Ch. 7 Sampling & Sampling Distribution

① Population = all items

Sample = subset of pop items

Sales Rep	Sales
Jo	185
Sioux	250
Chin	210
Sheliadawn	310
Gigi	298
Tyrone	402
Kip	370

Sales Rep	Sales
Chin	210
Kip	370
Jo	185

Pop mean = Mew = μ = 289.29

② Sample mean = X bar = Point estimator of

Sample Mean = 255

③ Sampling Error =

$$\boxed{\begin{array}{l}\mu - \text{X bar} \\ 289 - 255 = 34\end{array}}$$

Question is:

Is this sampling Error just "sampling Error" or does it represent a TRUE difference? Is the Point Estimator good enough? Reasonable?

④ This chapter we will build a Model so we can check whether or not our point estimator & the sampling Error is reasonable.

ⓐ The model is called:

Sampling Distribution ... more soon...

ⓑ In this chapter we will be given

① population mean = mew = μ

② pop. standard Deviation = sigma = σ

(Next chapter (ch. 8) we will see what to do if we don't have σ)

ⓒ First we have to talk about

Sampling ...

⑤ # Sampling

(a) why take samples?

① Some populations are impossible to check
All the fish in the sea?

② calculating off of population costs a lot!
General Mills hires firm to test new cereal: sample cost = $40,000
Pop. cost ≈ $ 1,000,000,000

③ contacting whole population is time consuming
Political polls can be conducted 1-2 days
contact whole pop would take years!

④ Destructive Nature of some tests
Test each bottle of wine!?!
Test each seed in bag of seeds!?

⑤ samples are usually adequate

Consumer price index from sample is excellent estimate for CPI constructed from pop.

(b) why do we need samples?

we select samples to answer research questions about population.

- Do Highline students think advising should be Manditory?
- Is machine filling accurately?

⑤ ⓒ If the population is big, how do we know where to go and get data?

- we must be careful!!
- we must make sure that sampled population is same as Target Population.

Sampled Population

population from which the sample is drawn

Target Population

Population we want to make inference about

Example 1: Goal is to gather data about students at Laney Community college

- If you took a sample from Laney college Registration list: Sampled Pop. = Target Pop.
- If you took a sample from people who ate at the Culinary Arts restaurant at Laney you would get:
  ① Students at Laney (Target Pop.)
  ② People eating who are Not Laney Students (Not Target Pop.)
- Sampled Population = People eating at culinary Arts restaurant.
- Target Pop. ≠ sampled pop.

Example 2:

calculating Seattle House Price Mean, but you accidentally used some data from Burien also.

Example 3: 1936 presidential poll based on phone directories and other middle/upper class lists.

⑥ Frame

ⓐ
- List of all elements in Population
- List of elements that sample selected from
- Frames are not always possible to create

ⓑ Frame that can be created:

- List of names of registered Students @ Laney
- List of Companies Listed in NYSE.
- Finite Lists

ⓒ Frames that can NOT be constructed:

- population too big (all fish in sea)
- pop. from an on going process like:
  - Machines filling boxes of cereal
  - Transactions occur at bank
  - Customers entering a store

  } Theoretically infinite cuz No upper Limit
- Considered infinite because we cannot construct a Frame.

☆ Sampled pop. = conceptual pop. of all boxes that could have been filled at a particular point in time. In this sense, it is considered infinite.

⑦ Sampling Methods

ⓐ Probability Sample

good cuz we have techniques to test reasonableness

Each possible sample has a known probability of selection and a random process is used to select elements for the sample.

- Examples:

① Simple Random Sample
② Random Sample

} covered in this class

③ Stratified Random Sampling
④ Cluster Sampling
⑤ Systematic Sampling

ⓑ Non-probability Sampling

① Convenience sampling
② Judgment Sampling

Don't have techniques to test reasonableness

⑦ ⓒ For Finite Population when you have Frame use:

Simple Random Sample (SRS)

SRS Method:

create random number } ① Add new column with RAND Excel function to Proper Data set.

Sort smallest n values to top } ② click in 1 cell in RAND column & click "A to Z" sort button 5 times (Brings smallest to top).

③ copy & paste top n records

RAND Excel Function

① Generates random 15 digit number between 0 and 1 $(0 <= \text{Number} < 1)$ using Uniform Distribution. (means each number has same Probability of being selected)

② RAND is a volatile function which means it recalculates after any action (like Enter)

③ F9 will tell RAND to recalculate.

