线性和非线性回归 (17) 线性相关 (16) Spearman 秩相关（当数据不是二元正态分布时） (16)

表 3. 两个变量相关性的检验方法及其前提假设（或检验多组数据的差异）

Number of treatments 处理组数量	Tests assuming normal distribution	假设数据符合正态分布的检验	Tests not assuming normal distributions	假设数据不符合正态分布的检验
Two treatments (independent samples) 两独立样本	Welch's <i>t</i> -test (12) Two-sample <i>t</i> -test (use when variance is equal in the two groups) (12)	Welch's <i>t</i> 检验 (12) 双样本 <i>t</i> 检验（两组方差相等时使用） (12)	Mann-Whitney <i>U</i> -test (Wilcoxon rank-sum test) (13)	<i>U</i> 检验（秩和检验） (13)
Two treatments (paired data) 两配对样本	Paired <i>t</i> -test (12)	配对 <i>t</i> 检验 (12)	Sign test (13)	符号检验 (13)
More than two treatments 超过两组设置	ANOVA (15)	方差分析 (15)	Kruskal-Wallis test (15)	Kruskal-Wallis 检验 (15)

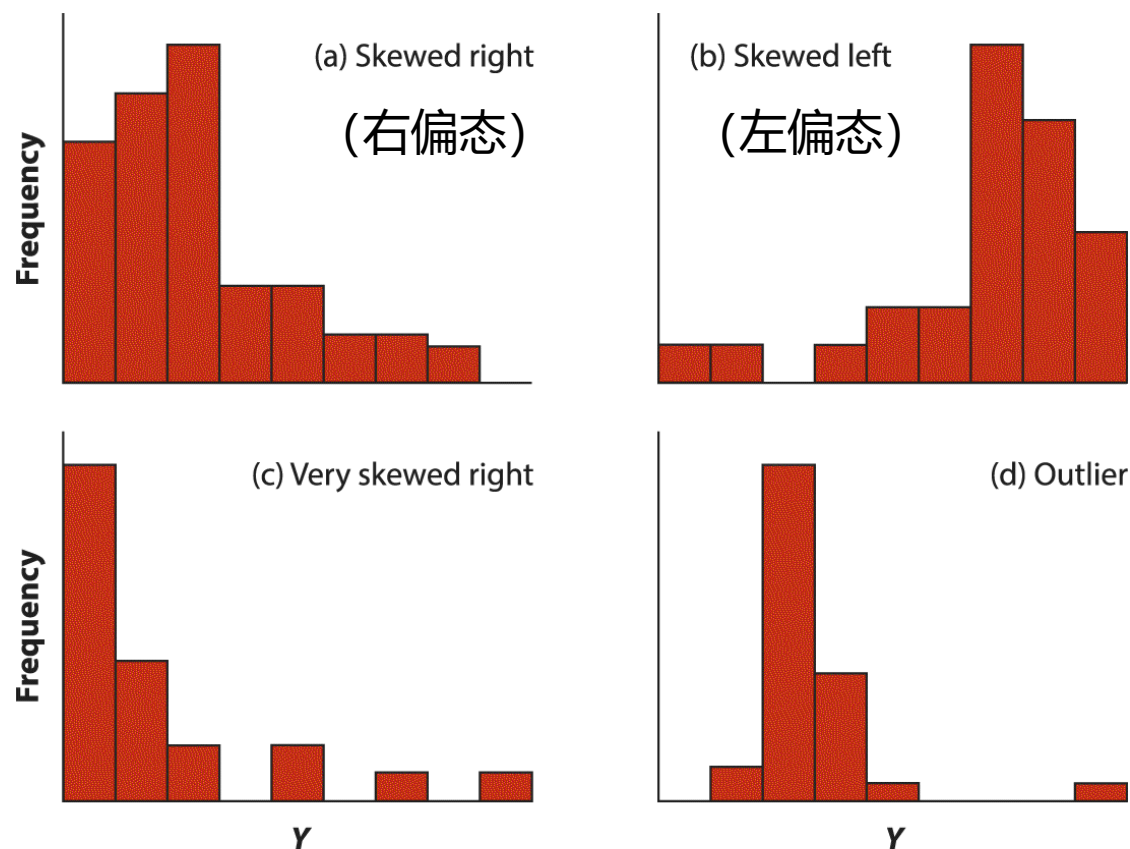
## 1.2 对正态性的检验

- 重要的前提假设
  - 到目前为止，我们学到的所有估计和检验总体均值的方法都假设数值变量近似呈正态分布。
- 检验正态性
  - 方法1：图像法
    - (1) 频数的直方图 (histogram)
    - (2) 正态分位数图 (normal quantile plot)
      - Q-Q Plot: Q表示quantiles
  - 方法2：正式的假设检验

