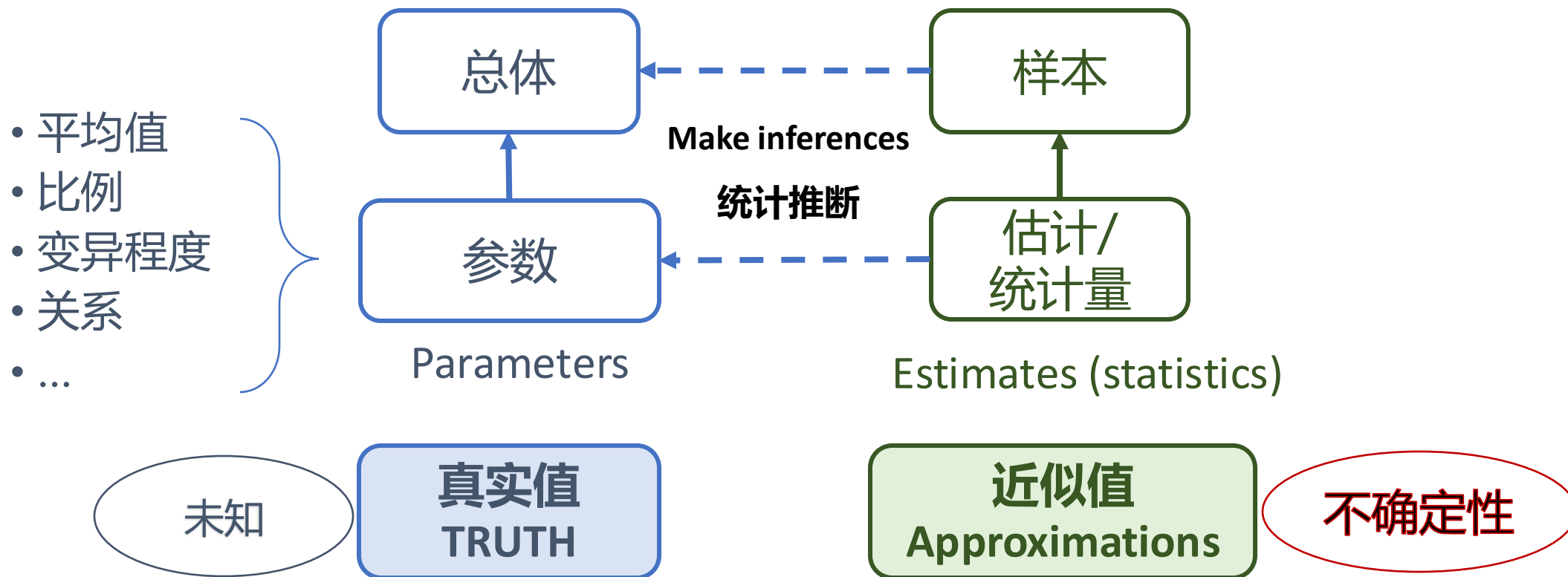


# Lecture 5 – 概率 Probability

- 内容大纲
  - 回顾 L04
  - 概率及其分布
    - 随机试验
    - 感兴趣的事件
  - 总结
  - 课堂讨论
  - R Lab

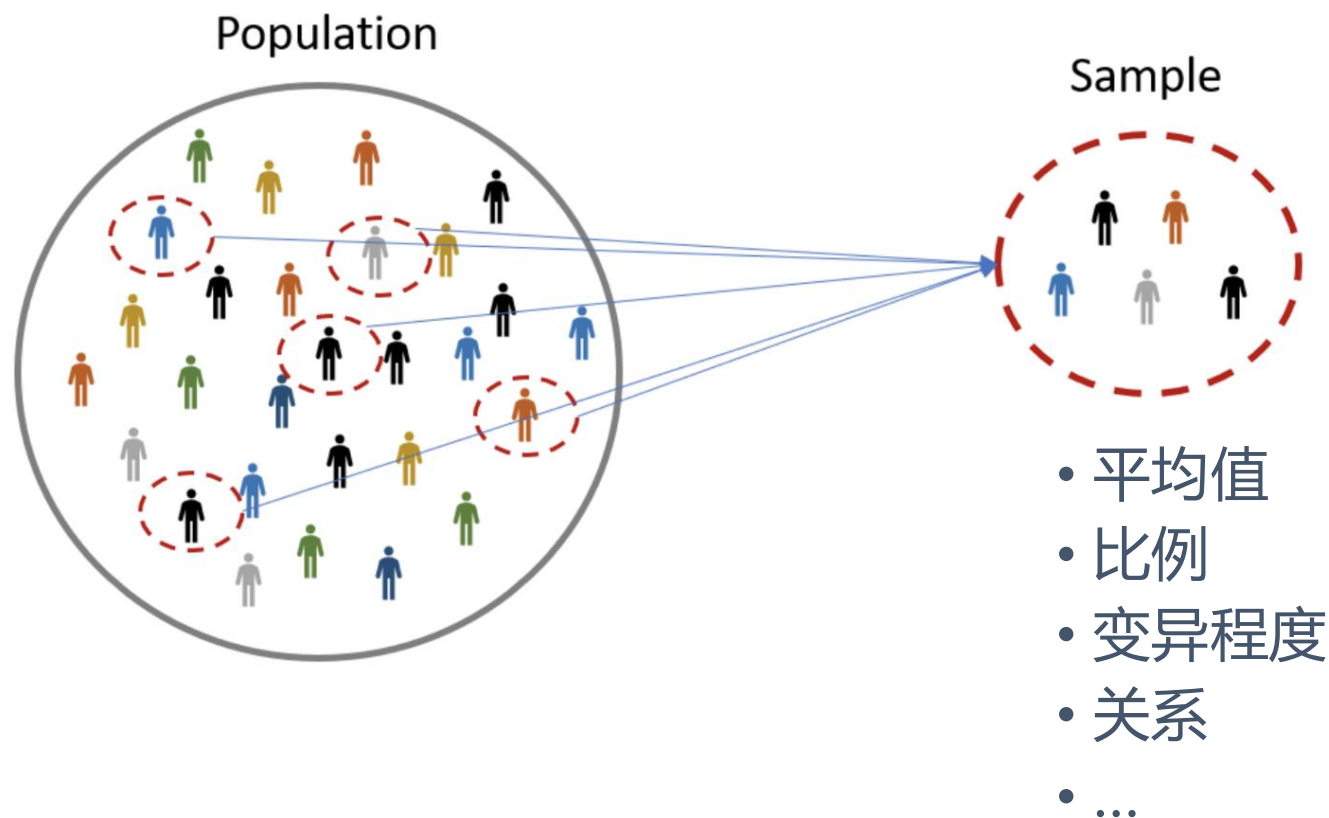
# 1. 回顾 L04

- 统计学的目的是使用样本对总体进行推断，并提供推断的准确性。



# 1. 回顾 L04

- 总体
- 样本
  - 通过随机抽样过程构建
  - 会受到偶然性的影响
- 估计值及其不确定性
  - “面对偶然性，我们在多大程度上能相信一个估计值？”