SRS definition from book:

A simple random sample of size n from a finite population of size N is a sample selected such that each possible sample of size n has the same probability of being selected.

⑦ ⓓ for infinite Population/ongoing Process use:

Random Sample

- select any n units in a random way
- MUST BE TRUE:

① Each element must be from same population (Sampled Population = Target Population)

② Each element is selected independently

(Don't select all items from similar group, similar attitudes - so you avoid bias!!)

*Think of 1936 presidential poll bias...

- For ongoing processes (like filling machines) Choose samples from same point in time so all elements are from same population

→ Now that we take a sample what about that sampling Error for our X bar !? !?

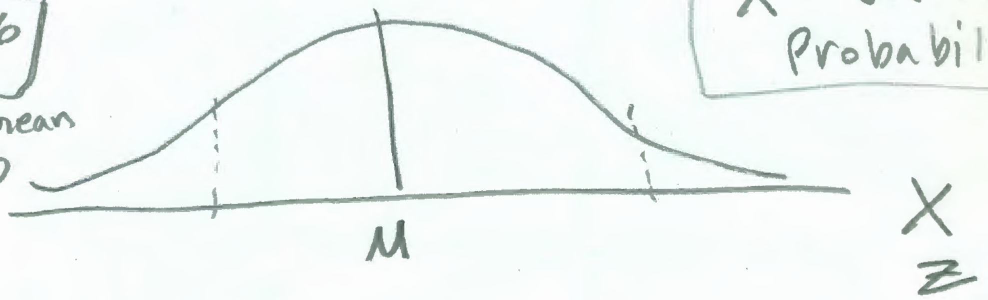
{Sampling Error} = μ - X bar = 289-255 = 34 ?

→ We will now treat X bar as a Random Variable.

