The Central Limit Theorem

Sec. 8.1: The Random Variable $\bar{X}$ and it's Distribution

Sec. 8.2: The Random Variable $\hat{p}$ and it's Distribution

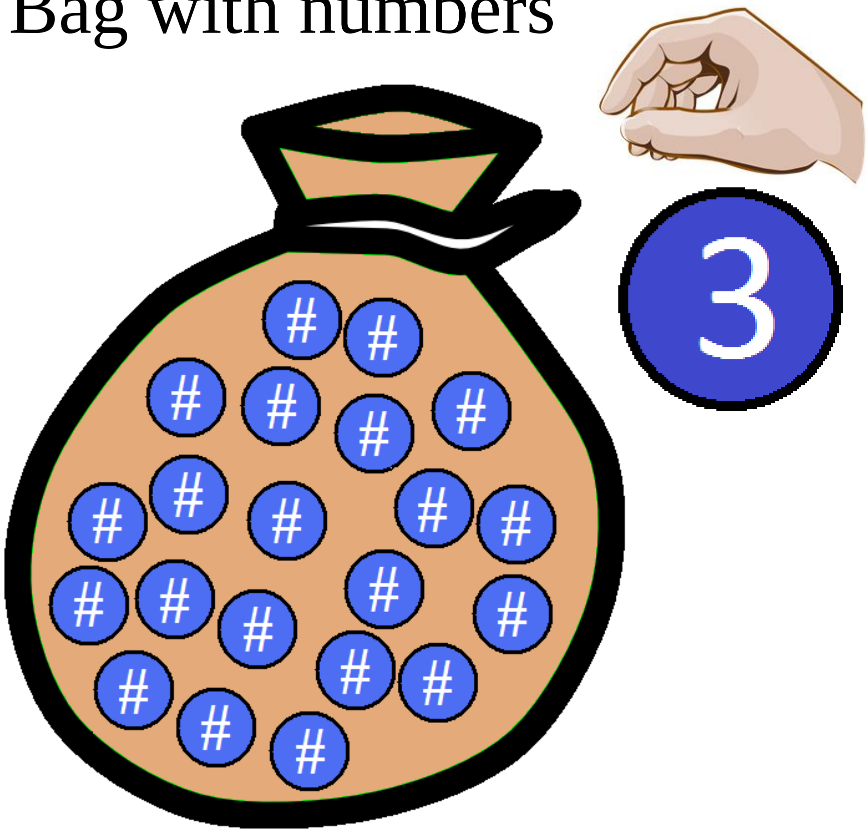
How Should You Think of a Random Variable?

Imagine a bag with numbers in it. You draw one number from the bag and the random variable X is the number that you drew from the bag.

- How did the bag with numbers get there? From some experiment in the background whose outcomes we relabeled with numbers. But the end product is the bag with numbers and that's what we care about.
- If you know the probability distribution of X then you can answer questions about what numbers can be drawn from the bag and what the chances are (probability) that various numbers are drawn from the bag.

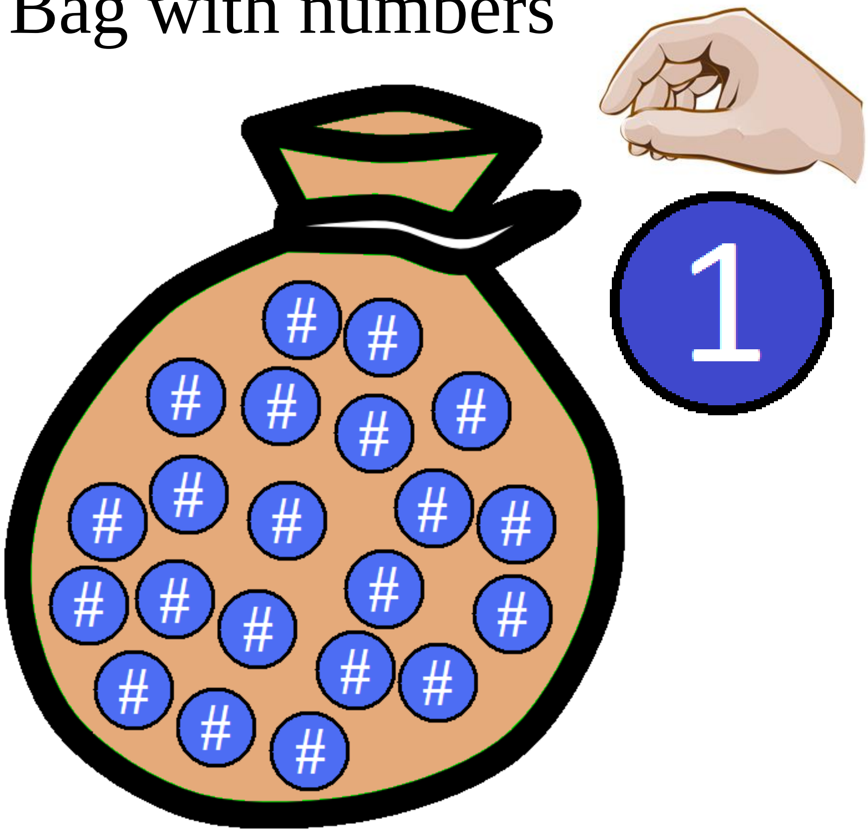
How Should You Think of a Random Variable?