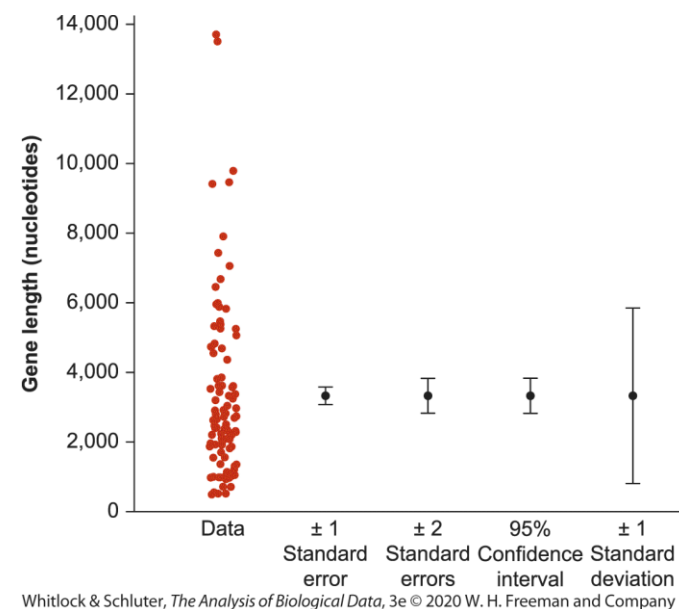


# 1. 回顾 L04

- 估计值的不确定性的量化/估计
  - 估计值都有一个抽样分布 (多次抽样的估计值的分布)
  - 估计值的标准误就是其抽样分布的标准差
    - 通常用一次抽样的 $s$ 来代替总体的 $\sigma$

$$\frac{\sigma}{\sqrt{n}} = \sigma_{\bar{Y}} \approx SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

- 置信区间是可能包含目标参数数值的一个范围
  - 通常使用95% CI
- 通常在图中添加误差线 (error bars) 以说明标准误或置信区间





## 2. Probability 概率

- 随机播放 1000首歌（有一首最爱）
  - 播放的第一首歌就是你最喜欢的歌的概率是？
- 抛一次硬币（一面图案一面数字）
  - 数字朝上的概率是
- 掷骰 (tóu) 子（掷多次）
  - 数字相加为 $x$ 的概率？



## 2.1 一次事件的概率 The probability of an event

- 随机试验 A random trail
  - 随机试验是一种过程或实验它有两种或两种以上可能的结果，而这些结果的出现是无法准确预测的。
  - 随机试验的每次重复只能观察到一种结果
- 随机事件 A random event
  - 随机试验的所有可能结果列表
    - 掷一个六面体的骰子
    - 掷一对六面体的骰子



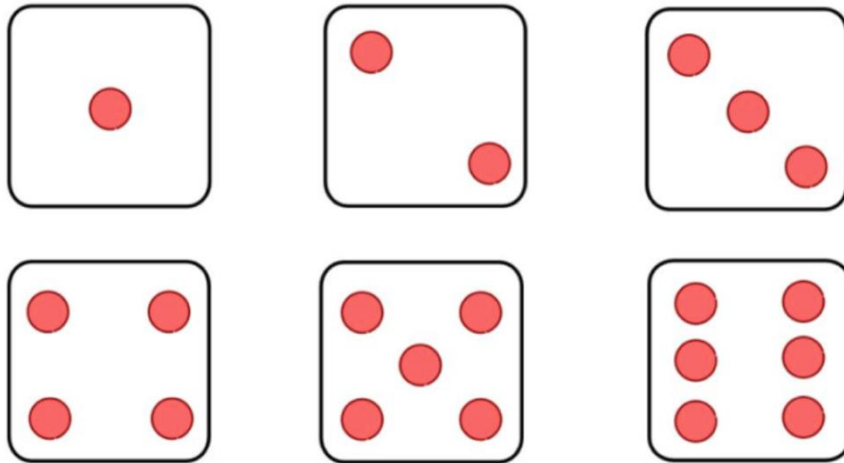
掷骰子 roll dice



## 2.1 一次事件的概率

- 随机事件

- 需要定义感兴趣的事件 (the event of interest)
- 事件是一次随机试验所有可能结果的任何潜在子集 (all possible outcomes)



掷骰子  
roll a six-side die

- 数字是5
- 数字是偶数
  - 包括?
- 数字比3大
  - 包括?
- 其它事件?

## 2.1 一次事件的概率

- 随机事件
  - 需要定义感兴趣的事件 (the event of interest)
  - 比如，骰子出现数字 “3”
- 概率 The probability
  - 我们一般是基于事件来定义概率
  - 一次事件的概率是指如果我们在相同的条件下反复进行随机试验，该事件发生次数的比例。
  - 概率介于 0 和 1 之间





## 2.1 一次事件的概率

- 概率

- 一次事件的概率是指如果我们在相同的条件下反复进行随机试验，该事件发生次数的比例。

- $\text{Pr}[A]$  means “the probability of event A.”

- $\text{Pr}[\text{rolling a four}] = 1/6$
    - $\text{Pr}[\text{rolling an even number}] = ?$
    - $\text{Pr}[\text{rolling a number} > 3] = ?$

- 数字是5
    - 数字是偶数
      - 包括?
    - 数字比3大
      - 包括?
    - 其它?

## 2.1 一次事件的概率

- 概率

- 一次事件的概率是指如果我们在相同的条件下反复进行随机试验，该事件发生次数的比例。
- $\Pr[A]$  means “the probability of event A.”
- 概率介于 0 和 1 之间  $[0, 1]$ 
  - 如果一个事件从不会发生:  $\Pr[A] = 0$
  - 如果一个事件总是发生:  $\Pr[A] = 1$

## 2.1 一次事件的概率

- 概率

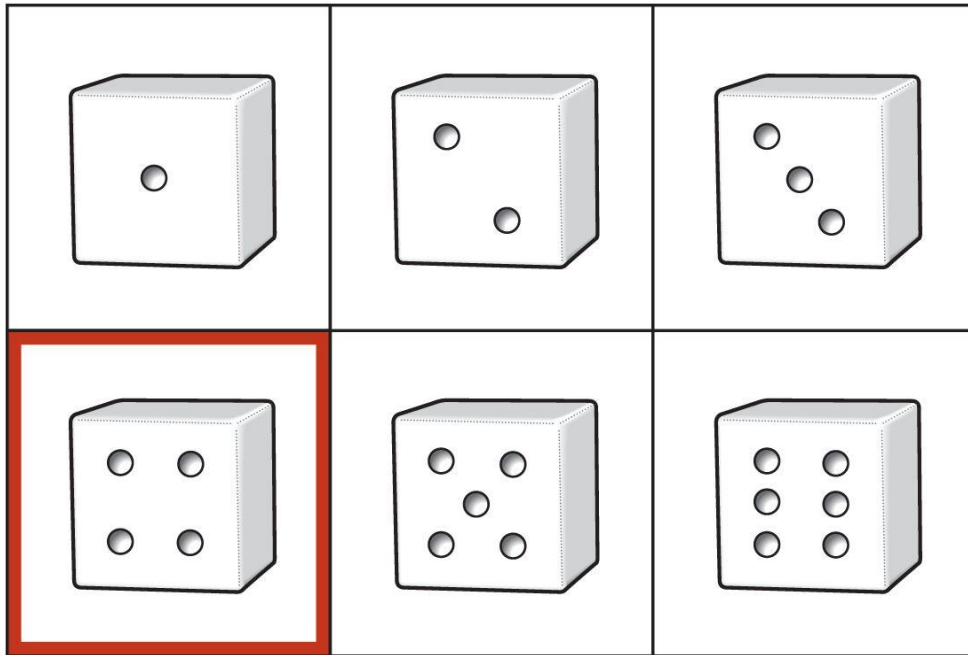
- 一次事件的概率是指如果我们在相同的条件下反复进行随机试验，该事件发生次数的比例。  $\text{Pr}[A]$
- 概率介于 0 和 1 之间  $[0, 1]$
- 抛硬币和掷骰子不是生物过程，但它们与生物学的相关性很高，因为它们模仿了取样过程 (the process of sampling)。
  - 随机抽样100个新生儿并计算其中男女的数量，就好比掷100次硬币并计算其中出现朝上为数字的数量: 2 possible outcomes