Chapter 6

μ = Pop mean
σ = Pop SD
X = X

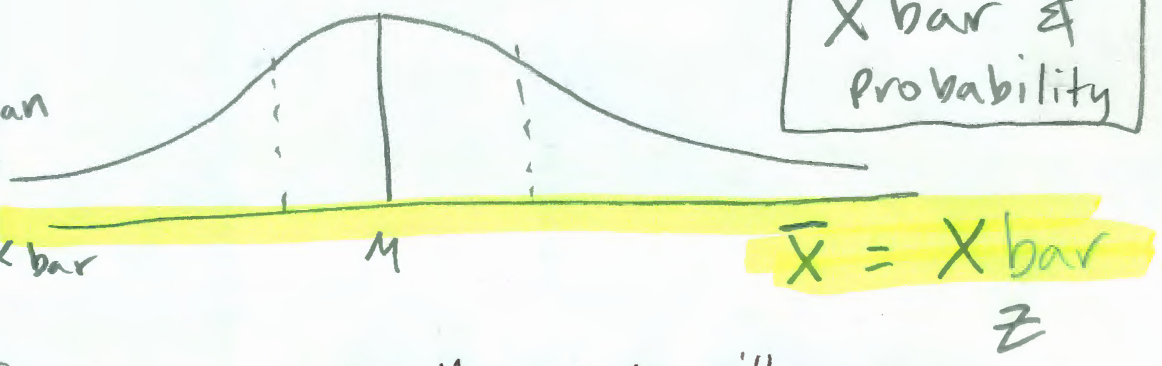
X- values & Probability



Chapter 7

μ = Pop mean
σ = Pop SD
X bar = X bar

X bar & Probability



Chapter 8

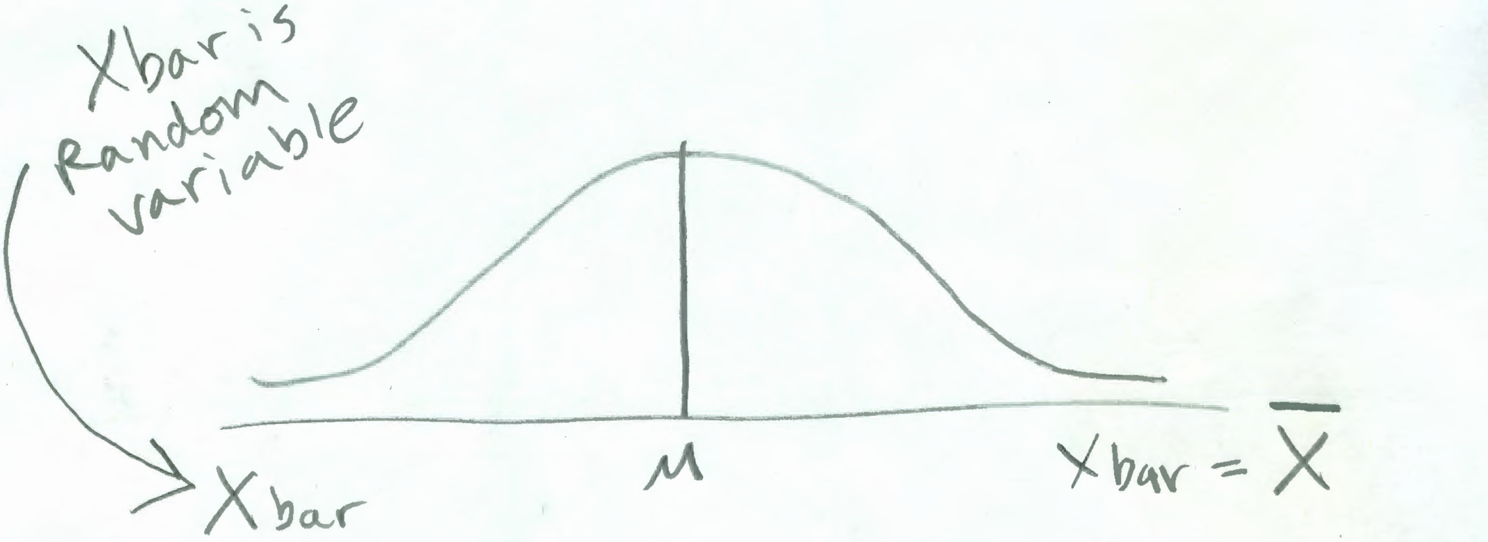
X bar
S

① Mew = Pop Mean = μ will not be KNOWN