Bag with numbers



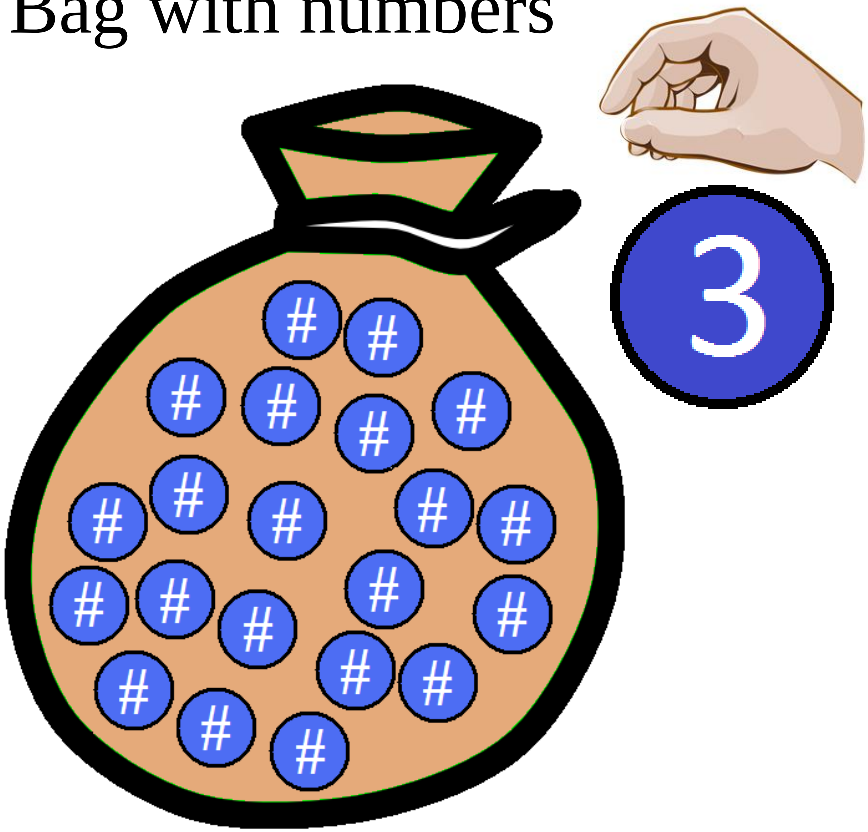
How Should You Think of a Random Variable?

Bag with numbers



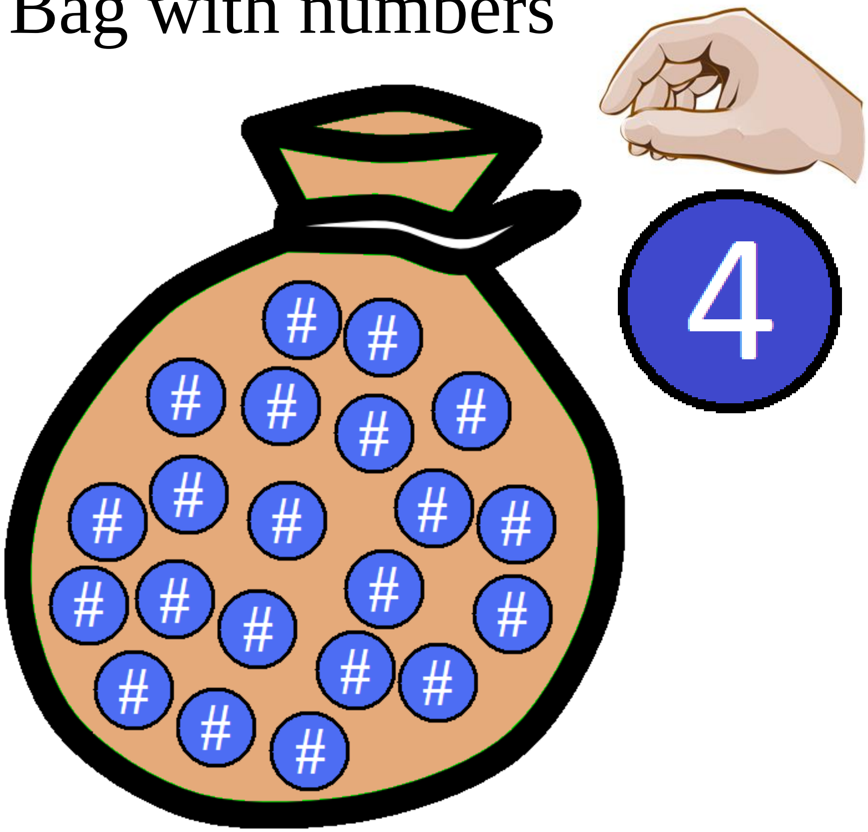
How Should You Think of a Random Variable?

Bag with numbers



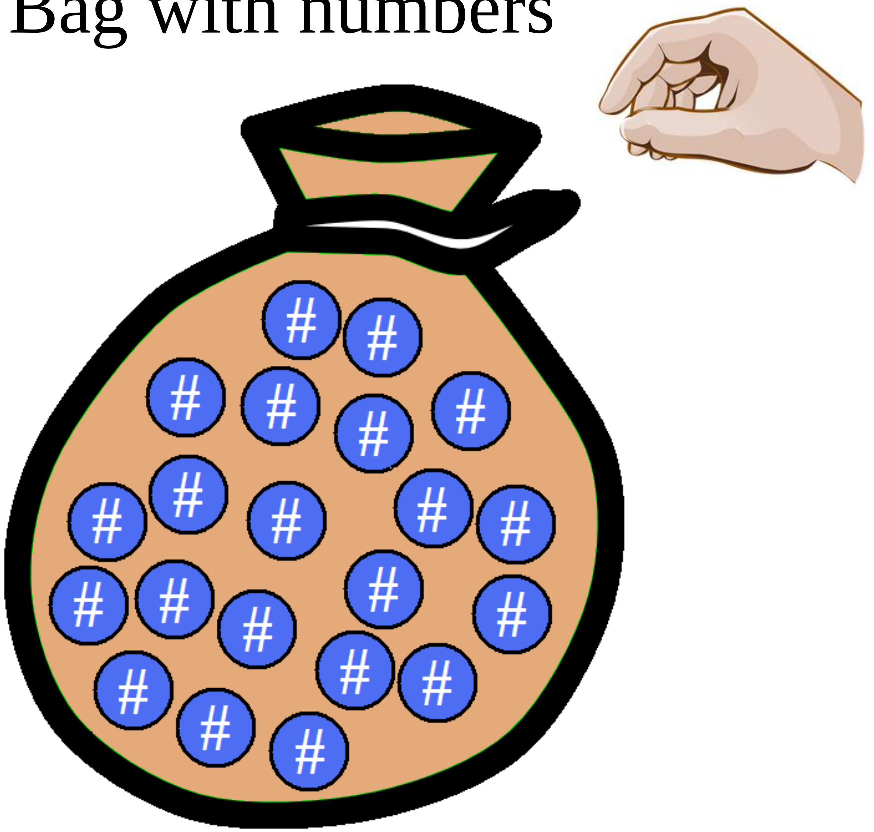
How Should You Think of a Random Variable?

Bag with numbers



How Should You Think of a Random Variable?

