

# Lecture 10 – Comparing two means

- Outline for today
  - Recall One sample  $t$ -test 单样本 $t$ 检验
  - Comparing two means 两个样本的均值比较 Ch12
    - Two sample  $t$ -test 双样本 $t$ 检验
  - Handling violations of assumptions 违背前提假设的处理 Ch13
    - Normality 正态性的检验
- Summary
- R Lab & Discussion

生物统计学

李 勤

生态与环境科学学院



# 1. 回顾 — 单样本t检验

- 针对连续数值变量的检验
  - 基于t-分布来比较随机样本的均值和零假设中的总体均值。
  - 零假设和备择假设
    - $H_0$ : The true mean equals  $\mu_0$  ( 总体均值为  $\mu_0$  ) ;
    - $H_A$ : The true mean does not equal  $\mu_0$  ( 总体均值不为  $\mu_0$  ) ;
  - 检验统计量为:  $t = \frac{\bar{Y} - \mu_0}{\text{SE}_{\bar{Y}}}$
  - P值: 对比观测 t 值与 Student's t-distribution

## 1.1 单样本 $t$ -test 示例

- 例子：人体温度；大样本  $n = 130$

- 1.  $H_0: \mu_0 = 98.6^\circ\text{F};$

$$H_A: \mu_0 \neq 98.6^\circ\text{F};$$

- 2.  $\bar{Y} = 98.25^\circ\text{F}; s = 0.733^\circ\text{F}$

$$t = \frac{\bar{Y} - \mu_0}{\text{SE}_{\bar{Y}}} = -5.44;$$

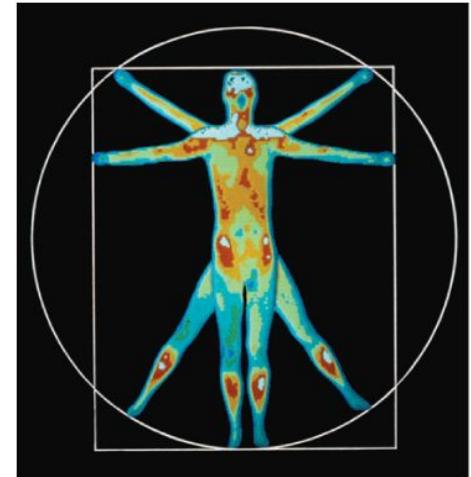
- 3. P值 = 0.000016 < 0.05

- 4. 结论：拒绝零假设；

- 均值的 95% 置信区间：

- $\bar{Y} - t_{0.05(2), df} \text{SE}_{\bar{Y}} < \mu < \bar{Y} + t_{0.05(2), df} \text{SE}_{\bar{Y}}$

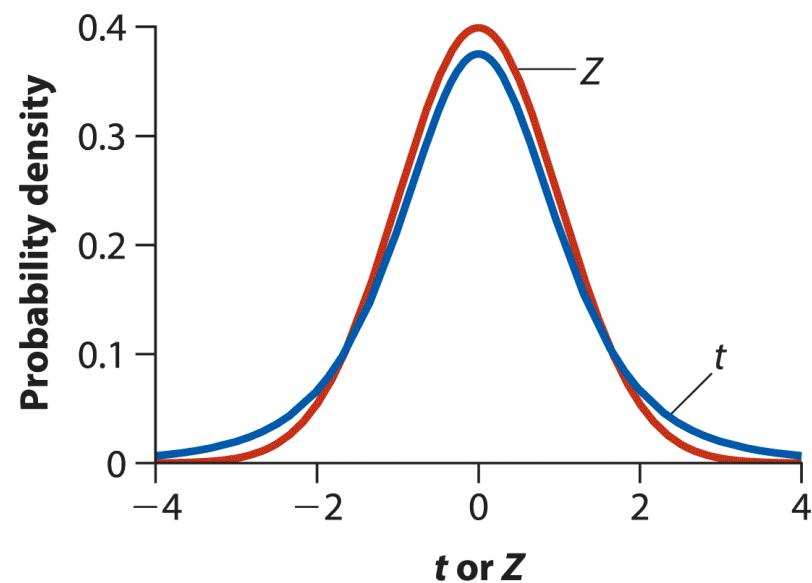
- $98.12^\circ\text{F} < \mu < 98.38^\circ\text{F}$  (范围: 0.26 °F)



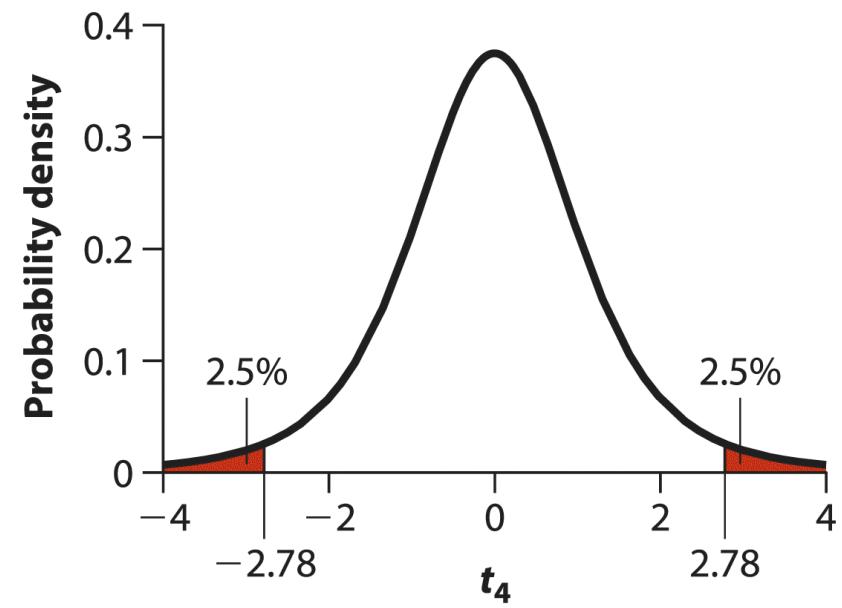
Dr. Ray Clark FRPS & Mervyn de  
Calcina-Goff FRPS/Science  
Source

## 1.2 单样本 $t$ -test 的前提假设

- 随机样本；
- 变量在总体中呈正态分布；



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



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

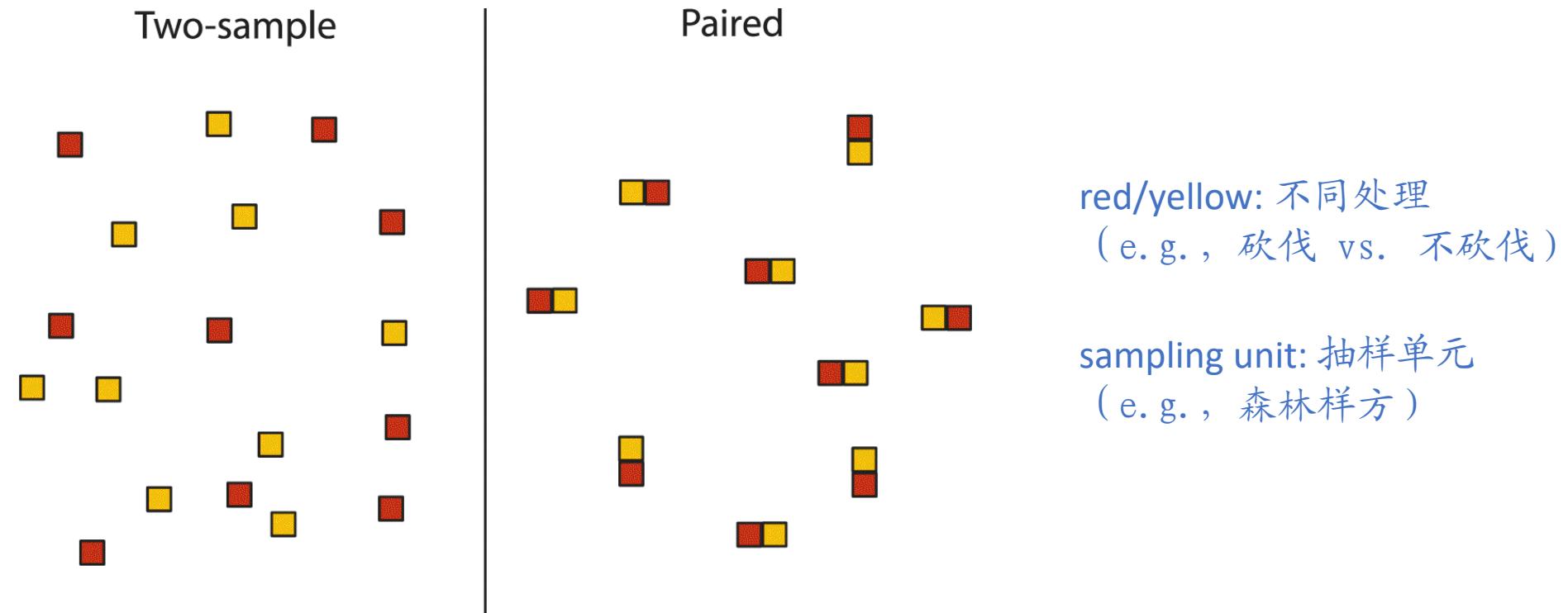
## 2. 两组数据的均值比较

- 生物数据通常用于比较不同组或处理的均值：
  - 接受新药治疗的患者是否在平均反应上显著优于接受旧药治疗的患者？
  - 学生是在夜间长时间学习后在测试中表现出更好的平均水平，还是在晚上得到充足休息后表现更好？
  - 在向阳处生长的植株是否比在遮阴处生长的植株更高？
- 比较两组数据间的方差