② Sigma = Pop. SD = σ will not be KNOWN & we will learN about T- distribution

To talk about X bar as Random variable & determine whether our Sample error of 34 is reasonable or not we need to learn about:

8

⑨ Sampling Distribution of X bar
(SD of X bar)



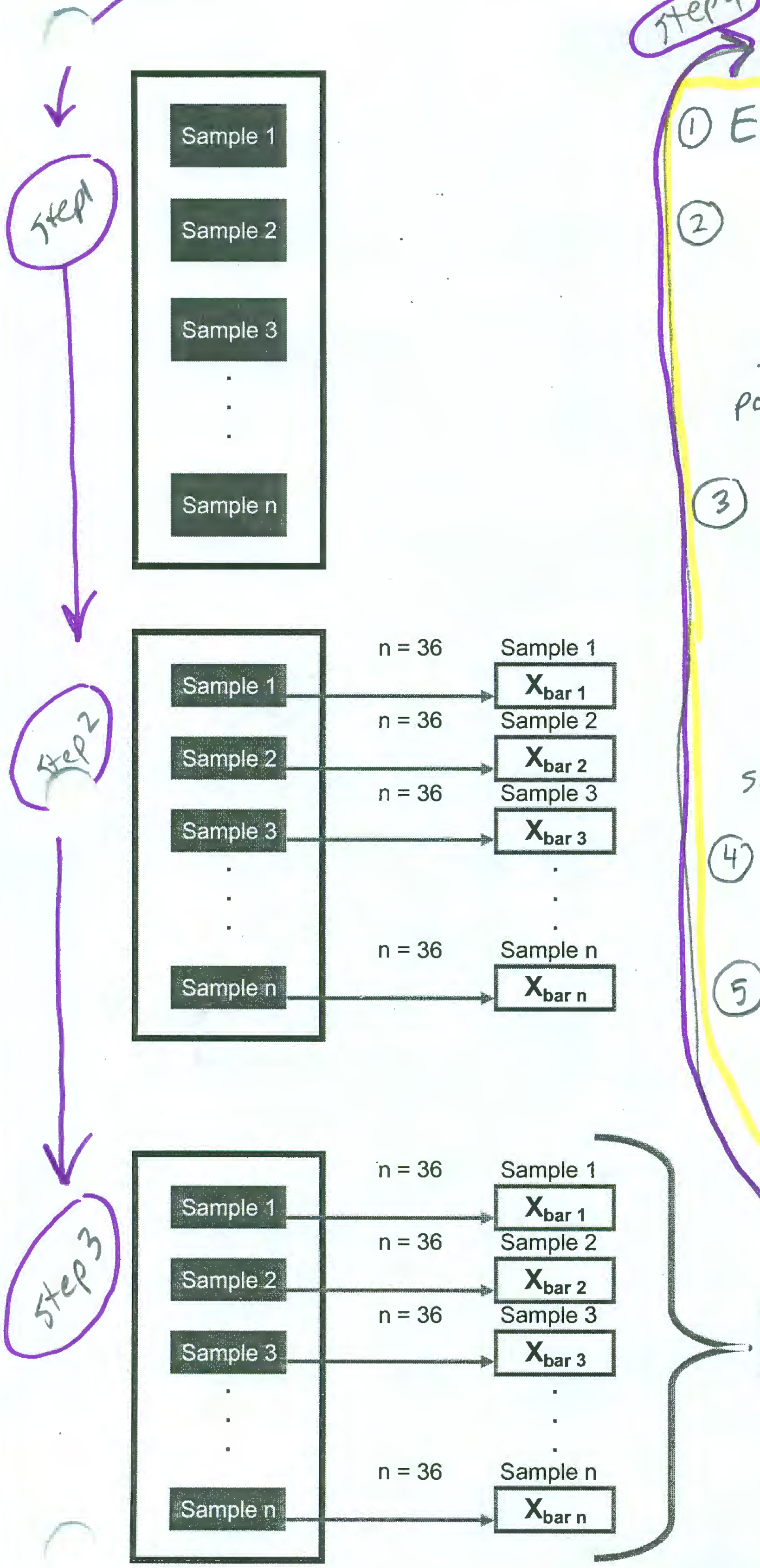
$$\mu = \mu_{\bar{x}} = E(\bar{x}) = E(\text{xbar})$$

