FISHER TRADITION AND PROBABILITY

LECTURE 5 – QUANTITATIVE METHODS I

ANNOUNCEMENTS

- Next Tuesday will be a homework help session in this room
- Come with questions from the homework, there is not a pre-planned agenda
- Homework will then be due on ***THURSDAY, September 21st***

LAST WEEK

- Understanding measures of central tendency
- Considered metrics for how well those central tendency measures reflected our data
- This week: Start examining whether our sample differs “significantly” from chance

“FISHER” TRADITION

WHAT FISHER SUGGESTED…

- Set up a statistical null hypothesis (note: null does NOT mean nil)
- Report the exact level of significance
- Do not use a "conventional" level, do not talk about accepting or rejecting hypotheses, do not pass and do not collect $200
- Use this procedure only if you know very little about the problem at hand

NEYMAN-PEARSON

- Set up two hypotheses, and design a study based on the ”rejection region” for each hypothesis
- If data is within rejection range for H1, accept H2. Otherwise accept H1. Note that accepting it doesn’t mean you believe in it, just that you act though it was so
- Utility is limited to situations where there is a clear difference in hypotheses, and when you can make a rationale decision about when to accept versus reject H1 and H2

A MERGER OF SORTS

- Unfortunately, these two ideas were melded together into something neither camp would be too excited by
- 1) Set up a null hypothesis, where null almost always means "chance"
- 2) Make a yes-no decision about that hypothesis
- 3) Repeat

BASIC PREMISE IS…

- We want to know what is the probability that we would get the values evidenced *(or those more extreme*) given our null hypothesis

A P-VALUE

- We want to know what is the probability that we would get the values evidenced *(or those more extreme*) given our null hypothesis
- $P(\text{Data} + \mid H_0)$

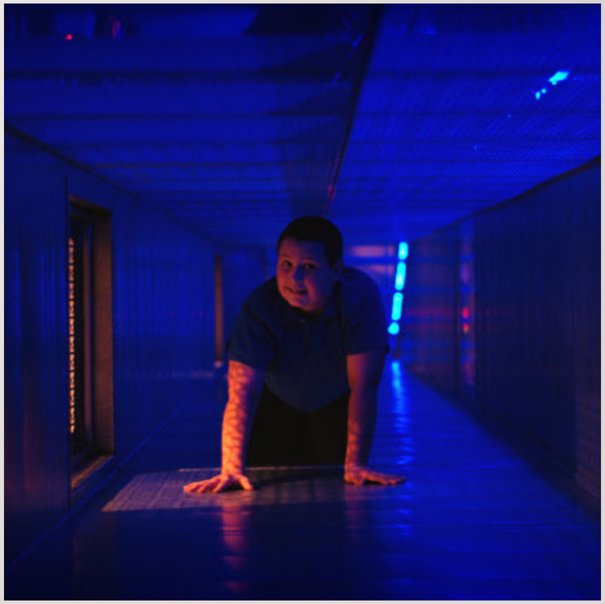
A P-VALUE

- We want to know what is the probability that we would get the values evidenced *(or those more extreme*) given our null hypothesis
- P(Data + | H_0)
  - Assumes, among other things, that the null is exactly true, that you have a random sample, and the scores are independent

BASICS OF PROBABILITY

THE MAGIC HOUSE

THE MAGIC HOUSE



FIRST, WE HAVE OUR SAMPLE SPACE

TWO IMPORTANT ASSUMPTIONS OF PROBABILITY

- P(S) = 1
- 0 is less than or equal to P(A) which is less than or equal to 1

P(MISS SCARLET) =

P(MISS SCARLET) = 1/6

P(MISS SCARLET) = N OF EVENTS / SAMPLE SIZE

P(FEMALE) = 3/6 = 1/2

COMPLEMENT

- The probability the event does not occur
- $1 - P(A)$

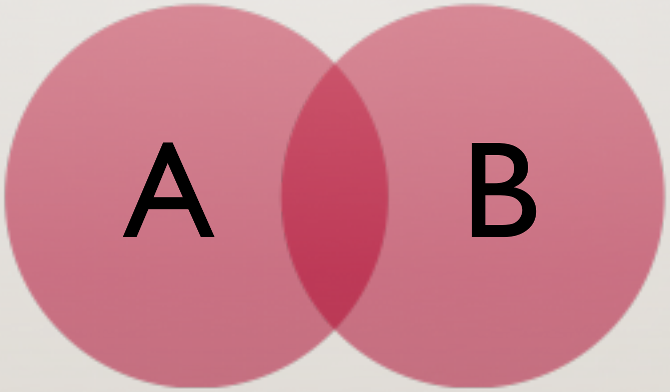
P(NOT FEMALE) = 3/6 = 1/2

UNIONS

- The possibility of A *or* B occurring
- All elements that are in one of A or B
- $P(A \cup B)$

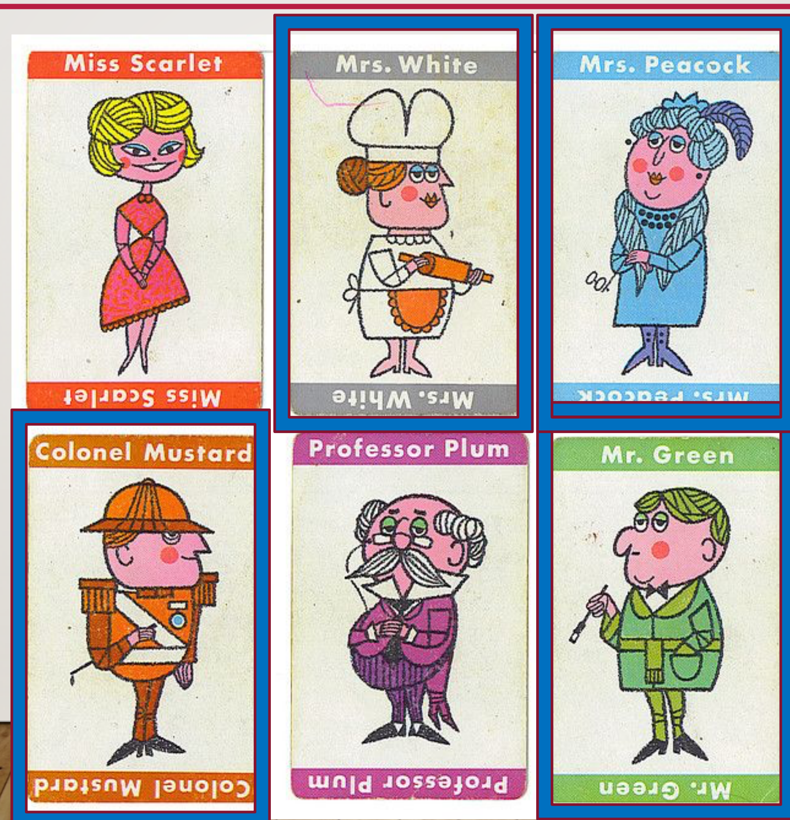
UNIONS

- The possibility of A *or* B occurring
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