## 2.1 比较不同组或处理的均值

- 不同的研究设计：配对样本 vs. 两个独立样本
  - 有什么不同？



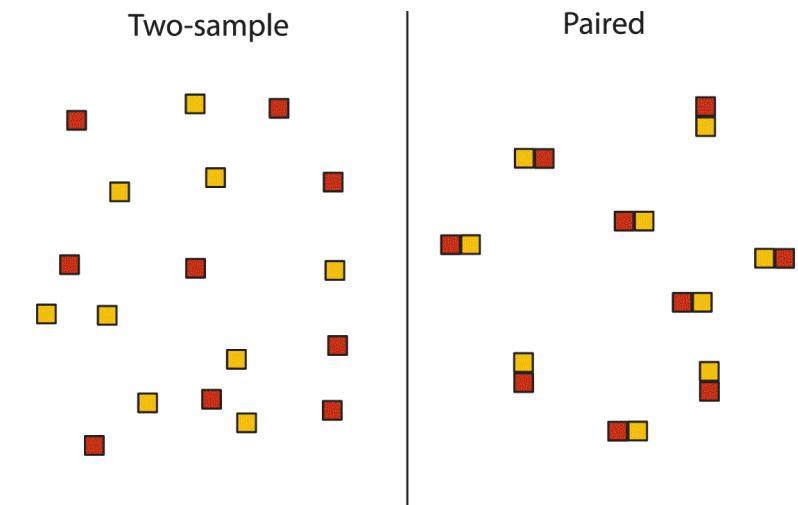
## 2.1 比较不同组或处理的均值

- 配对样本设计 (paired design)

- 两种处理都被应用于每个抽样单元;
- 同一单元内的两种处理的结果不是独立的;
  - 例如, 环境条件可能更相似;
  - 但同时, 这种相似性将处理之外的其它因素(干扰)最小化;
  - 因此, 更容易看到两种处理之间的差异;
  - 但不一定容易进行配对操作;

- 两个独立样本设计

- 每个处理组由独立的、随机抽取的单元组成;



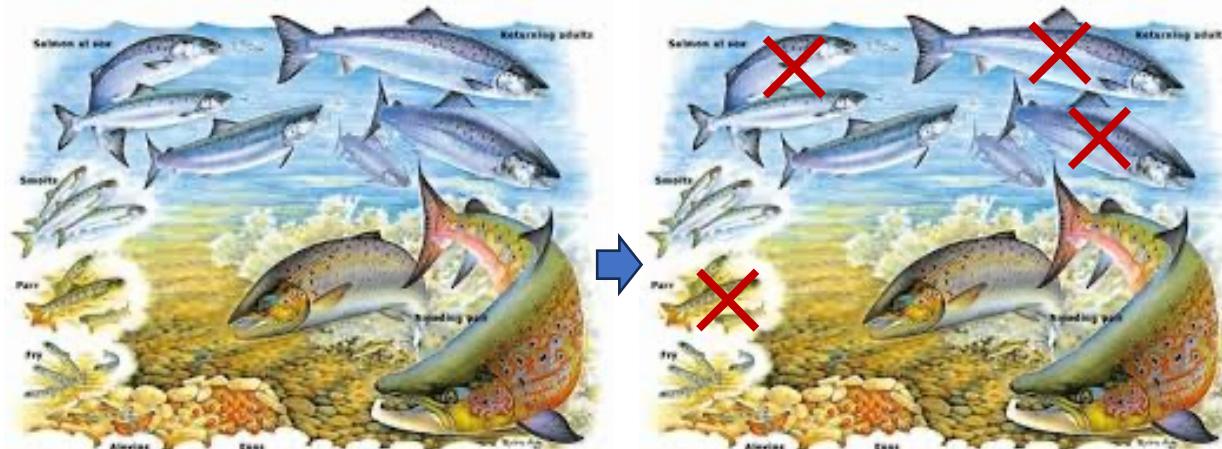
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

## 2.2 均值的配对比较 paired comparison of means

- 特点/优势:

- 减少了与处理本身无关的抽样单元之间的变异效应;
- 提高了估计的功效和精度;

- 配对比较的例子:



湖泊里的多样性的变化



防晒霜的效果



双胞胎研究

## 2.2 均值的配对比较 paired comparison of means

- 特点/优势：
  - 减少了与处理本身无关的抽样单元之间的变异性效应；
  - 提高了估计的功效和精度；
- 配对比较的例子
- 配对比较的量化：差异值
  - 通过计算每个抽样单元内不同处理之间的差异 → 成对的测量值被转换成单一的测量值；
  - 据此生成与随机抽样单元数量相同的数据点；
    - 如：20个个体被分成10对 → 10个对两种处理之间差异的测量值 → 样本大小为10；
  - 使用差异的均值来进行处理结果的估计；

## 2.2 基于配对数据估计平均差异 (mean difference)

- 配对比较的量化: 差异值
  - 使用差异的均值来进行处理结果的估计;
- 例子: 激素水平和免疫力
  - EXAMPLE 12.2: So macho it makes you sick?
  - 目标: 繁殖代价?
  - 是否较高激素水平会降低免疫力?
    - 雄性如果有更高的激素水平, 更可能吸引雌性;
    - 但是否在其它方面 (如抵御疾病) 付出代价呢?



Gerald A. DeBoer/  
Shutterstock.com

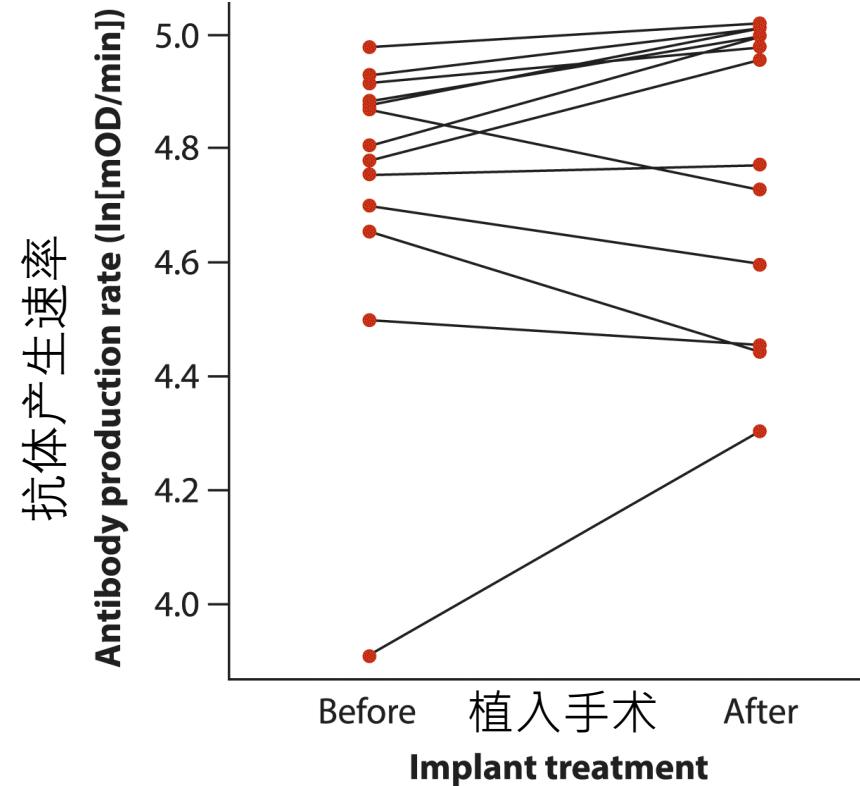
red-winged blackbird  
红翅黑鹂

## 2.2 基于配对数据估计平均差异 (mean difference)

- 配对比较的量化: 差异值
  - 使用差异的均值来进行处理结果的估计;
- 例子: 激素水平和免疫力
  - 目标: 是否较高激素水平会降低免疫力?  
(繁殖代价?)
  - 方法: 人为提高睾酮水平;  
测量植入前后鸟类血清中产  
生抗体的速率 (光学密度);
  - 数据: 植入前后的抗体产生  
速率的平均差异是多少?