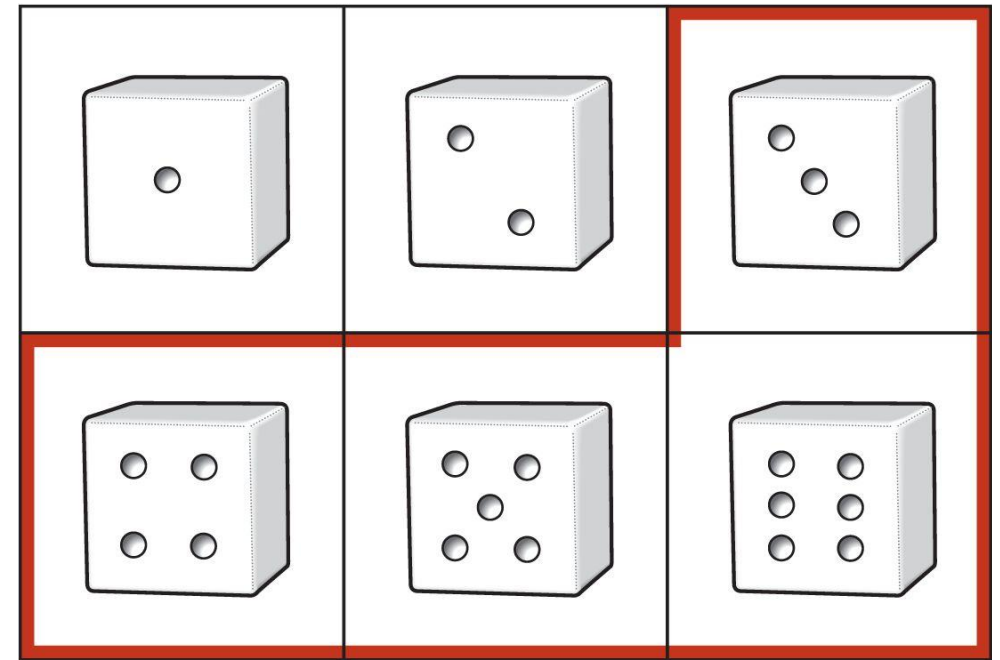
## 2.2 维恩图 Venn diagrams

- 用集合的方式展示可能事件的概率：总面积 = 1
  - 部分面积 = 某一事件概率



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

$\Pr[\text{rolling a four}] = ?$

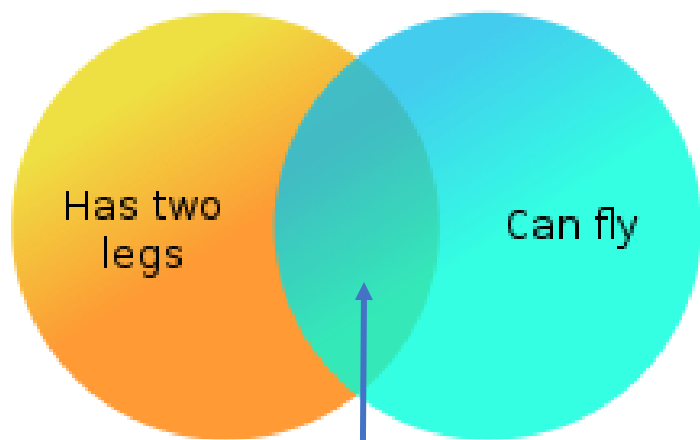


Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

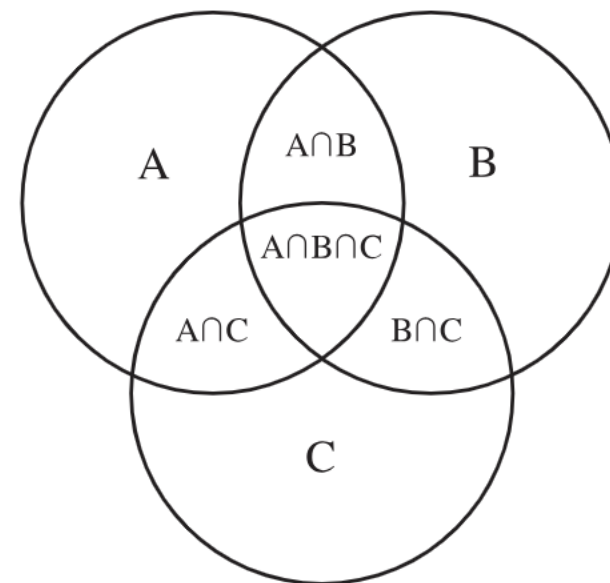
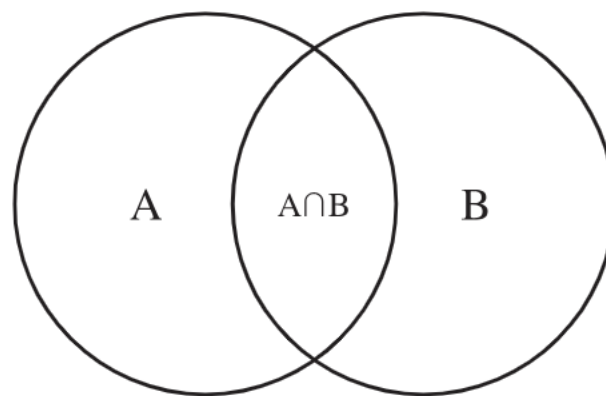
$\Pr[\text{rolling a number} > 2] = ?$