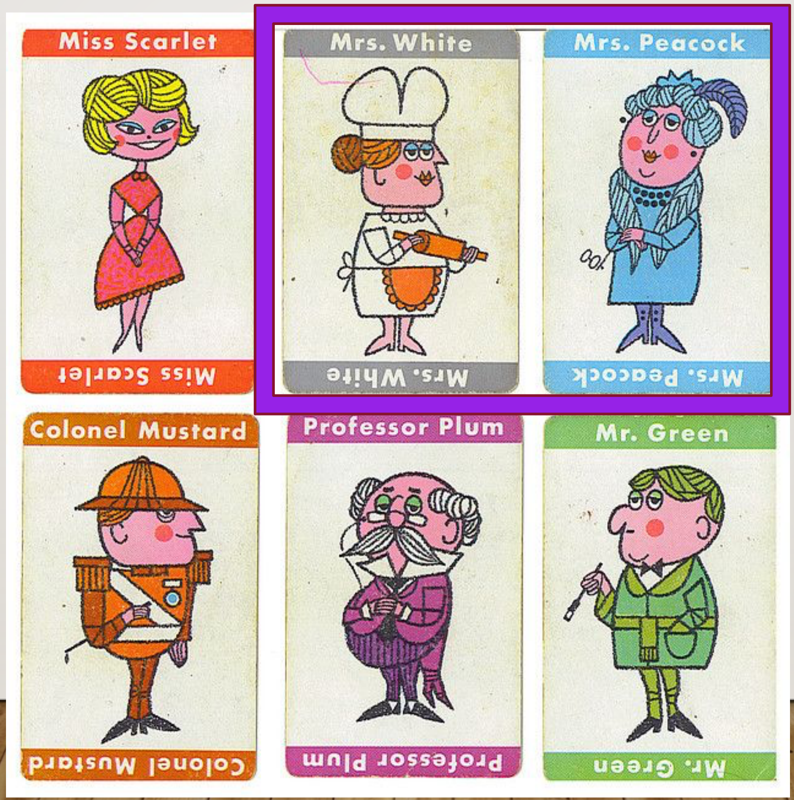


P(FEMALE OR HOLDING SOMETHING) = 3/6

P(FEMALE OR HOLDING SOMETHING) = 3/6 + 4/6



P(FEMALE OR HOLDING SOMETHING) = 3/6 + 4/6 – 2/6 = 5/6

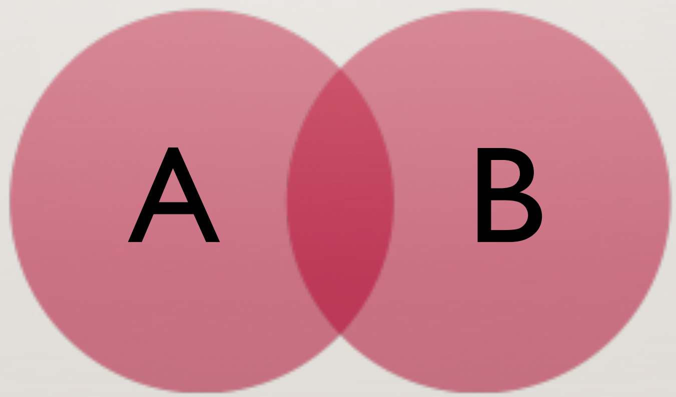


INTERSECTION

- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B \mid A)$

INTERSECTION

- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B \mid A)$



P(BAKER AND FEMALE) = 1/6

P(BAKER AND FEMALE) =
1/6 X 1 = 1/6

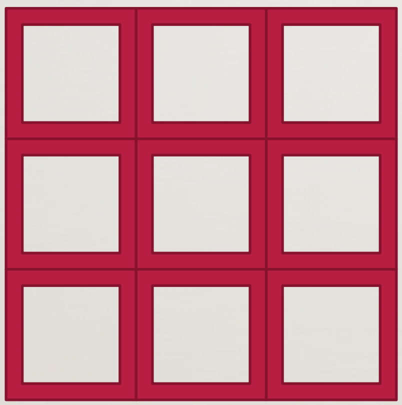
INTERSECTION

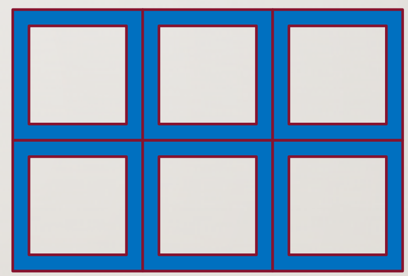
- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B \mid A)$
- $P(A \cap B) / P(A) = \mathbf{P(B \mid A)}$

INTERSECTION

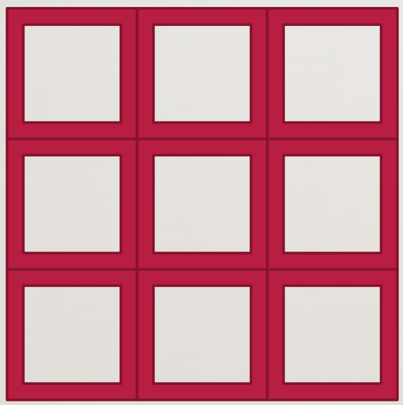
- The probability of A *and* B occurring
- $P(A \cap B) = P(A) \times P(B \mid A)$
- If A and B are independent, $P(A \cap B) = P(A) \times P(B)$
- ***Independence of observations is one of the criteria for having interpretable p-values***

INTERSECTION IN THE MAGIC HOUSE

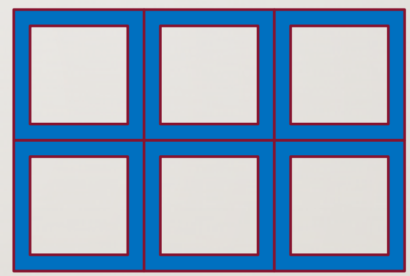




INTERSECTION IN THE MAGIC HOUSE



$P(A) = 1/9$

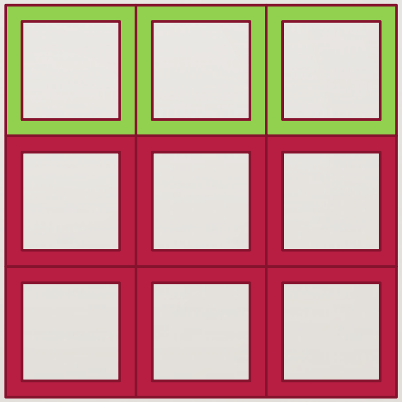


$P(B) = 1/6$

$1/9 \times 1/6 = 1/54$

GROUP WORK

- P(Green ∩ Male) =
- P(Red ∩ Miss Scarlet) =
- P(Green ∩ Mr. Green) =
- P(Title Starts with M ∪ Male) =



PERMUTATIONS

- Permutations refer to all the ways n objects can be arranged
- n = 2: AB BA
- n = 3: ABC ACB BCA BAC CAB CBA
- n = 4: ABCD ABDC ACBD ACDB ADBC ADCB BACD BADC
  BCAD BCDA BDAC BDCA CBAD CBDA CABD CADB CDBA
  CDAB DBCA DBAC DCBA DCAB DABC DACB

PERMUTATIONS

- In general, the number of permutations for n objects is:
  - $n! = n(n-1)(n-2)\ldots$

PERMUTATIONS

- However, if you are looking for the number of r choices from n…
  - n! / (n-r)!

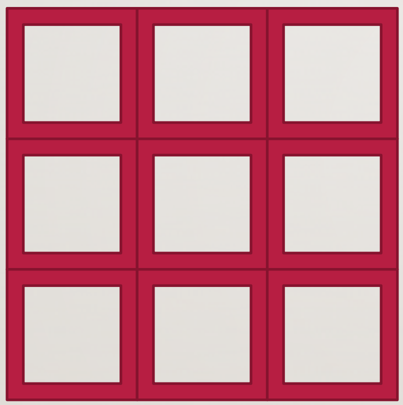
PERMUTATIONS

- However, if you are looking for the number of r choices from n…
  - n! / (n-r)!
- One more trick to the Magic House, we learned that we were supposed to find *the pair* of culprits of the nine possibilities at the end
  - 9! / (9-2)!
  - 72 possible permutations

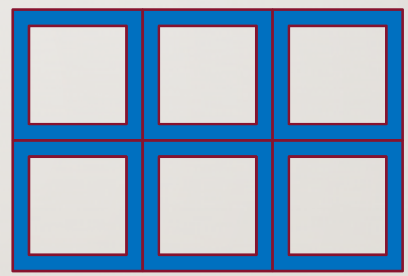
COMBINATIONS

- We don't really care if we got the first or second culprit in the "first" position
- When order doesn't matter, we can count the number of combinations
- $C = n! / (n-r)!r!$
- $C = 9! / (9-2)!2!$
- C = 36 possible combinations
    - 1/36 chance of getting the right two

COMBINATION AND INTERSECTION IN THE MAGIC HOUSE



P(Right 2) = 1/36

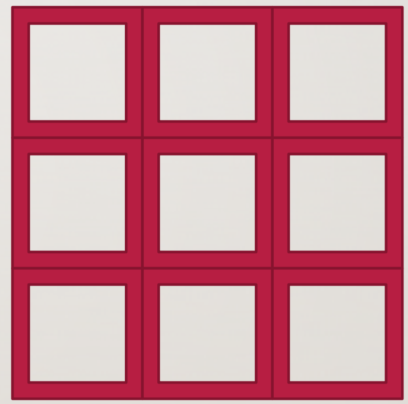


P(B) = 1/6

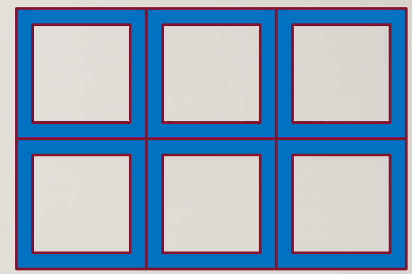
1/36 x 1/6 = 1/216

GROUP WORK

- What if you had to pick one culprit and two accomplices? P =
- What if you had to pick two and two? P =