Gerald A. DeBoer/  
Shutterstock.com



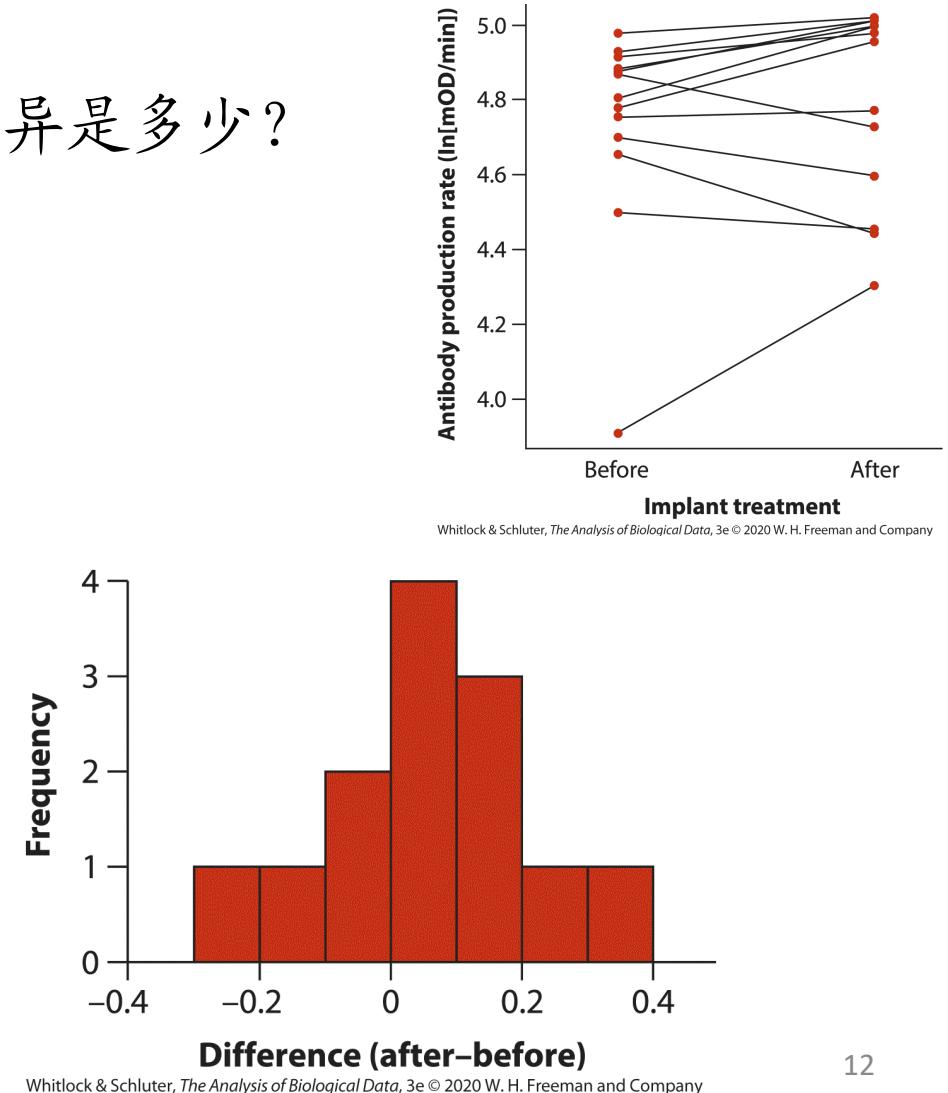
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

Hasselquist et al. (1999)

## 2.2 基于配对数据估计平均差异 (mean difference)

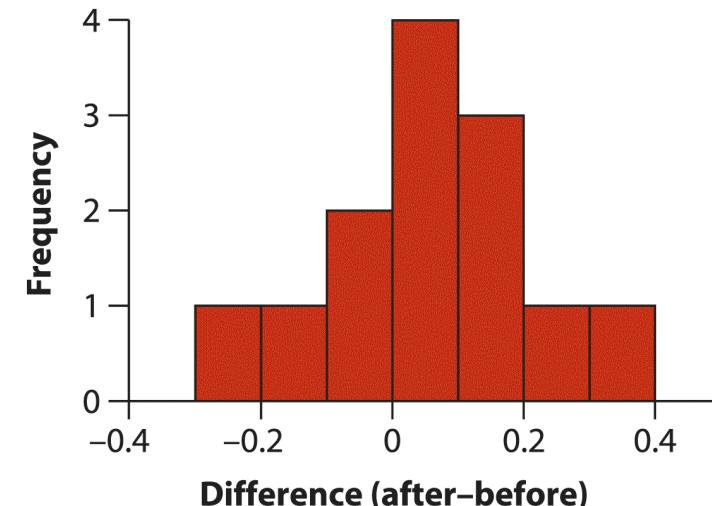
- 例子：激素水平和免疫力
  - 数据：植入前后的抗体产生速率的平均差异是多少？

blackbird	beforeImplant	afterImplant	logBeforeImplant	logAfterImplant	diff
1	105	85	4.65	4.44	-0.21
2	50	74	3.91	4.3	0.39
3	136	145	4.91	4.98	0.07
4	90	86	4.5	4.45	-0.05
5	122	148	4.8	5	0.2
6	132	148	4.88	5	0.12
7	131	150	4.88	5.01	0.13
8	119	142	4.78	4.96	0.18
9	145	151	4.98	5.02	0.04
10	130	113	4.87	4.73	-0.14
11	116	118	4.75	4.77	0.02
12	110	99	4.7	4.6	-0.1
13	138	150	4.93	5.01	0.08



## 2.2 基于配对数据估计平均差异 (mean difference)

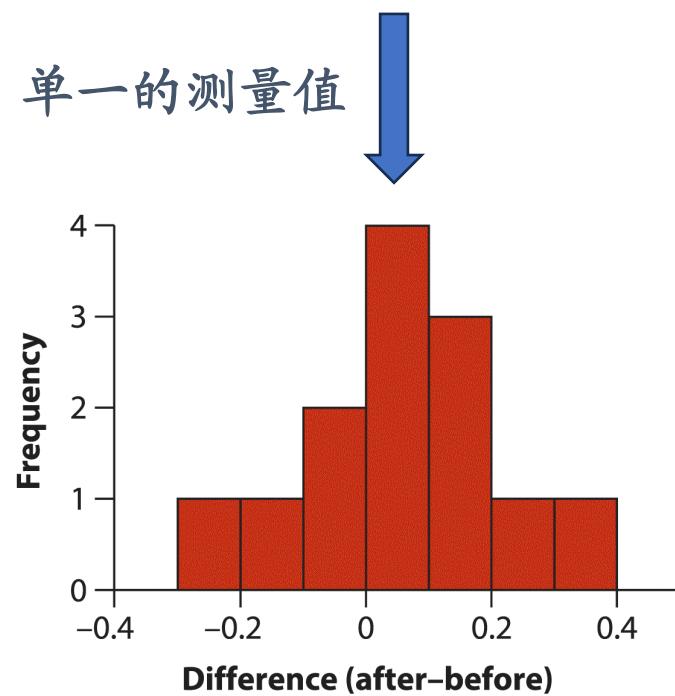
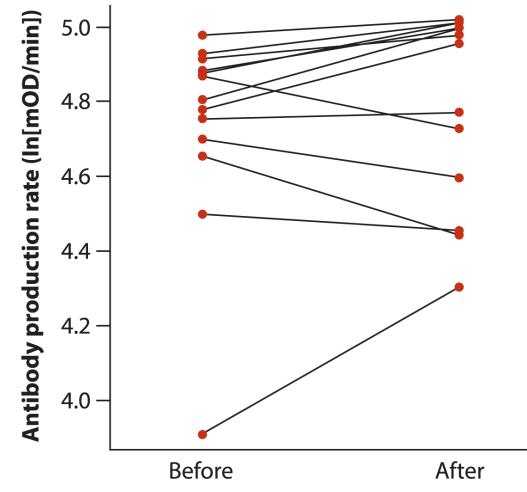
- 例子：激素水平和免疫力
  - 数据：植入前后的抗体产生速率的平均差异是多少？ $d$  ( $n = 13$ )
  - 差异的均值： $\bar{d} = 0.056$
  - 差异的标准差： $s = 0.159$
  - 差异的置信区间：（与单样本均值一致）
    - $\bar{d} - t_{\alpha(2), df} \text{SE}_{\bar{d}} < \mu_d < \bar{d} + t_{\alpha(2), df} \text{SE}_{\bar{d}}$
    - 其中  $\text{SE}_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{0.159}{\sqrt{13}} = 0.044$ ;  $t_{0.05(2), 12} = 2.18$
    - $-0.040 < \mu_d < 0.152$
    - unit (单位):  $\ln[\text{mOD}/\text{min}]$
  - 结果：95%CI的范围包含0值；



