September 12th – Probability and how it relates to statistics

Quantitative Thinking in the Life Sciences

Today

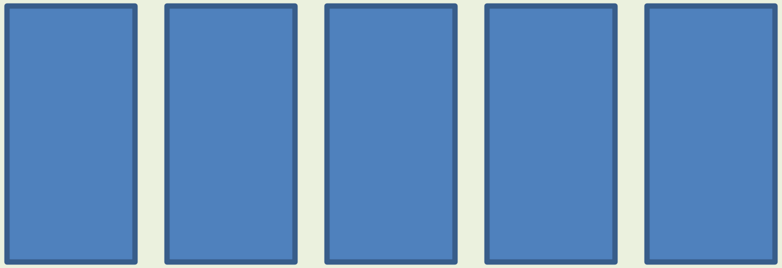
- Probability!
- More R fun!
  - revisiting assignment code to date
  - Answers to Hobb's exercise questions
  - expanding on the matrix and array functions

Housekeeping

- Next weeks class meets in Jeffords 326 (same time – different place)
- I will still be using uvm’s contact information (e-mails)

Probability

- Coin flip = ½
- Ace in a deck of cards = 4/52
- Each result is called an outcome or an event
- I think of these as outcome slots

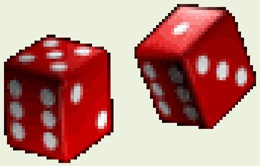


In R!!!

```
> library(animation)

> flip.coin(faces = 2, prob = NULL, border = "white",
   grid = "white", col = 1:2, type = "p", pch = 21, bg = "transparent",
   digits = 3)
```

Rolling two dice



- Two six-sided dice with sides numbered 1-6
- Likelihood of the dice landing on any of 6 numbers is equal
- All die rolls are independent

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Sum on dice

2: One possibility (1,1) probability = 1/36 options

3: Two possibilities (1,2) & (2,1) probability = 2/36 options

4: Three possibilities (1,3), (2,2) & (3,1) probability = 3/36 options

⋮

7: Six possibilities (1,6), (2,5), (3,4), (4,3), (5,2) & (6,1) probability = 6/36 options

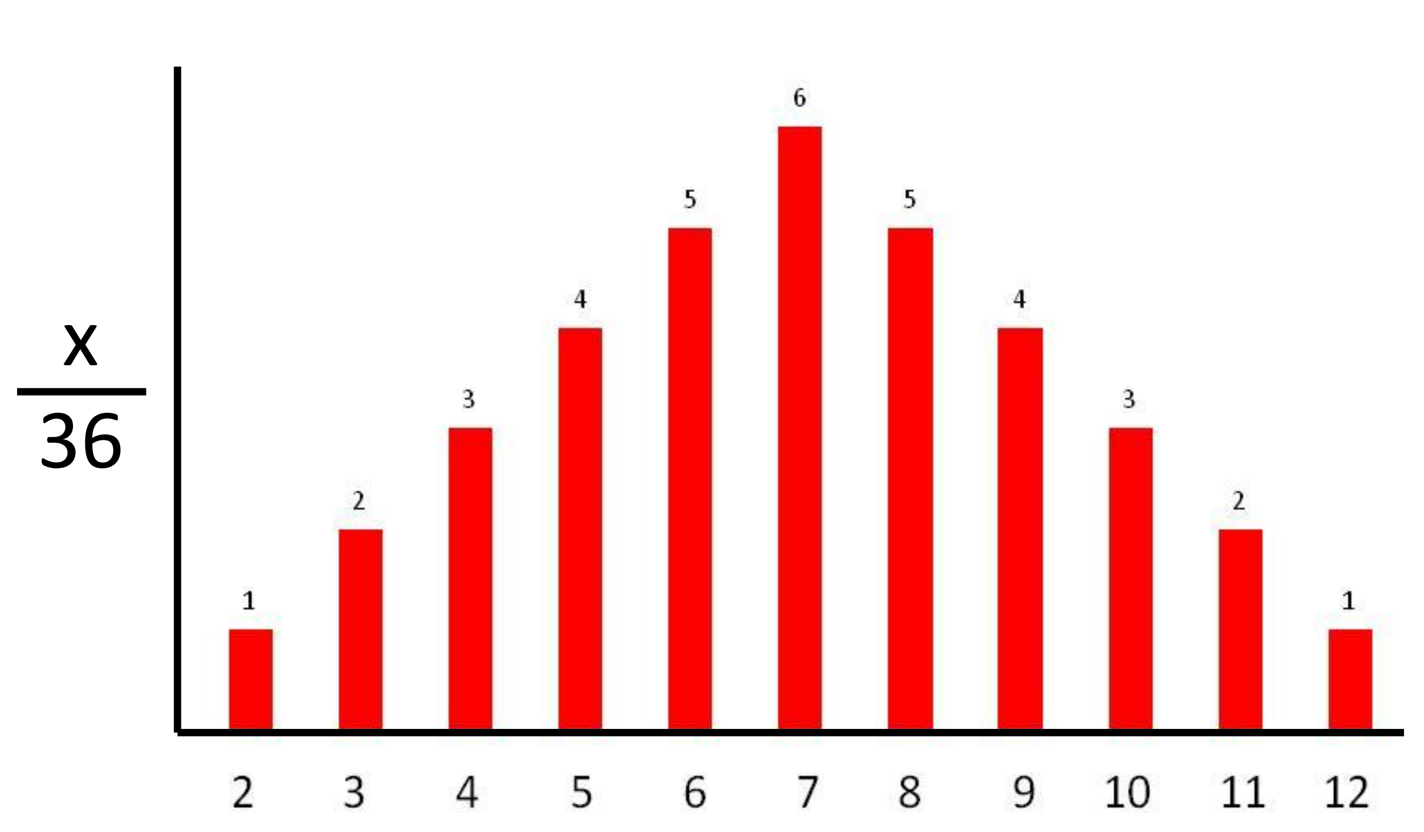
Sum on dice

2: One possibility: (1,1) probability = 1/36 options

3: Two possibilities: (1,2) & (2,1) probability = 2/36 options

4: Three possibilities: (1,3), (2,2) & (3,1) probability = 3/36 options

7: Six possibilities: (1,6), (2,5), (3,4), (4,3), (5,2) & (6,1) probability = 6/36 options



Go to R for a sweet animation!