Culprit



Accomplice

FIRST PROBLEM – ONE CULPRIT / TWO ACCOMPLICES

- 1/9
- P = 6! / (6-2)!2!
- 720 / 24(2) = 1/15
- 1/9 * 1/15 = 1/135

SECOND PROBLEM - TWO AND TWO

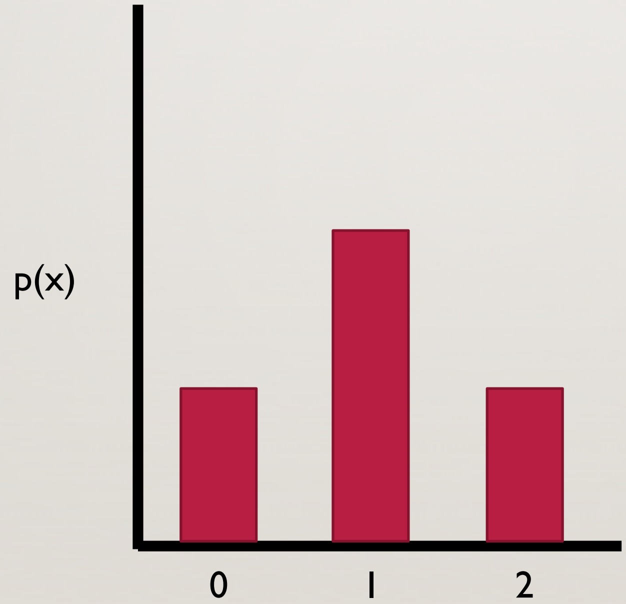
- P(Two Culprits) = 1/36
- P(Two Accomplices) = 1/15
- 1/36 * 1/15 = 1/540

DISCRETE PROBABILITY

- We can use these discrete probabilities to create a distribution function
- Consider number of heads when flipping a coin

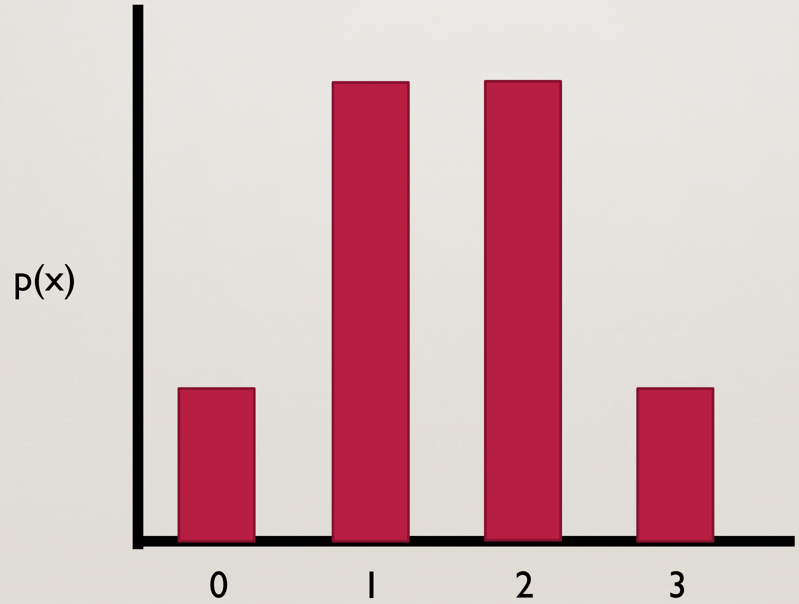
FLIP TWO TIMES

- TT, TH, HT, HH

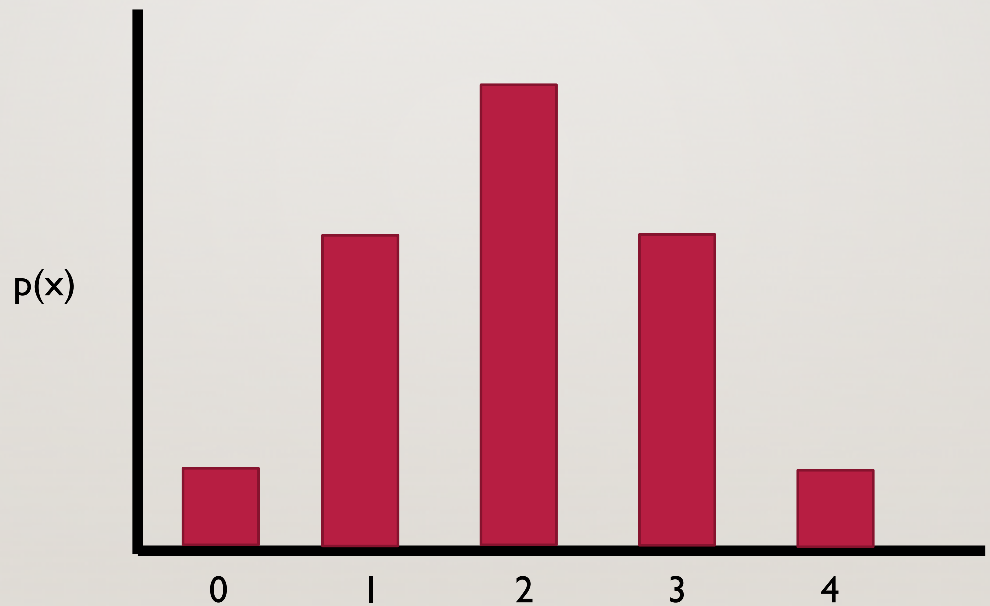


FLIP THREE TIMES

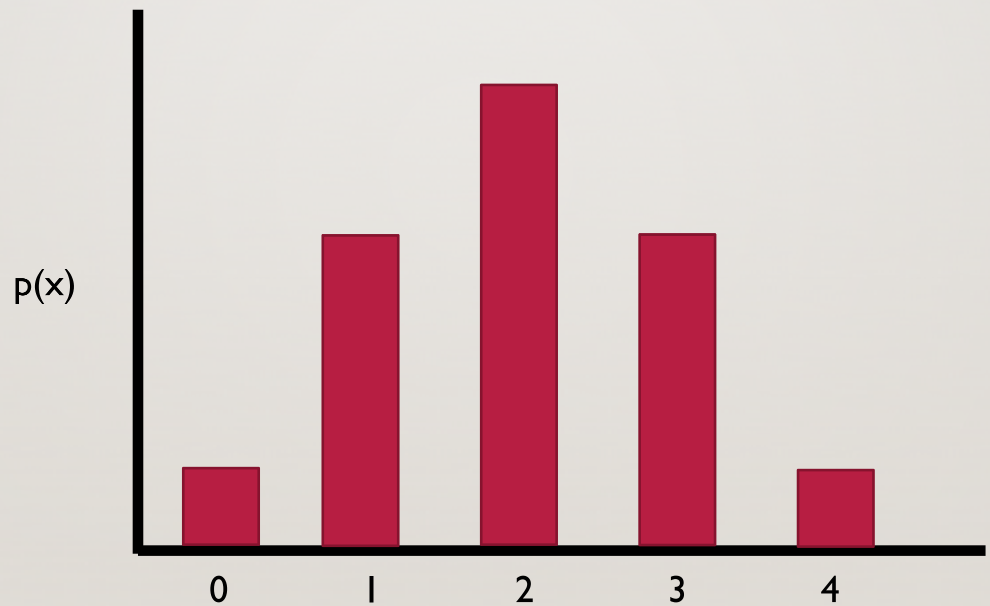
- TTT, TTH, THT, THH, HTT, HTH, HHT, HHH



FLIP FOUR TIMES



PROBABILITY MASS FUNCTION



COUNTING RULE: COMBINATIONS

- Total number of ways of selecting r distinct combinations of N objects, irrespective of order

$$\frac{N!}{r!(N-r)!} = \binom{N}{r}$$

- Called a binomial coefficient
- What are the odds of getting three heads out of five flips?

DEVELOPING A PROBABILITY DISTRIBUTION

- For *N* trials, only need to know two *parameters*: # of successes (r) and probability of a success (p)
  - $p^r q^{N-r}$

- For *N* trials, only need to know two *parameters*: # of successes (r) and probability of a success (p)
  - $p^r q^{N-r}$
  - $p^r(1-p)^{N-r}$
- This gives you the probability for a particular sequence with r correct

WHAT IS THE PROBABILITY OF GETTING THREE HEADS?

How many different sequences have 3 successes?