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

## 2.3 配对t检验 (Paired t-test)

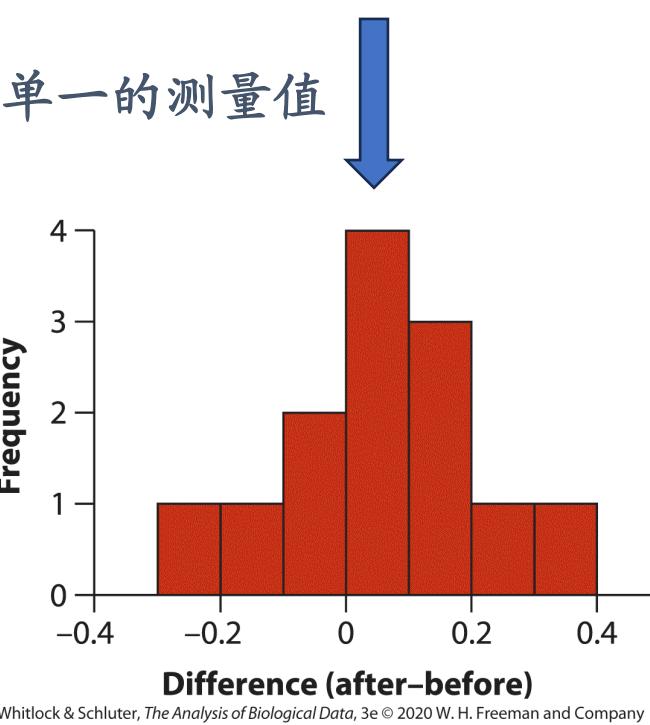
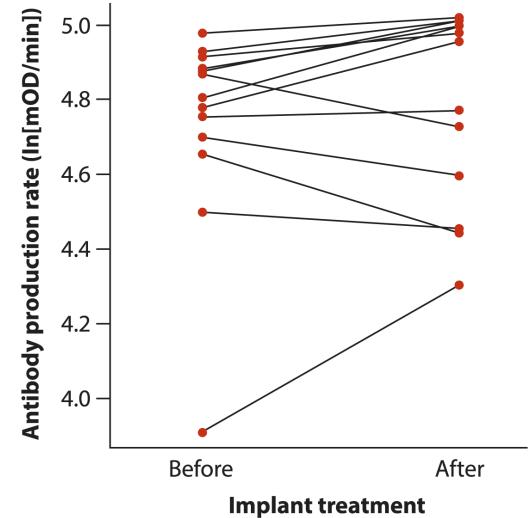
- 例子：激素水平和免疫力
  - 数据：抗体产生速率的差异？ $d$  ( $n = 13$ )
- 假设检验
  - $H_0$ : 睾酮植入后抗体产生速率的平均变化为零；  
(配对测量差异的均值等于一个特定值)
  - $H_A$ : 睾酮植入后抗体产生速率的平均变化不为零；
- $H_0: \mu_d = 0$ ;  $H_A: \mu_d \neq 0$ ;
- 此后的步骤与单样本t检验一致：



## 2.3 配对t检验 (Paired t-test)

- 例子：激素水平和免疫力
  - 数据：抗体产生速率的差异？ $d$  ( $n = 13$ )

- 假设检验
  - 1.  $H_0: \mu_d = 0$ ;  $H_A: \mu_d \neq 0$ ;
  - 2. 检验统计量:  $t = \frac{\bar{d} - \mu_{d0}}{\text{SE}_{\bar{d}}} = \frac{0.056 - 0}{0.044} = 1.27$
  - 3. 计算P值:
    - $df = n - 1 = 13 - 1 = 12$ ;
    - $P = \Pr[t_{12} < -1.27] + \Pr[t_{12} > 1.27] = 2 \Pr[t_{12} > 1.27]$
    - $P = 0.23$  (借助统计软件); 或  $t_{0.05(2),12} = -2.18$ ;
  - 4. 结论: 不拒绝 $H_0$ ;  $-0.040 < \mu_d < 0.152$ 。





## 2.3 配对t检验 (Paired t-test)

- 例子：激素水平和免疫力
  - 数据：抗体产生速率的差异？ $d$  ( $n = 13$ )
- 假设检验
  - 1.  $H_0: \mu_d = 0$ ;  $H_A: \mu_d \neq 0$ ;
  - 2. 检验统计量： $t = \frac{\bar{d} - \mu_{d0}}{\text{SE}_{\bar{d}}} = \frac{0.056 - 0}{0.044} = 1.27$
  - 3. 计算P值：
    - $df = n - 1 = 13 - 1 = 12$ ;
    - $P = \Pr[t_{12} < -1.27] + \Pr[t_{12} > 1.27] = 2 \Pr[t_{12} > 1.27]$
    - $P = 0.23$  (借助统计软件); 或  $t_{0.05(2),12} = -2.18$ ;
  - 4. 结论：不拒绝 $H_0$ ;  $-0.040 < \mu_d < 0.152$ 。
- 我们观察到的平均差异为  $0.056 \ln[\text{mOD/min}]$ , 即在睾酮植入后免疫系统功能略有提高。
- 但我们无法拒绝零假设 (即 睾酮对免疫力没有影响)。
- 置信区间包含一系列广泛的值。
- 基于这个数据集, 我们不拒绝零假设。
- 但希望进一步研究可以更精确地回答这个问题。

## 2.3 配对t检验 (Paired t-test)

- 例子：激素水平和免疫力
  - 数据：抗体产生速率的差异？
- 假设检验
  - 1.  $H_0: \mu_d = 0$ ;  $H_A: \mu_d \neq 0$ ;
  - 2. 检验统计量:  $t = 1.27$ ;
  - 3. 计算P值:  $P = 0.23$ ;
  - 4. 结论: 不拒绝 $H_0$ ;
  - $-0.040 < \mu_d < 0.152$ .

blackbird	beforeImplant	afterImplant	logBeforeImplant	logAfterImplant	diff
1	105	85	4.65	4.44	-0.21
2	50	74	3.91	4.3	0.39
3	136	145	4.91	4.98	0.07
4	90	86	4.5	4.45	-0.05
5	122	148	4.8	5	0.2

```
t.test(blackbird$d)
```

```
##  
##  One Sample t-test  
##  
## data: blackbird$d  
## t = 1.2714, df = 12, p-value = 0.2277  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -0.04007695  0.15238464  
## sample estimates:  
## mean of x  
## 0.05615385
```

## 2.3 配对t检验 (Paired t-test)

- 例子：激素水平和免疫力
  - 数据：抗体产生速率的差异？
- 假设检验
  - 1.  $H_0: \mu_d = 0; H_A: \mu_d \neq 0; \rightarrow \mu_1 - \mu_2$

blackbird	beforeImplant	afterImplant	logBeforeImplant	logAfterImplant	diff
1	105	85	4.65	4.44	-0.21
2	50	74	3.91	4.3	0.39
3	136	145	4.91	4.98	0.07
4	90	86	4.5	4.45	-0.05
5	122	148	4.8	5	0.2

```
t.test(blackbird$logAfterImplant, blackbird$logBeforeImplant, paired = TRUE)
```

```
##  
## Paired t-test  
##  
## data: blackbird$logAfterImplant and blackbird$logBeforeImplant  
## t = 1.2714, df = 12, p-value = 0.2277  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.04007695 0.15238464  
## sample estimates:  
## mean of the differences  
## 0.05615385
```

## 2.3 配对t检验 (Paired t-test)

- 假设检验

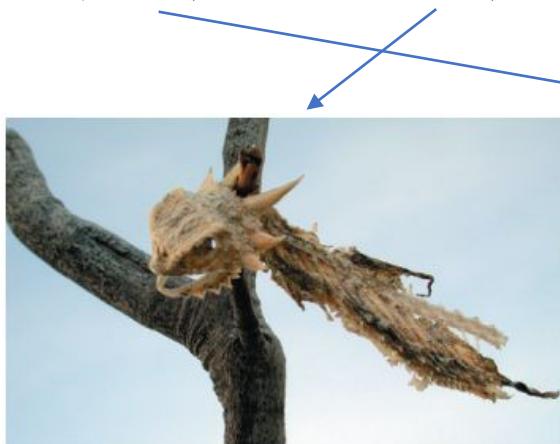
- 1.  $H_0: \mu_d = 0$ ;  $H_A: \mu_d \neq 0$ ; 其中 $\mu_d$ 为抽样单元内配对个体间的差异;
- 2. 检验统计量:  $t = \frac{\bar{d} - \mu_{d0}}{\text{SE}_{\bar{d}}}$
- 3. 计算P值:  $P = \Pr[t_{df} < -|\textcolor{blue}{t}|] + \Pr[t_{df} > |\textcolor{blue}{t}|]$
- 4. 结论: 是否拒绝 $H_0$ ?  $95\% \text{CI}_{\text{lower}} < \mu_d < 95\% \text{CI}_{\text{upper}}$ 是否包含0?

- 前提假设 Assumptions

- 样本为随机样本;
- 两组处理间的配对个体间的差异值呈正态分布;
  - 而对个体测量值的分布不做正态假设;

## 2.4 两个独立样本间的均值比较

- 两个独立样本间的均值比较 Two-sample comparison of means
  - 两个处理组是分别来自两个总体的独立样本;
  - Independent samples from two populations;
  - NOT paired;
- 例子：长刺还是被刺？ EXAMPLE 12.3: Spike or be spiked



