COMM215

First the Foundation, then Innovation

LESSON 6
SAMPLING DISTRIBUTION

SAMIE L.S. LY

REVISIT LESSON 5 PROBLEM # 6.1

Contents of a 32-ounce Bottle

The foreman of a bottling plant has observed that the amount of soda in each 32-ounce bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce.

a. If a customer buys one bottle, what is the probability that the bottle will contain more than 32 ounces?

Because the random variable is the amount of soda in one bottle, we want to find P(X>32), where X is normally distributed, μ=32.3, and $\sigma = .3$. Hence,

$$P(X > 32) = P\left(\frac{X-\mu}{\sigma} > \frac{32-32.2}{.3}\right)$$
$$= P(Z > -.67)$$
$$= 1 - P(Z < -.67)$$
$$= 1 - .2514 = .7486$$

3 SECTIONS

7.1 Sampling Error: What it is and Why it happens

7.2 The Sampling Distribution of the Sample Mean $\bar{x}$

7.3 The Sampling Distribution of the Sample Proportion $\bar{p}$

ALL ORANGES IN THE WORLD

SAMPLE # 2

SAMPLE # 1

SAMPLED POPULATION

FRAMED POPULATION

SELECTING A SAMPLE

- The reason we select a sample is to collect data to answer a research question about a population.

- The sample results provide only **ESTIMATES** of the values of the population.
  Sample Results ≈ Values of Population

- WHY? Because the sample contains only a **PORTION** of the population.

Bowerman, et al. (2017) pp. 327

SELECTING A SAMPLE

FINITE POPULATION

- SIMPLE RANDOM SAMPLE

INFINITE POPULATION

- RANDOM SAMPLE

Bowerman, et al. (2017) pp. 337

SELECTING A SAMPLE

FINITE POPULATION

- An organization's membership roster
- Credit card account numbers
- Students IDs in Concordia University

SIMPLE RANDOM SAMPLE

Sample size **n** from a finite population of size **N**

Each possible sample size **n** has the same probability of being selected.

SELECTING A SAMPLE

INFINITE POPULATION

- **Hard to** obtain a list of **ALL ELEMENTS** in the population.
- **Cannot** construct a **FRAME** for the population
- Hence, we **CANNOT USE RANDOM NUMBER** selection procedure.

THEN HOW?

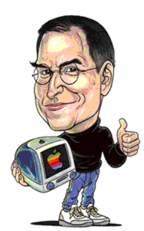
SELECTING A SAMPLE

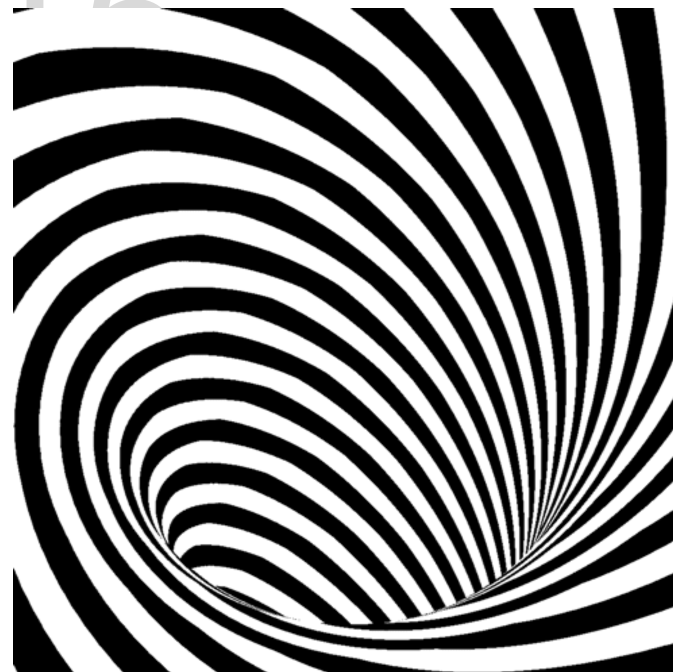
THEN HOW?

Populations are often generated by an ONGOING PROCESS where there is no upper limit on the number of units that can be generated.

INFINITE POPULATION

- Transactions occurring at a bank
- Customers entering a store





INFINITE POPULATION

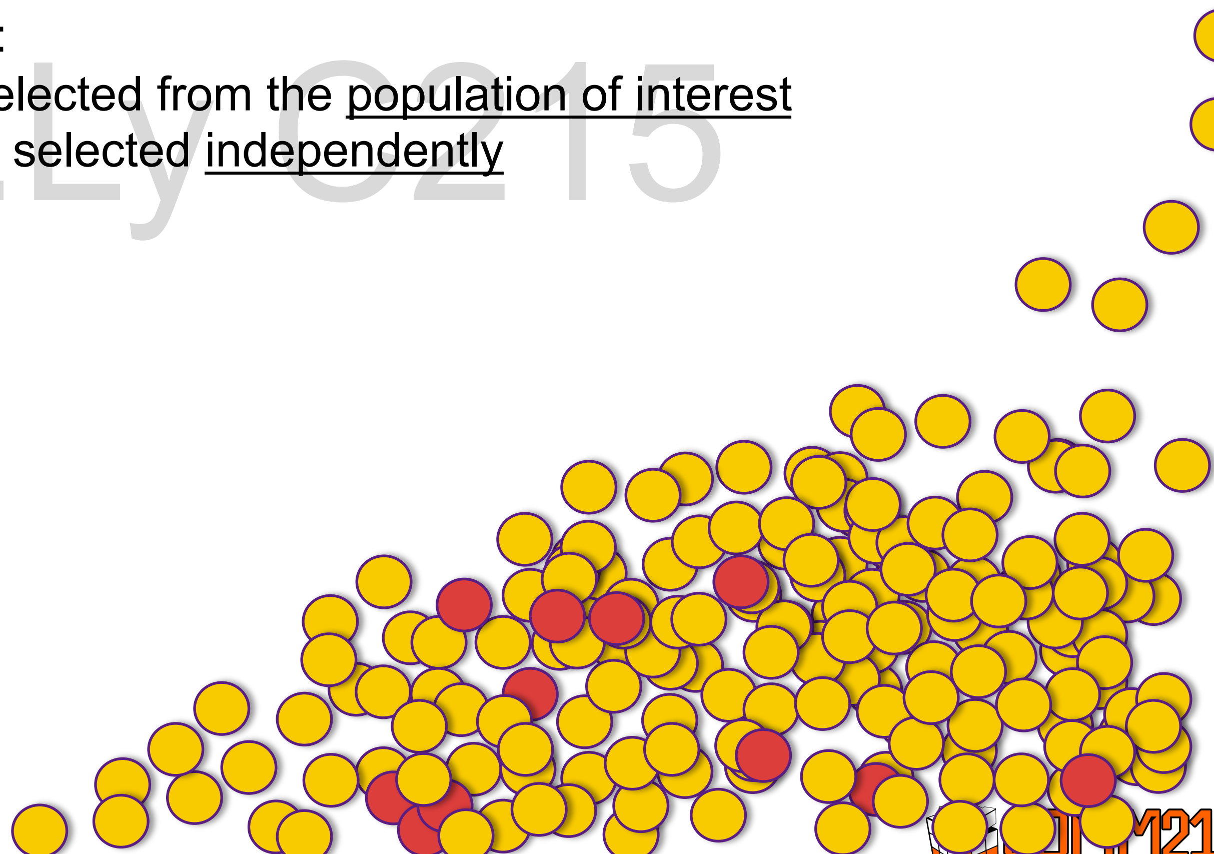
IN THIS CASE: ~~SIMPLE~~ RANDOM SAMPLE or SAMPLE

We select a random sample in order to make valid statistical inferences about the population from which the sample is taken.

RANDOM SAMPLE FROM AN INFINITE POPULATION

Follow 2 conditions:

- Each element selected from the <u>population of interest</u>
- Each element is selected <u>independently</u>



POINT ESTIMATION

POINT ESTIMATION is a form of statistical inference.

We refer to

$\bar{x}$ as the point estimator of the population mean μ.

s is the point estimator of the population standard deviation σ.

$\bar{p}$ is the point estimator of the population proportion p.

Bowerman, et al. (2017) pp. 329

Summary of Point Estimates

Obtained from a Random Sample

Population Parameter	Parameter Value	Point Estimator	Point Estimate
μ = Population mean	1090	$\bar{x}$ = Sample mean	1097
σ = Population std. deviation	80	s = Sample std. deviation	75.2
p = Population proportion	.72	$\bar{p}$ = Sample proportion	.68

TARGET VS SAMPLED

The <u>target population</u> is the population we want to make inferences about.