HHHTT, HTHTH, TTHHH, THTHH…

COIN FLIPS

- X = # of heads after 5 flips
- p = success = head
- $p(X = 3)$

COIN FLIPS

- X = # of heads after 5 flips
- p = success = head
- p(X = 3)

$$\frac{N!}{r!(N-r)!} = \binom{N}{r} = \binom{5}{3} =$$

COIN FLIPS

- X = # of heads after 5 flips
- p = success = head
- p(X = 3)

$$\frac{N!}{r!(N-r)!} = \binom{N}{r} = \binom{5}{3} =$$

$$\binom{5}{3} p^r q^{N-r} = 10(.5^3)(.5^2)$$

COIN FLIPS

- X = # of heads after 5 flips
- p = success = head
- $p(X = 2)$?

BINOMIAL DISTRIBUTION

$$\binom{N}{r} p^r q^{N-r}$$

BINOMIAL DISTRIBUTION

$$p(X = r); N, p = \binom{N}{r} p^r q^{N-r}$$

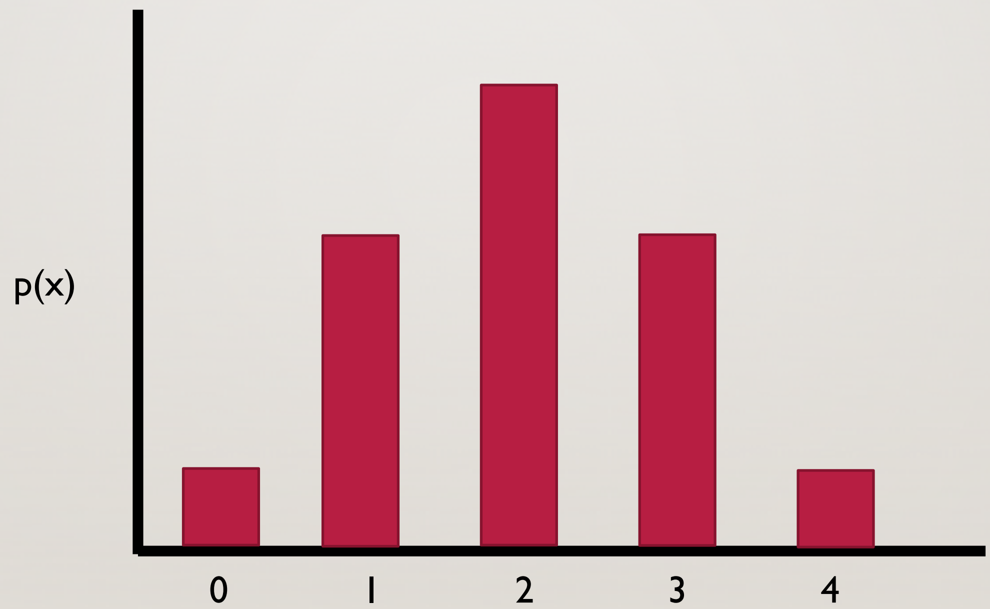
BINOMIAL DISTRIBUTION

$$p(X = r); N, p = \binom{N}{r} p^r q^{N-r}$$

FAMILY OF DISTRIBUTIONS

- By changing our parameters we can change the distribution

FLIP FOUR TIMES



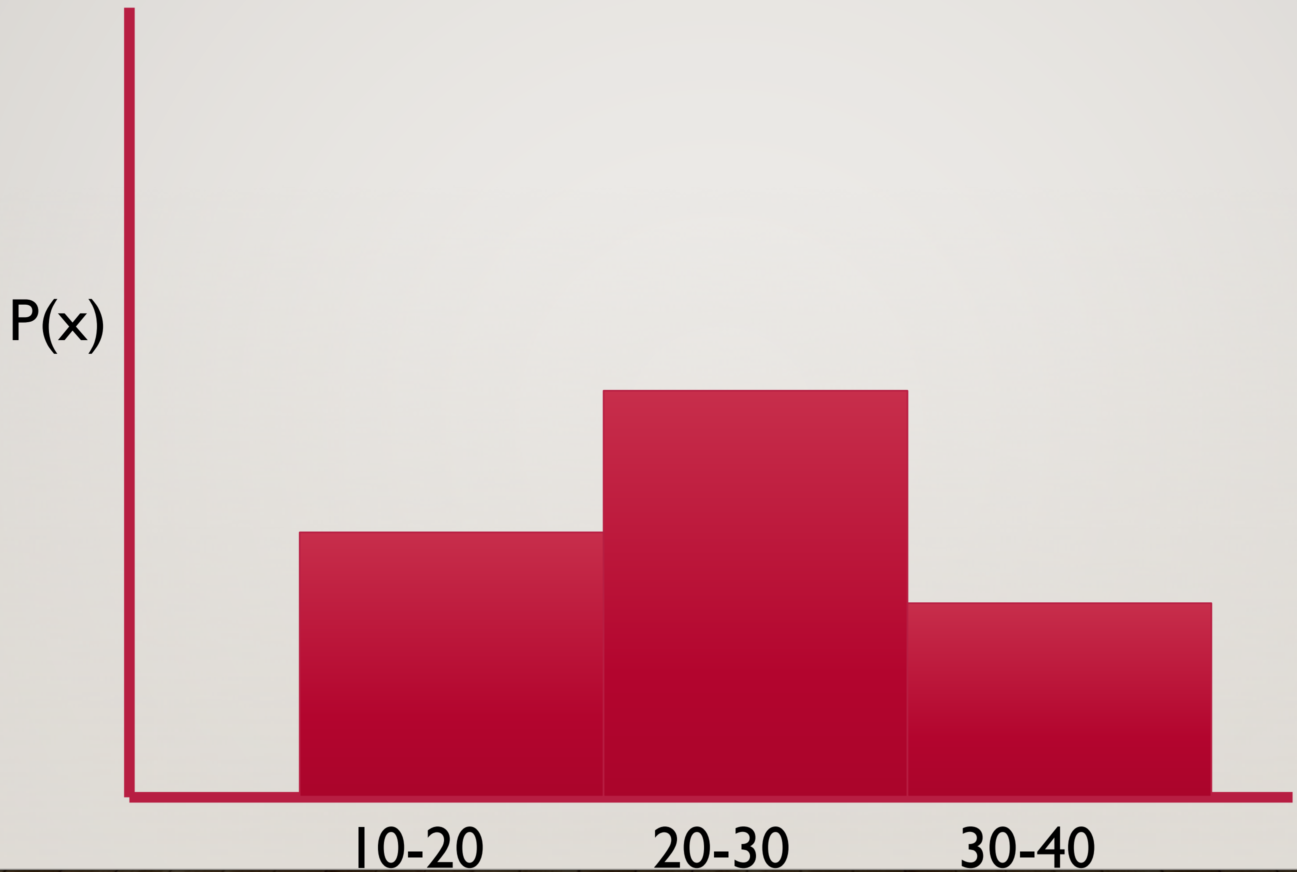
CENTRAL TENDENCY & DISPERSION

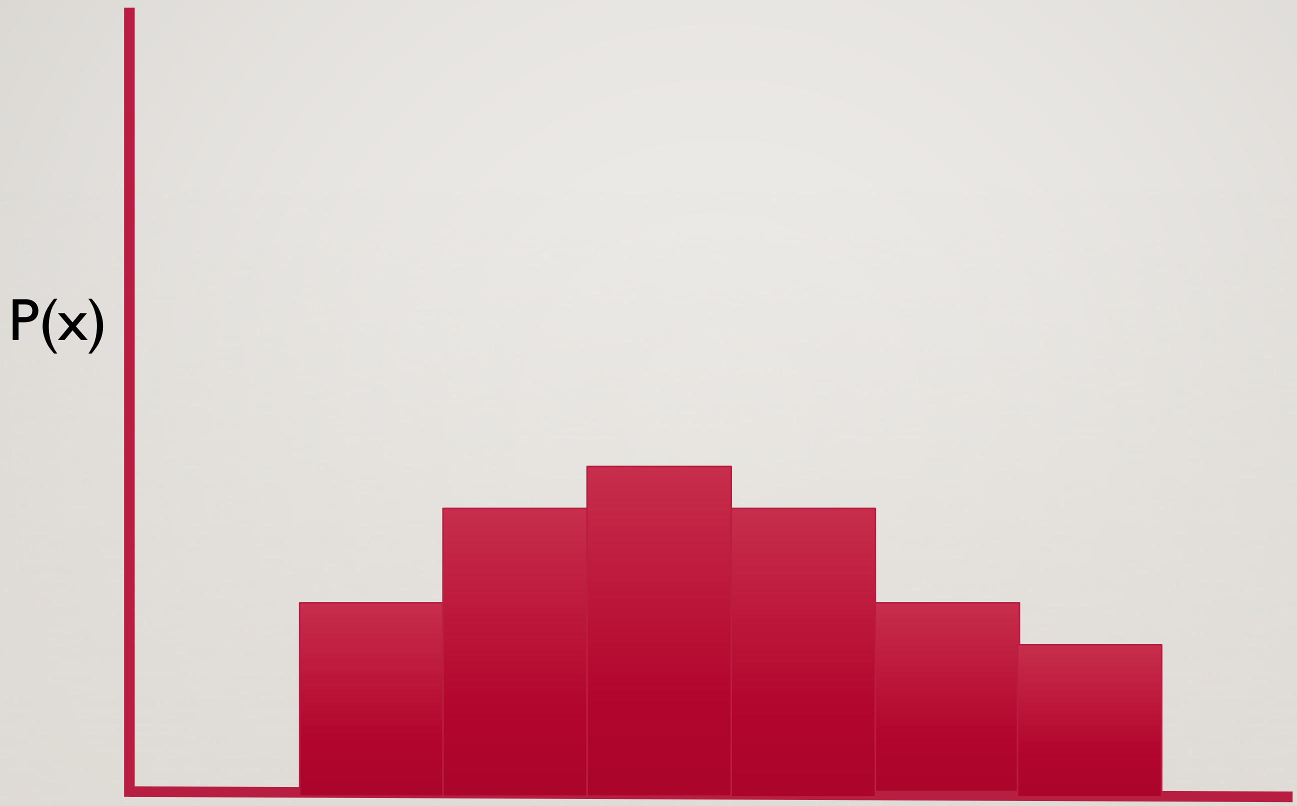
- Mean = np
- Variance = np(1 – p)
- Standard deviation = $\sqrt{np(1-p)}$