© Qin Li @AZ

Left and middle: © Butch Brodie; right: PaulWolf/iStock/Getty Images

角蜥 horned lizard

伯劳 Shrike

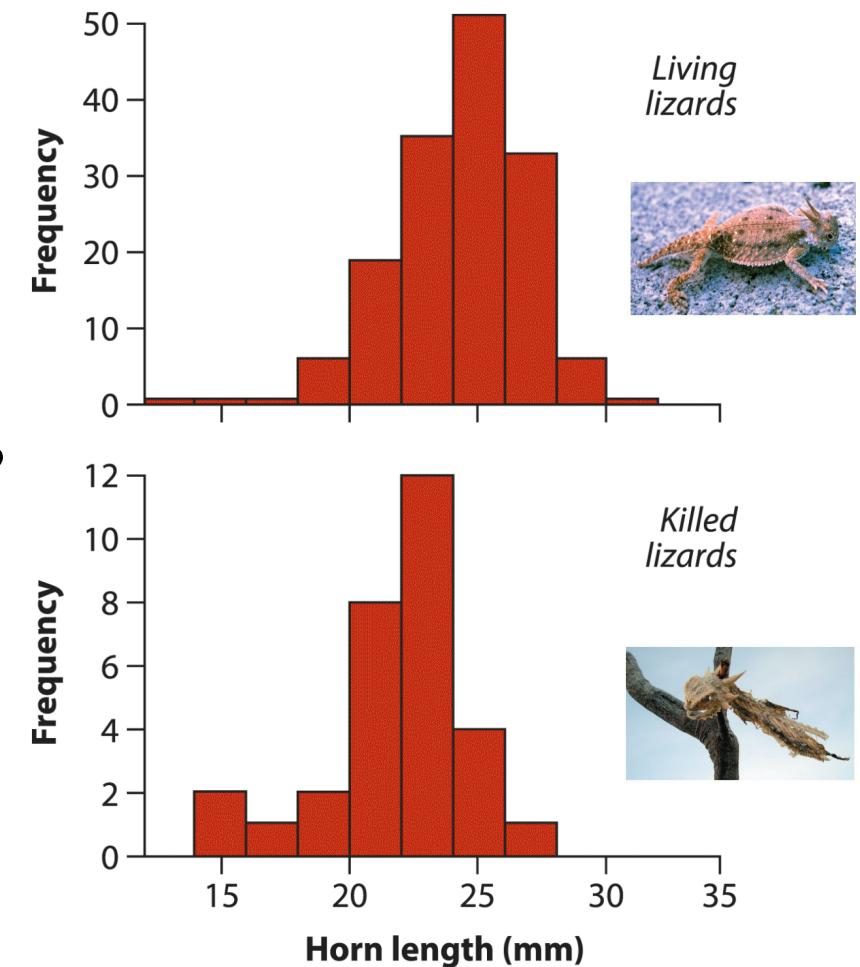


## 2.4 两个独立样本间的均值比较

- 两个独立样本（来自两个总体）
- 例子：长刺还是被刺？
  - 样本：154只健康活着的蜥蜴的角的长度  
vs. 30只被伯劳杀死的蜥蜴的角的长度
  - 问题：活着和死亡蜥蜴的角的平均长度相同吗？
    - 从图和表中可以初步判断吗？

### Summary statistics

Lizard group	Sample mean, $\bar{Y}$	Sample SD, $s$	Sample size, $n$
Living	$\bar{Y}_1 = 24.28$	2.63	154
Killed	$\bar{Y}_2 = 21.99$	2.71	30



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

## 2.4 两个独立样本间的均值比较

$n_1 = 154$

$n_2 = 30$



- 例子：活着和死亡的蜥蜴的角的平均长度相同吗？

$$\mu_1 ? = \mu_2$$

- $\bar{Y}_1 - \bar{Y}_2$  的置信区间的计算

- 若变量在两个总体中都是正态分布的，那么样本均值之差的抽样分布也呈正态分布；那么标准化的均值之差符合Student's  $t$ -distribution。

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

单样本： $SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$

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

配对样本： $SE_{\bar{d}} = \frac{s_d}{\sqrt{n}}$

- $s_p^2$  为两个样本的方差以自由度加权平均（假定两个样本各自的方差一样）；

- 标准化转换： $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{SE_{\bar{Y}_1 - \bar{Y}_2}}$

单样本： $t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$

- 自由度： $df = df_1 + df_2 = n_1 + n_2 - 2$

- 置信区间： $(\bar{Y}_1 - \bar{Y}_2) - t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2} < \mu_1 - \mu_2 < (\bar{Y}_1 - \bar{Y}_2) + t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2}$

## 2.4 两个独立样本间的均值比较

$$n_1 = 154 \quad n_2 = 30$$



- 例子：活着和死亡的蜥蜴的角的平均长度相同吗？

$$\mu_1 ? = \mu_2$$

- $\bar{Y}_1 - \bar{Y}_2$  的置信区间的计算

- 合并方差： $s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} = \frac{153(2.63^2) + 29(2.71^2)}{153 + 29} = 6.98$

- 均值差异的标准误： $SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{6.98 \left( \frac{1}{154} + \frac{1}{30} \right)} = 0.527$

- 自由度： $df = df_1 + df_2 = n_1 + n_2 - 2 = 154 + 30 - 2 = 182$

- 关键值： $t_{\alpha/2, df} = t_{0.05/2, 182} = 1.97$

- 置信区间： $(\bar{Y}_1 - \bar{Y}_2) - t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2} < \mu_1 - \mu_2 < (\bar{Y}_1 - \bar{Y}_2) + t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2}$

$$2.29 - 1.97 \times 0.527 < \mu_1 - \mu_2 < 2.29 + 1.97 \times 0.527$$

$$1.52 < \mu_1 - \mu_2 < 3.33$$

## 2.4 两个独立样本间的均值比较

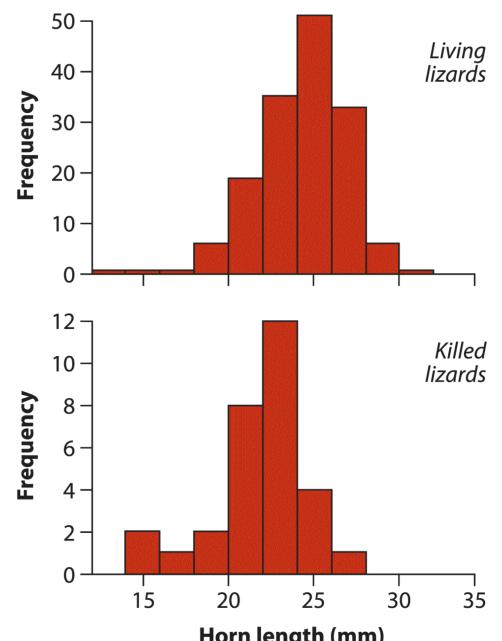
$$n_1 = 154 \quad n_2 = 30$$



- 双样本t检验 Two-sample t-test

- 是比较两个独立样本间数值变量均值的最简单方法。
- 它最常见的用途是检验两个总体均值是否相等的零假设；
- 或者等价地，它用来检验两个均值之差是否为零。
- 检验的假设和统计量
  - 零假设和备择假设： $H_0: \mu_1 = \mu_2 \quad (\mu_1 - \mu_2 = 0)$ ；  
 $H_A: \mu_1 \neq \mu_2 \quad (\mu_1 - \mu_2 \neq 0)$ ；
  - 检验统计量： $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}}$

$$\mu_1 ? = \mu_2$$



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

## 2.4 两个独立样本间的均值比较

$$n_1 = 154 \quad n_2 = 30$$

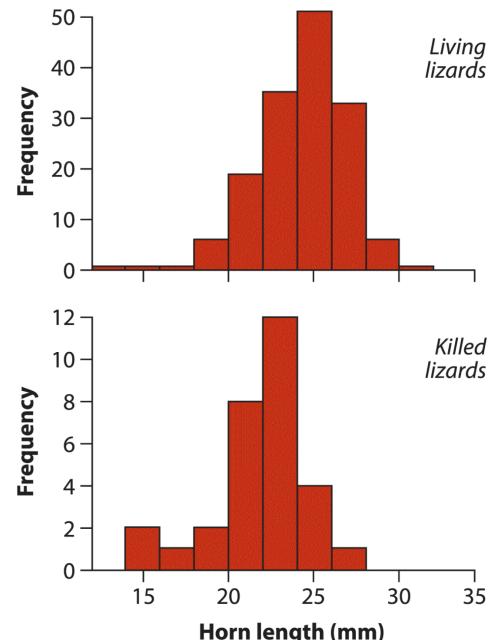


- 双样本t检验 Two-sample *t*-test: 蜥蜴的角

$$\mu_1 ? = \mu_2$$

- 零假设和备择假设

- $H_0: \mu_1 = \mu_2$ , 活着和被伯劳杀死的蜥蜴的角的平均长度没有差异;
- $H_A: \mu_1 \neq \mu_2$ , 活着和被伯劳杀死的蜥蜴的角的平均长度有差异;
- 检验统计量:  $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} = \frac{2.29}{0.527} = 4.35$



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

## 2.4 两个独立样本间的均值比较

$n_1 = 154$

$n_2 = 30$



- 双样本 $t$ 检验 Two-sample  $t$ -test: 蜥蜴的角

$$\mu_1 ? = \mu_2$$

- 零假设和备择假设

- $H_0: \mu_1 = \mu_2$ , 活着和被伯劳杀死的蜥蜴的角的平均长度没有差异;

- $H_A: \mu_1 \neq \mu_2$ , 活着和被伯劳杀死的蜥蜴的角的平均长度有差异;

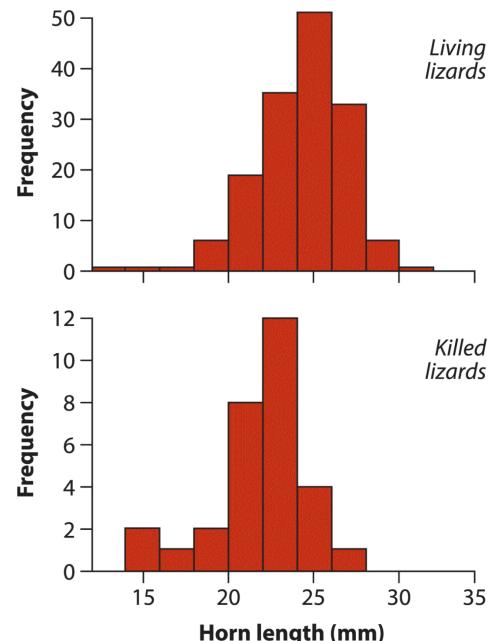
- 检验统计量:  $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\text{SE}_{\bar{Y}_1 - \bar{Y}_2}} = \frac{2.29}{0.527} = 4.35$