The <u>sampled population</u> is the population from which the sample is **actually taken**.

Whenever a sample is used to make inferences about a population, we should make sure that …

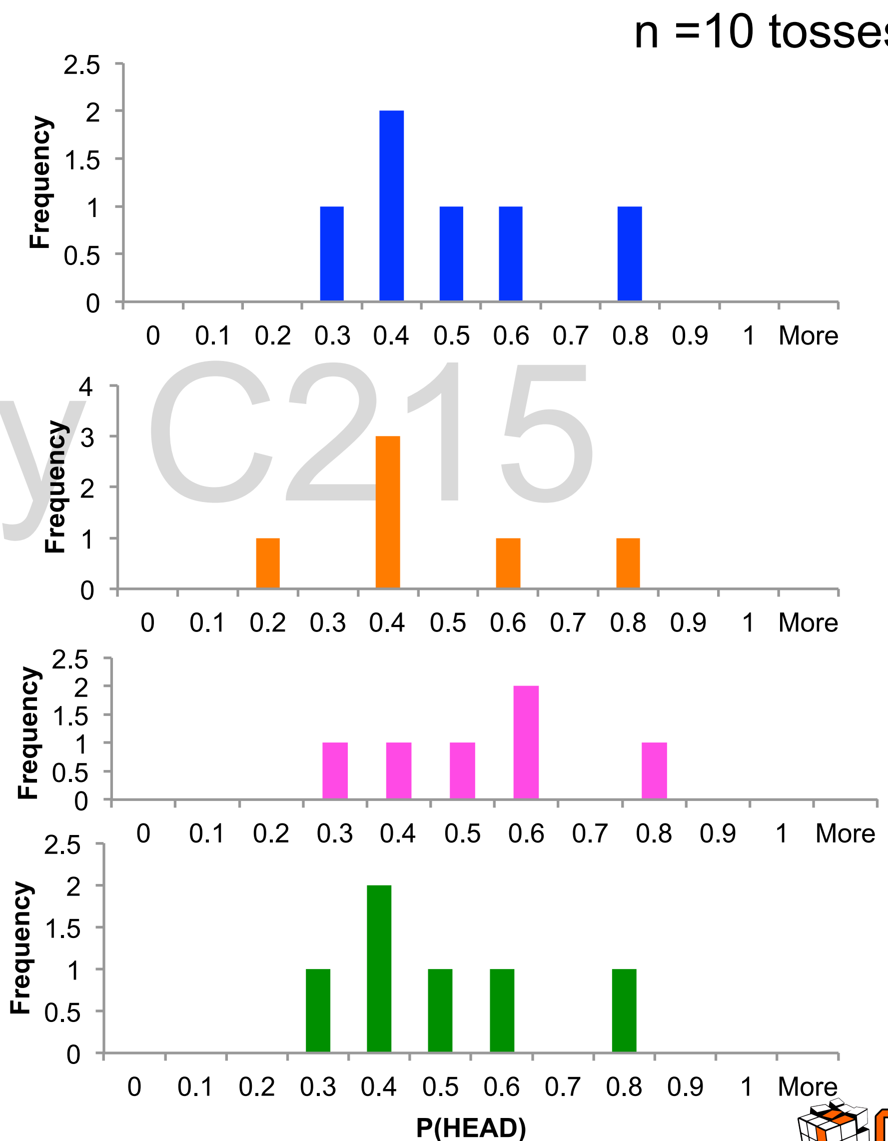
the **targeted population** and the **sampled population** are in close agreement.