## 1.2 对正态性的检验

- 方法1：图像法

- (1) 频数的直方图 (histogram)：正态分布 vs 偏态分布

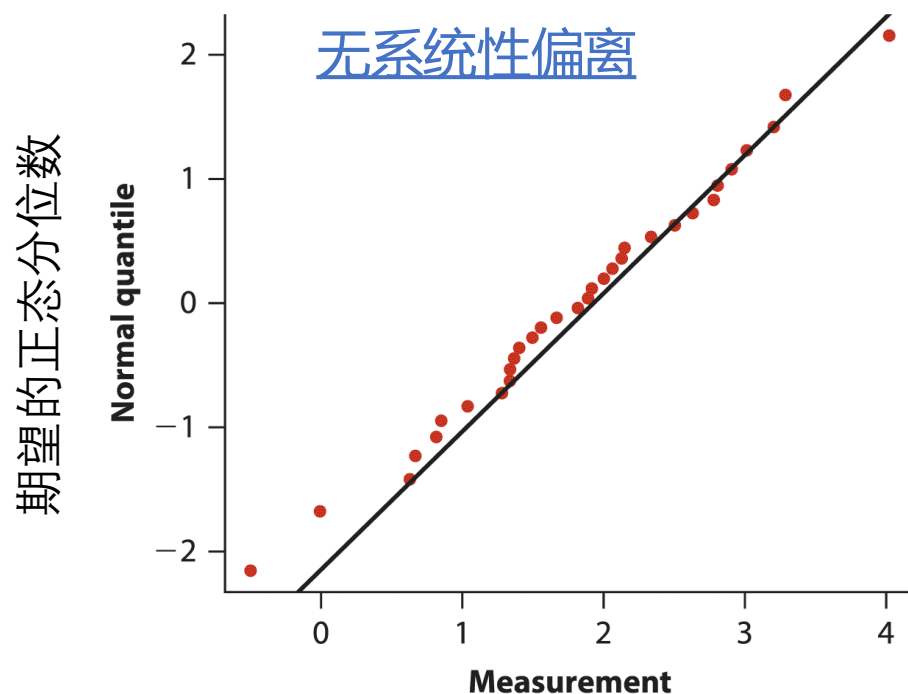


## 1.2 对正态性的检验

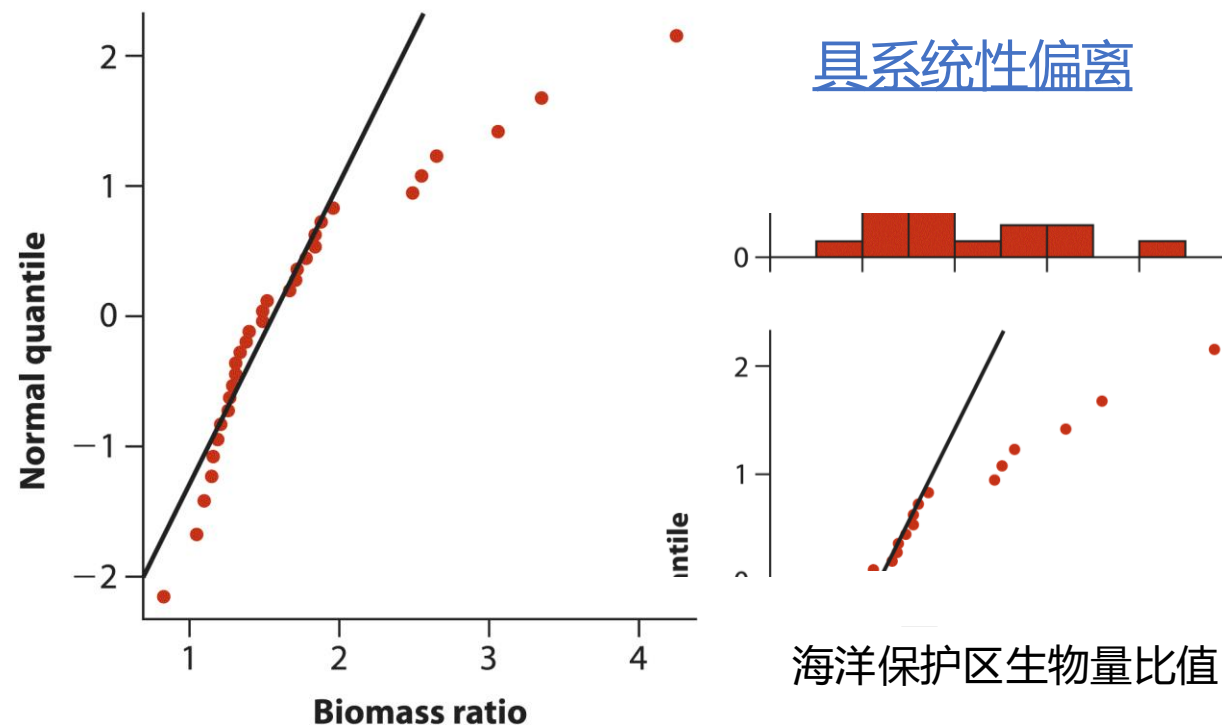
### • 方法1：图像法

- (2) 正态分位数图 (Q-Q Plot)：实际分布 vs 理论正态分布
  - 将样本中的每个观测值与标准正态分布中预期的相应分位数进行比较。

R代码:  
qqplot()  
qqline()



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



## 1.2 对正态性的检验

- 方法2：对正态性的假设检验

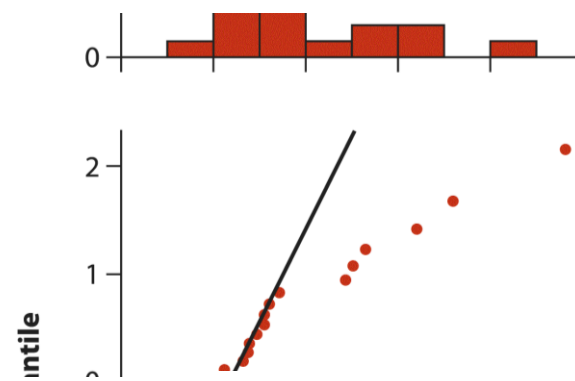
- 应谨慎使用（错误的安全感）；随着样本大小的增加，正态性假设对均值检验的重要性下降；

- Shapiro-Wilk 检验

- $H_0$ : 数据是从具有正态分布的总体中抽样得到的；
- $H_A$ : 数据是从不具有正态分布的总体中抽样得到的；

检验过程：（不讨论具体计算过程）

- 首先利用样本数据来估计总体的均值和标准差，然后检验样本数据与具有相同均值和标准差的正态分布的拟合程度。



海洋保护区生物量比值

```
shapiro.test(marine$biomassRatio)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  marine$biomassRatio  
## W = 0.81751, p-value = 8.851e-05
```

## 2. 违背前提假设时的处理方法

- 违背前提假设
  - 频数分布通常不是正态的
  - 样本间标准差也并不总是相等的
- 四个常见的方案
  - 1. 忽略对前提假设的违背
  - 2. 对数据进行转换
  - 3. 使用非参方法
  - 4. 使用置换检验