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

$$Z_{\bar{x}} = \frac{\text{Xbar} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
← sample Error
← standard Error

8 b

How to construct The Sampling Distribution Of The Sample Mean X_{bar}



step4

We will Discover

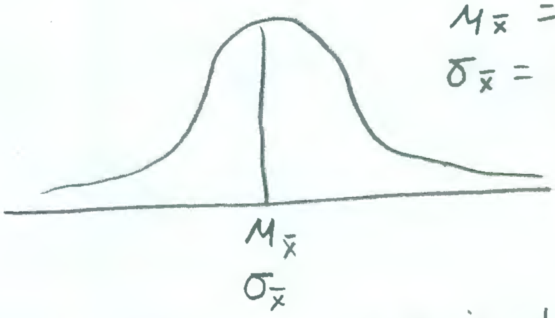
① $E(\bar{x}) = \mu_{\bar{x}} = \mu$

②

population spread is greater

③ $\mu_{\bar{x}} = \mu$

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



Sampling Distribution of $\bar{x}$ is less

④ Sampling Distribution of $\bar{x}$ is Normally Distributed (if n big)

⑤

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

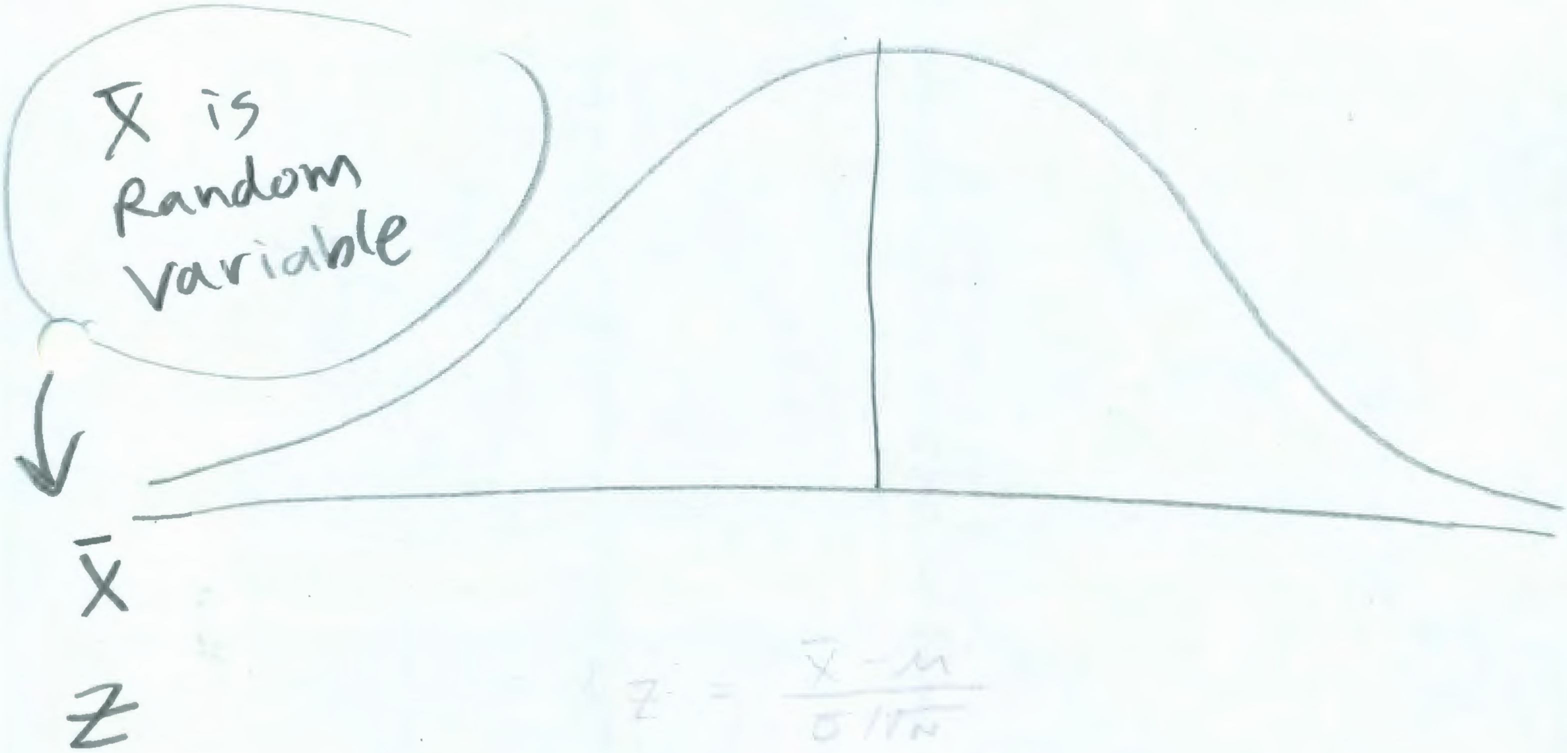
we will look at Example in Excel ↓

⑧ Ⓒ

Sampling Distribution of $\bar{X}$ ($SD\bar{X}$)

(P.1/2)

probability Distribution of all possible values of sample mean $\bar{X}$



$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

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

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

We can now take a sample & compare it to $SD\bar{X}$ & see if our sample seems reasonable or Not.

9 Expected value of $\overline{X}$ of SDo$\overline{X}$

or

Mean of Sampling Distribution of $\overline{X}$

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

$$\underbrace{\mu_{\overline{X}} = \frac{\text{Sum of all possible Sample Means}}{\text{Total number of Samples}}}$$

if we are able to select all possible samples of a particular size from a given population, then the mean of Sampling Distribution of $\overline{X}$ is equal to population mean.

well do an example in Excel.

⑩

Standard Deviation of Sampling Distribution of $\bar{X}$
or
Standard Error

Infinite Population

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

Finite Population

Finite population correction factor

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$$

when $\frac{n}{N} \leq 0.05$

then simply:

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

text assumes: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ unless stated otherwise

why:

usually populations are very large & sample size very big, so correction factor close to 1 (No affect)

⑪ Z for Sampling Distribution of $\bar{X}$

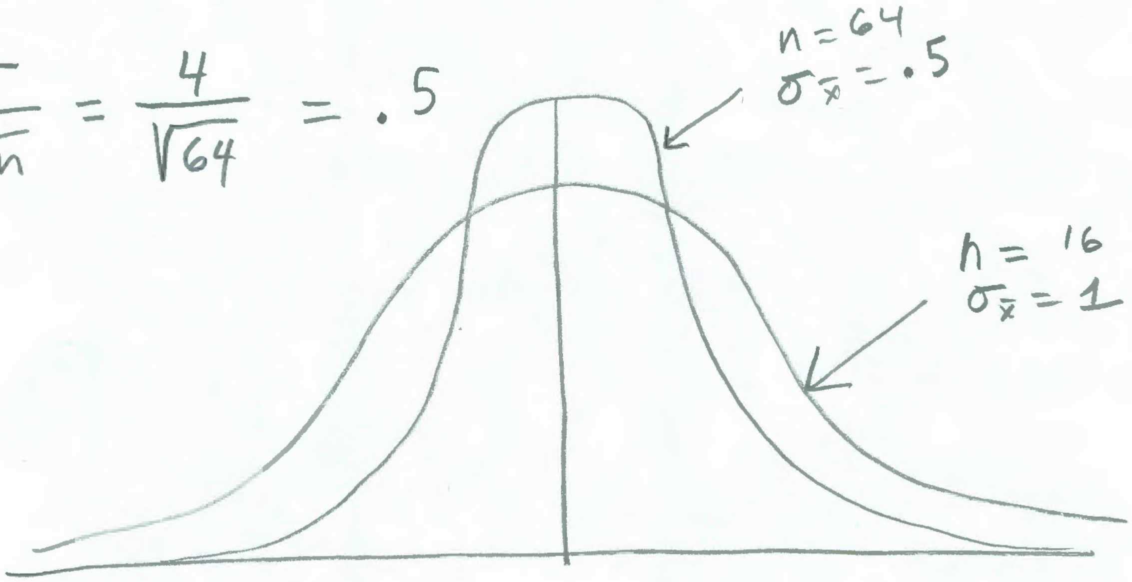
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \begin{matrix} \leftarrow \text{Sampling error} \\[1em] \leftarrow \text{Standard Error} \end{matrix}$$