Traditional frequentist statistics

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Testing to see if a series of values are probable given assumptions about reality. Our model assumes that:

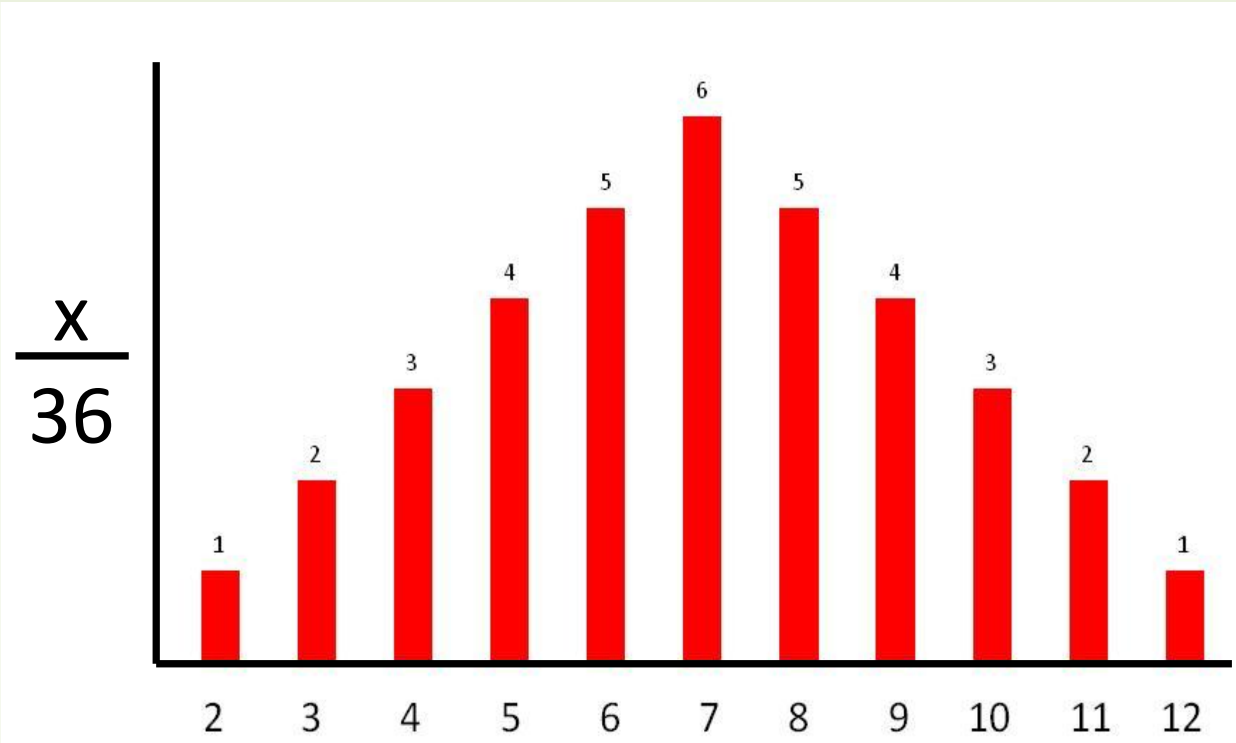
- Two six-sided dice with sides numbered 1-6
- Probability of the dice landing on any of 6 sides/numbers is equal
- All die rolls are independent

Go to R for another sweet animation!

What if you tested this "model" with 10 rolls of the dice and found that all of sum values were between 2 and 4?

- Probability given the model is true = 1/6^10 = 0.00000001654
  - Not too likely
  - Reject the model
- Could we say that:
  - The dice were loaded?
  - Dice were actually only three-sided (what is a three-sided die?)