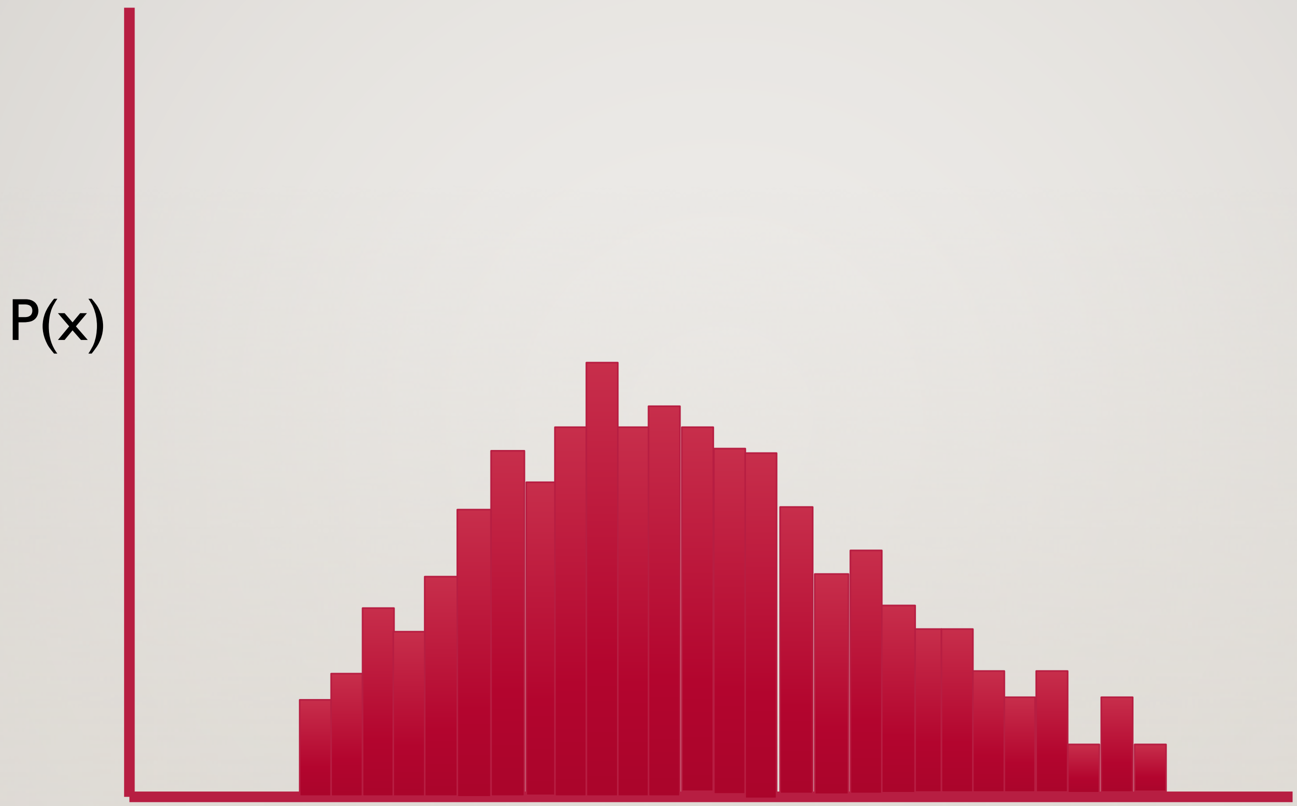
CONTINUOUS VARIABLES

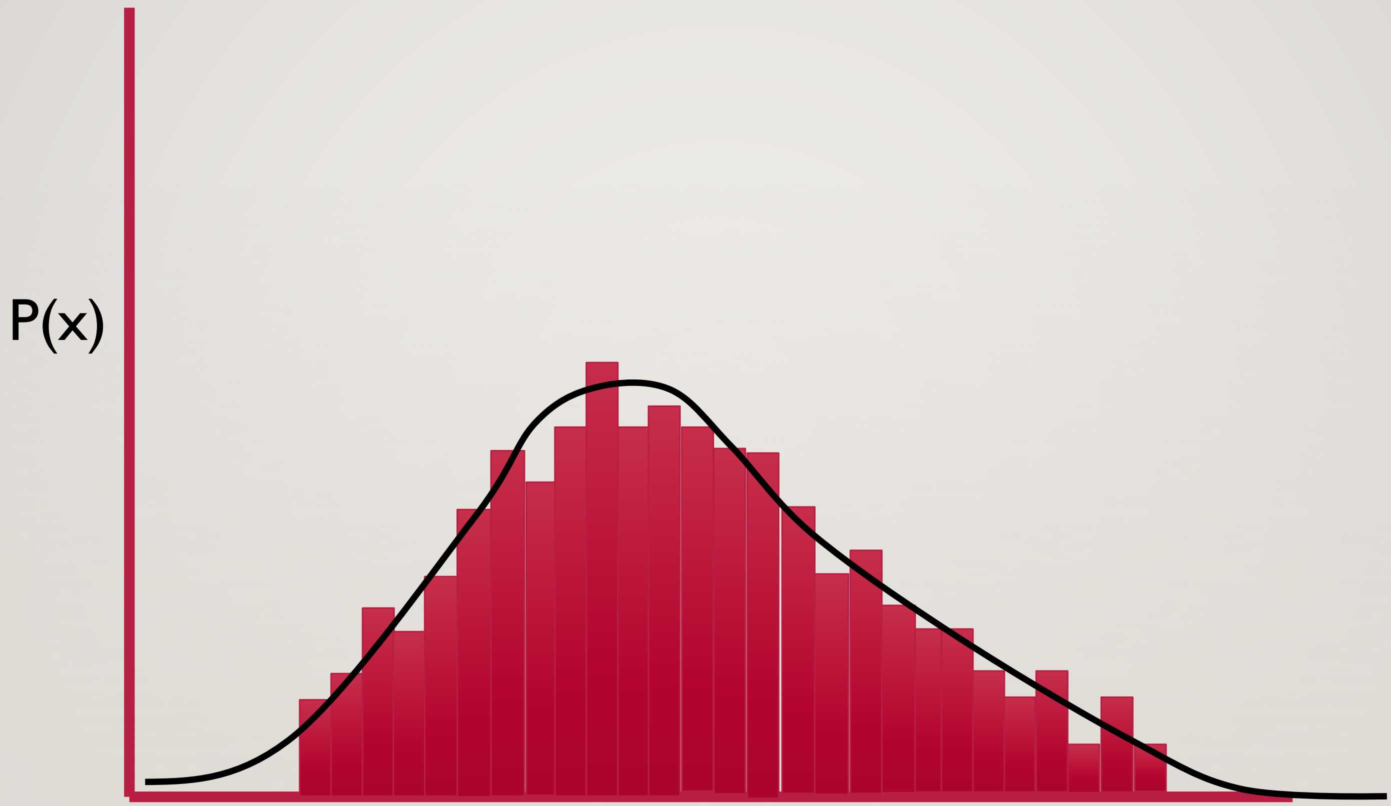
CONTINUOUS RANDOM VARIABLES

- Age