## 2.2 维恩图

- 用集合的方式展示可能事件的概率：总面积 = 1
  - 部分面积 = 某一事件概率

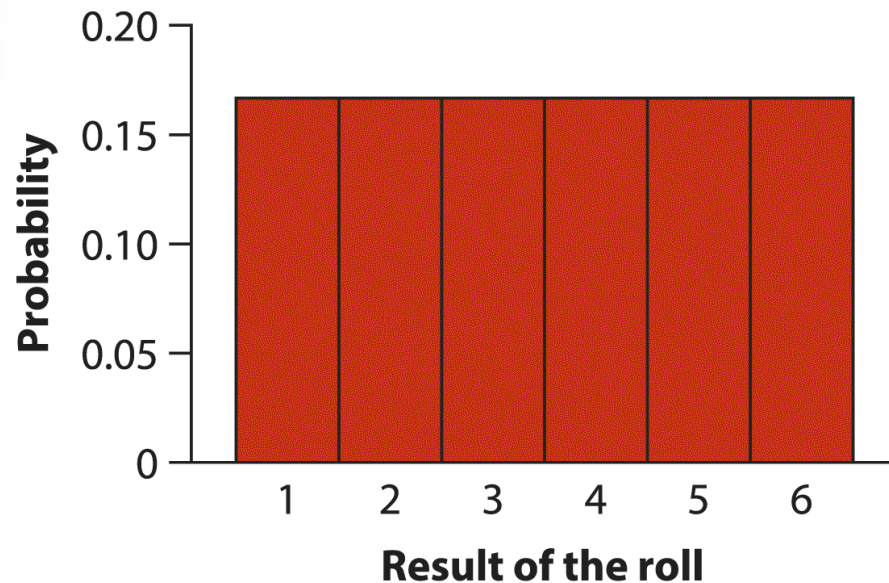


$\text{Pr}[2 \text{ legs \& can fly}] = ?$

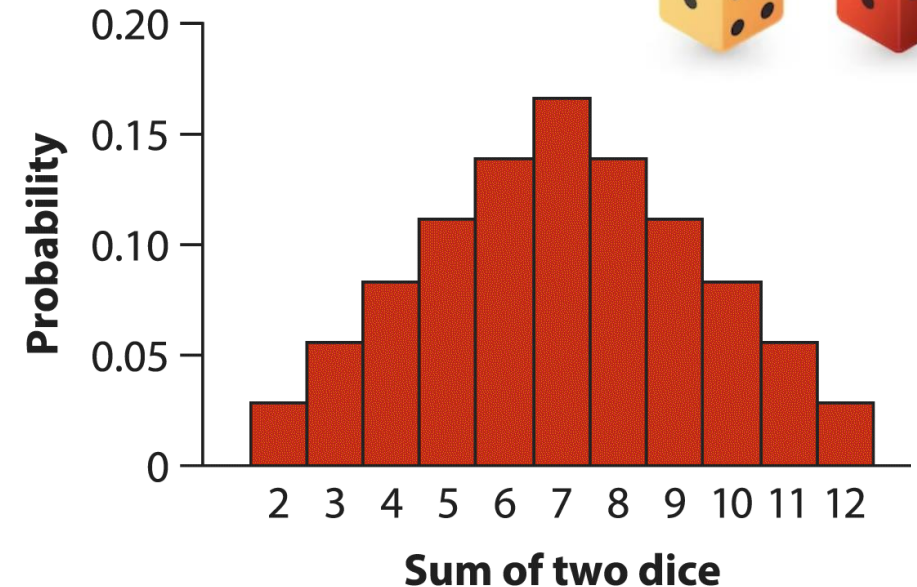


## 2.3 概率分布 Probability distribution

- 离散概率分布 Discrete probability distributions
  - 分类变量/离散数值变量



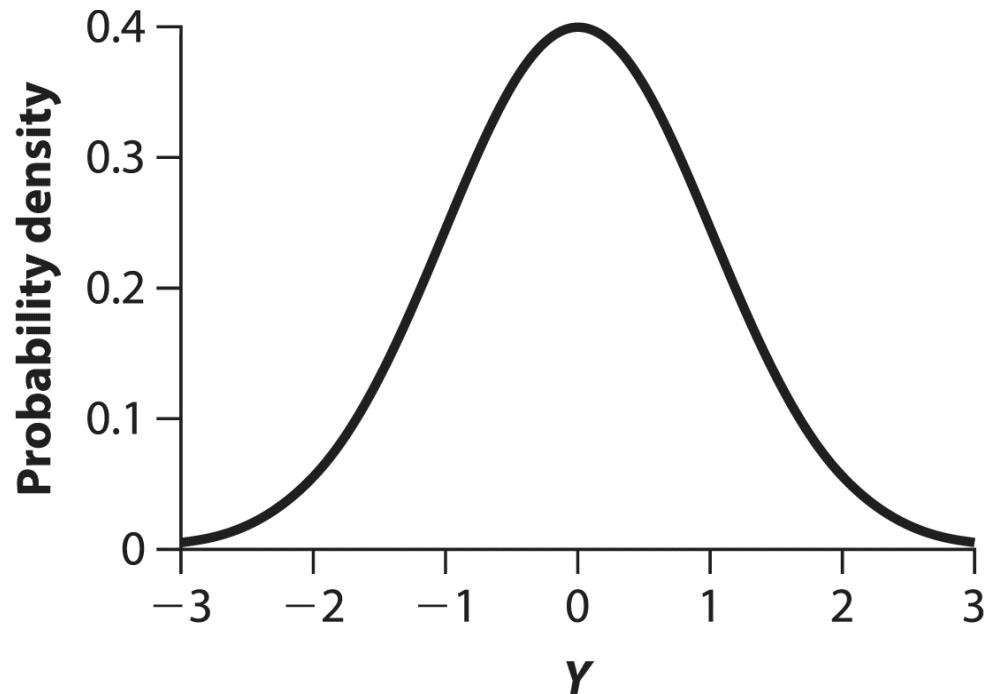
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company



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

## 2.3 概率分布

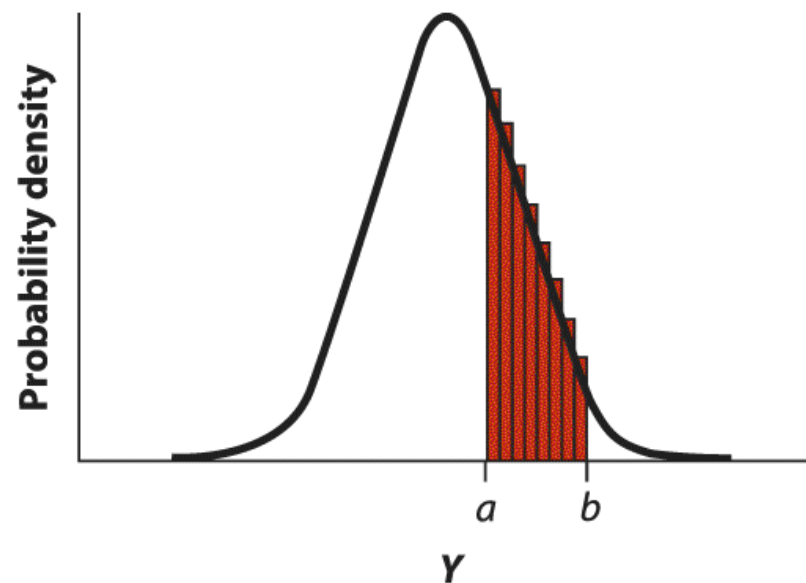
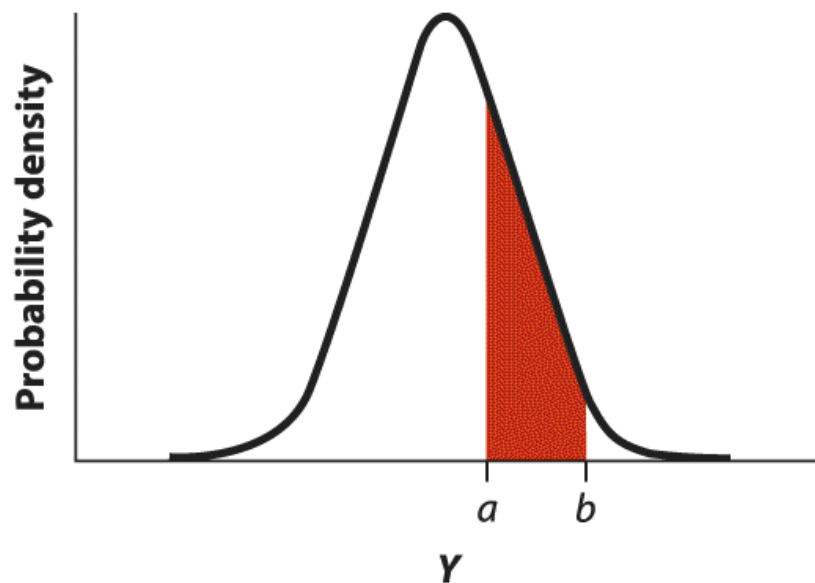
- 连续概率分布 Continuous probability distributions
  - 连续数值变量
  - Y-axis: 概率密度 (曲线高度/probability density)



## 2.3 概率分布

- 连续概率分布

- 连续数值变量
- Y-axis: 概率密度 (曲线高度/probability density)
- Y 值在一定范围内的概率:  $\Pr[a < Y < b]$