## 2. 违背前提假设时的处理方法

- 四个常见的方案

- 1. 忽略对前提假设的违背

- 检验方法具有一定稳健性，对没有严格满足前提假设情况也适用；

- 2. 对数据进行转换

- 转换后的数据满足前提假设；

- 3. 使用非参方法

- 非参方法不要求正态性假设；

- 4. 使用置换检验

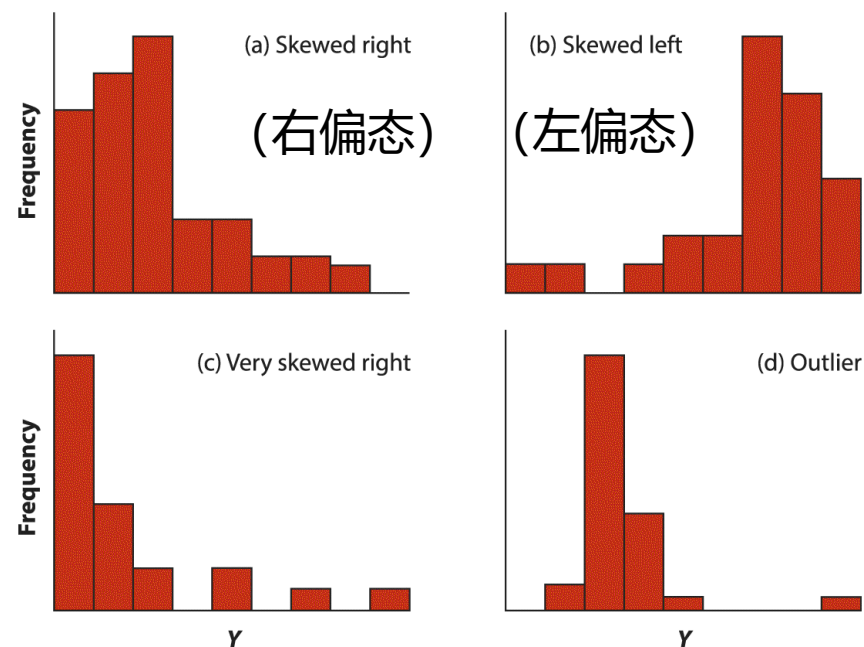
- 计算机通过反复和随机重新排列一个变量的数据来生成检验统计量的零分布；

## 2.1 忽略对前提假设的违背

- 因为检验方法具有一定稳健性
  - 对没有严格满足前提假设情况也适用；
- (1) 对正态性的违背
  - 尽管计算置信区间和 $t$ 检验均要求数据呈正态分布，但有时这些方法可以用于分析不呈正态分布的数据；
  - 方法的稳健性来自中心极限定理：当变量不呈正态分布时，样本均值的分布在样本量大时仍然近似正态；
  - 前提是对正态性的违背不太严重；
  - 稳健性仅适用于针对均值的统计方法，而不适用于诸如用于检验方差的 $F$  检验之类的方法（后者对于违反正态性的假设不具有稳健性）；

## 2.1 忽略对前提假设的违背

- (1) 对正态性的违背
  - 计算置信区间和 $t$ 检验时：样本量大时可以忽视对正态性的违背；
  - 样本量的大小取决于数据分布的形状；
  - 例如：
    - 两个分布同(a)：至少  $n = 30$ ；
    - (a) 和 (b)：至少  $n = 500$ ；
    - 如果更加偏态如(c)(d)，则不适用；
      - 无法得出准确的结论；
      - 采用其它方法；



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

## 2.1 忽略对前提假设的违背

- (2) 对方差相等的违背

- 在中等样本量（每组大于30）的情况下，即使两组标准差相差三倍，只要两组的样本量大致相等，双样本 $t$ 检验时仍能表现良好；
- 如果两个样本的样本量差异很大，或者标准差的差异超过三倍，那么应该使用 Welch's  $t$ -检验；
- 如果正态性假设也不成立，那么最好尝试数据转换、置换检验或其它方法（Ch19）；

## 2.2 对数据进行转换

- 转换后的数据可以满足前提假设;
  - 目标：找到一个合适的数值尺度使得测量值间的差异保持相似的解释;
- 常见的数据转换方法:  $Y \rightarrow Y'$  (Y-prime)

### (1) 对数转换

the Log transformation

$$Y' = \ln[Y]$$

$$Y' = \ln[Y + 1]$$

back-transform  
转换回原尺度:  
(计算95%CI时)

$$Y = e^{Y'}$$

### (2) 反正弦转换

the arcsine transformation

$$p' = \arcsin[\sqrt{p}]$$

$$p = (\sin(p'))^2$$

### (3) 平方根转换

the square-root transformation

$$Y' = \sqrt{Y + 1/2}$$

$$Y = Y'^2 - 1/2$$

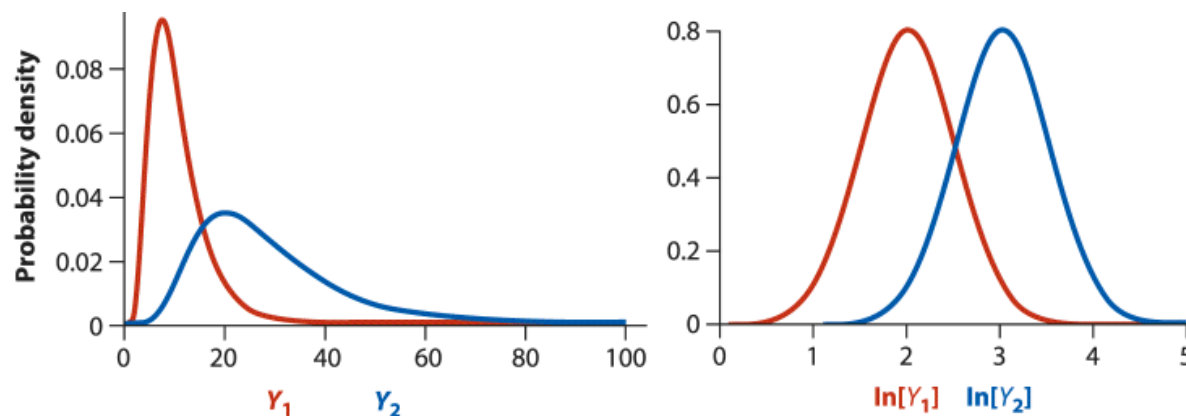
## 2.2 对数据进行转换

- 对数据进行转换 Data transformations  $Y \rightarrow Y'$

- (1) 对数转换:  $Y' = \ln[Y]$

- 常用场景:

- 测量值是变量的比率或乘积;
    - 数据的频率分布向右倾斜 (即右侧有长尾);
    - 数据跨越了多个数量级;
      - E.g., 生物量 biomass