Bag with numbers

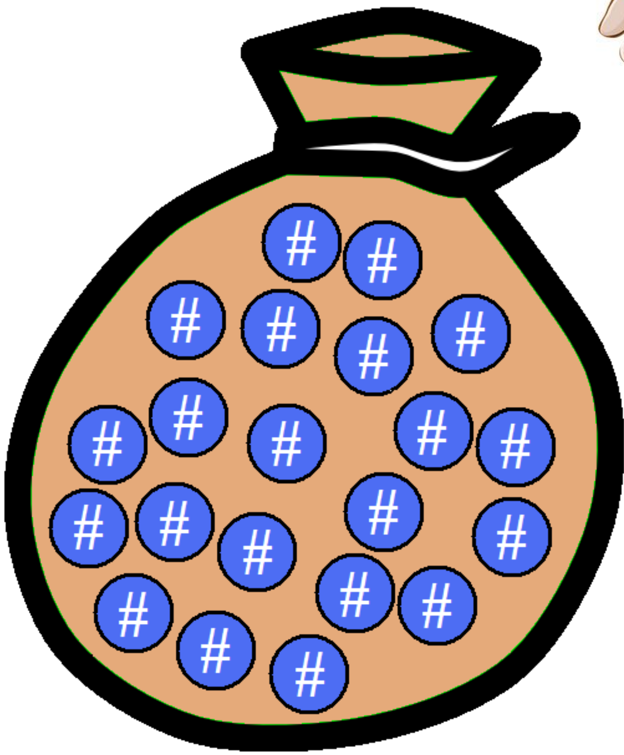


The experiment in the background could be…

- Betting on the draw of a card from a deck
- Counting the total number of heads when you flip a coin 4 times
- or something else

How Should You Think of a Random Variable?

Bag with numbers



If you know the probability distribution, say…

x	0	1	2	3	4
P(X=x)	0.05	0.15	0.10	0.50	0.20

Then I know

- The possible numbers that can be drawn from the bag
- The probabilities for the different numbers that can come out of the bag
- The average (EV), standard deviation and variance of all numbers in the bag

How Should You Think of a Random Variable?

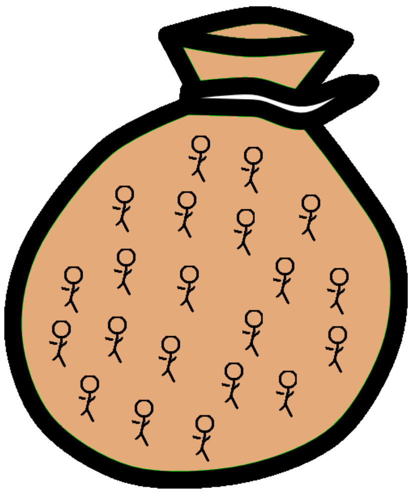
Note about random variables:

A random variable is not a fixed quantity, it varies (variable).

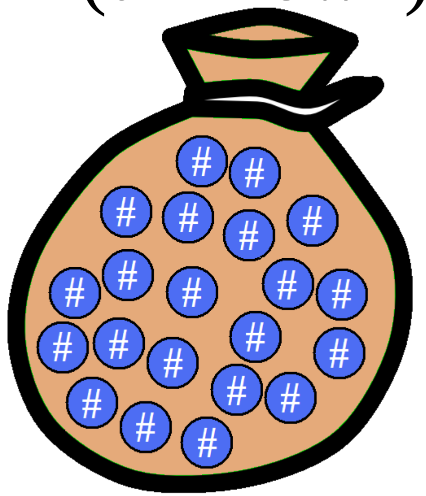
Every time you perform the experiment (or draw a number from the bag), you get a new value for the random variable.

<u>Recall:</u>

Population



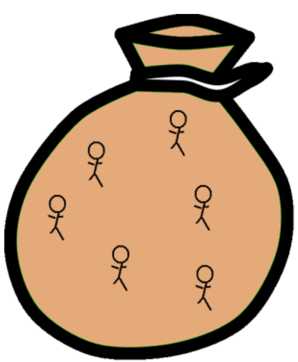
Population Data (unknown)



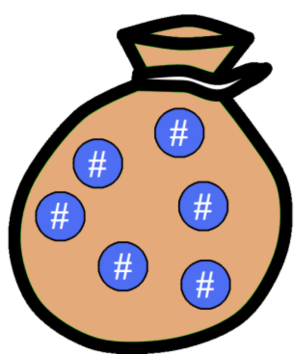
Population Parameters (unknown)

$$p \quad \mu \quad \sigma \quad \sigma^2$$

Sample



Sample Data (known)



Sample Statistics (known)

$$\hat{p} \quad \bar{x} \quad s \quad s^2$$

These are Random Variables!

Sec. 8.1: How is $\bar{X}$ a Random Variable?

Original Bag

- Start off with a bag of numbers (population data). The random variable X is the number you get if you draw 1 number from the bag.
- The population parameter μ = the average of all of the numbers in the entire original bag, which is usually unknown, is known in this section. σ will also be known

<u>Sec. 8.1</u>: How is $\bar{X}$ a Random Variable?

New Bag

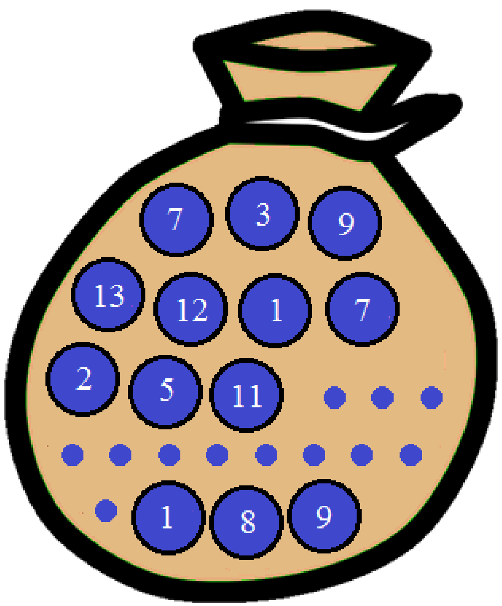
- Form a new bag of numbers as follows: Take a sample of numbers of size *n* from the original bag (population data) and calculate the average of the numbers in your sample. This is $\bar{X}$. Do this for all possible samples of size *n* to form your new bag. The random variable $\bar{X}$ is the number you get when you draw 1 number from this new bag.
- You can think of getting a value of $\bar{X}$ in 2 different ways:

1. Draw a sample of size *n* from the original bag and calculate the average, OR
2. Draw a single number from the new $\bar{X}$ bag

<u>Sec. 8.1</u>: How is $\bar{X}$ a Random Variable?