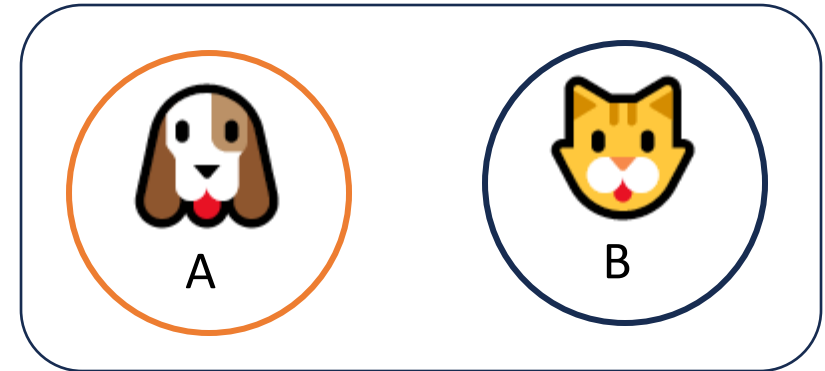
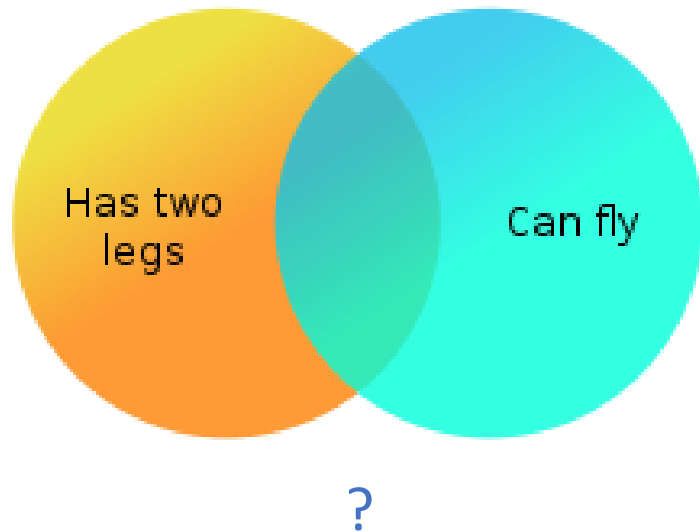




## 2.4 互斥事件 Mutually exclusive events

- 互斥事件

- 事件A和事件B不能同时发生:  $\Pr[A \text{ and } B]=0$



## 2.4 互斥事件

- 事件A和事件B不能同时发生:  $\Pr[A \text{ and } B]=0$

- 事件A或事件B发生的概率?

- 互斥事件的概率加法公式:

- $\Pr[A] + \Pr[B]$

- 血型是O的概率?

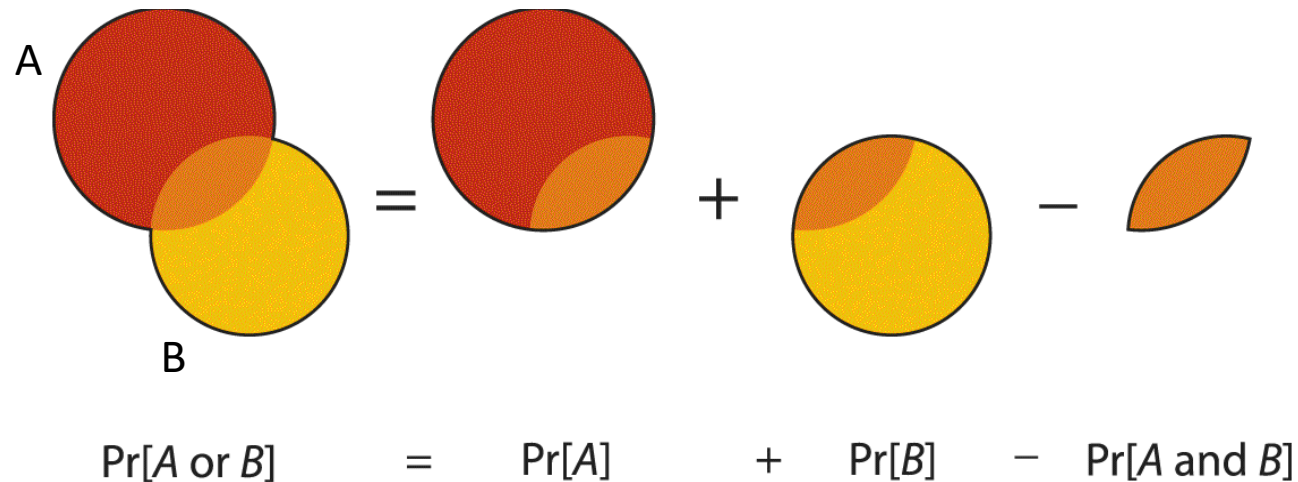
$$\Pr[\text{O+}] + \Pr[\text{O-}] = 0.374 + 0.066$$

- 血型是Rh+的概率?

0.374 O+	0.357 A+	0.085 B+	0.034 AB+
0.066 O-	0.063 A-	0.015 B-	0.006 AB-

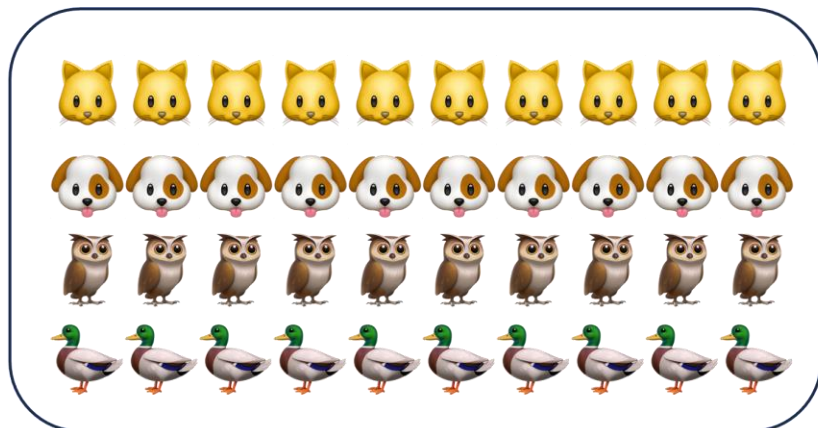
## 2.4 互斥事件

- 事件A和事件B不能同时发生:  $\Pr[A \text{ and } B]=0$
- 互斥事件的概率加法公式:  $\Pr[A] + \Pr[B]$
- 扩展的一般性加法公式:  $\Pr[A] + \Pr[B] - \Pr[A \text{ and } B]$ 
  - 当A和B不一定是互斥事件时



## 2.5 独立事件 Independent events

- 事件A和事件B的发生不依赖于彼此
- 独立事件的概率乘法表公式:  $\Pr[A \& B] = \Pr[A] \times \Pr[B]$ 
  - 抽两次小动物



$$\Pr[\text{🐱} \& \text{🦉}] = \Pr[\text{🐱}] \times \Pr[\text{🦉}] = \frac{1}{4} \times \frac{1}{4}$$