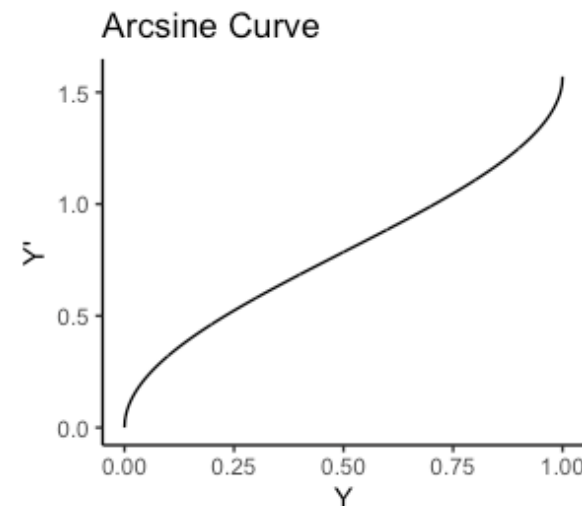




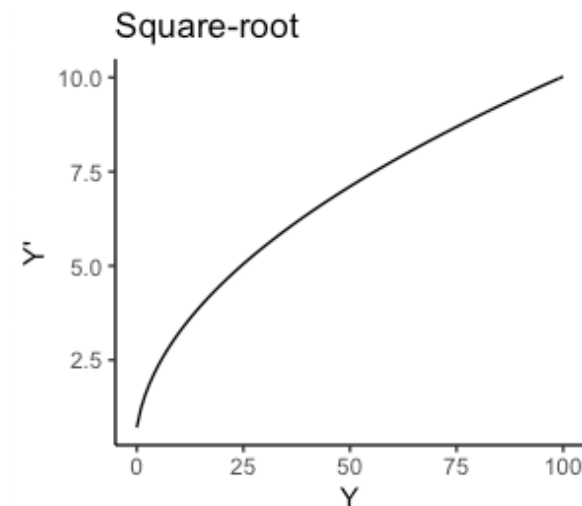
## 2.2 对数据进行转换

- 对数据进行转换  $Y \rightarrow Y'$

- (2) 反正弦转换:  $p' = \arcsin[\sqrt{p}]$ 
  - 常用场景: 比例数据  $p \in [0,1]$ ;
  - 因为比例往往不呈正态分布 (尤其当均值接近0或接近1时) ;
  - 而且因为均值不同的组往往没有相等的标准差;
  - 反正弦转换通常同时解决了这两个问题。



- (3) 平方根转换:  $Y' = \sqrt{Y + 1/2}$ 
  - 常用场景: 数据是计数数据时;
  - 例如产卵量、细菌菌落数等时;
  - 像对数转换一样有助于在均值较高的组也具有较大标准差的情况下使标准差趋于相等。



## 2.3 非参方法

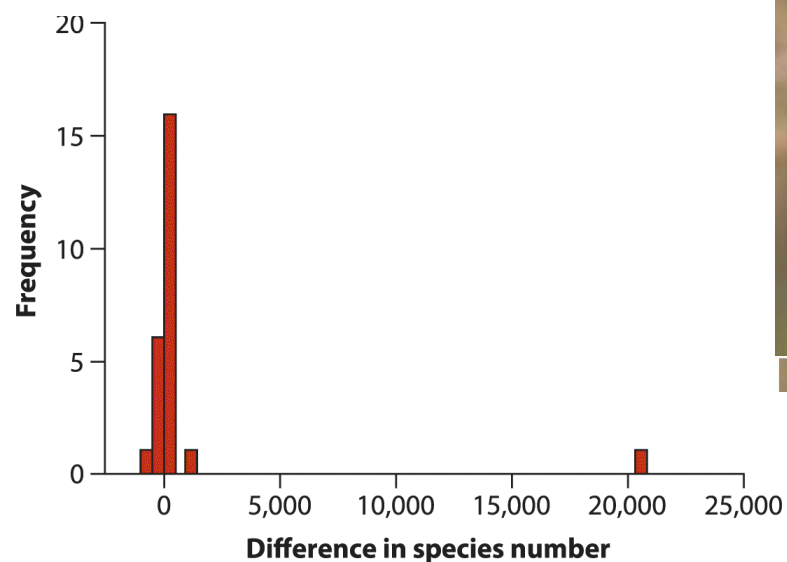
- 参数方法 parametric test
  - 对抽样数据的概率分布进行了假设;
  - 到目前为止, 我们学到的所有检验方法都是参数方法;
- 非参方法 nonparametric test
  - 对数据的概率分布进行的假设较少;
  - 适用于数据的频数分布不是正态分布 (例如存在异常值) 的情况;

## 2.3 非参方法

- 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
- 替代方差分析ANOVA的非参方法
  - (4) 基于秩的Kruskal-Wallis检验

## 2.3 非参方法

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



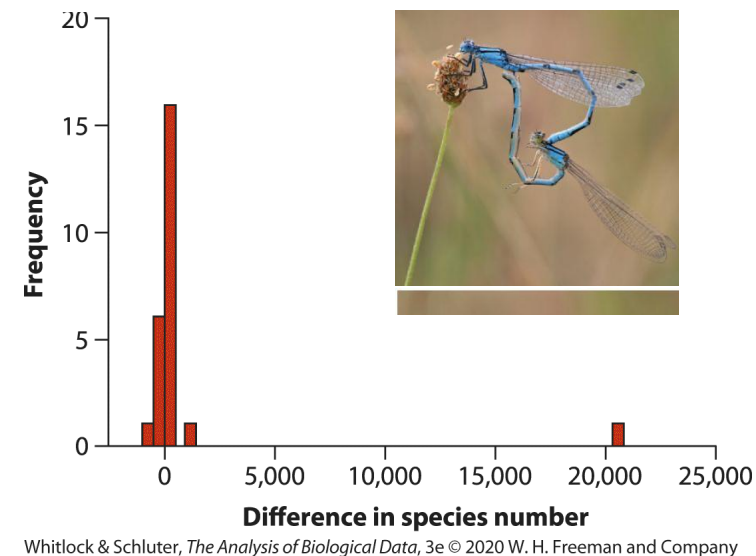
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company



	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	+

## 2.3 非参方法

- (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;