$\bar{p}$ 10 X 20 X

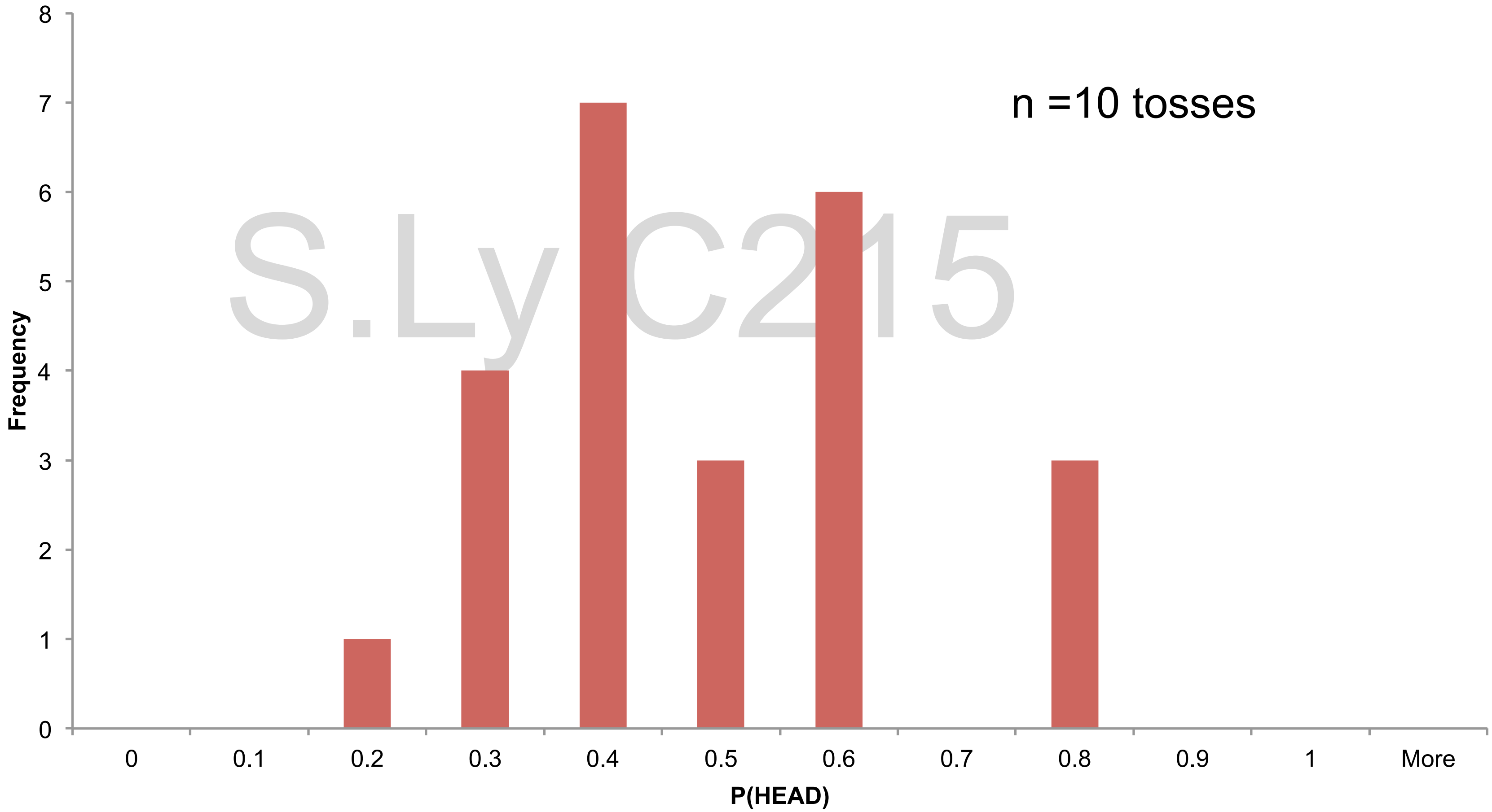
$\approx$ ***population p***

1 GROUP OF 6

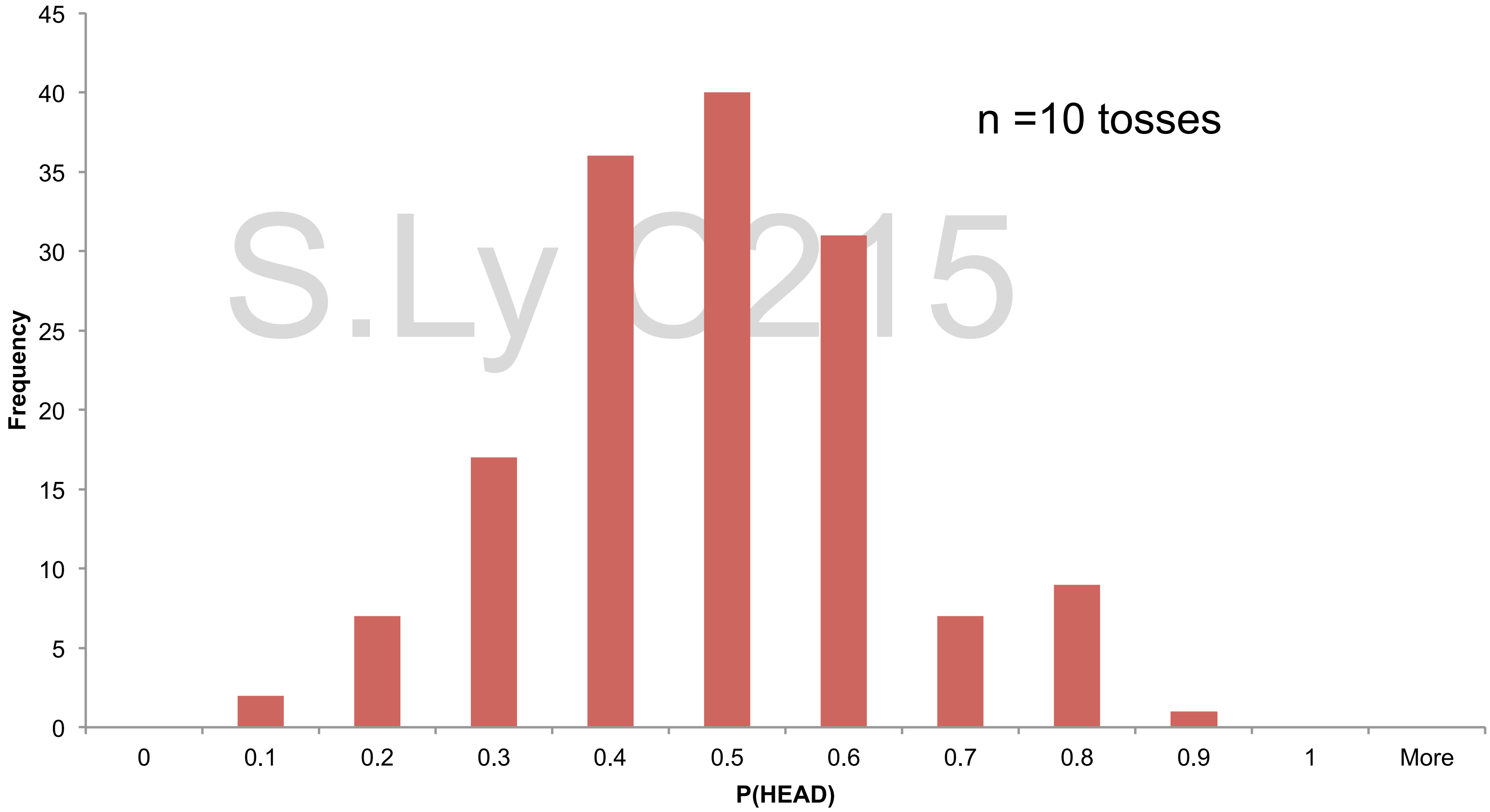
Bin	Frequency
0	0
0.1	0
0.2	0
0.3	1
0.4	1
0.5	1
0.6	2
0.7	0
0.8	1
0.9	0
1	0
More	0