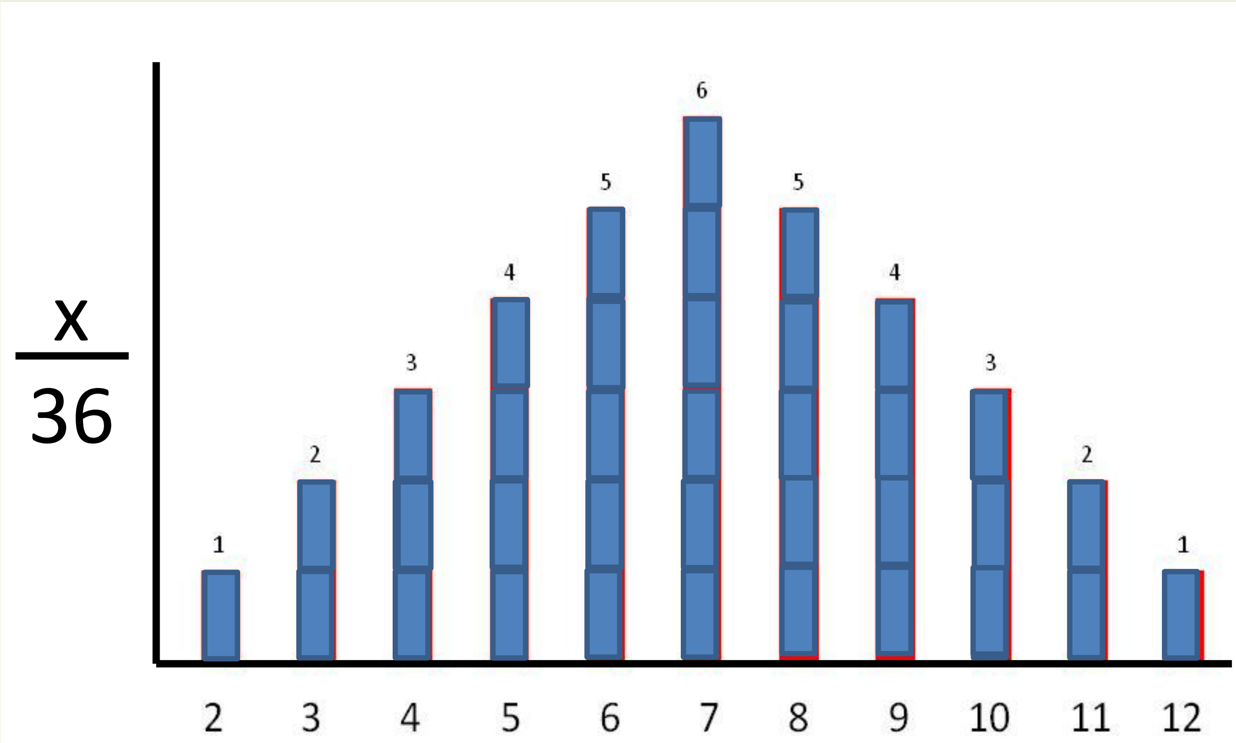
Probability space



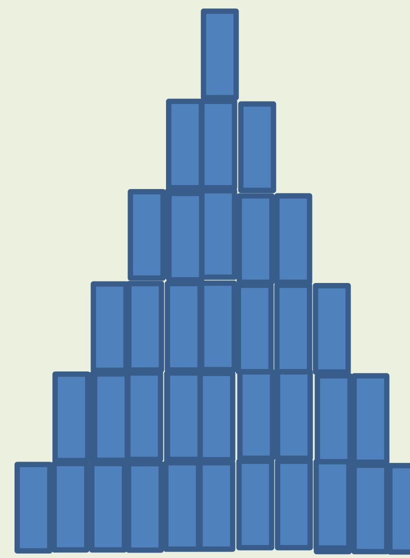
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36 + 5/36 + 4/36 + 3/36 + 2/36 + 1/36 **= 1**

Probability space



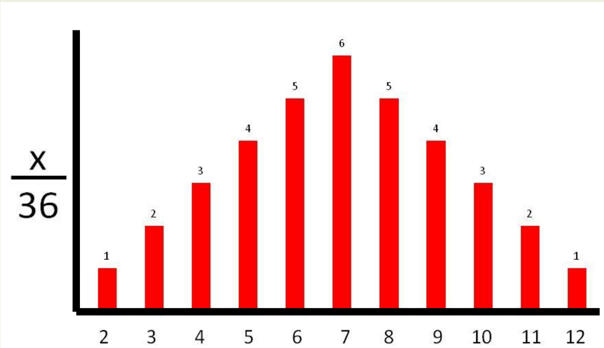
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)



1/36 + 2/36 + 3/36 + 4/36 + 5/36 + 6/36
+ 5/36 + 4/36 + 3/36 + 2/36 + 1/36 **= 1**

$$\sum_{space} probs = 1$$

Probability space



(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Roll a third die:

1, +

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

2, +

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

3, +

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

4, +

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

5, +

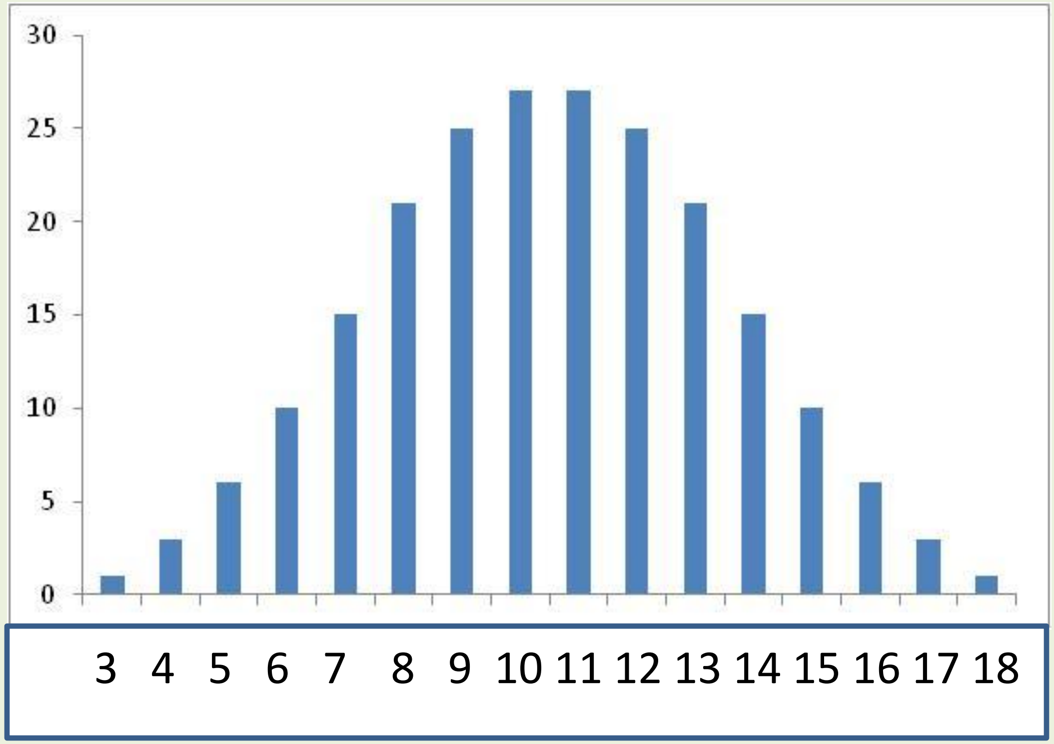
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