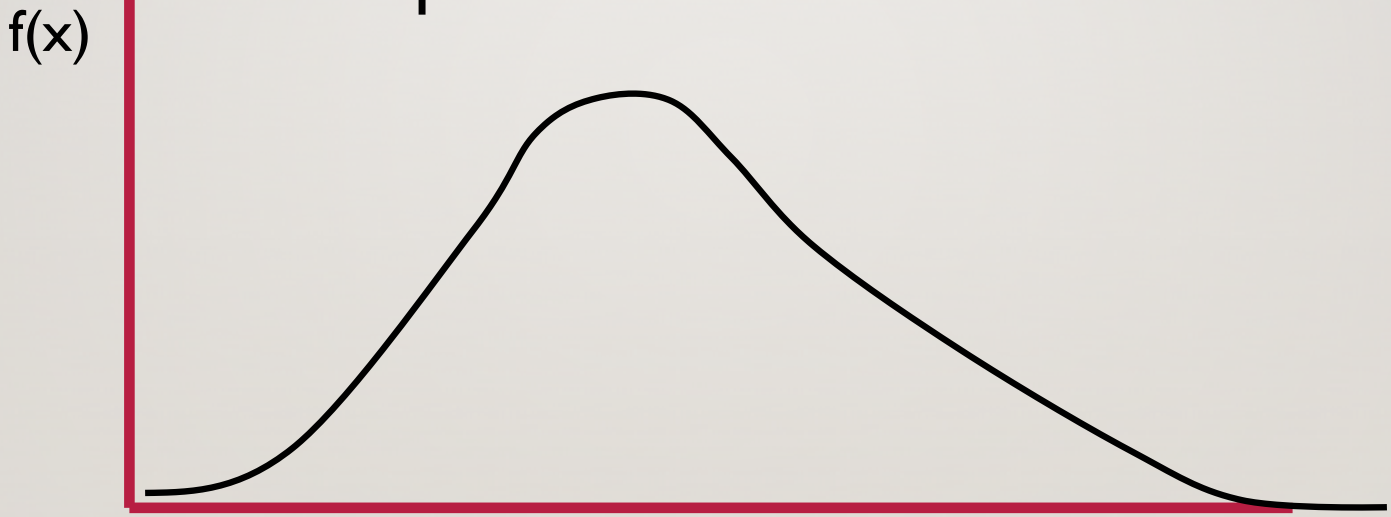

P(x)
10-20
20-30
30-40


P(x)

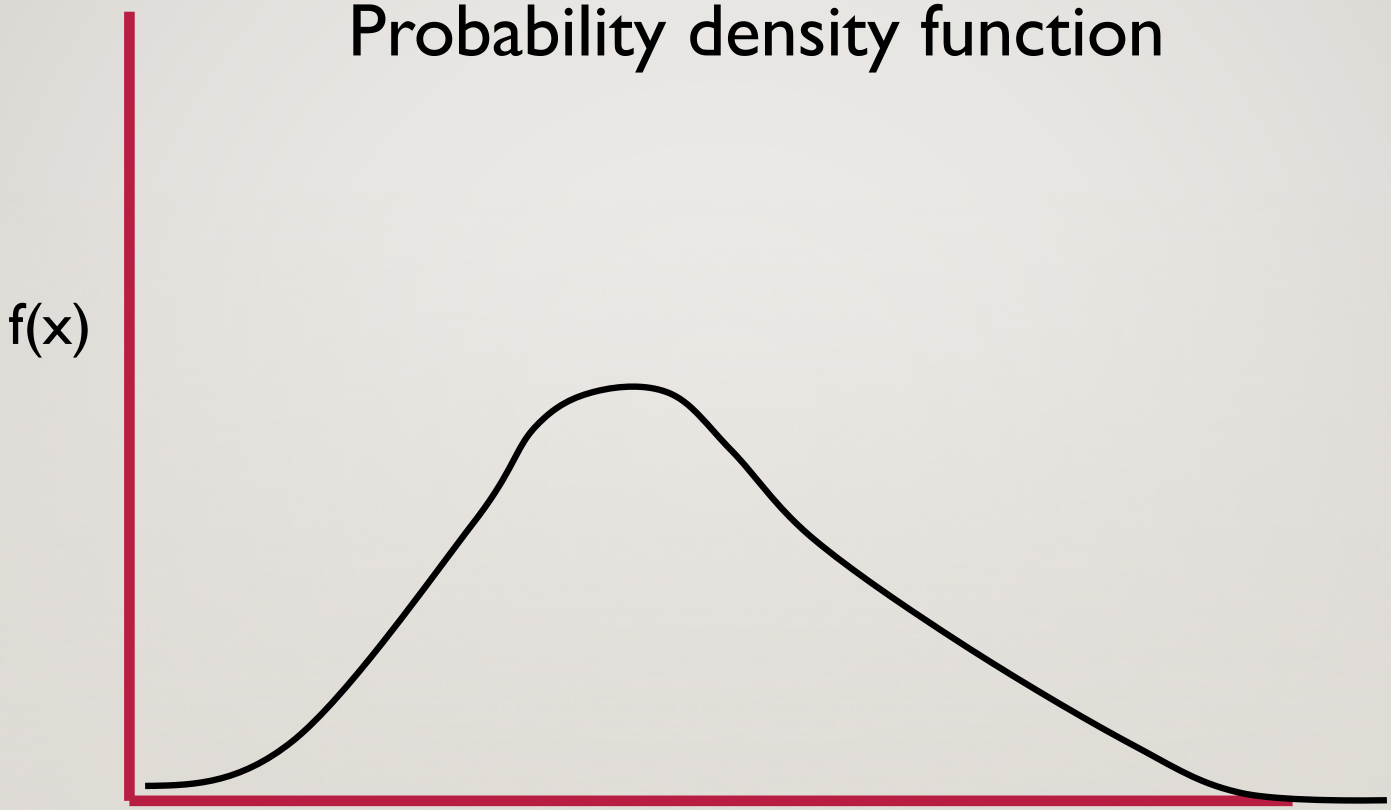

P(x)

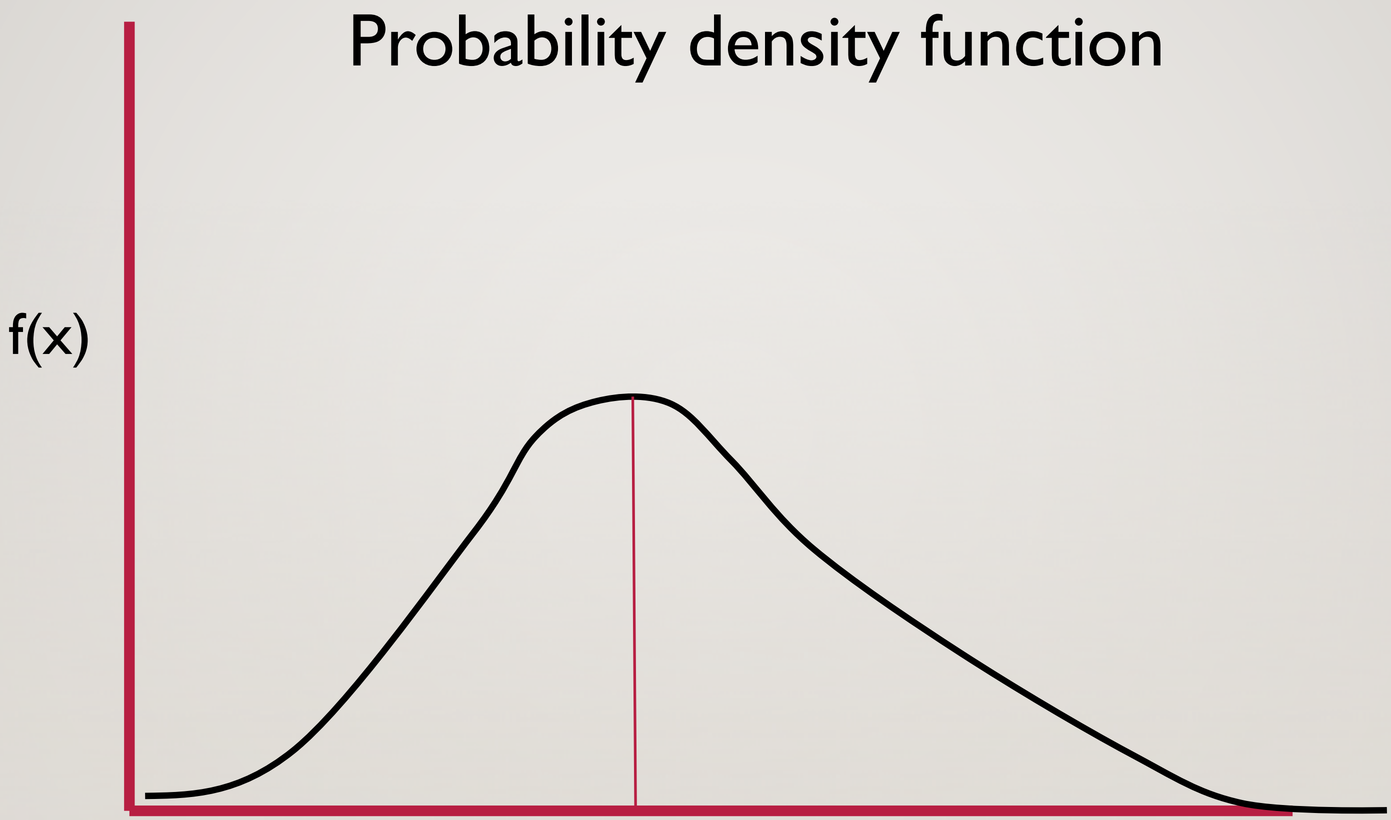

P(x)