Picture of the situation

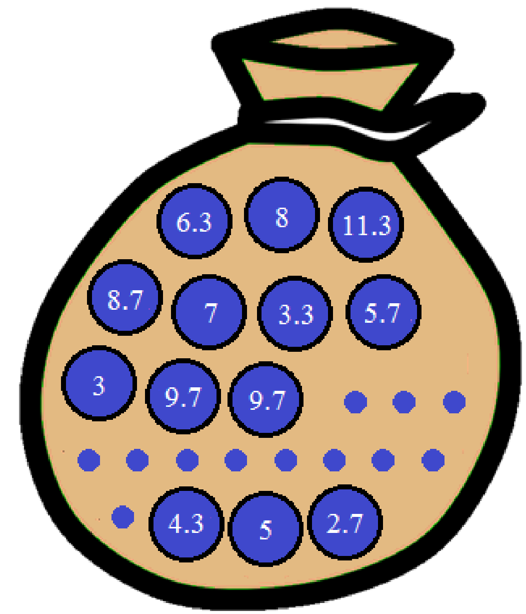
X



- Take a sample of size n
- Calculate $\bar{X}$ (the average of the numbers in sample)
- Put answer in bag on right

- Do this for <u>all</u> possible samples of size n

$\bar{X}$



- Original bag of numbers (population data)
- μ = the average of all numbers in the entire original bag <u>is known in this section</u> (usually unknown). σ is also known

- New bag has all sample averages in it
- Taking a sample of size n and calculating $\bar{X}$ is equivalent to drawing 1 number from the new bag

Sec. 8.1: How is $\bar{X}$ a Random Variable?

Notes:

- There are 3 different ways to draw a sample of size *n* from a bag:
  1. Draw one by one with replacement
  2. Draw one by one without replacement
  3. Draw all *n* at once

  When we draw samples, we will always think of them as drawn the first way (with replacement), but as long as the condition $n \leq 0.05N$ is met, all probabilities regarding $\bar{X}$ will be essentially the same.

Sec. 8.1: The Probability Distribution of $\bar{X}$

What does the probability distribution of $\bar{X}$ look like?

Ex: Suppose a bag contains the numbers 1, 2, 3, 4, 5, 6 in it, each appearing only once in the bag. Let

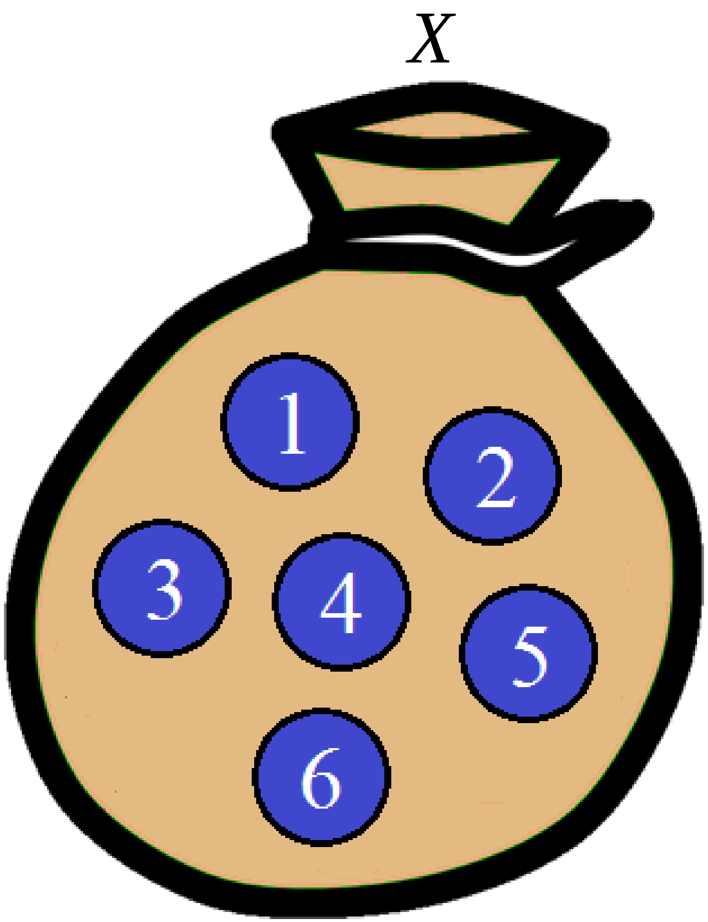
X = The number you get when you draw 1 number from this original bag

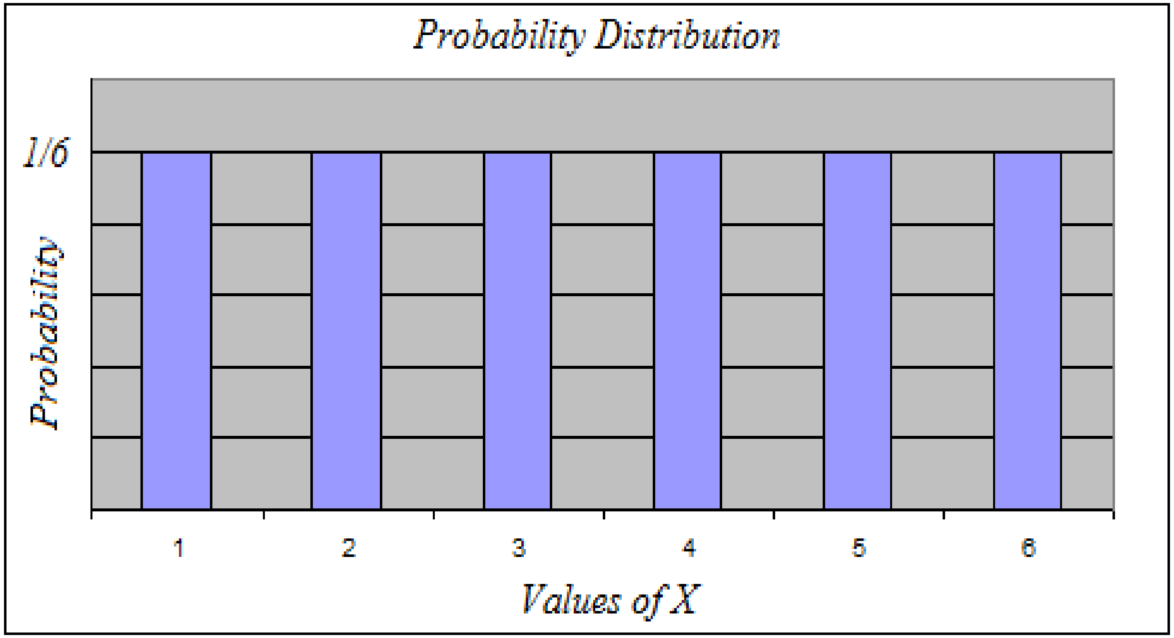
$\overline{X_2}$ = The average of a sample of size 2 drawn from the original bag (or the number you get when you draw 1 number from the $\overline{X_2}$ bag