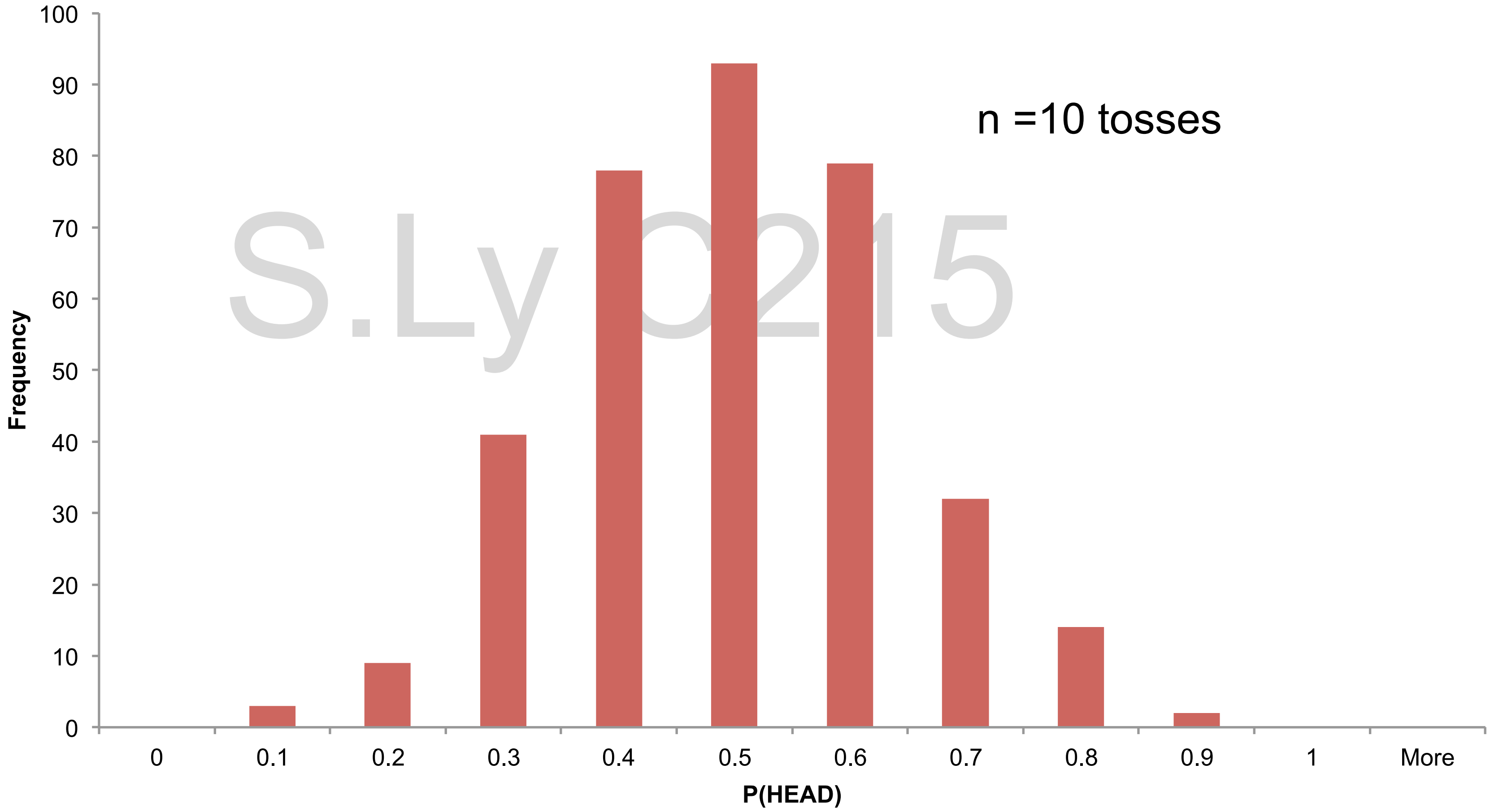
4X6 = 24 SAMPLES



25X6 = 150 SAMPLES

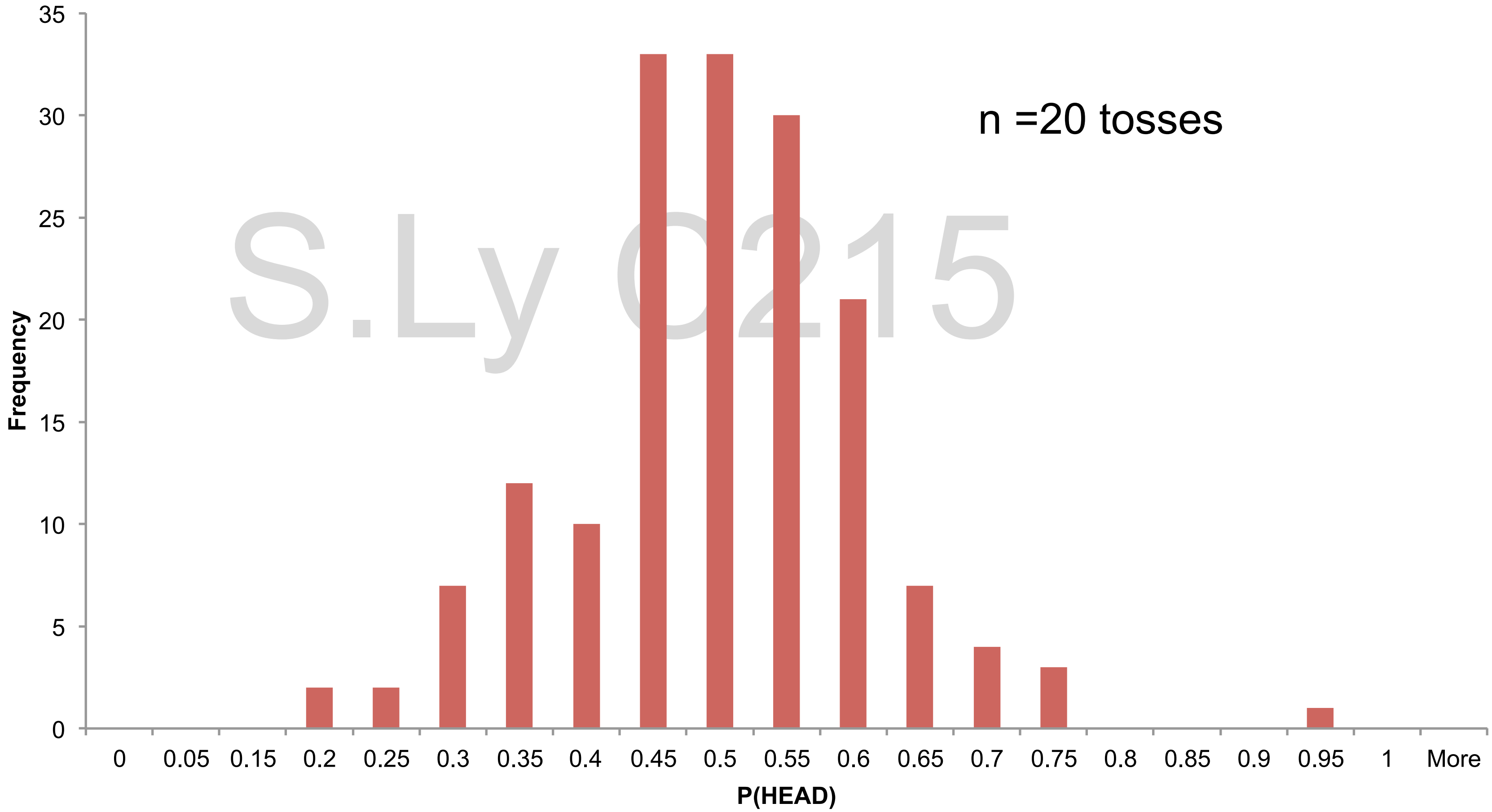


351 SAMPLES

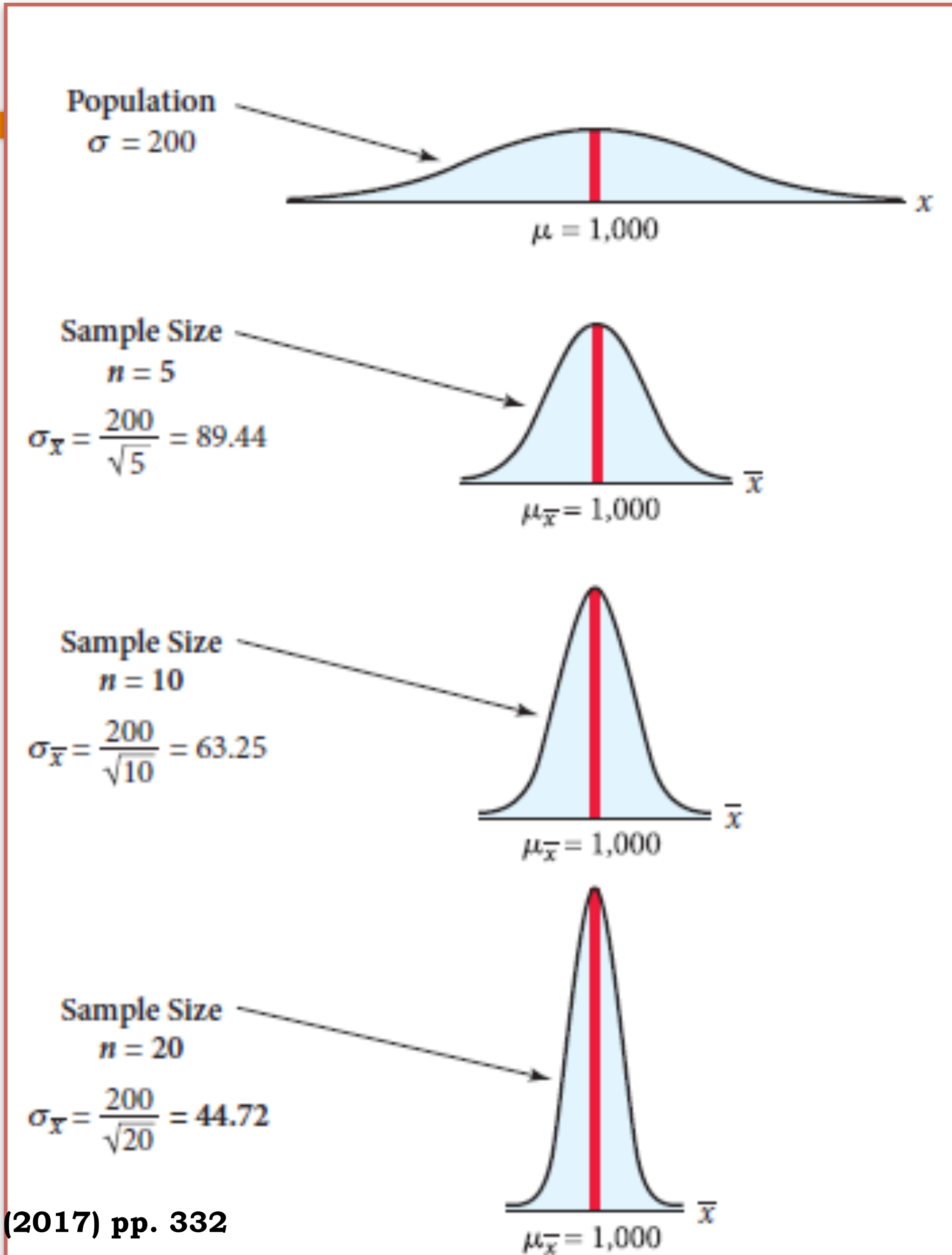


n =10 tosses
Frequency
100
90
80
70
60
50
40
30
20
10
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
More
P(HEAD)

165 SAMPLES



n =20 tosses
Frequency
35
30
25
20
15
10
5
0
0 0.05 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 More
P(HEAD)
S.Ly C215



Bowerman, et al. (2017) pp. 332

SAMPLING DISTRIBUTION OF $\bar{x}$

SAMPLING DISTRIBUTION OF $\bar{x}$

EXPECTED VALUE

FORM OF SAMPLING DISTRIBUTION OF $\bar{x}$

STANDARD DEVIATION

SAMPLING DISTRIBUTION OF $\bar{x}$

The <u>sampling distribution of $\bar{x}$</u> is the probability distribution of all possible values of the sample mean.

- Expected Value of $\bar{x}$

$$E(\bar{x}) = \mu$$

where: μ = the population mean

When the expected value of the point estimator equals the population parameter, we say the point estimator is <u>unbiased</u>.

Bowerman, et al. (2017) pp. 331

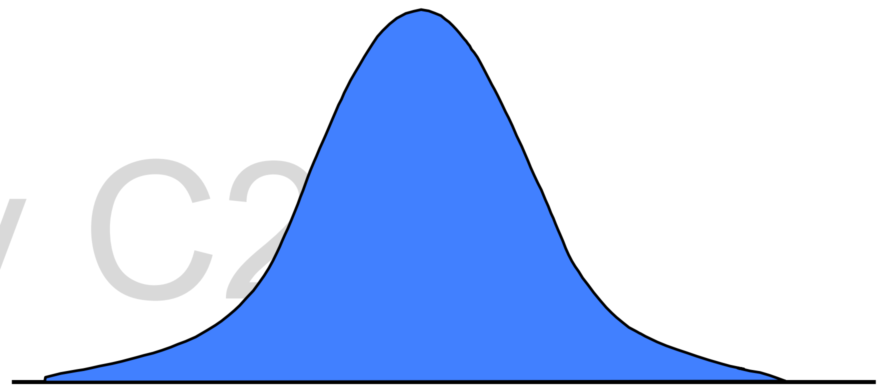
SAMPLING DISTRIBUTION OF $\bar{x}$

- When the population has a normal distribution, the sampling distribution of $\bar{x}$ is normally distributed for any sample size.
- In most applications, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution whenever the sample size is 30 or more.
- In some cases where the population is highly skewed or outliers are present, sample of size 50 may be needed.
- The sampling distribution of $\bar{x}$ can be used to provide probability information about how close the sample mean $\bar{x}$ is to the population mean μ.