6, +

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

= 216 possibilities, each with an equal occurrence probability

Probability space for 3 dice = 216 possibilities



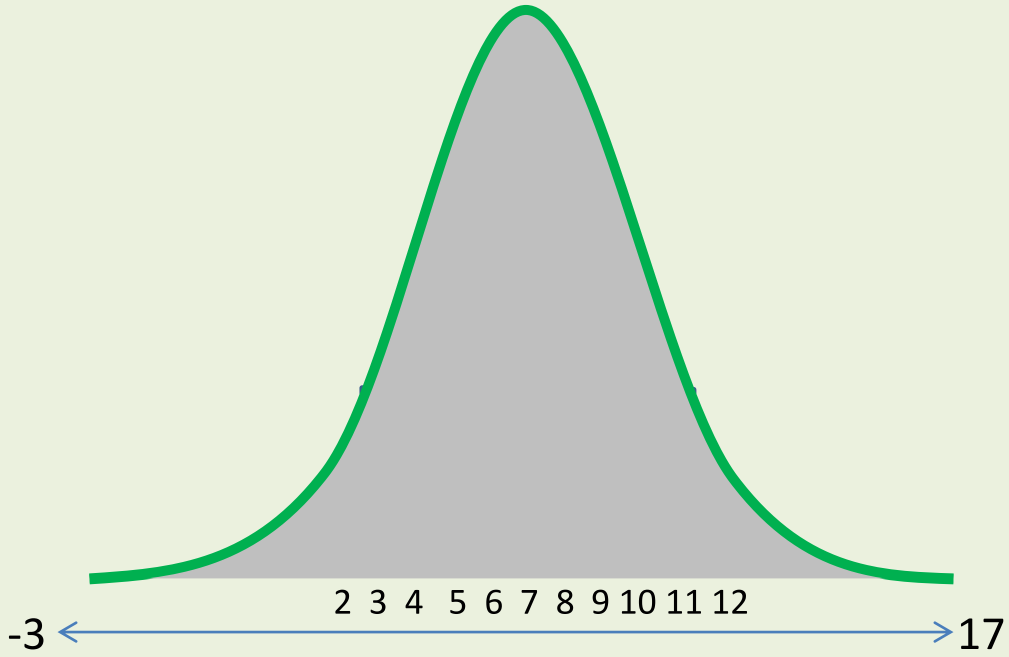
Probability space for rolling x dice

Dice	Combinations	Probability of any one combination	Range of values: Sum of dice
1 Die	6	0.167	Sum of dice: 1-6
2 Dice	36 	0.0278	Sum of dice: 2-12
3 Dice	216 	0.00463	Sum of dice: 3-18
4 Dice	1296	0.000772	Sum of dice: 4-24
5 Dice	7776	0.000129	.
6 Dice	46656	0.0000214	.
7 Dice	279936	0.00000357	.
8 Dice	1679616	0.000000595	
9 Dice	10077696	0.0000000992	
10 Dice	60466176	0.0000000165	
11 Dice	362797056	0.00000000276	
12 Dice	2176782336	0.000000000459	
13 Dice	13060694016	0.0000000000766	
14 Dice	78364164096	0.0000000000128	Sum of dice: 14-82