Animal 1  
or  
Animal 2

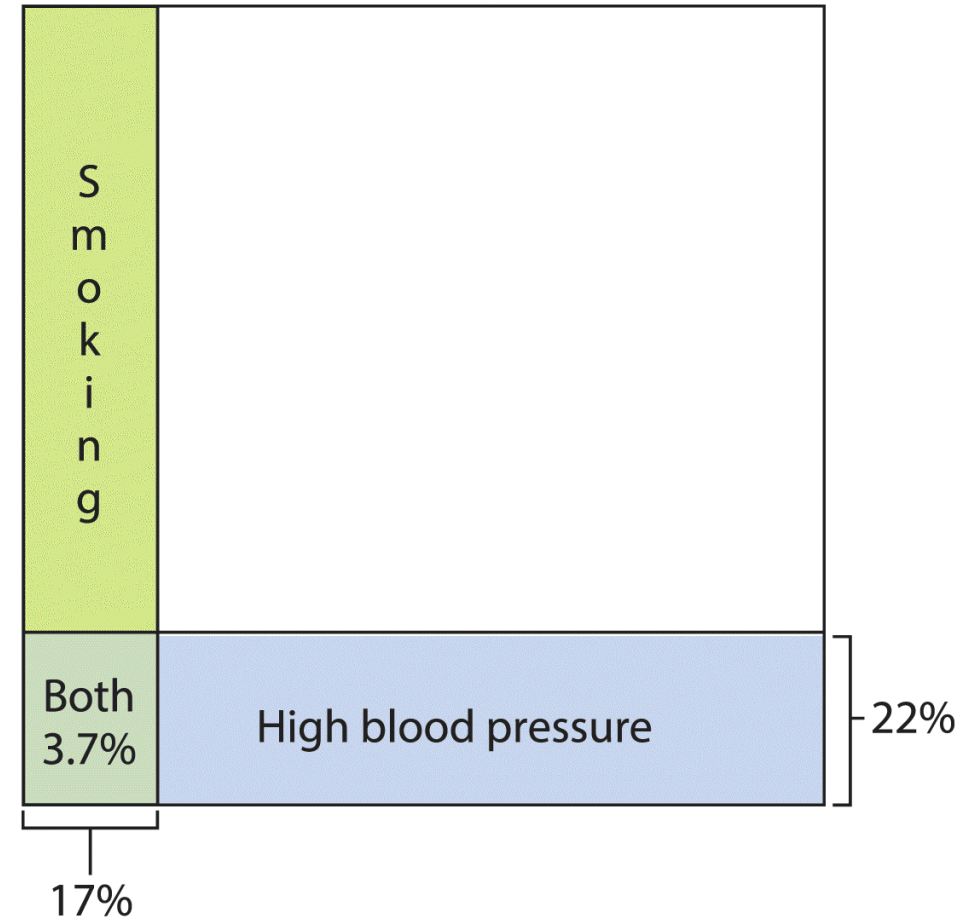


Animal 2  
or  
Animal 1



## 2.5 概率的 “and” versus “or”

- 互斥事件 – 加法公式
  - $\Pr[A \text{ or } B] = \Pr[A] + \Pr[B]$
- 独立事件 – 乘法公式
  - $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B]$
- Q: 抽烟或高血压的概率?