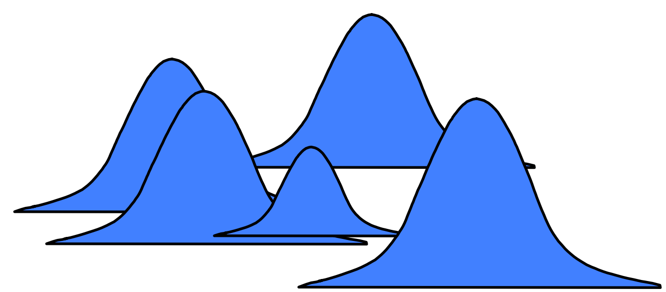
Bowerman, et al. (2017) pp. 335

- When the population has a normal distribution, the sampling distribution of $\overline{x}$ is normally distributed for any sample size.

IF POPULATION IS

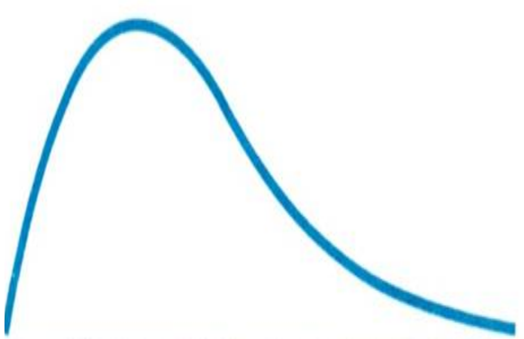


THEN SAMPLES are

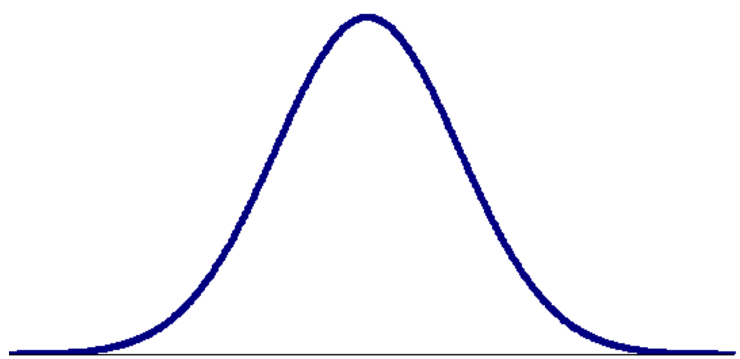


Bowerman, et al. (2017) pp. 335

- In most applications, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution whenever the sample size is 30 or more.
- In some cases where the population is highly skewed or outliers are present, sample of size 50 may be needed.



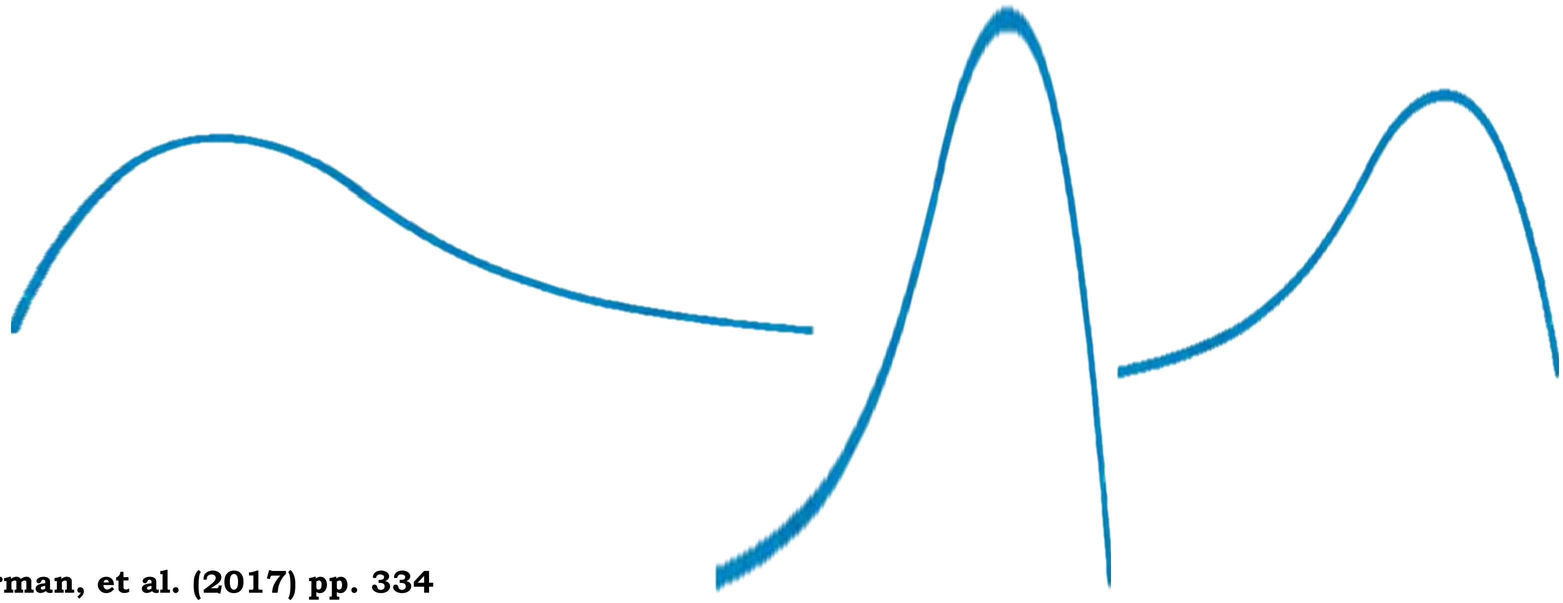
n = 50



n ≥ 30

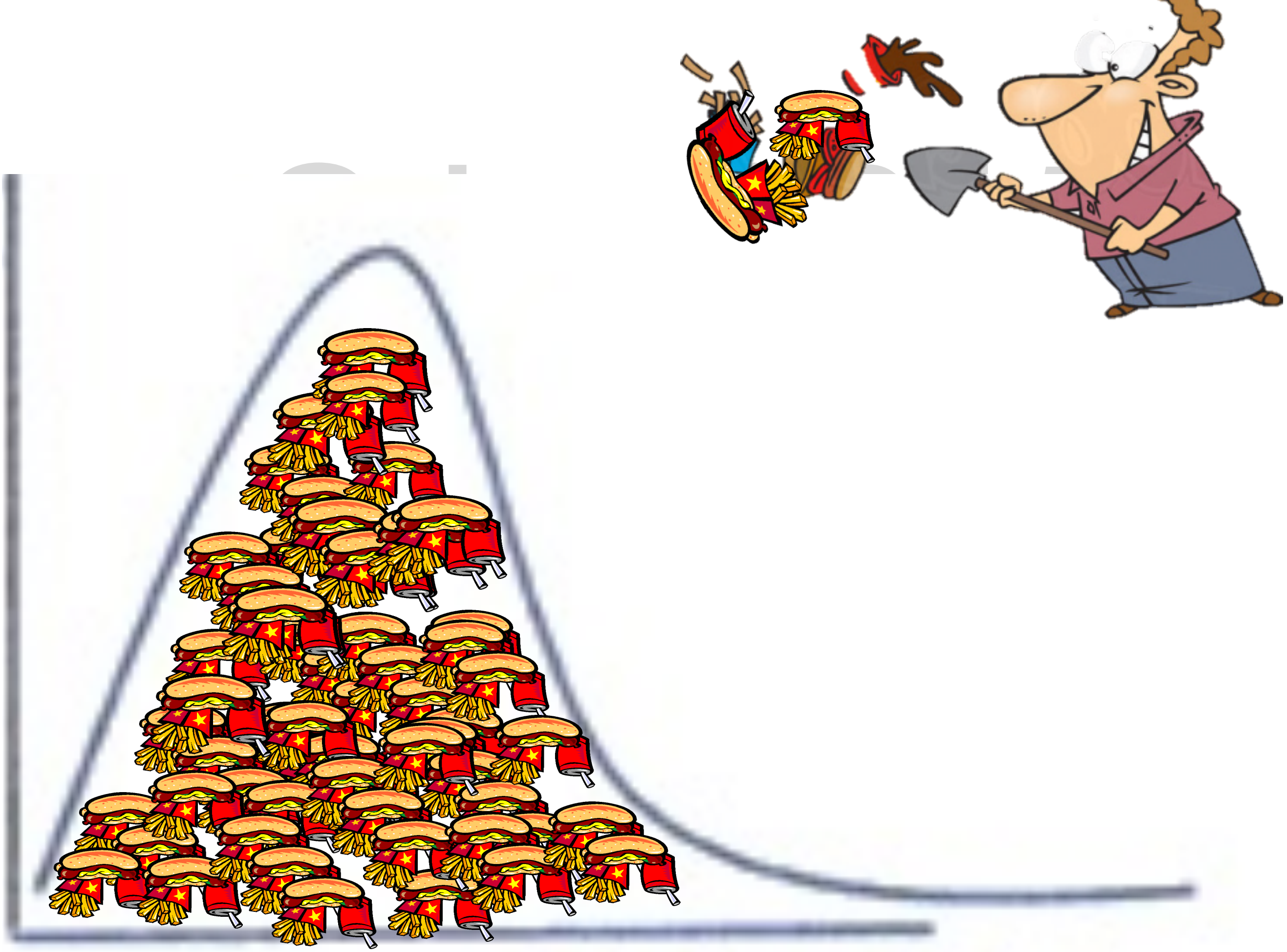
CENTRAL LIMIT THEOREM

When the population from which we are selecting a random sample **does not** have a normal distribution, the central limit theorem is helpful in identifying the shape of the sampling distribution of $\bar{x}$.