$\overline{X_3}$ = The average of a sample of size 3 drawn from the original bag (or the number you get when you draw 1 number from the $\overline{X_3}$ bag

Sec. 8.1: The Probability Distribution of $\bar{X}$

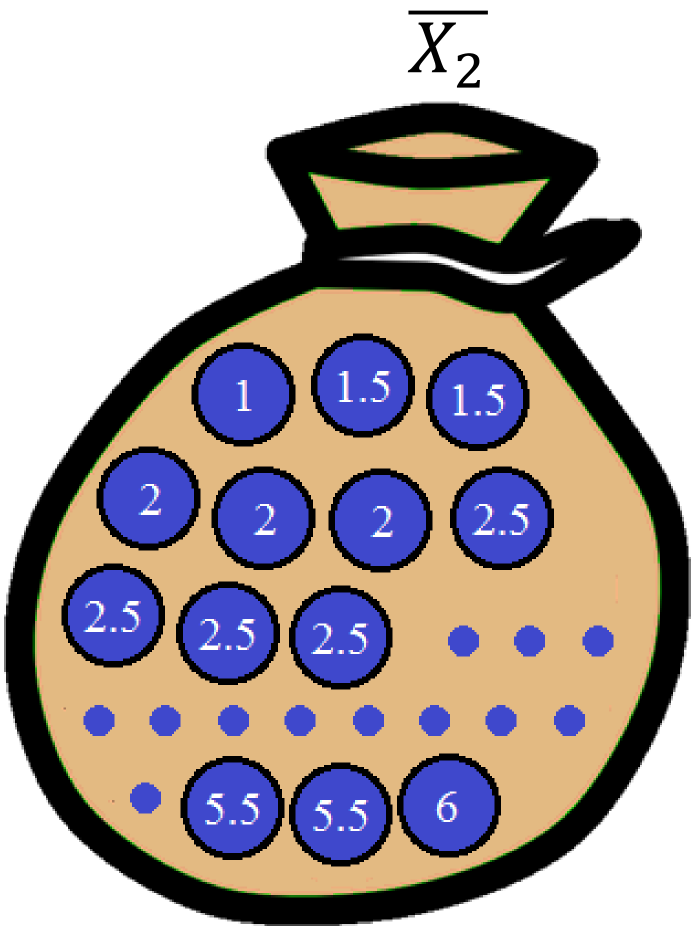
What does the probability distribution of $\bar{X}$ look like?

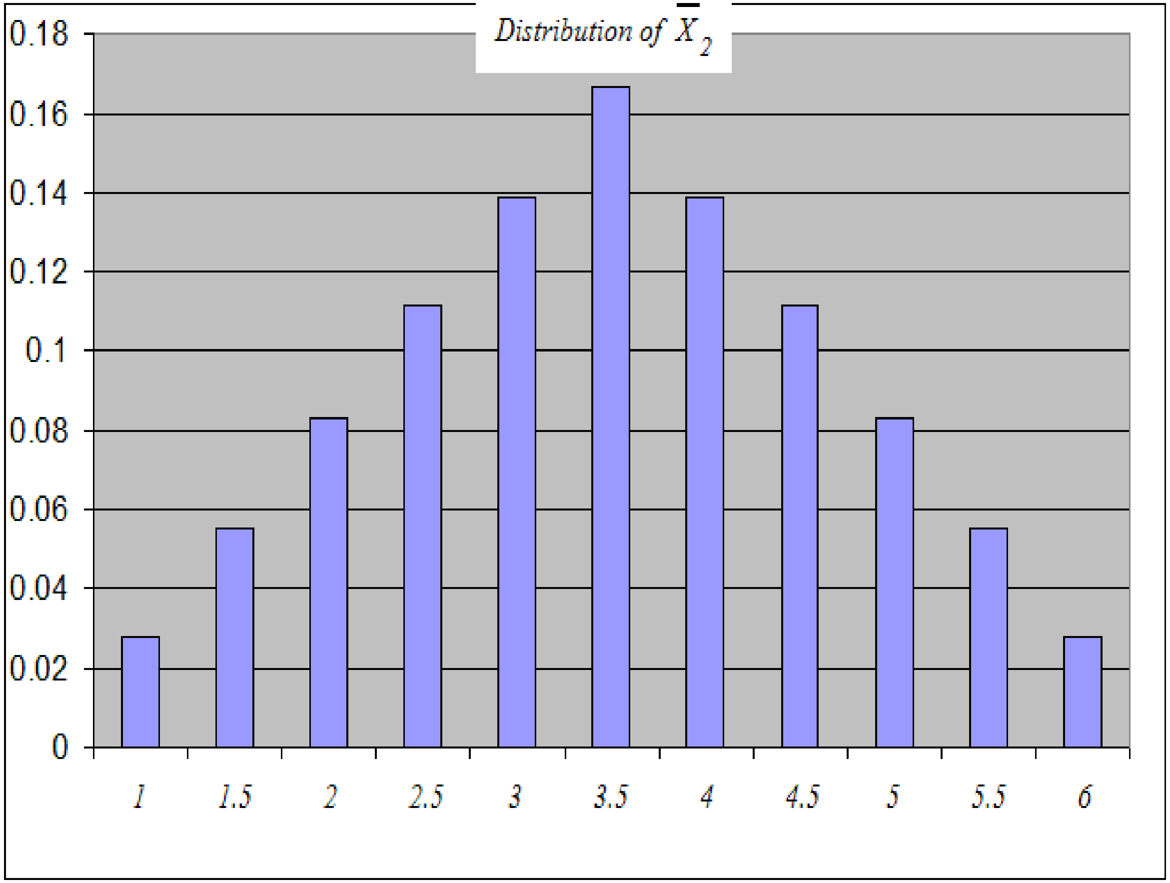




Sec. 8.1: The Probability Distribution of $\bar{X}$

What does the probability distribution of $\bar{X}$ look like?