## 2.3 非参方法

- (1) 符号检验 Sign test

- **EXAMPLE 13.4: Sexual conflict and the origin of new species**

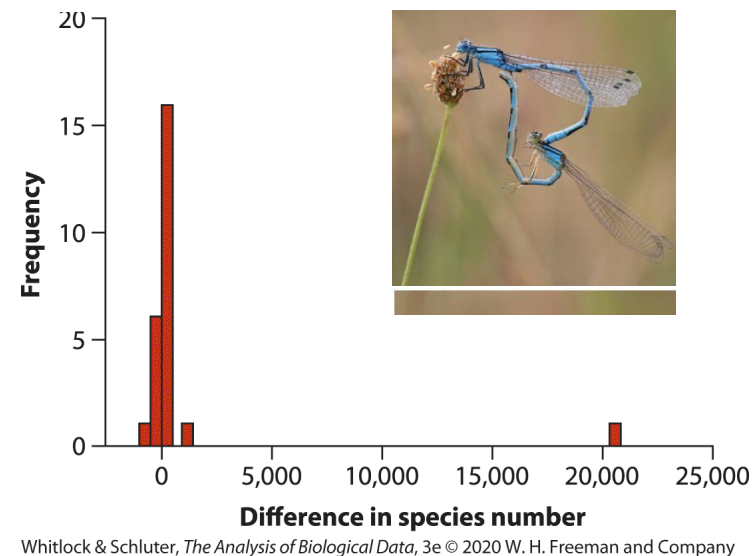
- 交配多次 vs. 交配一次 → 是否促进物种的演化?
- 25个数据点: 昆虫类群配对组数
- "+" 共计18组; "-" 共计7组;

- 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)

- $P\text{值} = 2 \Pr[X \leq 7] = 2 \sum_{i=0}^7 \binom{25}{i} 0.5^i 0.5^{25-i} = 2 \times 0.02164 = 0.043$
- 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。



$$\Pr[X] = \binom{n}{x} p^x (1-p)^{n-x}$$

## 2.3 非参方法



- (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
```

## 2.3 非参方法

- (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

- (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
```

## 2.3 非参方法

- (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）；

## 2.3 非参方法

- (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

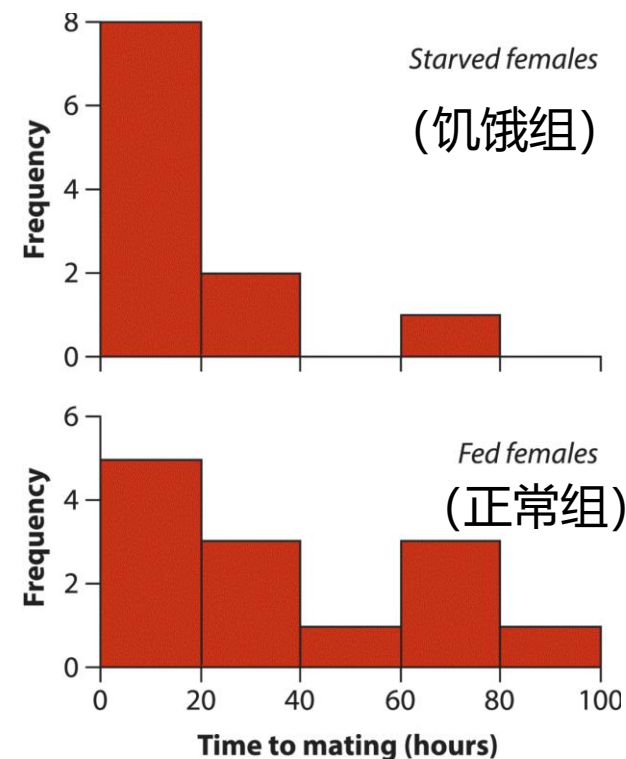


© David H. Funk

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

## 2.3 非参方法

- (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。



## 2.3 非参方法

- (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:
  - $R_1 = 3+4+7+9+10+11+12+13+14+17+21 = 121$
  - $U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 11 \times 13 + 11 \times 12 / 2 - 121 = 88$
  - $U_2 = n_1n_2 - U_1 = 11 \times 13 - 88 = 55$ 
    - (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

## 2.3 非参方法

- (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
```

## 2.3 非参方法

- (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：

- $$Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1) / 3}}$$

## 2.3 非参方法

- (4) Kruskal-Wallis检验：替代 ANOVA
  - 基于秩的方差分析
    - 在有两组的情况下，相当于Mann-Whitney  $U$  检验
  - 检验统计量：  $H$ 
    - 基于每组的秩和  $R_i$
    - $H$ 的抽样分布近似于自由度为  $k - 1$  的  $\chi^2$  分布
    - 当  $H > \chi^2_{\alpha, k-1}$  时，拒绝零假设