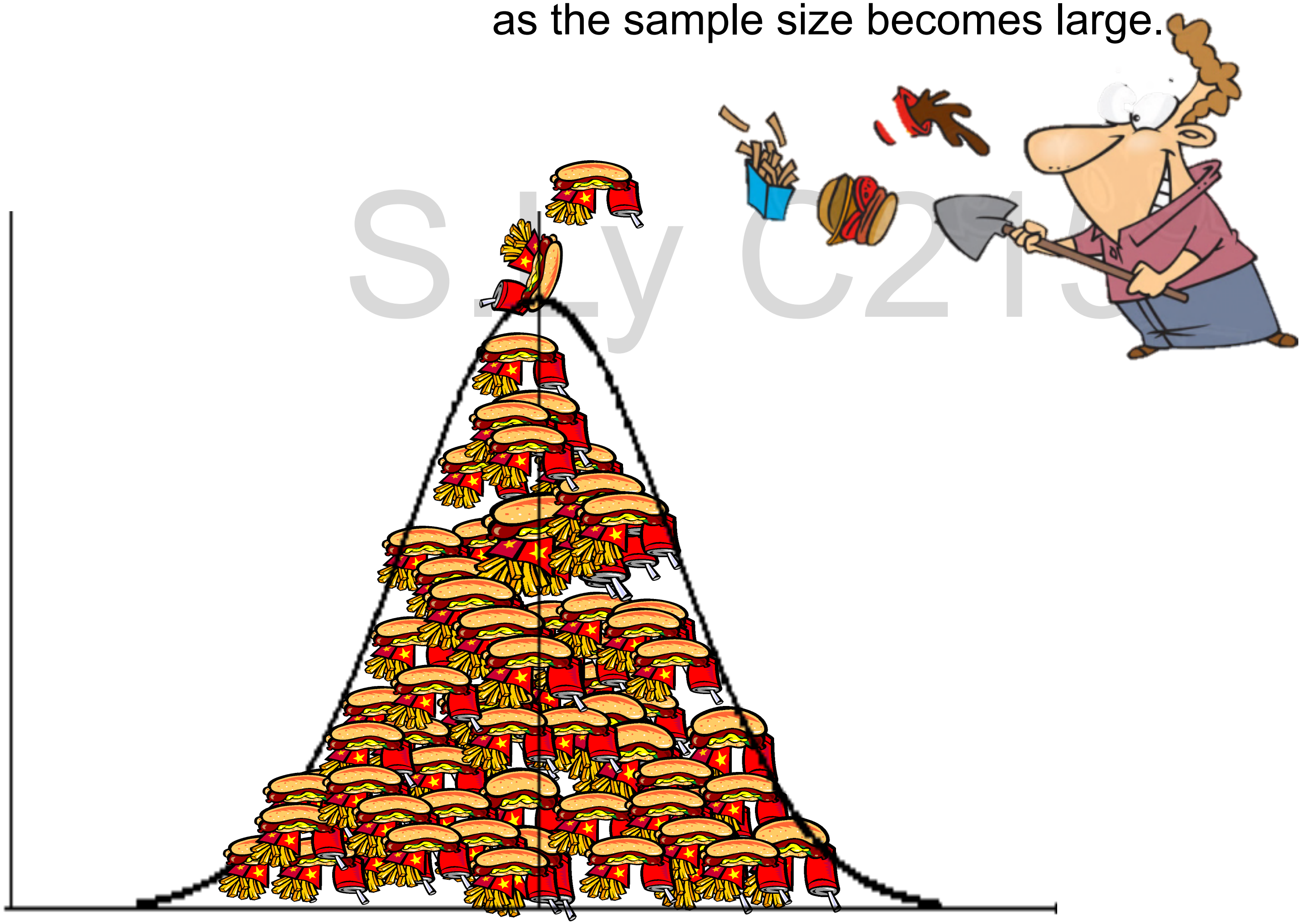


Bowerman, et al. (2017) pp. 334

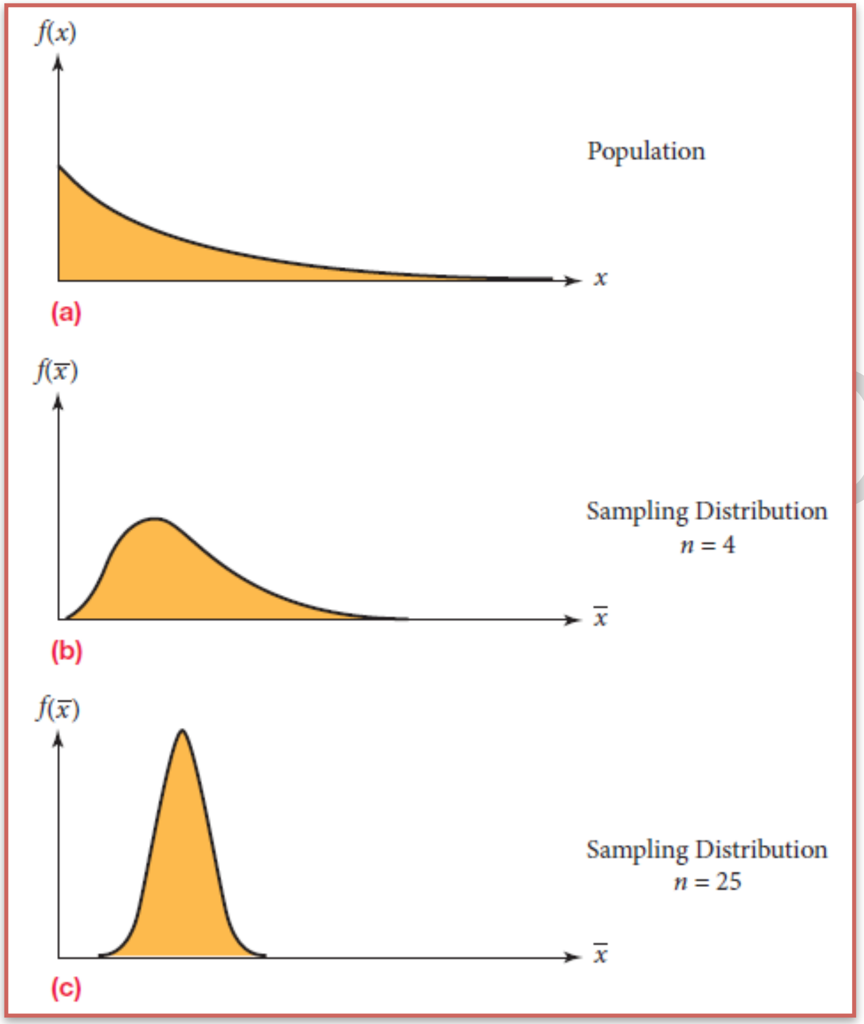
In selecting random samples of size *n* from a population, the sampling distribution of the sample mean $\overline{x}$ can be approximated by a normal distribution as the sample size becomes large.