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

## 2.6 概率树 Probability trees

- 用于计算多次随机试验所产生的事件组合的概率

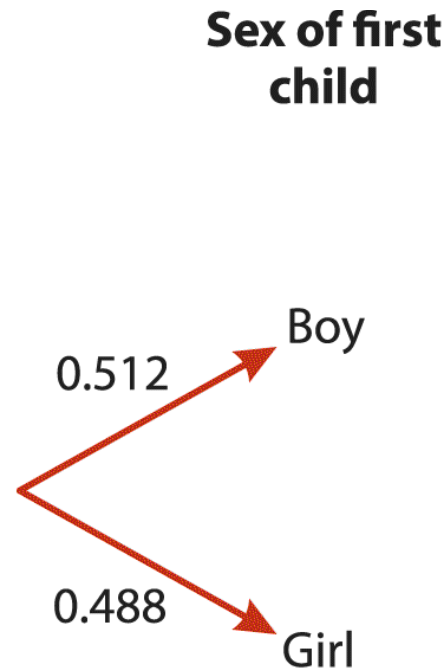
- 二胎家庭的子女

- 性别

- 男女概率不一

- 顺序

- 女女
    - 男男
    - 男女
    - 女男

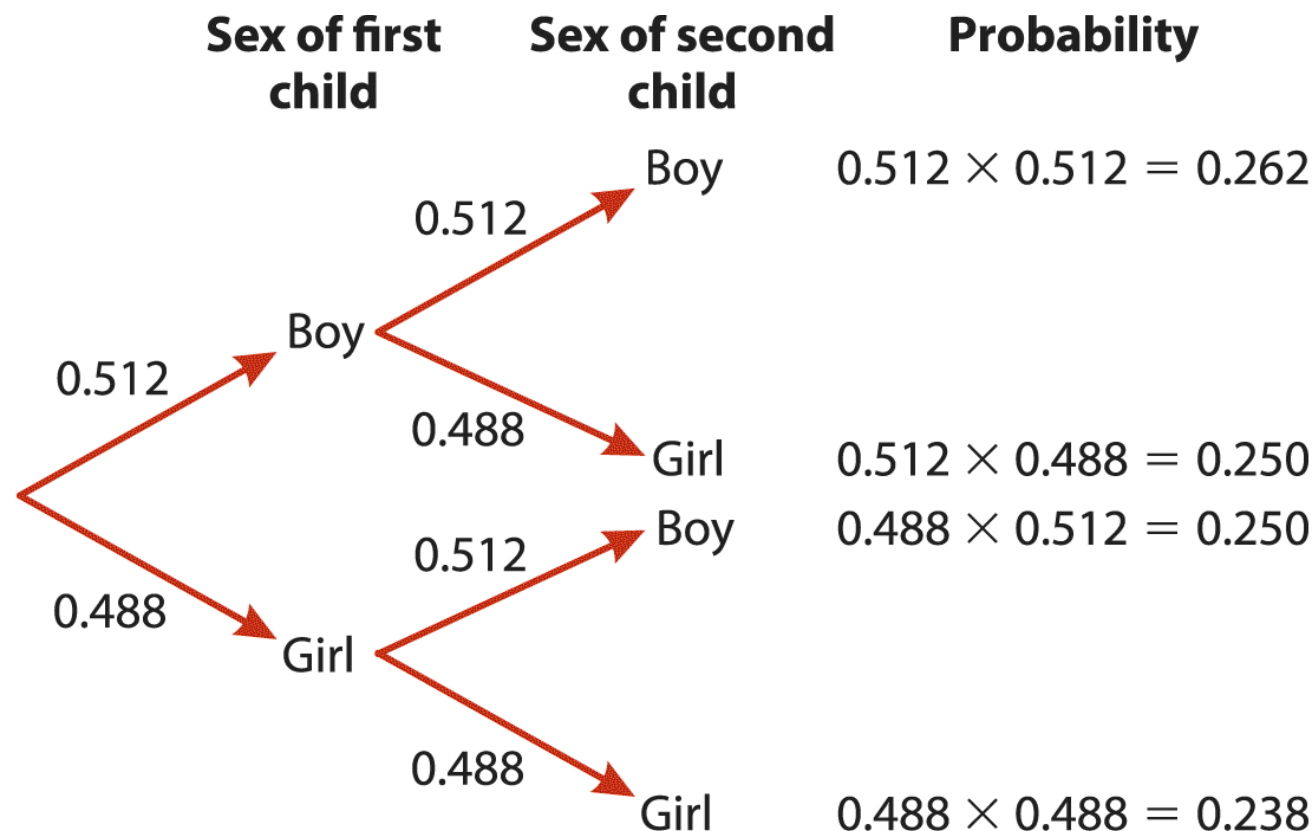


## 2.6 概率树 Probability trees

- 用于计算多次随机试验所产生的事件组合的概率

- 二胎家庭的子女

- 性别
  - 男女概率不一
- 顺序
  - 女女
  - 男男
  - 男女
  - 女男



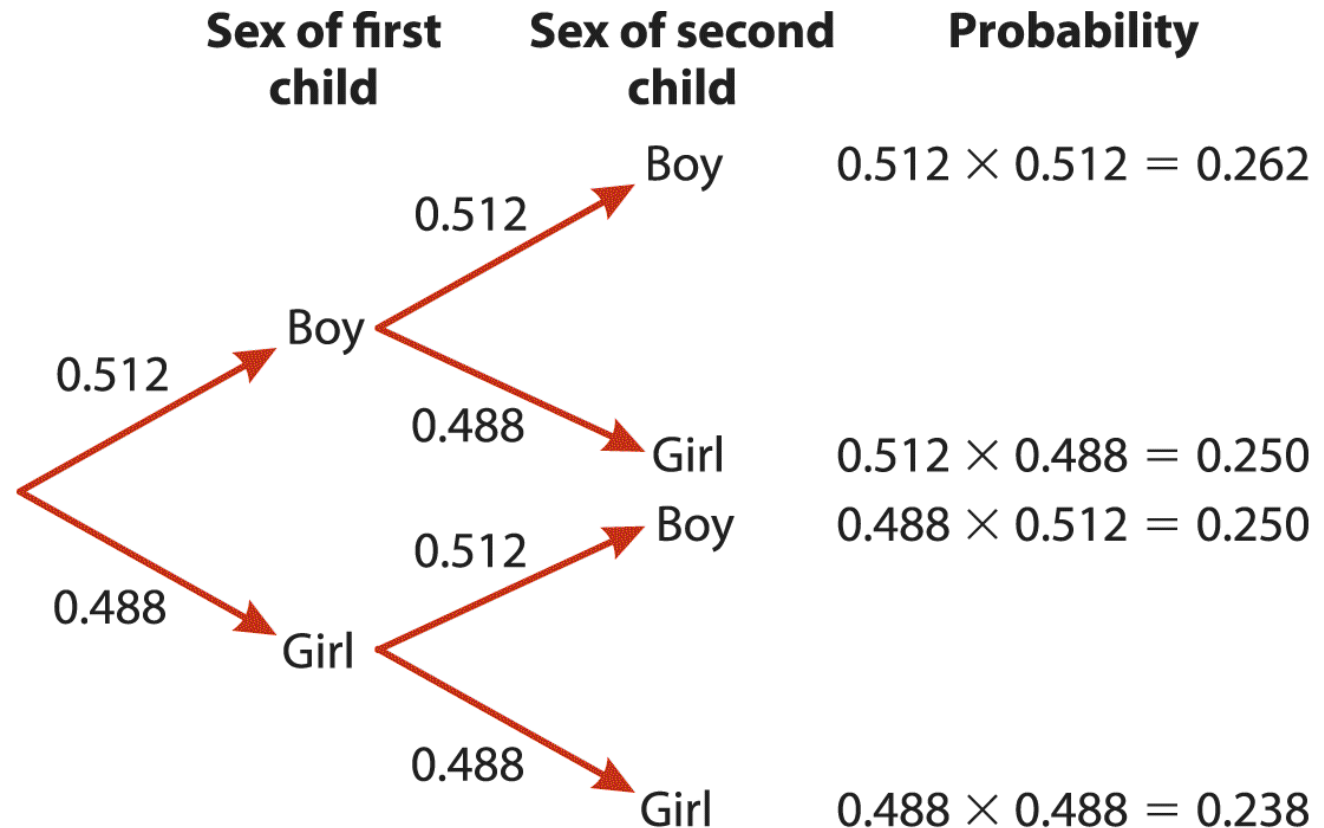
Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company



## 2.6 概率树

- 用于计算多次随机试验所产生的事件组合的概率
- 二胎家庭的子女

- 计算概率:
  - 至少一个女孩?
  - 至少一个男孩?
  - 孩子是相同性别?



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



## 2.6 其它内容

- 非独立事件 Dependent event
  - 如果事件不是独立的，那么它们就被称为依存事件。
  - 两个从属事件同时发生的概率由一般乘法法则给出：
    - $\Pr[A \& B] = \Pr[A] \Pr[B | A]$

### 3. 总结

- 概率是生物学中的一个重要概念
  - 对一个总体进行随机取样即是一种随机试验 (random trial), 其结果受概率规则的制约。
  - 随机试验是一种过程或实验, 它有两种或两种以上可能的结果, 而这些结果的发生是无法准确预测的。
  - 事件的概率 (the probability of an event) 是指在相同条件下反复进行随机试验时, 事件发生的次数比例。
  - 概率分布描述了随机试验所有可能结果的概率。

# 3. 总结

- 概率运算

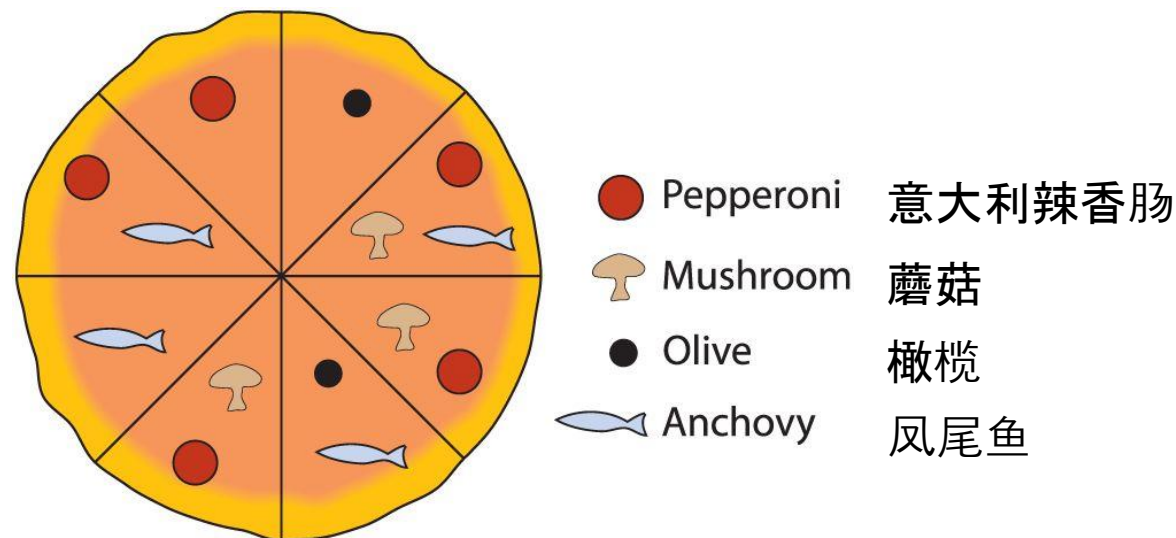
- 如果事件A和B不会同时发生，它们为互斥事件；那么A或B发生的概率参考加法公式： $\Pr[A \text{ or } B] = \Pr[A] + \Pr[B]$
- 扩展的一般加法公式为  $\Pr[A \text{ or } B] = \Pr[A] + \Pr[B] - \Pr[A \& B]$
- 如果事件A或B的发生不依赖于彼此，它们互为独立事件；那么A和B同时发生的概率参考乘法公式： $\Pr[A \& B] = \Pr[A] \times \Pr[B]$



## 4. 课堂讨论

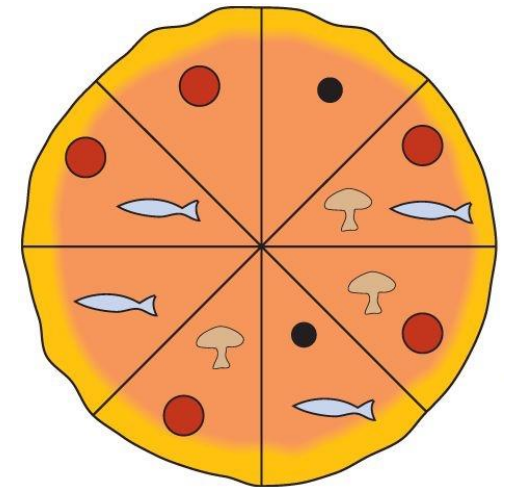
- 外卖的披萨，分为大小相等的8片；每片披萨上可能有意大利香肠、蘑菇、橄榄、或凤尾鱼。试想一下，随机拿起一片披萨（即拿起八片披萨中任何一片的概率为  $1/8$ ）；

- 请根据今天的概率知识回答：



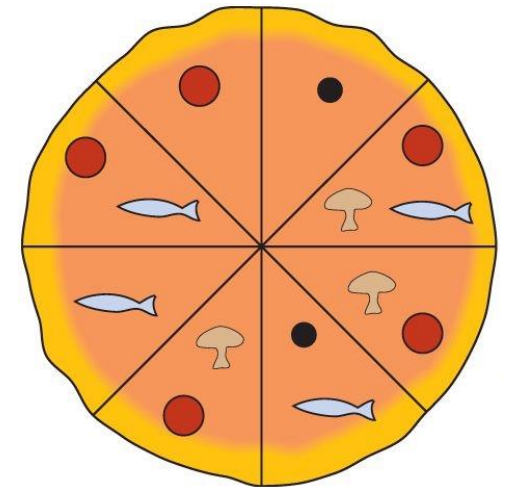
## 4. 课堂讨论

- 1. 你的披萨上有意大利辣香肠的概率是多少？
- 2. 你的披萨上既有意大利辣香肠又有凤尾鱼的概率是多少？
- 3. 你的披萨上有意大利辣香肠或凤尾鱼的概率是多少？
- 4. 这块比萨上的意大利辣香肠和凤尾鱼是互斥的吗？
- 5. 这块比萨上的橄榄和蘑菇是互斥的吗？