```
kruskal.test(shift ~ treatment, data = circadian)
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
## data:  shift by treatment  
## Kruskal-Wallis chi-squared = 9.4231, df = 2, p-value = 0.008991
```



## 2.3 非参方法

- (4) Kruskal-Wallis检验：替代 ANOVA
  - 前提假设
    - 随机样本 & 变量的分布在各总体中的形状是一样的；
  - 注意事项
    - 样本量很小时，该检验的功效低；样本量大时，该检验与ANOVA功效一致；

## 2.3 非参方法

- 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）。

## 2.3 非参方法

- 关于I类错误:

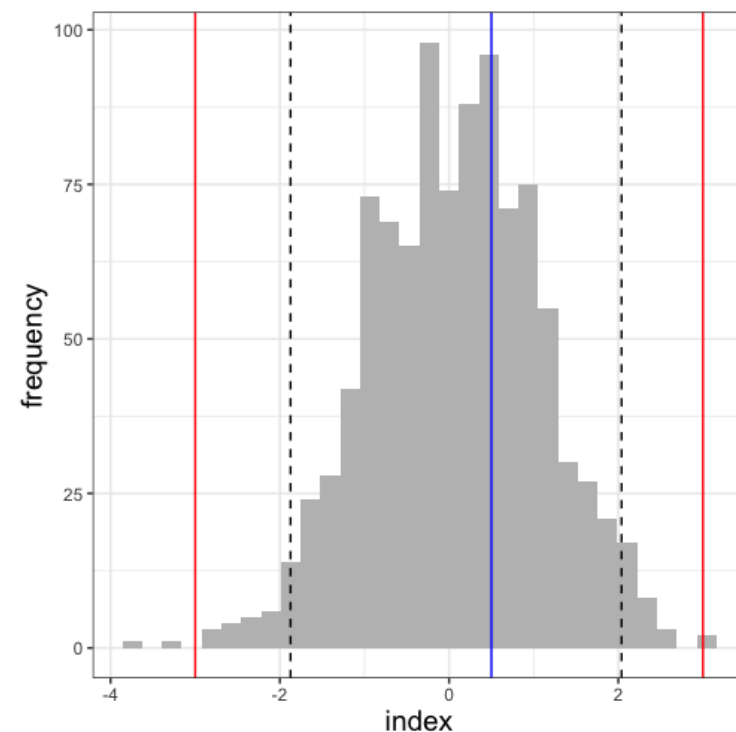
- 满足特定检验的前提假设时, 无论是参数检验还是非参数检验, 出现I类错误 (即拒绝true  $H_0$ ) 的概率都为  $\alpha$ ;
- 当不符合参数检验的前提假设时, 例如测量值的分布严重偏态, 那么I类错误率就会大于设定的  $\alpha$  值;
- 因此, 在前提假设被严重违反时不使用参数检验的主要原因就是过高的I类错误率;
- 非参数检验只要满足限制性较小的前提假设, 其I类错误率等于  $\alpha$ ;

## 2.3 非参方法

- 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。

## 2.4 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)
    - 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)
- 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。



## 2.4 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
starved	1.9
starved	2.1
starved	3.8
starved	9
starved	9.6
starved	13
starved	14.7
starved	17.9
starved	21.7
starved	29
starved	72.3

mean: 17.73

feedingStatus	timeToMating
fed	1.5
fed	1.7
fed	2.4
fed	3.6
fed	5.7
fed	22.6
fed	22.8
fed	39
fed	54.4
fed	72.1
fed	73.6
fed	79.5
fed	88.9

mean: 35.98

## 2.4 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

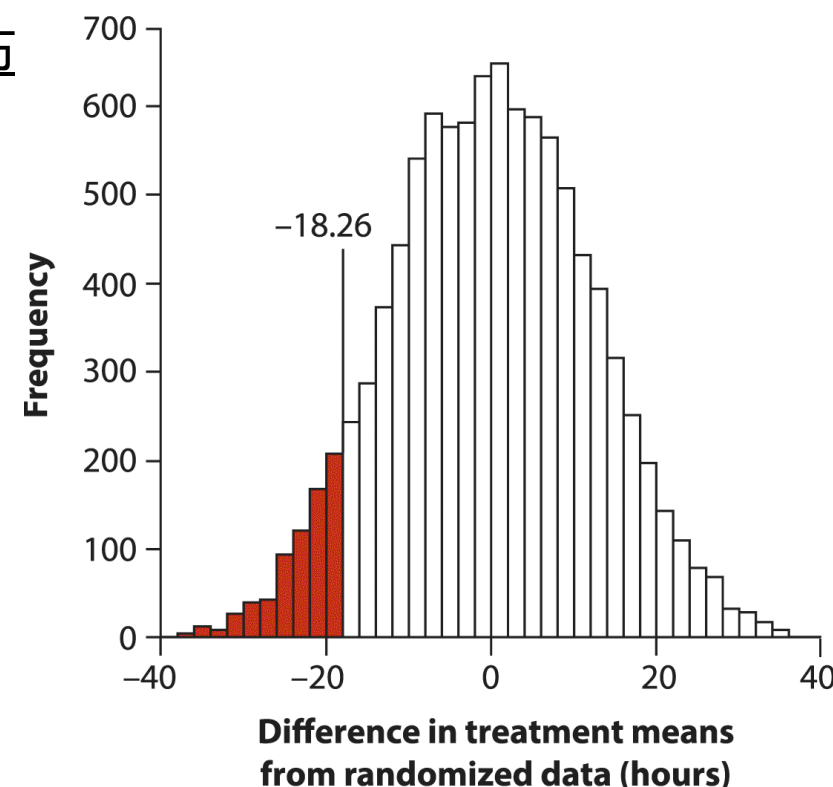
	feedingStatus	permSample		feedingStatus	permSample
1	starved	1.7	12	fed	1.9
2	starved	17.9	13	fed	1.5
3	starved	22.6	14	fed	2.1
4	starved	73.6	15	fed	9.6
5	starved	72.1	16	fed	9
6	starved	14.7	17	fed	54.4
7	starved	13	18	fed	2.4
8	starved	5.7	19	fed	79.5
9	starved	88.9	20	fed	3.6
10	starved	21.7	21	fed	22.8
11	starved	29	22	fed	72.3
			23	fed	3.8
			24	fed	39

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

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

## 2.4 置换检验

- 例子： [EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets](#) 蟋蟀的性食性
- 步骤：
  - (3) 1000个  $\bar{Y}_1 - \bar{Y}_2 \rightarrow$  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 & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company



## 2.4 置换检验

- 例子: [EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets](#) 蟋蟀的性食性
- 步骤:
- R code

进行随机化操作的关键

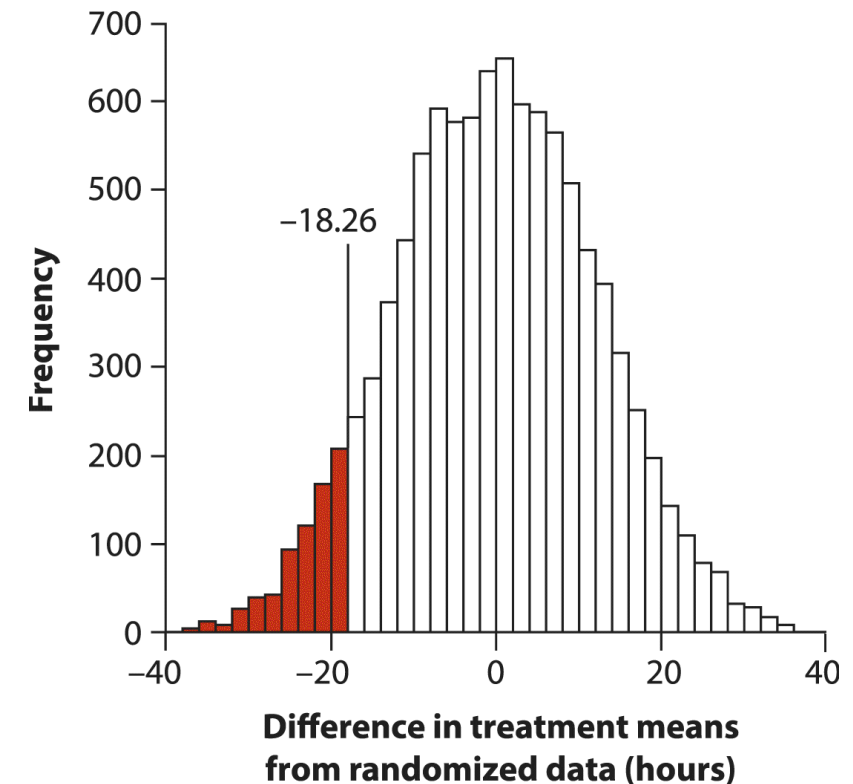
```
permResult <- vector() # initializes
for(i in 1:nPerm){
  if(i%%100) print()
  # step 1: permute the times to mating
  cannibalism$permSample <- sample
(cannibalism$timeToMating, replace = FALSE)
  # step 2: calculate difference between means
  permSampleMean <- as.data.frame(summarize
(group_by(cannibalism,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