- P值计算:  $P = 2 \Pr[t > 4.35] = 0.000023 < 0.05$

(关键值:  $t_{\alpha(2), df} = t_{0.05(2), 182} = 1.97$ )

- 结论: 拒绝 $H_0$ ; 基于这些数据, 我们认为被伯劳捕食的蜥蜴的角长与活着的蜥蜴的角长存在差异。  $\mu_1 > \mu_2$

- 观察性研究: 可能伯劳会回避或无法捕捉那些角比较长的蜥蜴;
- 为了推断因果关系, 我们需要进行一项操纵蜥蜴角长的受控实验。



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

## 2.4 两个独立样本间的均值比较

$n_1 = 154$



$n_2 = 30$



- 双样本t检验 Two-sample  $t$ -test: 蜥蜴的角

$$\mu_1 ? = \mu_2$$

- 零假设和备择假设

检验统计量:  $t = 4.35$

- $H_0: \mu_1 = \mu_2$ ,  $H_A: \mu_1 \neq \mu_2$ ;  $P < 0.05$ ; 结论: 拒绝  $H_0$ ;  $\mu_1 > \mu_2$

```
t.test(squamosalHornLength ~ Survival, data = lizard, var.equal = TRUE)
```

```
##  
## Two Sample t-test  
##  
## data: squamosalHornLength by Survival  
## t = 4.3494, df = 182, p-value = 2.27e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.253602 3.335402  
## sample estimates:  
## mean in group living mean in group killed  
## 24.28117 21.98667
```

## 2.4 两个独立样本间的均值比较

- 双样本t检验的前提假设 Assumptions
  - 每个样本都是从其总体中随机抽取的；
  - 数值变量在每个总体中都呈正态分布；
  - 数值变量的标准差和方差在两个总体中相同；
- 双样本t检验的稳健性 (robust, 鲁棒性)
  - 即使 $s_1$ 和 $s_2$ 差三倍以上，适度样本量时（如 $n_1 \approx n_2 \geq 30$ ）依然适用；
  - 稍微偏离正态分布依然适用；
  - 当两组的样本量非常不同，或者样本量小于30且标准差有差异，则应该使用 Welch's 近似t检验 (the Welch's approximate  $t$ -test);

## 2.4 两个独立样本间的均值比较

- 双样本的Welch's近似t检验 (the Welch's approximate  $t$ -test)

- 适用于样本量小且两个样本的方差不一样时;
- 相关统计量的两个修正 (modification)

- $SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)}$

- $df = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left( \frac{s_1^2}{n_1} \right)^2 + \left( \frac{s_2^2}{n_2} \right)^2}$   
$$[ \frac{\left( \frac{s_1^2}{n_1} \right)^2}{n_1-1} + \frac{\left( \frac{s_2^2}{n_2} \right)^2}{n_2-1} ]$$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$$

双样本 $t$ 检验

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$df = df_1 + df_2 = n_1 + n_2 - 2$$

```
t.test(proportionSurvived ~ troutTreatment, data = chinook, var.equal = FALSE)
```

```
##  
## Welch Two Sample t-test  
##
```

## 2.5 两个独立样本间的标准差/方差比较

- 检验来自不同总体的变量的变异性是否存在差异（第 12.7 节）
  - 比较:  $\sigma_1^2 ?= \sigma_2^2$
  - (1) F 检验 (比值):  $F = s_1^2/s_2^2 ?= 1$

### *F*-test of equal variances

Warning: Remember that the *F*-test is not robust to departures from assumption of normality.

R doesn't put the larger variance on top, as in our Quick Formula Summary, but the result is the same.

```
var.test(proportionSurvived ~ troutTreatment, data = chinook)
```

```
##  
## F test to compare two variances  
##  
## data: proportionSurvived by troutTreatment  
## F = 0.082202, num df = 5, denom df = 5, p-value = 0.01589  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.01150267 0.58745010  
## sample estimates:  
## ratio of variances  
## 0.08220245
```

## 2.5 两个独立样本间的标准差/方差比较

- 检验来自不同总体的变量的变异性是否存在差异（第 12.7 节）
  - 比较:  $\sigma_1^2 \stackrel{?}{=} \sigma_2^2$
  - (1) F 检验 (比值):  $F = s_1^2/s_2^2 \stackrel{?}{=} 1$
  - (2) Levene 检验 (方差齐性检验)
    - $W$ : 量化数据点离均值的绝对偏差, 服从  $F$ -分布;