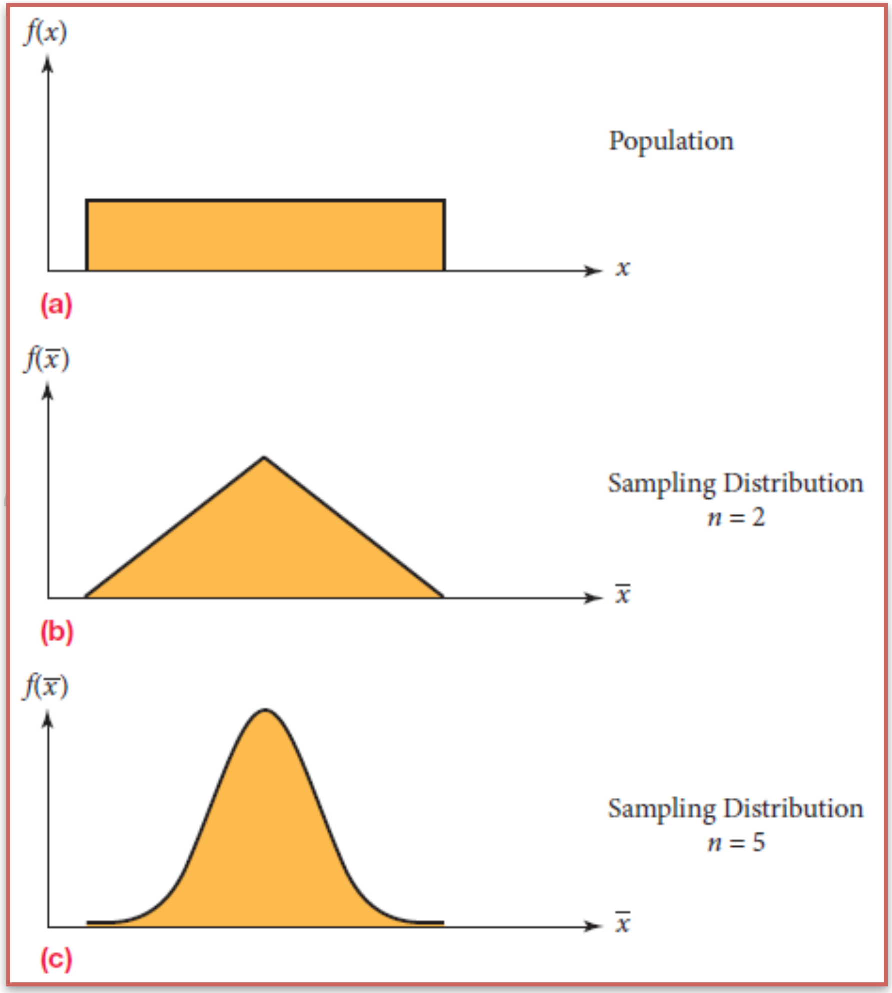


In selecting random samples of size *n* from a population, the sampling distribution of the sample mean $\overline{x}$ can be approximated by a normal distribution as the sample size becomes large.



CENTRAL LIMIT THEOREM

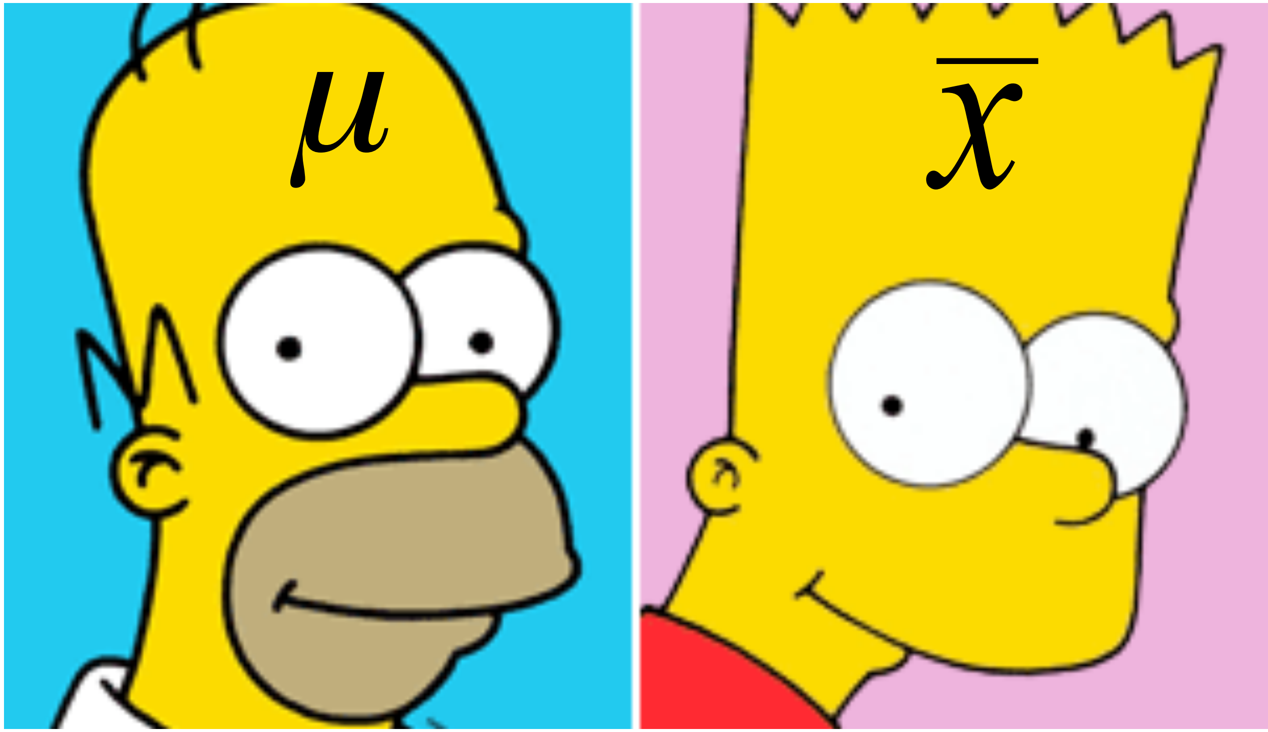




Bowerman, et al. (2017) pp. 335

The sampling distribution of $\overline{x}$ can be used to provide probability information about how close the sample mean $\overline{x}$ is to the population mean μ.

POPULATION **SAMPLE**



SAMPLING DISTRIBUTION OF $\overline{x}$

STANDARD DEVIATION OF $\bar{x}$

We will use the following notation to define the standard deviation of the sampling distribution of $\bar{x}$.

$\sigma_{\bar{x}}$ = the standard deviation of $\bar{x}$

σ = the standard deviation of the population

n = the sample size

N = the population size

SAMPLING DISTRIBUTION OF $\bar{x}$

Finite Population

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Infinite Population

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- A finite population is treated as being infinite if $n/N \leq .05$.
- $\sqrt{(N-n)/(N-1)}$ is the finite population correction factor.
- $\sigma_{\bar{x}}$ is referred to as the standard error of the mean.

PROBLEM # 6.2

A random variable is normally distributed with mean *μ=$ 1,500* and standard deviation *σ =$ 100.* Determine the standard error of the sampling distribution of the mean for simple random samples with the following sample sizes:

a. n = 16

b. n= 400

a. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{16}} = 25$

b. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{400}} = 5$

PROBLEM #6.3

A normally distributed population has a mean of 40 and a standard deviation of 12. What does the central limit theorem say about the sampling distribution of the mean if samples of size 100 are drawn from the population?

The sampling distribution of the mean is normal with a mean of 40 and a standard deviation of $12/\sqrt{100} = 1.2$.

PROBLEM # 6.4

What is the difference between a probability distribution and a sampling distribution?

A sampling distribution is a probability distribution. It is for a statistic (which is a random variable) with a sample size specified.

PROBLEM # 6.5

What is the difference between a standard deviation and a standard error?

The standard deviation of a sampling distribution is referred to as the standard error.

PROBLEM # 6.1

Contents of a 32-ounce Bottle

The foreman of a bottling plant has observed that the amount of soda in each 32-ounce bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce.

b. If a customer buys a carton of four bottles, what is the probability that the mean amount of the four bottles will be greater than 32 ounces?

Now we want to find the probability that the mean amount of four filled bottles exceeds 32 ounces. That is, we want $P(\bar{X} > 32)$. From our previous analysis and from the central limit theorem, we know the following:

1. $\bar{X}$ is normally distributed
2. $\mu_{\bar{x}} = \mu = 32.2$
3. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.3}{\sqrt{4}} = .15$

Hence,

$$P(\bar{X} > 32) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{32 - 32.2}{.15}\right)$$

$$= P(Z > -1.33) = 1 - P(Z < -1.33) = 1 - .0918 = .9082$$

COMM215

First the Foundation, then Innovation

SAMPLING DISTRIBUTION OF $\bar{p}$

SAMPLING DISTRIBUTION OF $\bar{p}$

EXPECTED VALUE

STANDARD DEVIATION

FORM OF SAMPLING DISTRIBUTION OF $\bar{p}$

SAMPLING DISTRIBUTION OF $\bar{p}$

The sampling distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion $\bar{p}$.