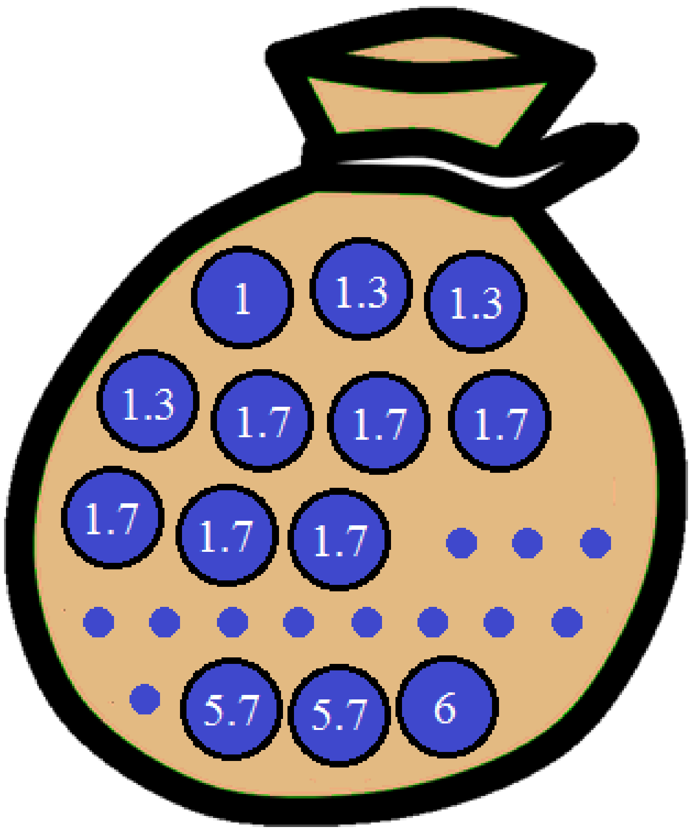


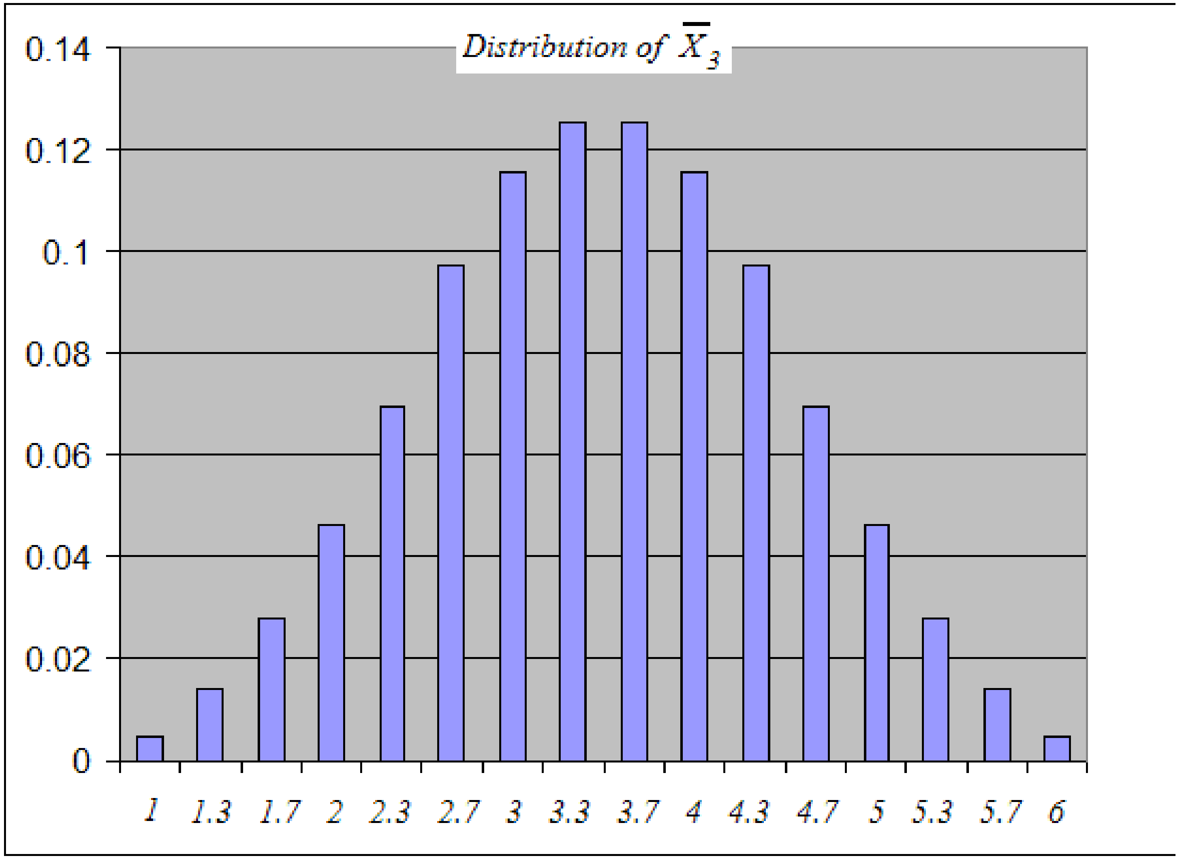


Sec. 8.1: The Probability Distribution of $\bar{X}$

What does the probability distribution of $\bar{X}$ look like?