12 Relationship between sample size and the Sampling Distribution of $\bar{X}$ (sample mean)

* As sample size n increases, the standard Error $\frac{\sigma}{\sqrt{n}}$ decreases

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{16}} = 1$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{64}} = .5$$



This means:

* The bigger the n, the higher the probability that the sample mean falls within a specified distance of the population mean

P. 17

Leading up to the Central Limit Theorem:

- If all samples of a particular size are selected from any population, the sampling distribution of the sample mean X_{bar} is approximately a normal distribution. This approximation improves with larger samples → See Next Page

13

Central Limit Theorem:

- In selecting random samples of size n from a population, the sampling distribution of the sample mean X_{bar} can be approximated by a normal distribution as the sample size becomes large
    - If population distribution is symmetrical but not normal, the distribution will converge toward normal when $n > 10$
    - Skewed or thick-tailed distributions converge toward normal when $n > 30$
    - Heavily skew distributions converge $n > 50$

Use of Central Limit Theorem:

- We can reason about the Sampling Distribution of Xbar with absolutely no information about the shape of the original distribution from which the sample is taken
- This means that:
    - We can take one sample and compare it to the Standard Normal Curve (NORM.S.DIST) or Normal Curve (NORM.DIST) to see if our sample result is reasonable or not.
    - If it is reasonable, the process or claim is reasonable
    - If it is not reasonable, the process or claim is not reasonable