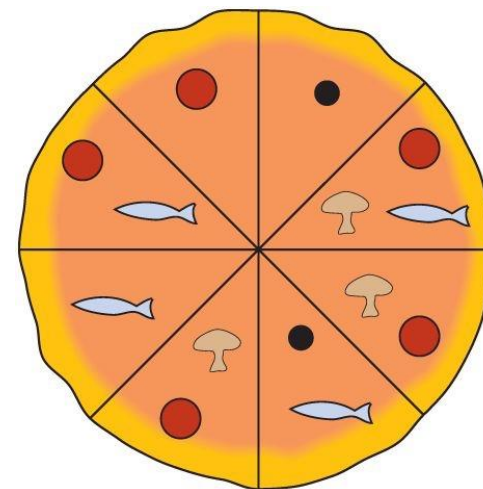
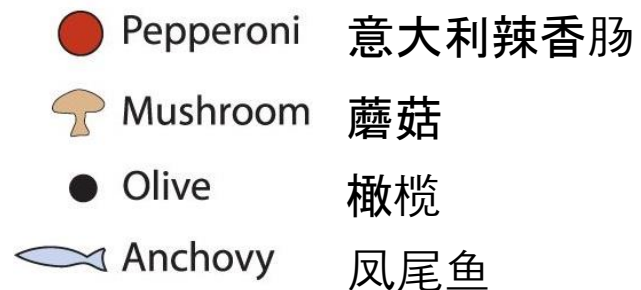
## 4. 课堂讨论

- 6. 在选择一片比萨时，蘑菇和凤尾鱼是独立的吗？
- 7. 如果我从这个比萨中挑选一片，并告诉你上面有橄榄，那么它也有凤尾鱼的概率有多大？
- 8. 如果我从这块披萨中挑选一块，并告诉您上面有凤尾鱼，那么它也有橄榄的概率是多少？



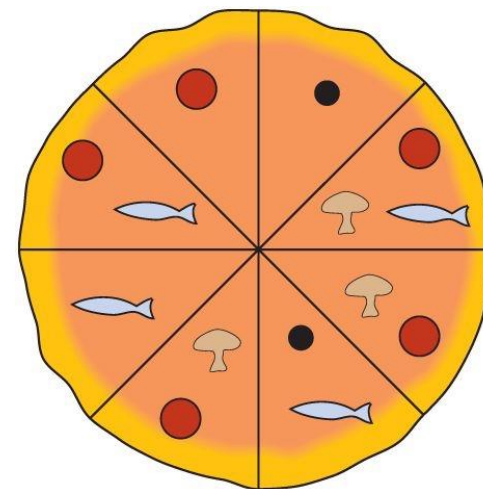
## 4. 课堂讨论

- 9. 你的七个朋友每人随机选择一片披萨，吃完后不告诉你他们吃了什么配料。那么，剩下的最后一块有橄榄的概率是多少？
- 10. 你随机选择两片披萨，它们都有橄榄的概率是多少？
  - (提示：选掉第一片后，从剩下的披萨片中选择一片的概率会发生变化)



## 4. 课堂讨论

- 11. 随机选择的一片披萨上没有意大利辣香肠的概率是多少？
- 12. 画一个同样分为8片的，且蘑菇、橄榄、凤尾鱼和意大利辣香肠相互排斥的比萨。

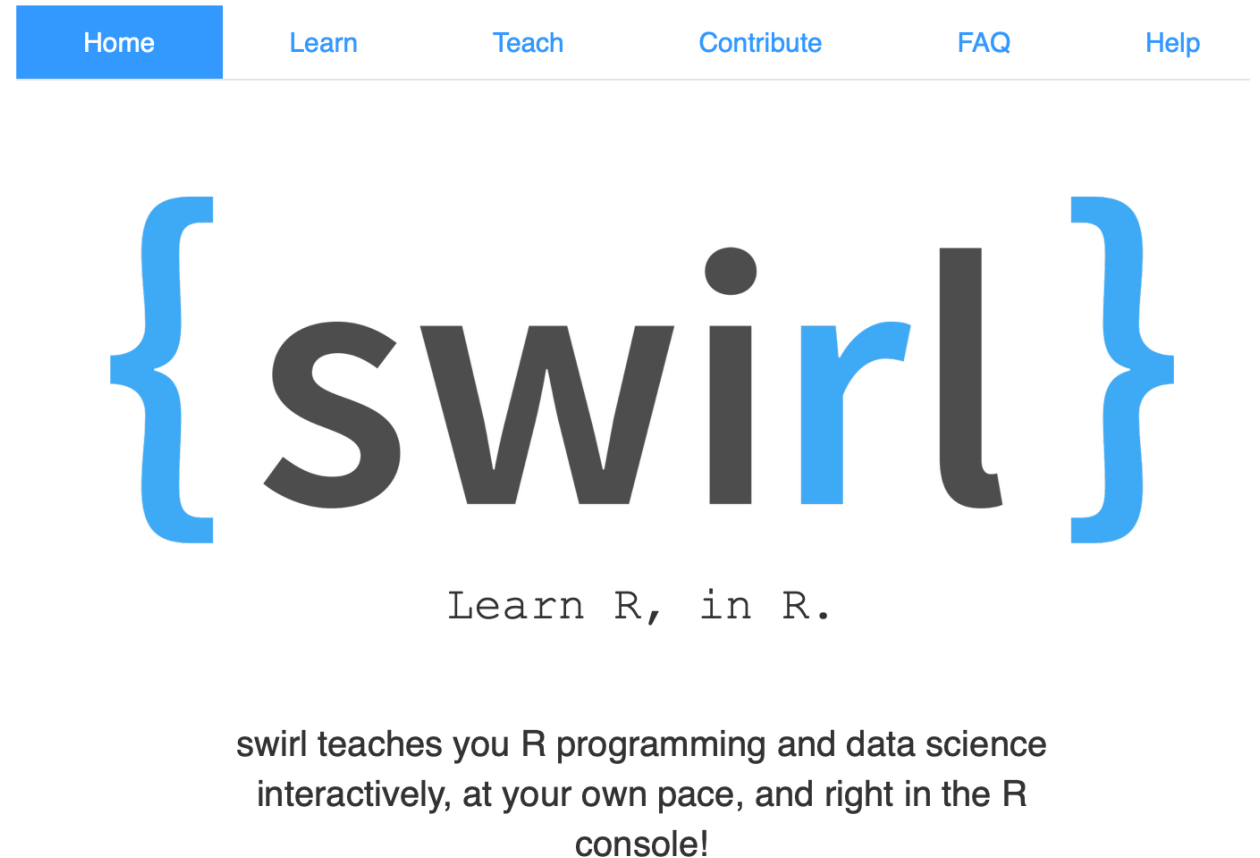




## 5. R-tips

- <https://swirlstats.com>
- <https://swirlstats.com/students.html>

```
> install.packages("swirl")  
> library(swirl)  
> swirl()
```



# 5. R-tips

- The tidyverse is an opinionated [collection of R packages](https://www.tidyverse.org) designed for data science.
- <https://www.tidyverse.org>

```
> install.packages("tidyverse")  
> library(tidyverse)  
> %>%
```