## 2.4 置换检验

- 例子： [EXAMPLE 13.5: Sexual cannibalism in sagebrush crickets](#) 蟋蟀的性食性

如果重复以上步骤，会得出相同的P值吗？



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

## 2.4 置换检验

- 置换检验的前提假设很少，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。

### 3. 补充资料

- 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



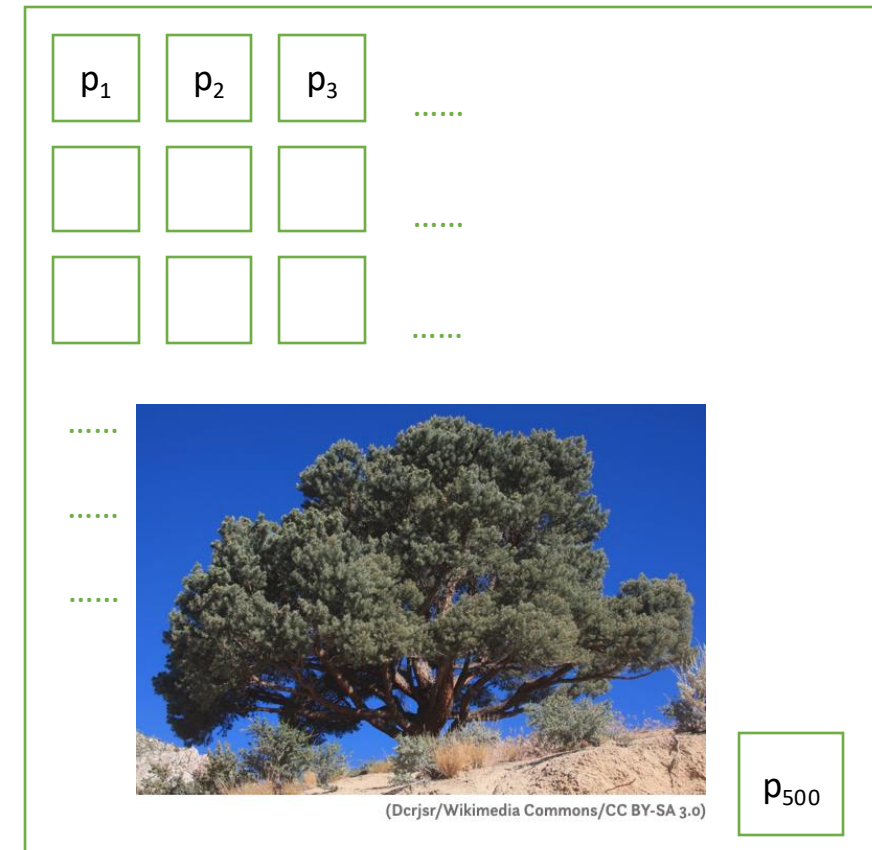
### 3.1 正确的抽样单元 sampling unit

- 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?



# 3.1 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

存活比例

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

## 3.1 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  $\chi^2$  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

# 3.1 正确的抽样单元

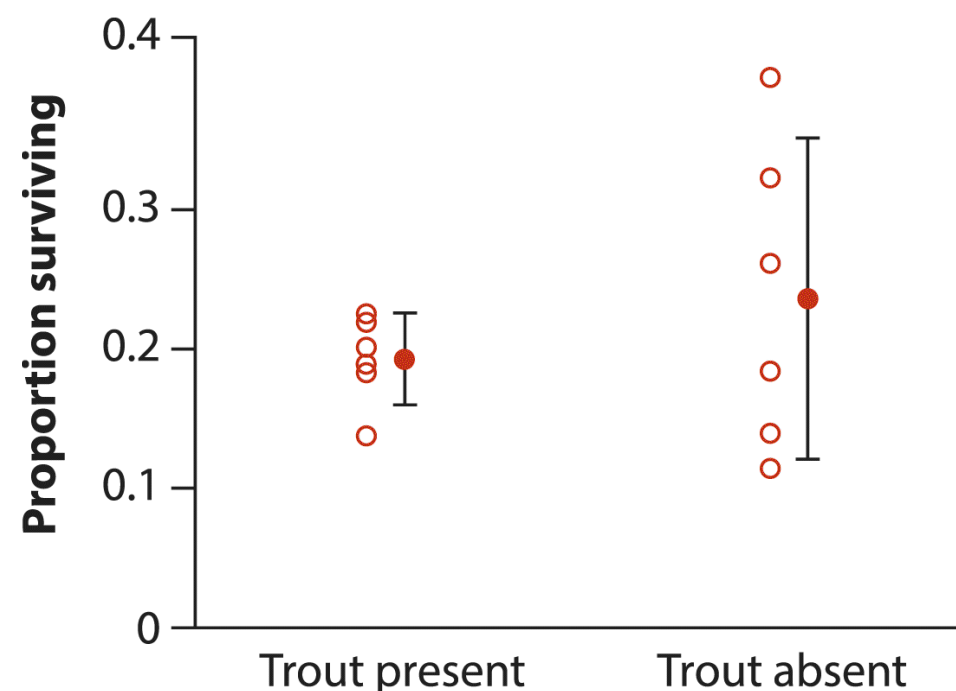
- 抽样单元间的独立性 (the independent units of replication)
  - 正确判断重复的抽样单元才能保证结论的准确性;