$\overline{X_3}$



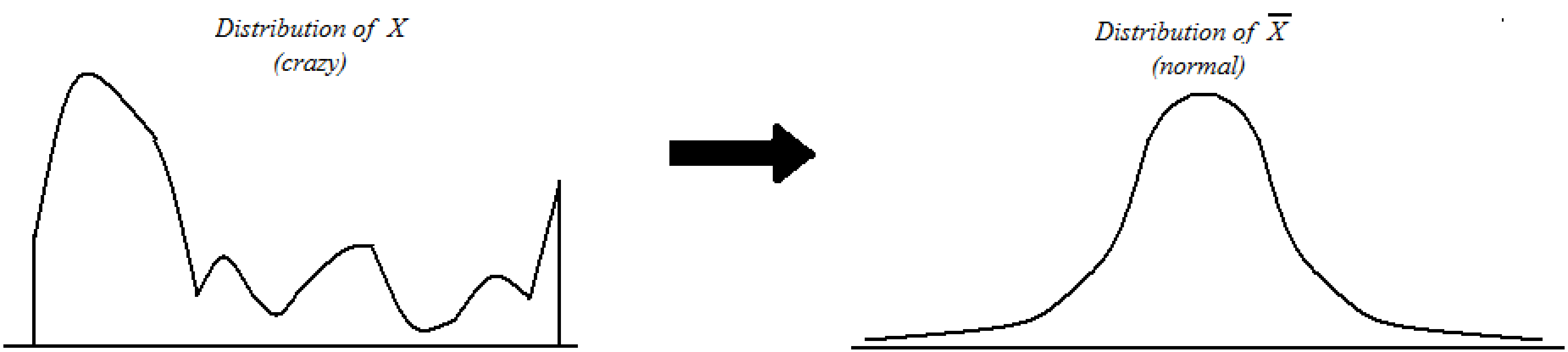


Sec. 8.1: The Probability Distribution of $\bar{X}$

Central Limit Theorem (for $\bar{X}$)

Suppose $n \leq 0.05N$

- If X is any random variable whatsoever and $n \geq 30$, then $\bar{X}$ has a normal distribution.
- If X has a normal distribution to begin with, then $\bar{X}$ has a normal distribution no matter what n is.
- $\mu_{\bar{X}} = \mu_X$ and $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}$



Ex 1: A bag of numbers has a normal distribution with a mean of 10 and a standard deviation of 4. Find

a) the probability that a single number drawn from the bag is less than 21
b) the probability that a single number drawn from the bag is larger than 5
c) the probability that a single number drawn from the bag is between 4.5 and 11.3
d) the probability that in a sample of size 20, the average is less than 9.5
e) the probability that in a sample of size 20, the average is at least 11.3
f) the probability that in a sample of size 20, the average is between 10.4 and 11.7

Ex 2: The amount of credit card debt among all Americans ages 18 to 24 has a mean of \$2,982 and a standard deviation of \$315. What is the probability that in a randomly selected group of 36 Americans between the ages of 18 and 24 that

a) their average credit card debt is at most \$3050?
b) their average credit card debt is more than \$2850?
c) their average credit card debt is between \$2900 and \$3000?

Sec. 8.2: How is $\hat{p}$ a Random Variable?

Original Bag

- Start off with a bag of numbers consisting of 0's and 1's only (population data). Think of 0's as standing for no's and 1's as standing for yes's. The random variable X is the number you get if you draw 1 number from the bag.
- The population parameter p = the percentage of 1's in the entire original bag, which is usually unknown, is known in this section.

Sec. 8.2: How is $\hat{p}$ a Random Variable?

New Bag

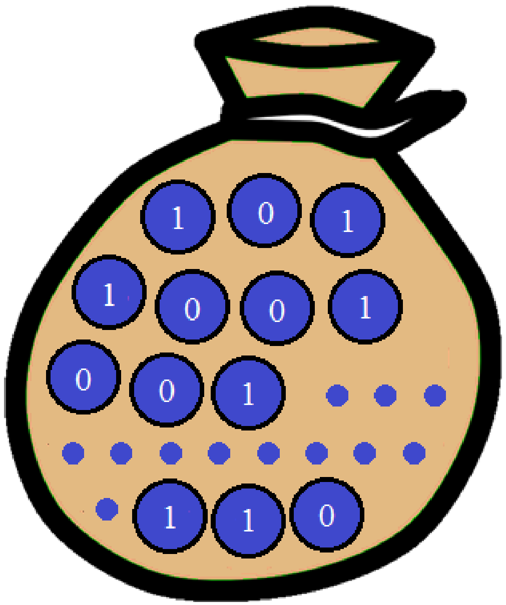
- Form a new bag of numbers as follows: Take a sample of 0's and 1's of size *n* from the original bag (population data) and calculate the percentage of 1's in your sample. This is $\hat{p}$. Do this for all possible samples of size *n* to form your new bag. The random variable $\hat{p}$ is the number you get when you draw 1 number from this new bag. This new bag does not contain 0's and 1's in it, but instead has many different percentages in it.
- You can think of getting a value of $\hat{p}$ in 2 different ways:

  1. Draw a sample of size *n* from the original bag and calculate the percentage of 1's in the sample, OR
  2. Draw a single number from the new $\hat{p}$ bag

Sec. 8.2: How is $\hat{p}$ a Random Variable?

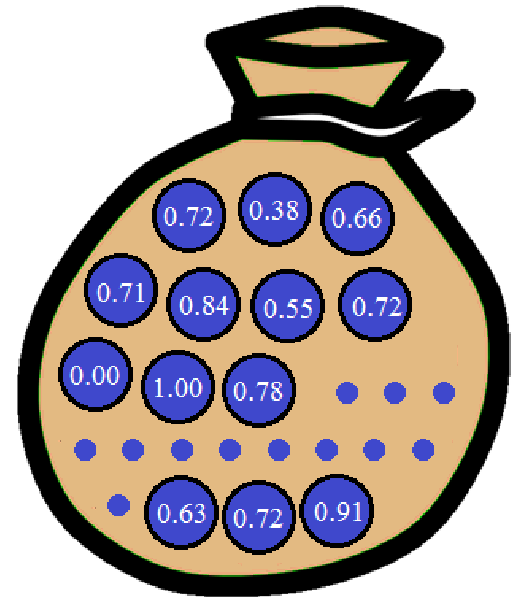
Picture of the situation

X



- Take a sample of size n
- Calculate $\hat{p}$ (the percentage of 1's in sample)
- Put answer in bag on right
- Do this for all possible samples of size n

$\hat{p}$



- Original bag of 0's and 1's (population data, yes=1, no=0)
- p = the percentage of 1's in the entire original bag is known in this section (usually unknown)

- New bag has all sample percentages in it instead of 0's and 1's
- Taking a sample of size n and calculating $\hat{p}$ is equivalent to drawing 1 number from the new bag

Sec. 8.2: How is $\hat{p}$ a Random Variable?

Notes:

- There are 3 different ways to draw a sample of size *n* from a bag:

  1. Draw one by one with replacement
  2. Draw one by one without replacement
  3. Draw all *n* at once

  When we draw samples, we will always think of them as drawn the first way (with replacement), but as long as the condition $n \leq 0.05N$ is met, all probabilities regarding $\hat{p}$ will be essentially the same.