Combinations * Probability of occurrence of each = 1

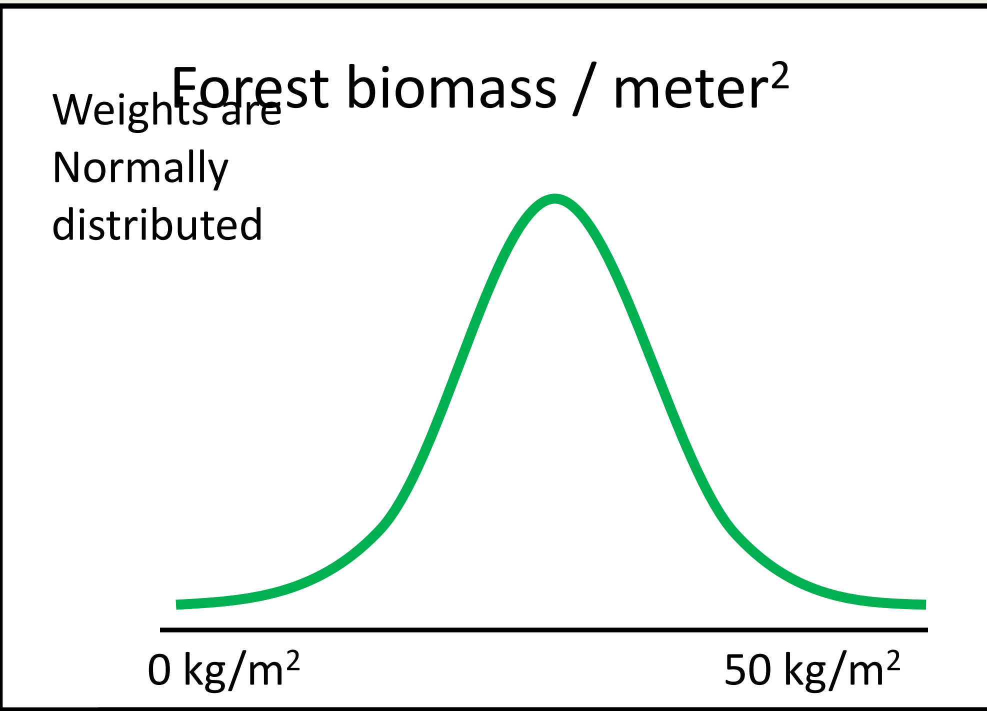
78364164096 * 0.0000000000128 = 1

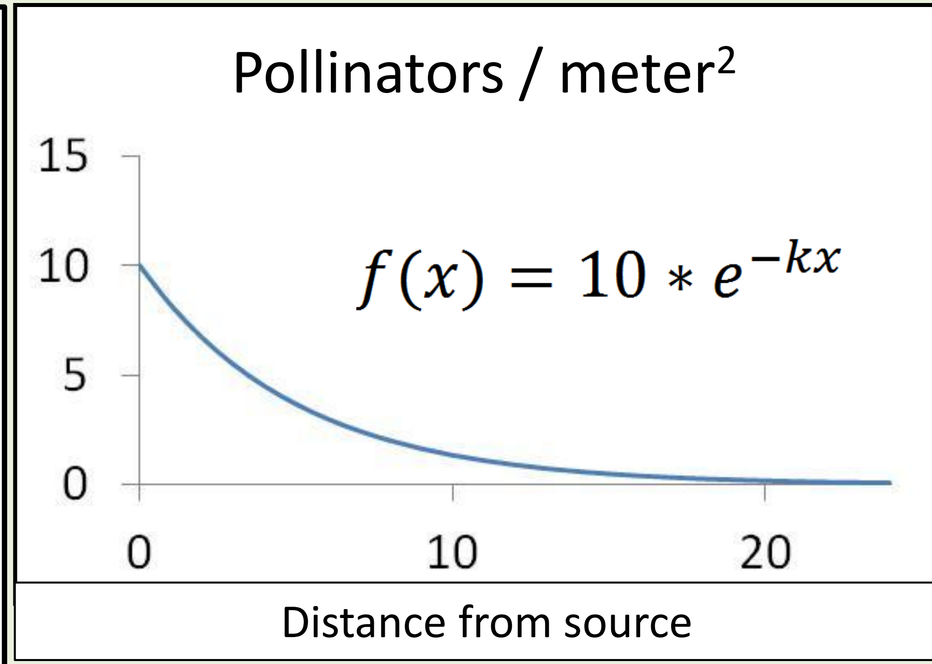
Discrete to continuous probability

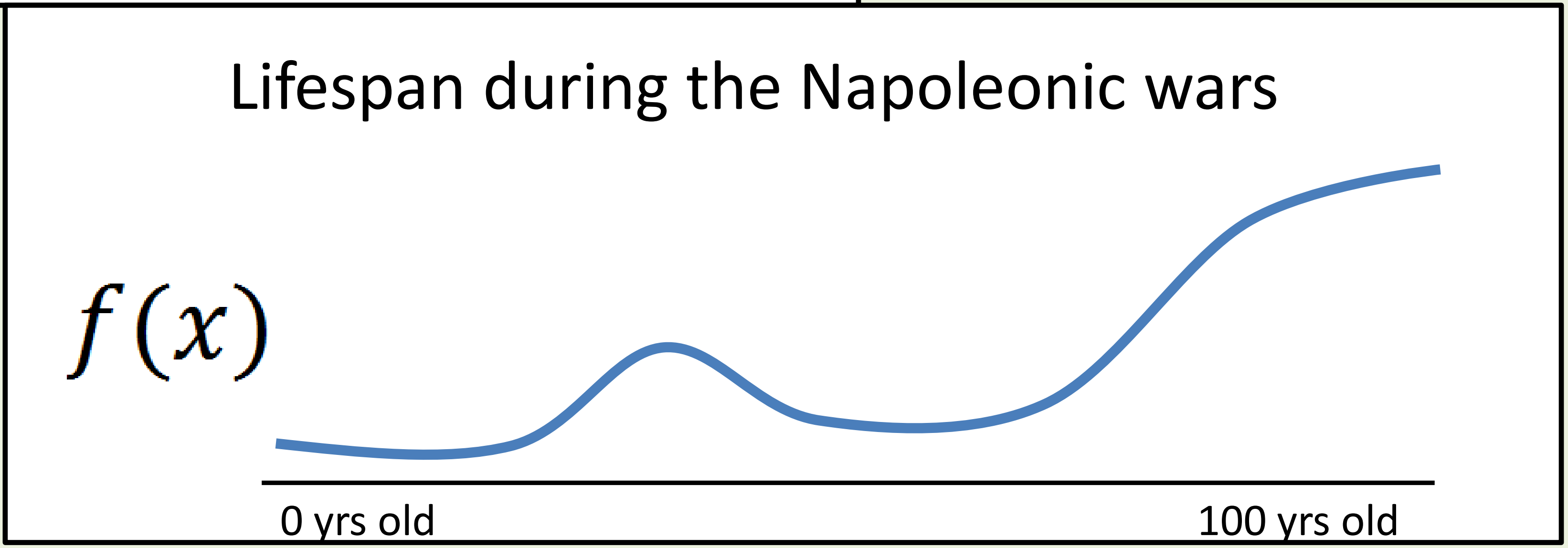


Area under the curve is the continuous probability space

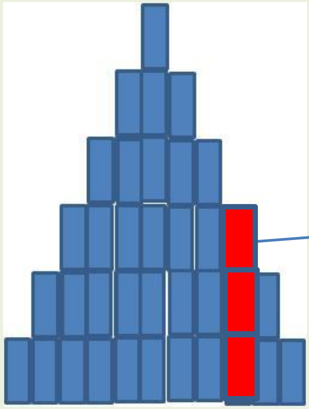
- Total area is equal to 1
- All the possible values are under the curve