Page 320 in text

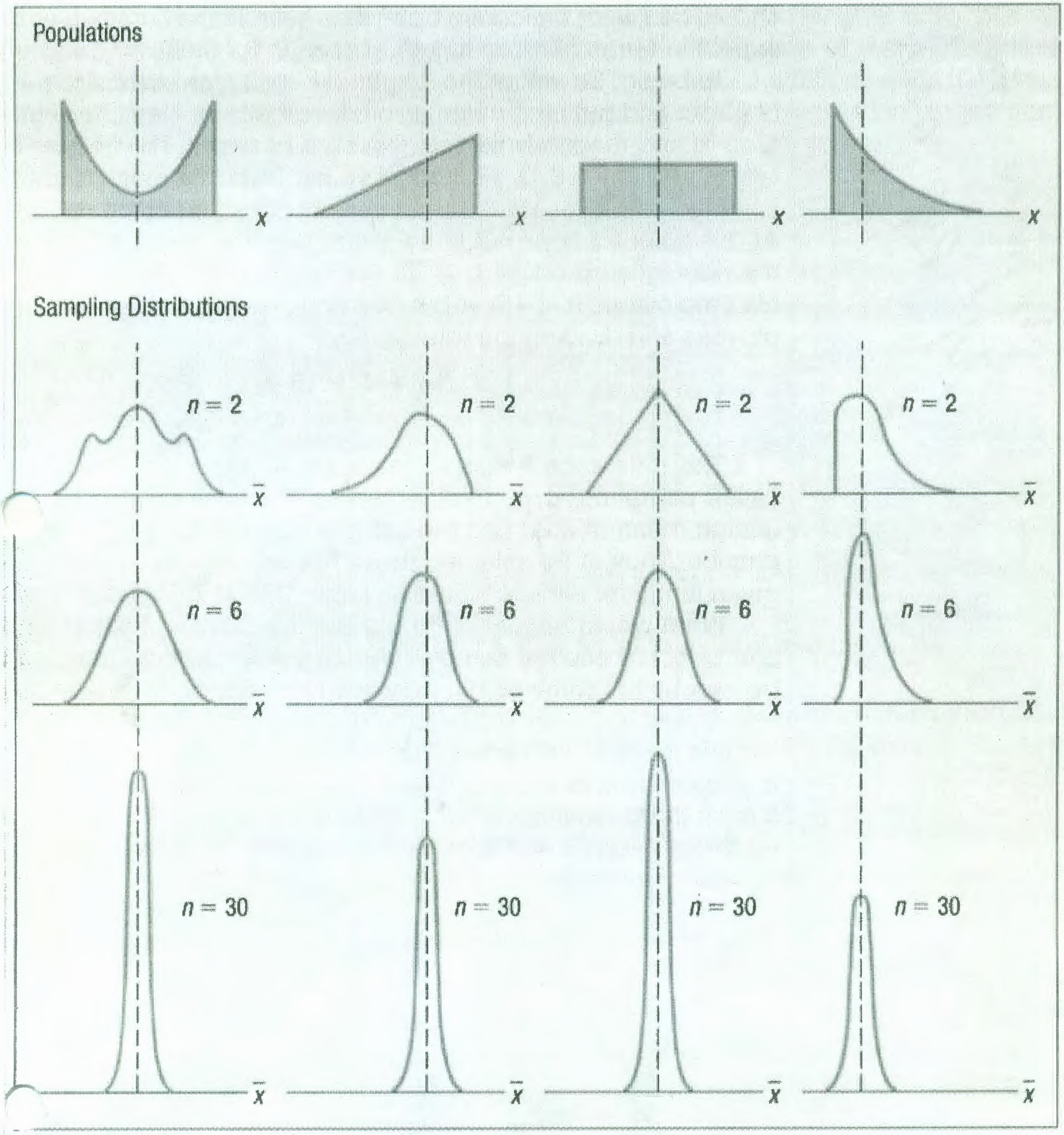


CHART 8–2 Results of the Central Limit Theorem for Several Populations

Business Decisions Example 1

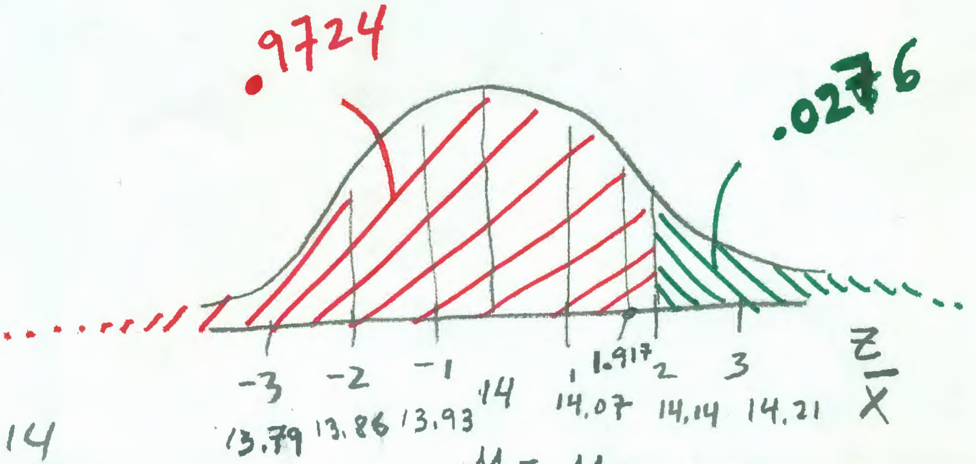
- History for a food manufacturer shows the weight for a Chocolate Covered Sugar Bombs (popular breakfast cereal) is:
  - $\mu = 14$ oz.
  - $\sigma = .4$ oz.
- If the morning shift sample shows:
  - $X_{bar} = 14.14$ oz.
  - $n = 30$
- Is this sampling error reasonable, or do we need to shut down the filling operations?