Probability density function (PDF)
a function that describes the relative likelihood for the random variable to take a particular



Probability density function

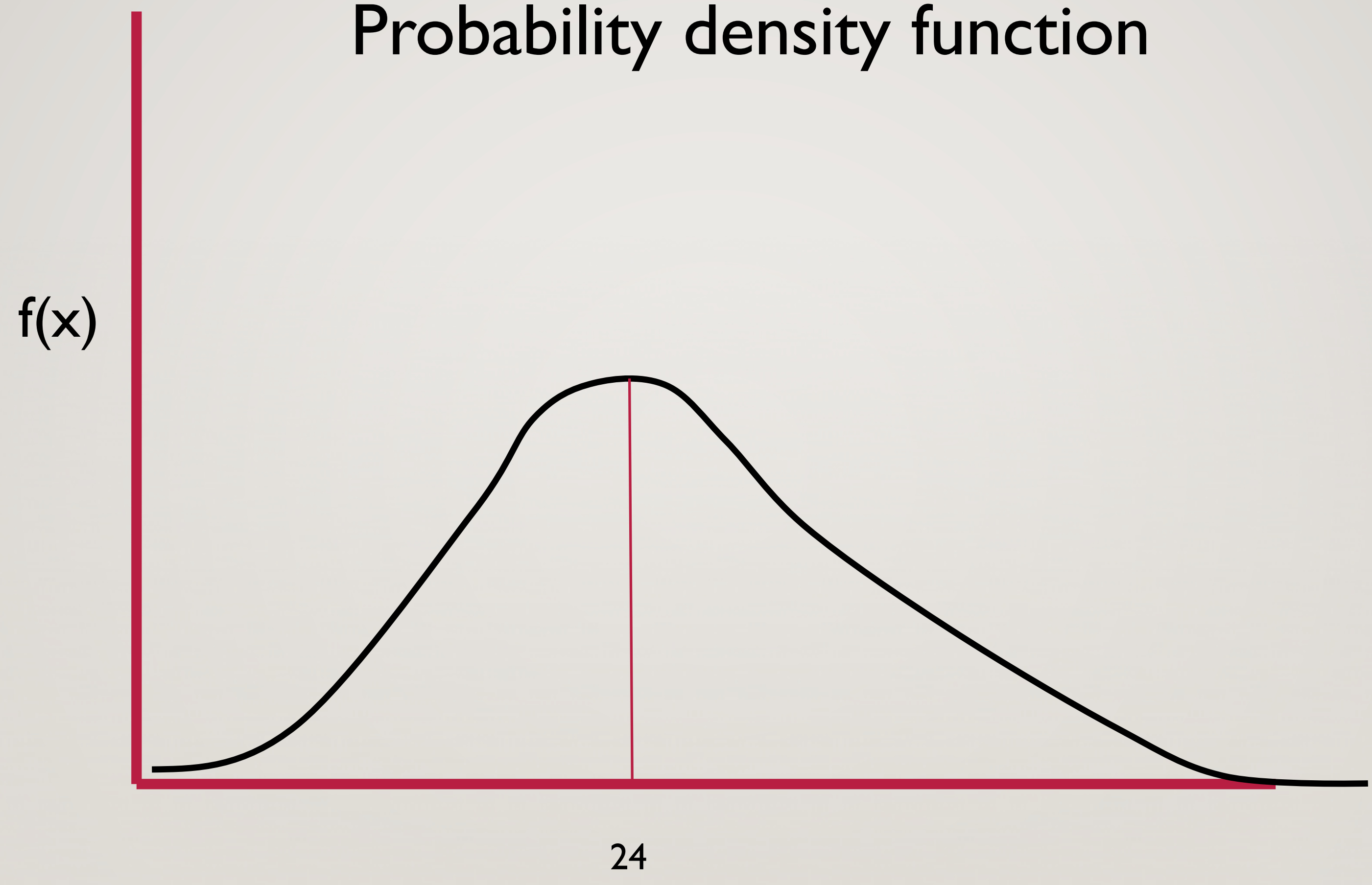



Probability density function
f(x)