Forest biomass / meter²
Weights are
Normally
distributed
0 kg/m²
50 kg/m²


Pollinators / meter²
15
10
5
0
0
10
20
$f(x) = 10 * e^{-kx}$
Distance from source

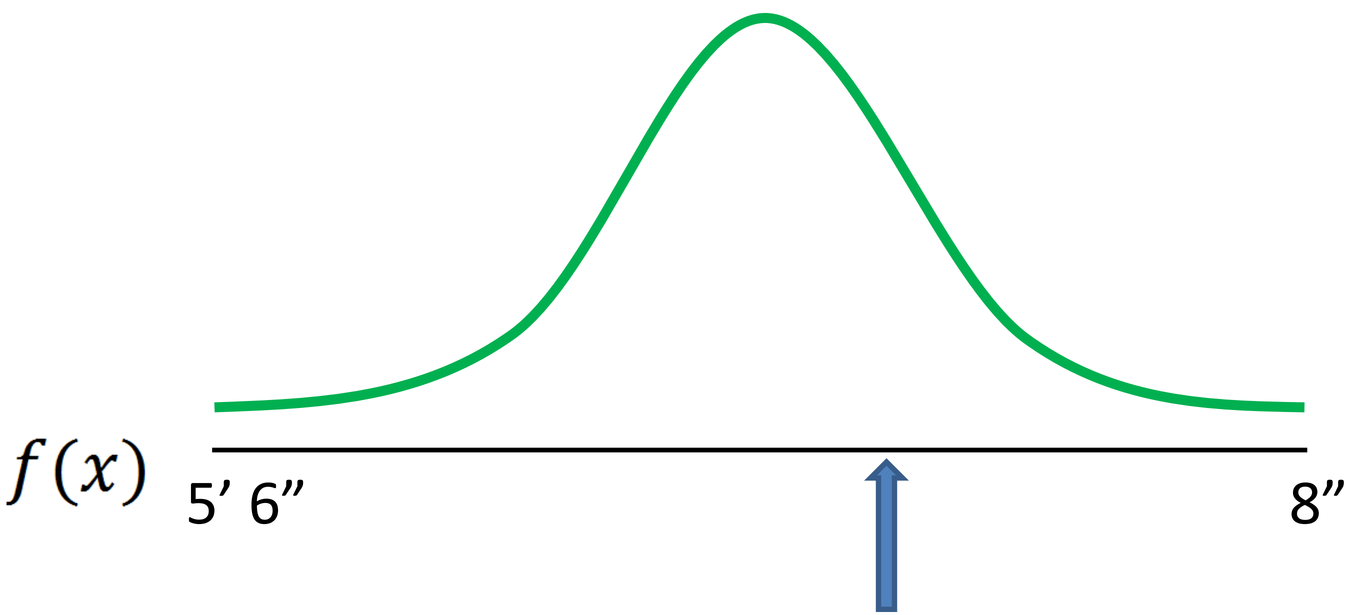

Lifespan during the Napoleonic wars
$f(x)$
0 yrs old
100 yrs old

Pitfall!



Probability of having dice add to 10 = 3/36

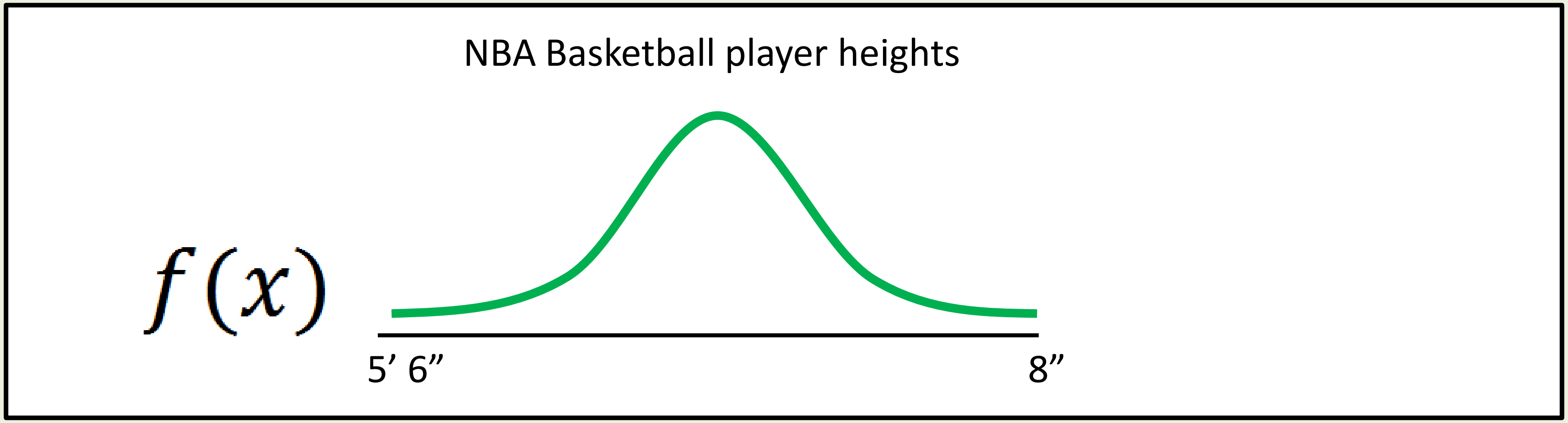
NBA Basketball player heights



What is the probability of measuring a player with a height of 6’8.0121352245662 3” ?