- 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



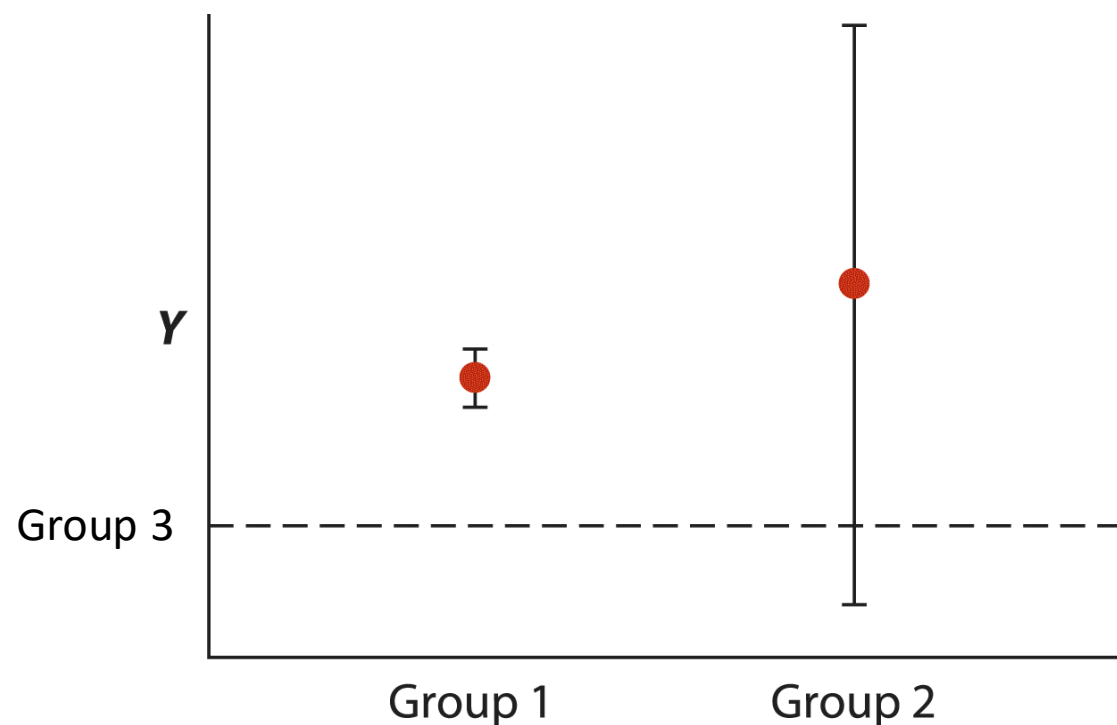
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

Welch's approximate ( $P=0.39$ )



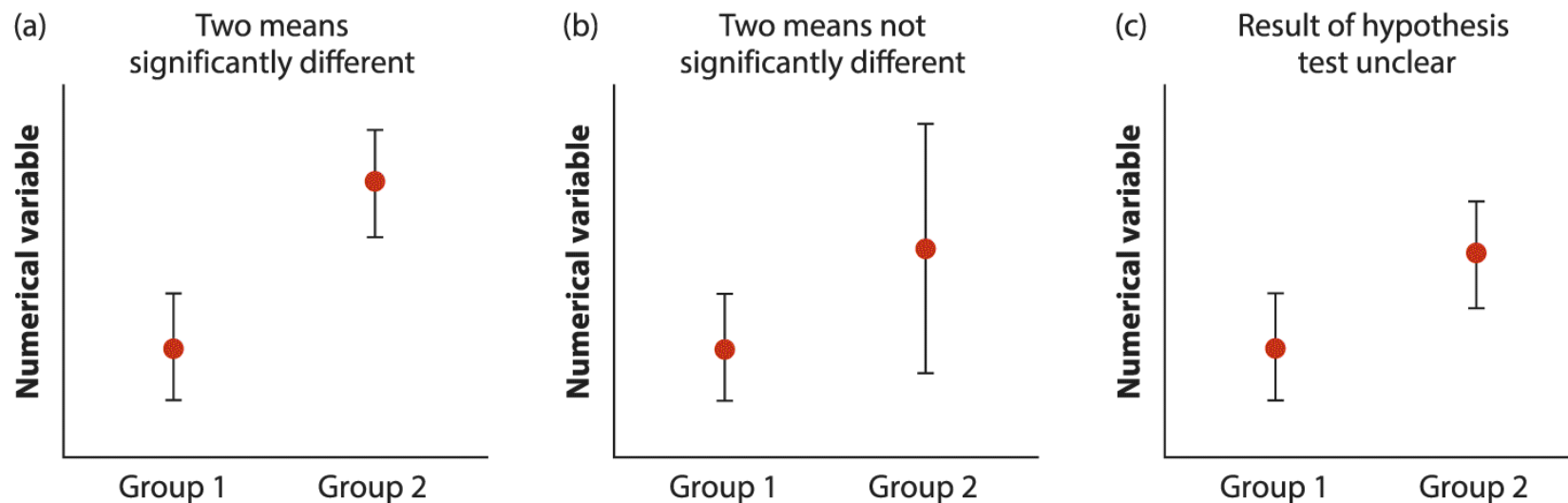
## 3.2 间接比较的谬误

- **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



### 3.3 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。



## 4. 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；

## 5. R coding

- [https://whitlockschluter3e.zoology.ubc.ca/RExamples/Rcode\\_Chapter\\_13.html](https://whitlockschluter3e.zoology.ubc.ca/RExamples/Rcode_Chapter_13.html)
  - `qqnorm()`
  - `shapiro.test()`
  - `binom.test()` # sign test
  - `wilcox.test()`
  - `sample(X, replace = FALSE)` # for permutations