### Levene's test

This test of equal variances is more robust than the F test. The `leveneTest()` function is in the `car` package.

```
leveneTest(chinook$proportionSurvived, group = chinook$troutTreatment,  
           center = mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)  
##          Df F value    Pr(>F)  
## group     1 10.315 0.009306 **  
##          10  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 2.5 两个独立样本间的标准差/方差比较

- 检验来自不同总体的变量的变异性是否存在差异（第 12.7 节）
  - 比较：  $\sigma_1^2 \stackrel{?}{=} \sigma_2^2$
  - (1) F 检验（比值）：  $F = s_1^2/s_2^2 \stackrel{?}{=} 1$
  - (2) Levene 检验（方差齐性检验）
- 快速公式摘要（第 12.9 节）提供了公式，大多数计算机统计程序都可以执行这些检验。
- 但要注意：F 检验对偏离变量在总体中呈正态分布的假设非常敏感。因此，它并不适用于大多数数据，因为真实数据往往显示出一定的正态偏离。
- Levene 检验是对偏离正态总体的假设更为稳健，但它的功效较低。

### 3. 选择检验方法及违背前提时的处理方法

- 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

### 3.1 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)

### 3.1 假设检验前的判断及方法选择

一个变量

一组数据

( 频数 / 比例 )

( 均值 / 差异 )

Data type	Goal	Test
Categorical 类型数据	Use frequency data to test whether a population proportion equals a null hypothesized value	Binomial test (7) $\chi^2$ Goodness-of-fit test with two categories (use if sample size is too large for the binomial test) (8)
Numerical 数值数据	Use frequency data to test the fit of a specific population model	$\chi^2$ Goodness-of-fit test (8)
	Test whether the mean equals a null hypothesized value when data are approximately normal (possibly only after a transformation) (13)	One-sample <i>t</i> -test (11)
	Test whether the median equals a null hypothesized value when data are not normal (even after transformation)	Sign test (13)
	Use frequency data to test the fit of a discrete probability distribution	$\chi^2$ Goodness-of-fit test (8)
	Use data to test the fit of the normal distribution	Shapiro-Wilk test (13)

### 3.1 假设检验前的判断及方法选择

#### 两个变量的相关性

解释变量

Type of explanatory variable

Type of response variable	Categorical	Categorical	Numerical
R	Categorical	Contingency analysis (9)	Logistic regression (17)
响应变量		$t$ -tests, ANOVA, Mann–Whitney $U$ -test, etc.	Linear and nonlinear regression (17)
	Numerical	[See Table 3 for more details.]	Linear correlation (16) Spearman's rank correlation (when data are not bivariate normal) (16)

### 3.1 假设检验前的判断及方法选择

一个变量：两（多）组数据间的差异

Number of treatments	(满足正态分布的假设)	(不满足正态分布的假设)
	Tests assuming normal distribution	Tests not assuming normal distributions
Two treatments (independent samples)	Welch's <i>t</i> -test (12)  Two-sample <i>t</i> -test (use when variance is equal in the two groups) (12)	Mann–Whitney <i>U</i> -test (13)
Two treatments (paired data)	Paired <i>t</i> -test (12)	Sign test (13)
More than two treatments	ANOVA (15)	Kruskal–Wallis test (15)

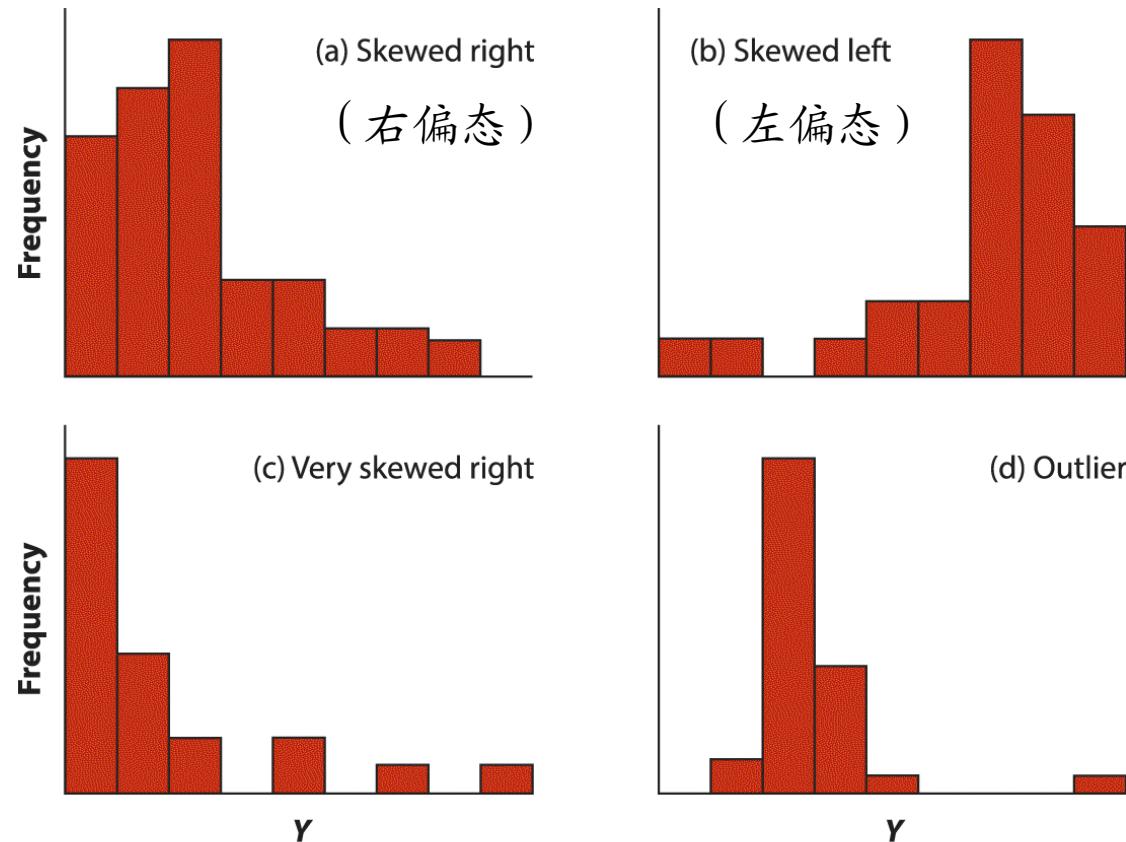
## 3.2 对正态性的检验

- 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

## 3.2 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

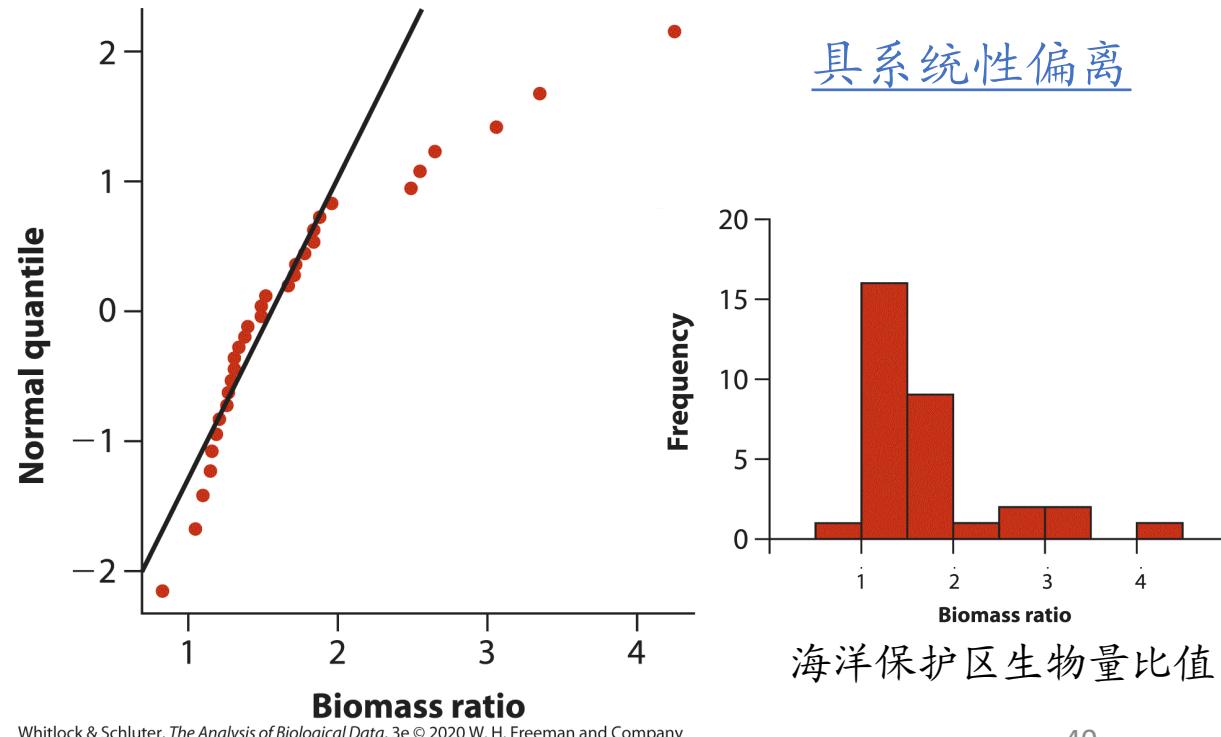
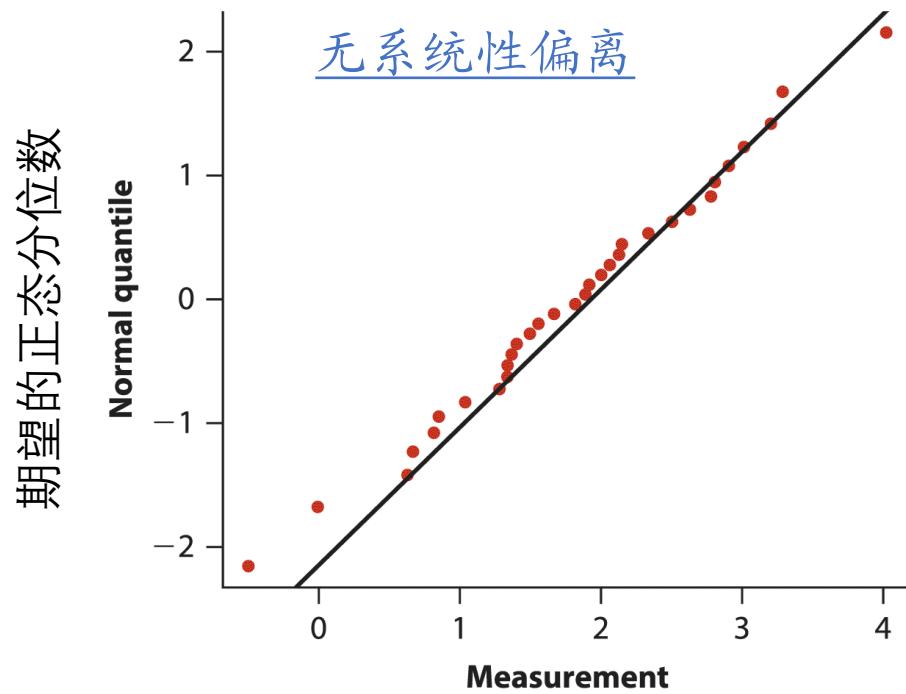


## 3.2 对正态性的检验

- 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()

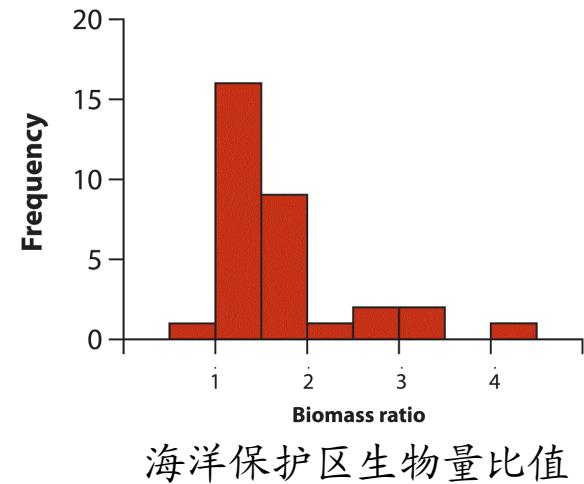


## 3.2 对正态性的检验

- 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。



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

### 3.3 违背前前提假设时的处理方法

- 违背前前提假设
  - 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. 忽略对前前提假设的违背 Ignore the violations of assumptions
    - 检验方法具有一定稳健性，对没有严格满足前前提假设情况也适用；