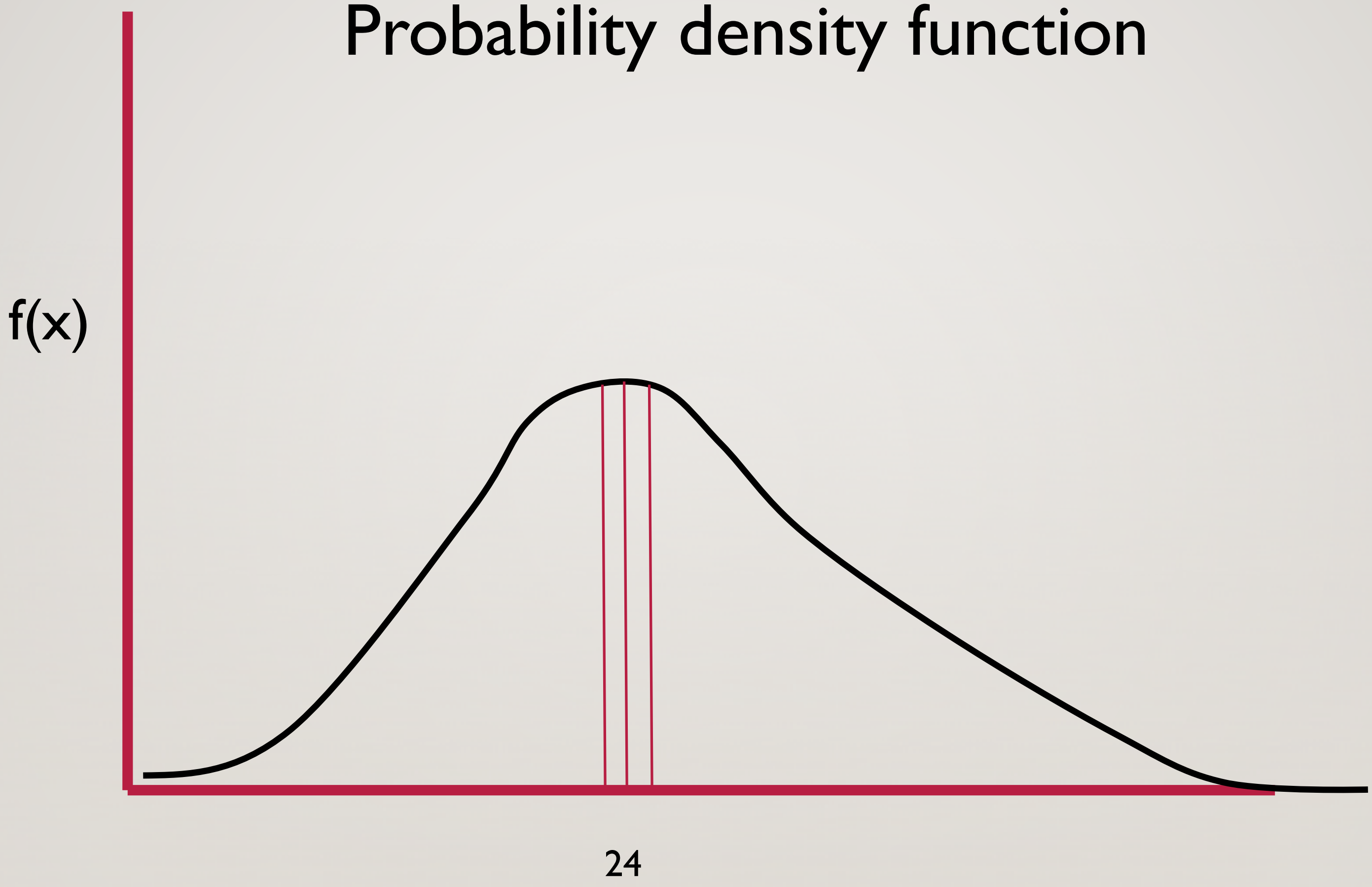
PROBABILITY DENSITY FUNCTION

- $p(23.5 \leq X \leq 24.5)$

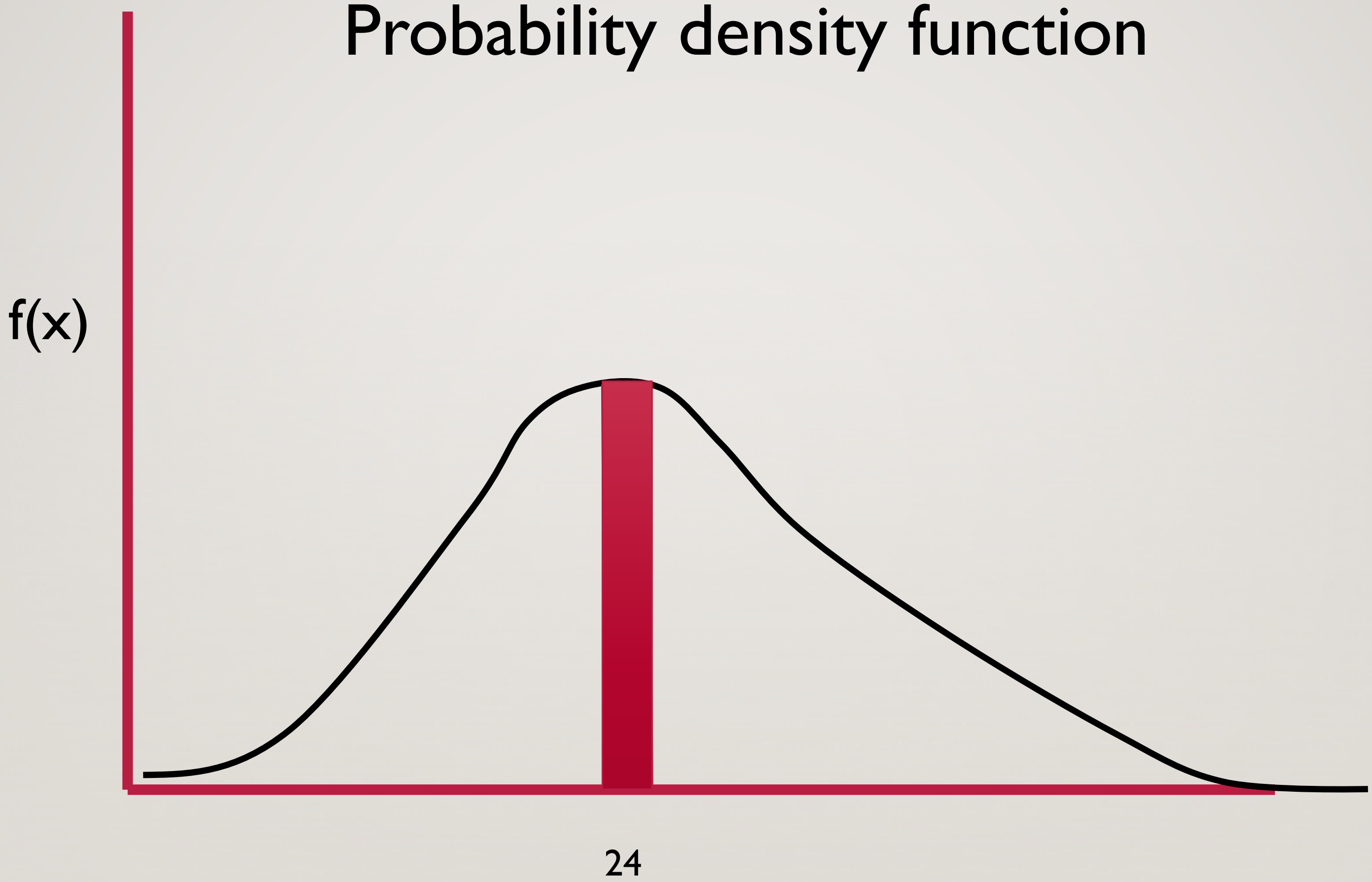
Probability density function



Probability density function



Probability density function

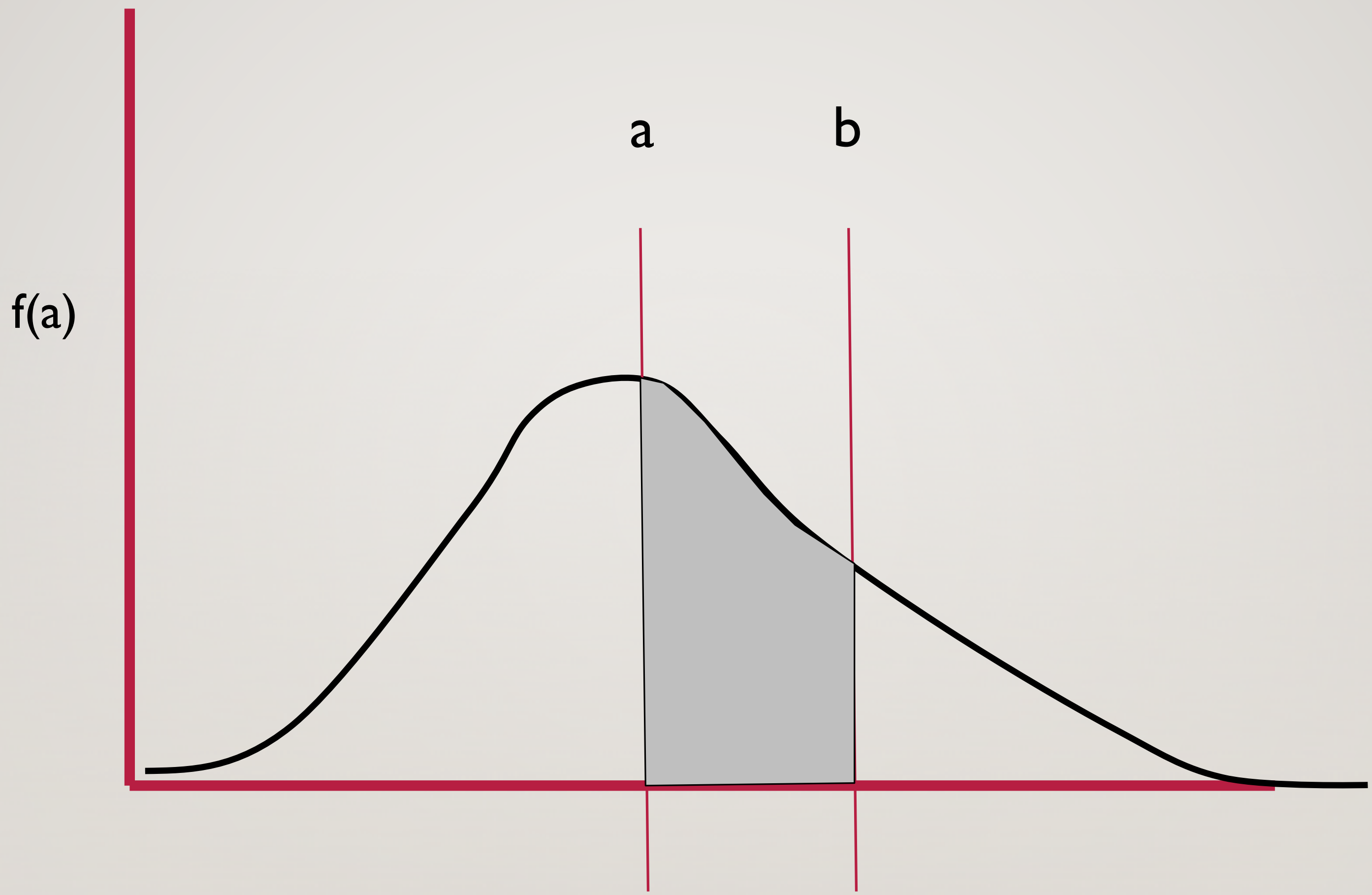


PROBABILITY DENSITY FUNCTION

- $p(23.5 \le X \le 24.5)$

$$p(a \le X \le b) = \int_a^b f(x)\,dx$$

- Probability = area under the curve


a
b
f(a)