Answer = 0

Calculus!

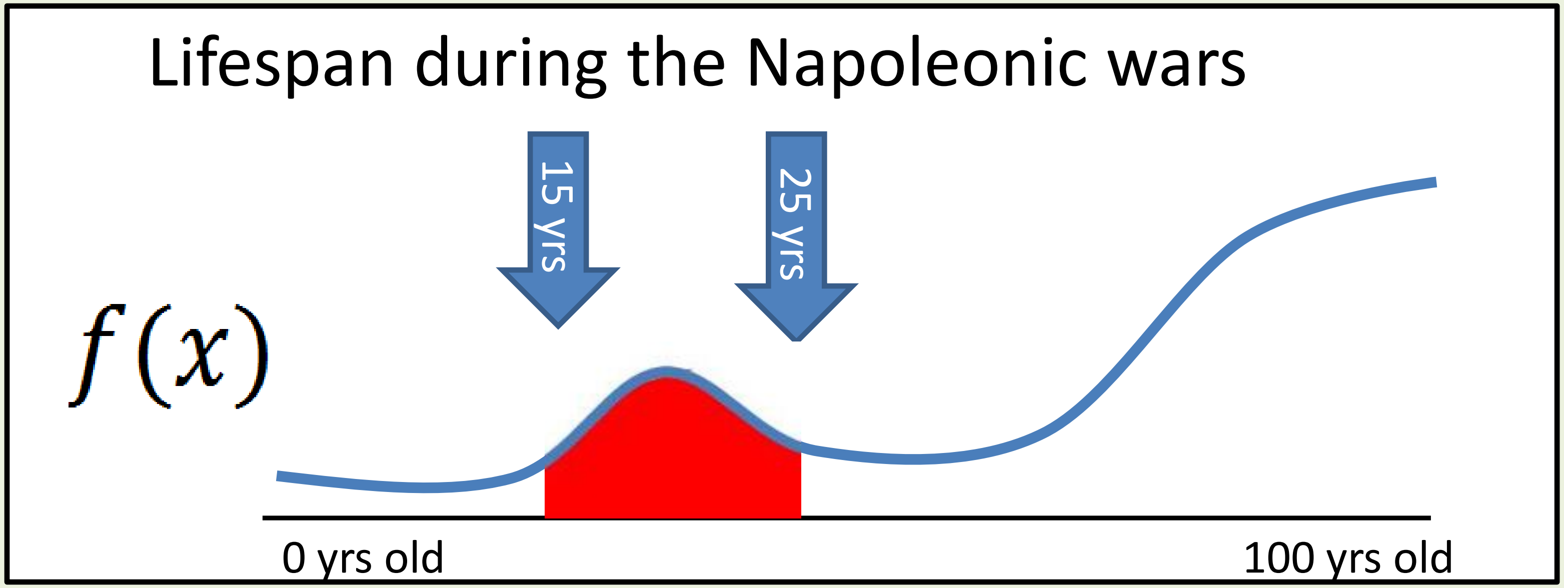


- If we know the function, $f(x)$, we can calculate the area as: $$\int f(x)dx$$

- Because all possibilities are under the curve:

$$\int_{space} f(x)\,dx = 1$$

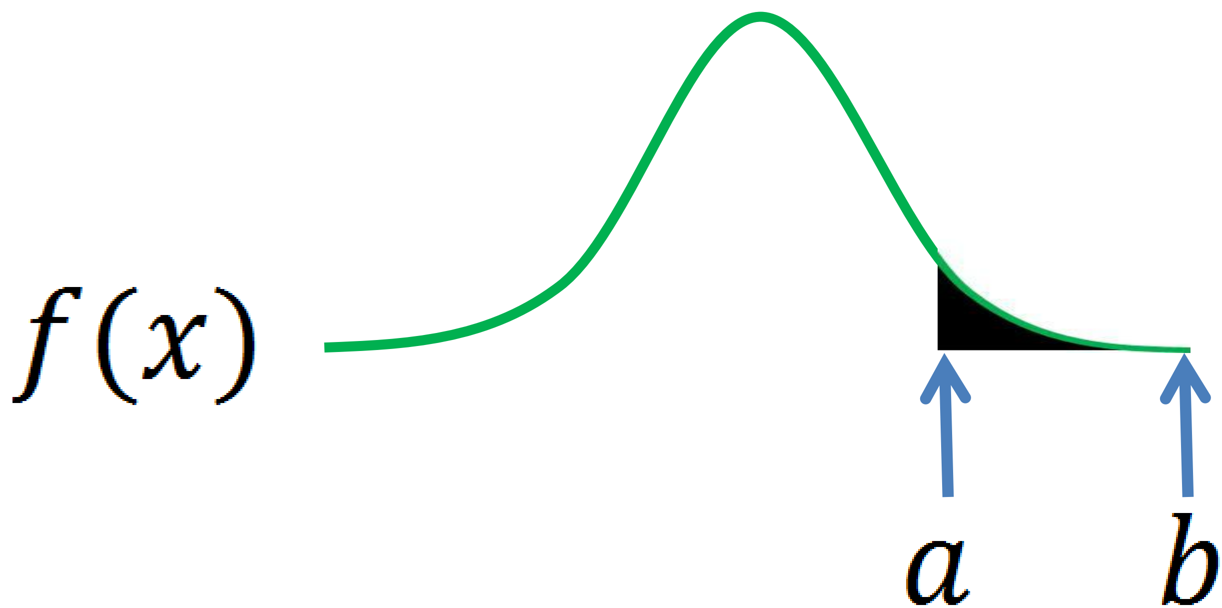
Probability example



$$\int_{15}^{25} f(x)dx$$ = probability of dying between 15 and 25 years old

Hypothesis testing – frequentist approach

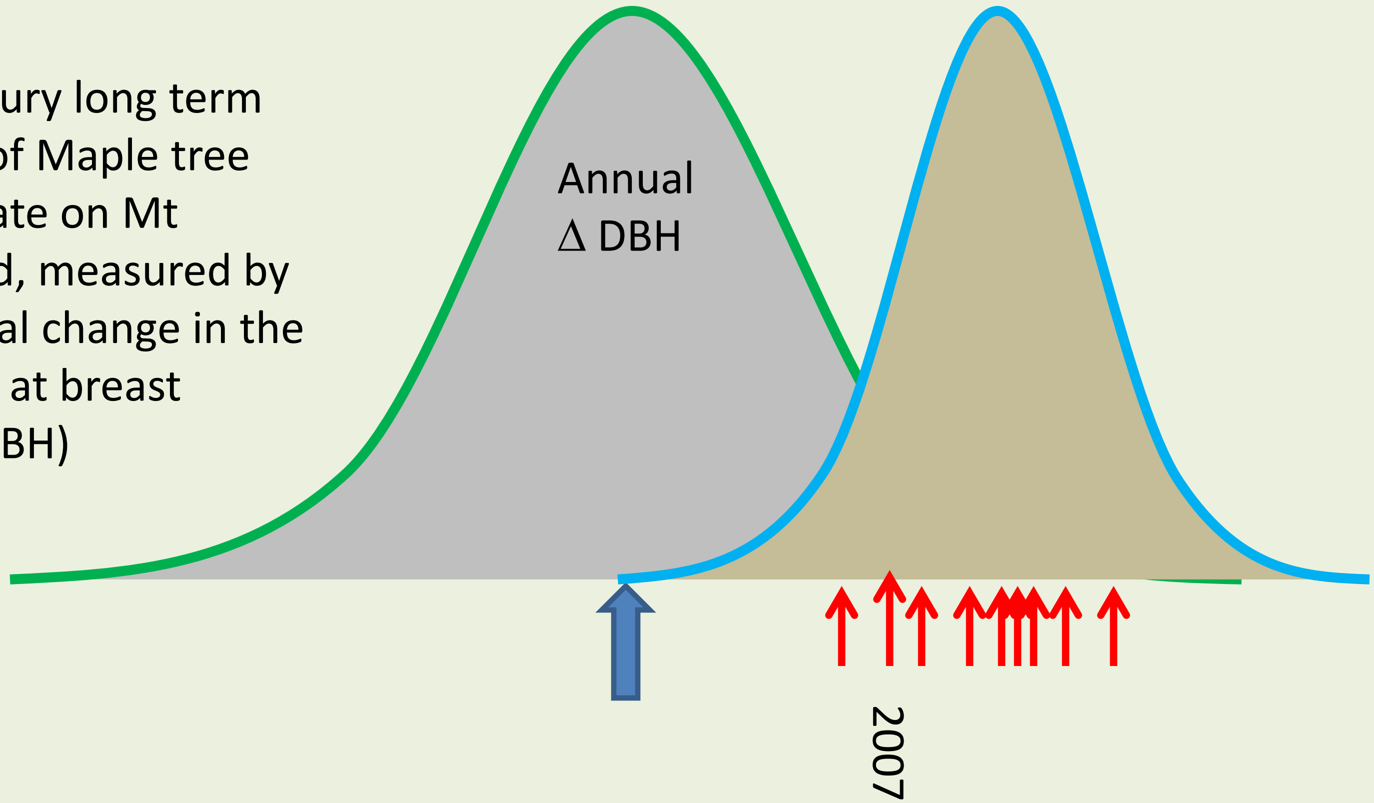
The **p-value** is the probability of obtaining a test statistic *at least* as extreme as the one that was actually observed, assuming that the null hypothesis is true.



$$\int_a^b f(x)dx = \text{prob(test statistic)}$$

Back to the conceptual

20th Century long term average of Maple tree growth rate on Mt Mansfield, measured by the annual change in the diameter at breast height (DBH)



The **p-value** is the probability of obtaining a **test statistic** at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

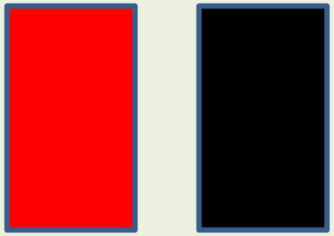
The “test statistic” is the quantification of the probability of observing your data

Assignment # 3

- On courses tab
  - http://www.uvm.edu/~scmerril/Courses.html
- Part 1: Manuscript format
  - Design a probability experiment
    - Short, concise, possibly even terse!
      - Introduction
      - Methods
      - Results
      - Discussion and conclusions
- Chapter 3 R code: Matrices, arrays and programming

Part 1:
Probability experiment

- Three cards
  - Black on one side, red on the other side
  - Black on both sides
  - Red on both sides



- Question: If you draw a card randomly from the three cards and look at one side, what is the probability that the other side is the same color?
  - e.g., if you draw a card and see a red side, what is the probability that the other side will be red?

Endless fun with R!

- Questions from last week?
  - assigning an object and then calling it out!
  - pmin()
- require() vs library()
  - "The other reason I use require is that it keeps me from referring to packages as libraries, a practice that drives the R-cognoscenti up the wall. The library is the directory location where the packages sit."
    - DWin Stackoverflow user