- $\hat{p}$ is a kind of $\bar{x}$ (explained on board). Therefore central limit applies to $\hat{p}$. But instead of the condition $n \geq 30$, we have the condition $npq \geq 10$. $(p + q = 1)$

Sec. 8.2: The Probability Distribution of $\hat{p}$

Central Limit Theorem for $\hat{p}$

As long as the conditions

1. $n \le 0.05N$
2. $npq \ge 10$ $\qquad (p + q = 1)$

are satisfied, $\hat{p}$ has a NORMAL DISTRIBUTION with

$$\mu_{\hat{p}} = p \quad \text{and} \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Ex 3: A bag of numbers contains only 0's and 1's in it (think of 0's as no's and 1's as yes's). The percentage of 1's in the bag is 72%. If a sample of 60 numbers is drawn from the bag,

a) what is the probability that the percentage of 1's in the sample is less than 65%?
b) what is the probability that the percentage of 1's in the sample is more than 68%?
c) what is the probability that the percentage of 1's in the sample is between 73% and 78%

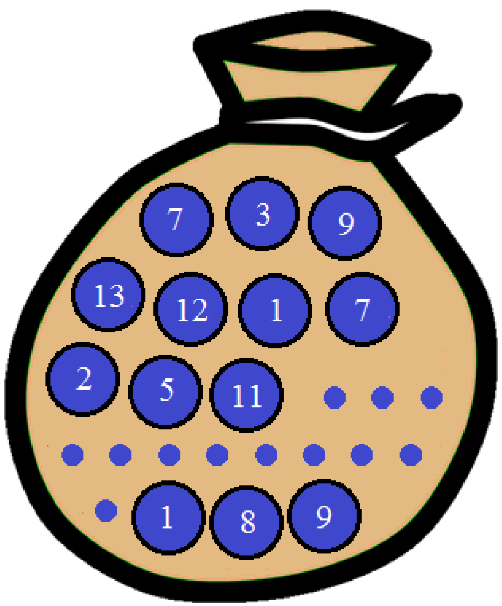
Ex 4: 56% of the students who enter America's colleges and universities graduate within six years. If 250 college freshman are randomly selected from American colleges and universities,

a) what is the probability that less than 64% of them will graduate within 6 years?
b) what is the probability that at least 60% of them will graduate within 6 years?
c) what is the probability that between 48% and 55% of them will graduate within 6 years?

How is s a Random Variable?

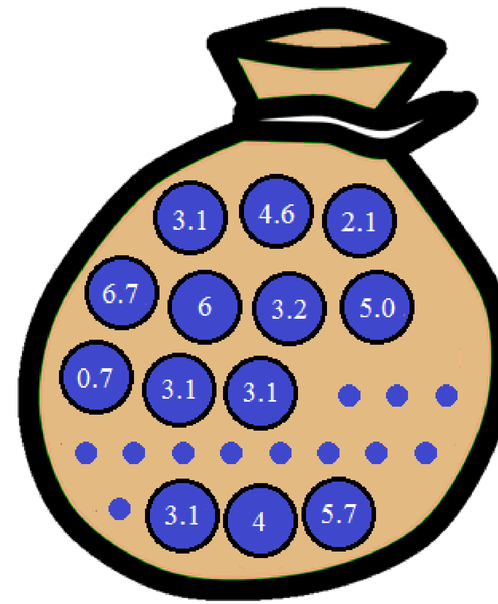
Picture of the situation

X



- Take a sample of size n
- Calculate s (the standard deviation of the numbers in sample)
- Put answer in bag on right
- Do this for <u>all</u> possible samples of size n

s



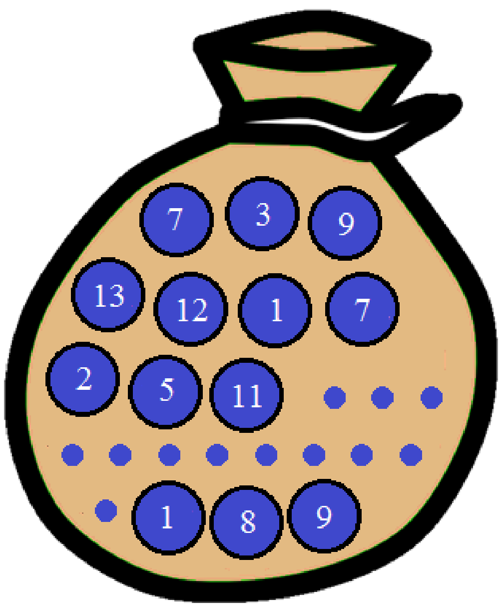
- Original bag of numbers (population data)

- New bag has all sample standard deviations in it
- Taking a sample of size n and calculating s is equivalent to drawing 1 number from the new bag

How is s^2 a Random Variable?

Picture of the situation

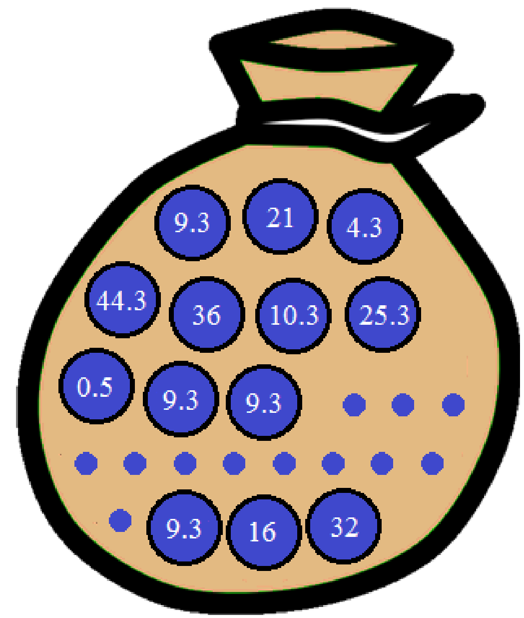
X



- Take a sample of size *n*
- Calculate s^2 (the variance of the numbers in sample)
- Put answer in bag on right

- Do this for <u>all</u> possible samples of size *n*

s^2



- Original bag of numbers (population data)

- New bag has all sample variances in it
- Taking a sample of size *n* and calculating s^2 is equivalent to drawing 1 number from the new bag