  - 2. 对数据进行转换 Transform the data
    - 转换后的数据满足前前提假设；
  - 3. 使用非参方法 Use a nonparametric method
    - 非参方法不要求正态性假设；
  - 4. 使用置换检验 Use a permutation test
    - 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；

### 3.3-1 忽略对前提假设的违背

- 1. 忽略对前提假设的违背 **Ignore the violations of assumptions**
  - 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 **Violations of normality**
    - 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（Box 和 Andersen 1955）；
    - 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）。

### 3.3-1 忽略对前提假设的违背

- 1. 忽略对前提假设的违背 Ignore the violations of assumptions

- 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 Violations of normality

- 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：

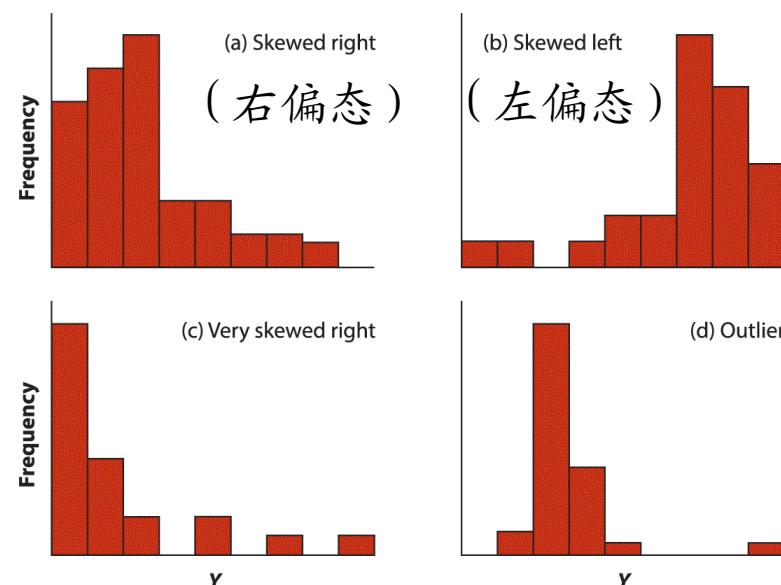
- 两个分布同(a): 至少  $n = 30$ ;

- (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；



### 3.3-1 忽略对前提假设的违背

- 1. 忽略对前提假设的违背 Ignore the violations of assumptions
  - 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) 对相等方差的违背 Unequal standard deviations
    - 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)。

## 3.3-2 对数据进行转换

- 2. 对数据进行转换 Data transformations

- 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

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

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

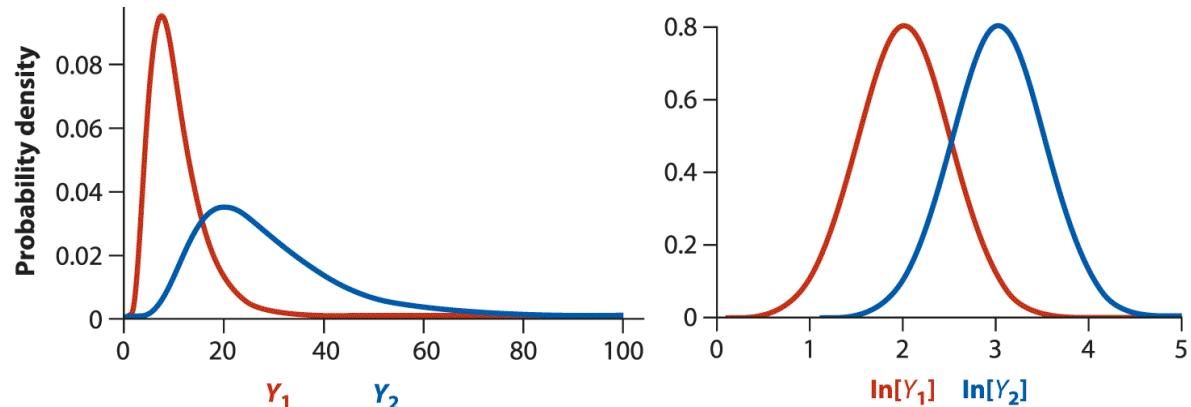
### 3.3-2 对数据进行转换

- 2. 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

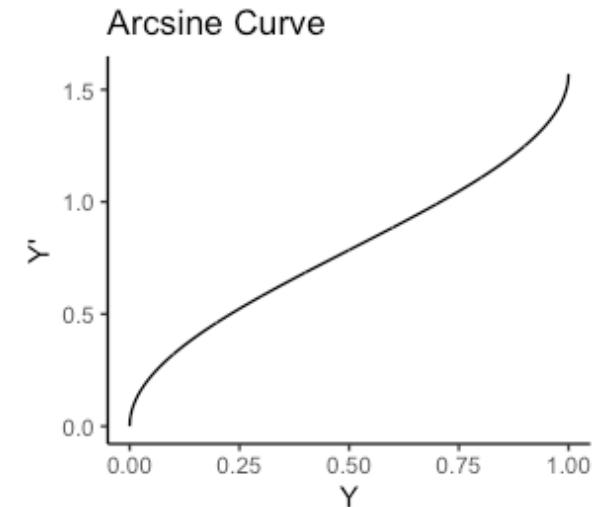


### 3.3-2 对数据进行转换

- 2. 对数据进行转换  $Y \rightarrow Y'$

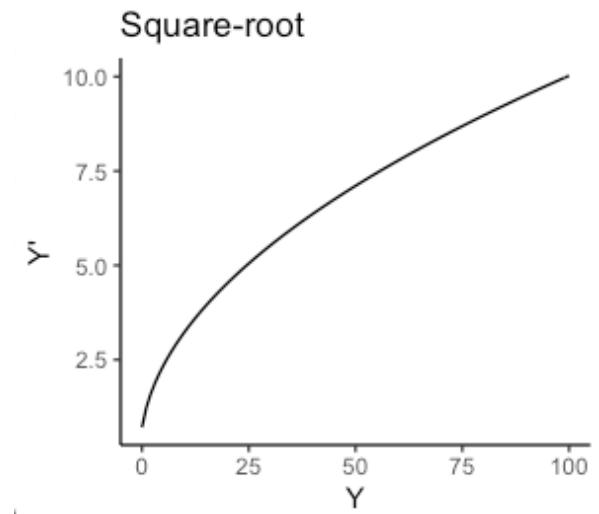
- (2) 反正弦转换:  $p' = \arcsin[\sqrt{p}]$

- 常用场景: 比例数据  $p \subset [0,1]$ ;
  - 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。



## 4. 小结 – 双样本 $t$ 检验及正态性检验

- 双样本  $t$  检验

- (1) 配对  $t$  检验: 对两个总体的均值差异 ( $\mu_d$ ) 进行检验;
  - 均值差异的置信区间和配对  $t$  检验都假设配对是随机样本;
  - 配对间的差异呈正态分布;
  - 检验对正态性的轻微偏差是稳健的;
- $H_0: \mu_d = 0; H_A: \mu_d \neq 0;$
- 检验统计量为:  $t = \frac{\bar{d} - \mu_{d0}}{\text{SE}_{\bar{d}}};$
- 置信区间:  $\bar{d} - t_{\alpha(2), df} \text{SE}_{\bar{d}} < \mu_d < \bar{d} + t_{\alpha(2), df} \text{SE}_{\bar{d}}$

## 4. 小结 – 双样本 $t$ 检验及正态性检验

- 双样本  $t$  检验

- (2) 独立双样本  $t$  检验: 对两个总体的均值进行检验 ( $\mu_1 ?= \mu_2$ ) ;
  - 均值差异的置信区间和双样本  $t$  检验都假设两个总体的变量在正态分布下具有相同的方差; 因此合并样本方差重新进行估计 (否则 Welch's  $t$  test) ;
  - 检验对正态性的轻微偏差是稳健的;

- 合并方差:  $s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$ ; 均值差异的标准误:  $SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$

- $H_0: \mu_1 = \mu_2$ ;  $H_A: \mu_1 \neq \mu_2$ ;
- 检验统计量为:  $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{SE_{\bar{Y}_1 - \bar{Y}_2}}$ ;
- 置信区间:  $(\bar{Y}_1 - \bar{Y}_2) - t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2} < \mu_1 - \mu_2 < (\bar{Y}_1 - \bar{Y}_2) + t_{\alpha/2, df} SE_{\bar{Y}_1 - \bar{Y}_2}$

## 4. 小结 – 双样本 $t$ 检验及正态性检验

- 正态性检验

- (1) 图像法: 正态分布 vs 偏态分布

- 频数的直方图

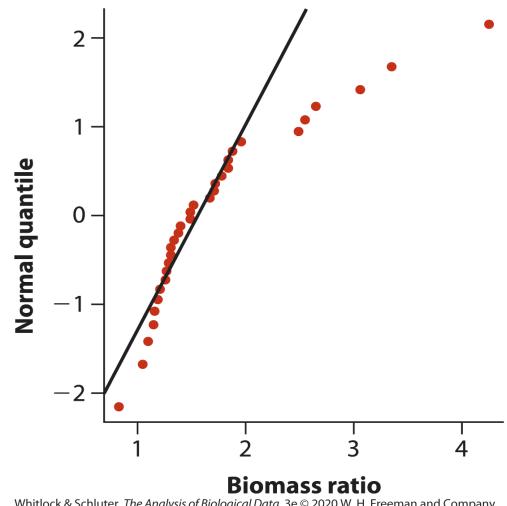
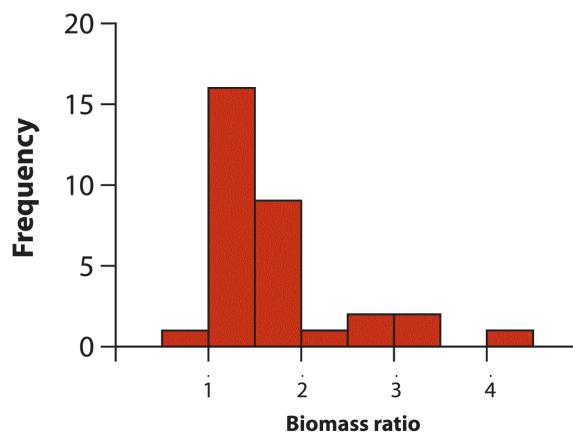
- Q-Q Plot

- (2) Shapiro-Wilk 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；

- 不满足正态性的常用方法: 数据转化



# Next Lecture – Ch13 & Ch14

- 四个常见的方案来处理对前提假设的违背
  - 1. 忽略对前提假设的违背 Ignore the violations of assumptions
    - 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 Transform the data
    - 转换后的数据满足前提假设；
  - 3. 使用非参方法 Use a nonparametric method
    - 非参方法不要求正态性假设；
  - 4. 使用置换检验 Use a permutation test
    - 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；
- 试验设计