- Expected Value of $\bar{p}$

$$E(\bar{p}) = p$$

where:

p = the population proportion

PROBLEM # 6.6

A simple random sample of size 100 is selected from a population with p=.40

a. What is the expected value of $\bar{p}$?

$$E(\bar{p}) = p = .40$$

SAMPLING DISTRIBUTION OF $\bar{p}$

- Standard Deviation of $\bar{p}$

Finite Population

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

Infinite Population

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- $\sigma_{\bar{p}}$ is referred to as the standard error of the proportion.
- $\sqrt{(N-n)/(N-1)}$ is the finite population correction factor.

PROBLEM # 6.6

A simple random sample of size 100 is selected from a population with p=.40

b. What is the standard error of $\bar{p}$?

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.40(.60)}{100}} = .0490$$

FORM OF SAMPLING DISTRIBUTION OF $\bar{p}$

The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever the sample size is large enough to satisfy the two conditions:

$$np \geq 5 \quad \text{and} \quad n(1 - p) \geq 5$$

. . . because when these conditions are satisfied, the probability distribution of x in the sample proportion, $\bar{p}$ = x/n, can be approximated by normal distribution (and because n is a constant).

PROBLEM # 6.6

A simple random sample of size 100 is selected from a population with p=.40

c. Show the sampling distribution of $\bar{p}$.

d. What does the sampling distribution of $\bar{p}$ show?

Normal distribution with $E(\bar{p}) = .40$ and $\sigma_{\bar{p}} = .0490$

It shows the probability distribution for the sample proportion $\bar{p}$.

LESSON 6 IN A NUTSHELL

Population Mean

μ

Sample Mean

$\bar{x}$

Standard Error of the Mean

$\sigma_{\bar{x}}$

Population Standard Deviation σ

Sample Proportion

$\bar{p}$

Standard Error of the Proportion

$\sigma_{\bar{p}}$

Population Proportion

p

Sample Standard Deviation s

LESSON 6 IN A NUTSHELL

Normal Probability Distribution

<u>The sample size has been specified.</u>

SAMPLING DISTRIBUTION OF $\bar{x}$

If it is <u>NOT NORMALLY DISTRIBUTED</u>,

with a large enough sample n > 30, **"Central Limit Theorem"**, it will approximately be NORMALLY DISTRIBUTED.

SAMPLING DISTRIBUTION OF $\bar{p}$

$$np \geq 5 \qquad n(1-p) \geq 5$$

PROBLEM # 6.8

Given a normal population whose mean is 50 and whose standard deviation is 5,

a. Find the probability that a random sample of 4 has a mean between 49 and 52.

$$P(49 < \overline{X} < 52) = P\left(\frac{49-50}{5/\sqrt{4}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{52-50}{5/\sqrt{4}}\right)$$

$$= P(-.40 < Z < .80)$$
$$= P(Z < .80) - P(Z < -.40)$$
$$= .7881 - .3446 = \underline{.4435}$$

PROBLEM # 6.8

Given a normal population whose mean is 50 and whose standard deviation is 5,

b. Find the probability that a random sample of 16 has a mean between 49 and 52.

$$P(49 < \overline{X} < 52) = P\left(\frac{49-50}{5/\sqrt{16}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{52-50}{5/\sqrt{16}}\right)$$

$$= P(-.80 < Z < 1.60)$$
$$= P(Z < 1.60) - P(Z < -.80)$$
$$= .9452 - .2119 = \underline{.7333}$$

PROBLEM # 6.13

a. In a binomial experiment with n=300 and p=0.5, find the probability that *P* is greater than 60%.

$$P(\hat{P} > .60) = P\left(\frac{\hat{P} - p}{\sqrt{p(1-p)/n}} > \frac{.60 - .5}{\sqrt{(.5)(1-.5)/300}}\right) = P(Z > 3.46) = 0$$

PROBLEM # 6.17

The manager of a restaurant in a commercial building has determined that the proportion of customers who drink tea is 14%. What is the probability that in the next 100 customers at least 10% will be tea drinkers?

$$P(\hat{P} > .10) = P\left(\frac{\hat{P} - p}{\sqrt{p(1-p)/n}} > \frac{.10 - .14}{\sqrt{(.14)(1-.14)/100}}\right) =$$

$$= P(Z > -1.15) = 1 - P(Z < -1.15)$$
$$= 1 - .1251 = .8749$$

MIDTERM REVIEW PROBLEMS

The size of servings of soft ice cream in a fast-food chain is monitored carefully so that the mean serving is 250g, with a standard deviation of only 7.5g. Treat the weights as being normally distributed.

a. Inspectors select 20 servings at random, and weigh them. What is the probability that the mean weight is either below 240g or above 260g?

b. At least how much ice cream is there in a serving that is among the heaviest 15% of servings?

PROBLEM # 6.18

For a population of five individuals, television ownership is as follows:

a. Determine the probability distribution for the discrete random variable, x = number of television sets owned. Calculate the population mean and standard deviation.

	x= Number of Television Sets Owned
Allen	2
Betty	1
Chuck	3
Dave	4
Eddie	2

Let x = number of TV sets owned. The probability distribution is given below.

x:	1	2	3	4
P(x):	0.2	0.4	0.2	0.2

$$\mu = 1(0.2) + 2(0.4) + 3(0.2) + 4(0.2) = 2.4$$

$$\sigma^2 = E(x-\mu)^2 = (1 - 2.4)^2(0.2) + (2 - 2.4)^2(0.4) + (3 - 2.4)^2(0.2) + (4 - 2.4)^2(0.2) = 1.04$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.04} = 1.0198$$

COMM215

First the Foundation, then Innovation

PROBLEM # 6.18

	x= Number of Television Sets Owned
Allen	2
Betty	1
Chuck	3
Dave	4
Eddie	2

b. For the sample size n = 2, determine the mean for each possible simple random sample from the five individuals.

c. For each simple random sample identified in part (b), what is the probability that this particular sample will be selected?

sample	$\bar{x}$ = mean of this sample	probability of selecting this sample
Allen, Betty	(2 + 1)/2 = 1.5	1/10
Allen, Chuck	(2 + 3)/2 = 2.5	1/10
Allen, Dave	(2 + 4)/2 = 3.0	1/10
Allen, Eddie	(2 + 2)/2 = 2.0	1/10
Betty, Chuck	(1 + 3)/2 = 2.0	1/10
Betty, Dave	(1 + 4)/2 = 2.5	1/10
Betty, Eddie	(1 + 2)/2 = 1.5	1/10
Chuck, Dave	(3 + 4)/2 = 3.5	1/10
Chuck, Eddie	(3 + 2)/2 = 2.5	1/10
Dave, Eddie	(4 + 2)/2 = 3.0	1/10

COMM215
First the Foundation, then Innovation

PROBLEM # 6.18

d. Combining the results of parts (b) and (c), describe the sampling distribution of the mean.

The sampling distribution of the mean (n = 2) is:

$\bar{x}$:	1.5	2.0	2.5	3.0	3.5
$P(\bar{x})$:	0.2	0.2	0.3	0.2	0.1

$$\mu_{\bar{x}} = 1.5(0.2) + 2(0.2) + 2.5(0.3) + 3(0.2) + 3.5(0.1) = 2.4$$

$$\sigma_{\bar{x}}^2 = E(\bar{x} - \mu)^2 = (1.5 - 2.4)^2(0.2) + (2 - 2.4)^2(0.2) + (2.5 - 2.4)^2(0.3) + (3 - 2.4)^2(0.2) + (3.5 - 2.4)^2(0.1) = 0.39$$

and $\sigma_{\bar{x}} = \sqrt{0.39} = 0.624$

PROBLEM # 6.14

The probability of success on any trial of a binomial experiment is 25%. Find the probability that the proportion of successes in a sample of 500 is less than 22%.

$$P(\hat{P} < .22) = P\left(\frac{\hat{P} - p}{\sqrt{p(1-p)/n}} > \frac{.22 - .25}{\sqrt{(.25)(1-.25)/500}}\right) = P(Z < -1.55) = .0606$$

PROBLEM # 6.8

Given a normal population whose mean is 50 and whose standard deviation is 5,

c. Find the probability that a random sample of 25 has a mean between 49 and 52.

$$P(49 < \bar{X} < 52) = P\left(\frac{49-50}{5/\sqrt{25}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{52-50}{5/\sqrt{25}}\right)$$

$$= P(-1.00 < Z < 2.00)$$
$$= P(Z < 2.00) - P(Z < -1.00)$$
$$= .9772 - .1587 = .8185$$