32

① variables

$\mu = 14$ oz.
$\sigma = .4$ oz
$\bar{X} = 14.14$ oz
$n = 30$

② Draw Picture



$\mu = \mu_{\bar{x}} = 14$ oz

$$\underbrace{\frac{\sigma}{\sqrt{n}} = \frac{.4}{\sqrt{30}} = .07303}_{\text{Standard Deviation of sample Means "Standard Error" Because Distr. of } \bar{X} \text{ is less spread out}}$$

③ Calculate Z

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{30}}} = \frac{14.14 - 14}{.07303}$$

$$Z_{14.14} = 1.917$$

④ Conclude

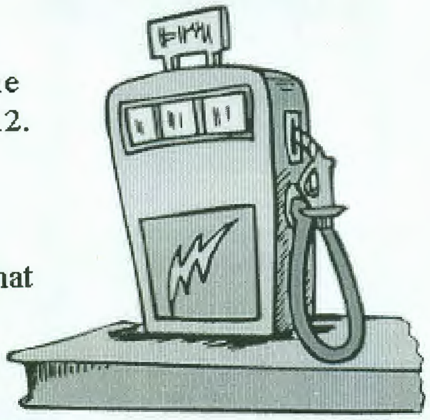
The probability associated with $\bar{X} = 14.14$ oz. or greater is .0276. This is low. It is unlikely that we could have taken a sample of 14.14 & had the sample error $(14.14 - 14 = .14)$ occur by chance ....↗

④ conclude continued...

Because it is unlikely that the sample error is due to chance, the 14.14 probably represents a machine that is filling too much.

Shut down and Fix

Suppose the mean selling price of a gallon of gasoline in the United States is $3.12. (μ) Further, assume the distribution is positively skewed, with a standard deviation of $0.98 (●). What is the probability of selecting a sample of 35 gasoline stations (n = 35) and finding the sample mean within $.33?



① Variables

$\mu = \$3.12$

sigma = $\sigma = \$.98$

$n = 35$

{distance on either side of μ} = .33

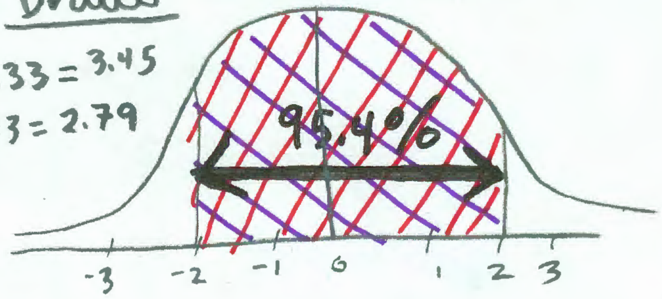
{Standard error = SD of Distribution of sample means} $= \frac{\sigma}{\sqrt{n}} = \frac{.98}{\sqrt{35}} = .16565$

② Draw

est. $\bar{X}_1 = 3.12 + .33 = 3.45$

est. $\bar{X}_2 = 3.12 - .33 = 2.79$

±2 = 95.4%



Z

X

$\mu = \mu_{\bar{x}} = 3.12$

$\frac{\sigma}{\sqrt{n}} = .16565$

③ Calculate $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$z_{3.45} = \frac{3.45 - 3.12}{.16565} = 1.99 \approx 2$

$z_{2.79} = \frac{2.79 - 3.12}{.16565} = -1.99 \approx -2$

④ The probability of selecting a sample of 35 gas stations & finding the sample mean within \$.33 of \$3.12 is .954.

↳ alternative ways of stating answer →

Alternative ways to state Answer:

① "simple random sample of 35 gas stations has a .954 probability of providing a sample mean $\overline{x}$ that is within \$.33 of the population mean of \$3.12." ✓

OR

② ".046 probability that the sampling error will be more than ± \$.33."

The sampling Distribution can be used to provide probability information about how close the sample mean is to the population mean μ

⑭ Sample proportion

$\bar{p} = \frac{x}{n}$ = sample proportion = Random Variable.

x = the number of elements in the sample that possess the characteristic of interest

n = sample size

Note: x is a binomial variable
Bi = 2
Nomial = Nominal variable

⑮ Sampling Distribution of $\bar{p}$

① The sampling Distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion $\bar{p}$.

② The sampling Distribution of $\bar{p}$ can be approximated by a Normal distribution whenever :

$$n * p \geq 5$$

and

$$n * (1 - p) \geq 5$$

⑯ Expected Value of $\bar{p}$

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

$E(\bar{p})$ = Expected value of $\bar{p}$ = unbiased estimator

p = population proportion

17

"standard error of the proportion"

Standard Deviation of Sampling Distribution of $\bar{p}$

Finite population

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

$*$ if $\frac{n}{N} \leq .05$ use : $\sigma_{\bar{p}} = \sqrt{\frac{p*(1-p)}{n}}$

Infinite pop. or process or Not feasible to List all Elements

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

Example:

If $p = .55$, $n = 30$

and you want to find probability of finding $\bar{p}$ within a margin of error of .05:

$n*p = .55 * 30 = 16.5$

$n*(1-p) = .45 * 30 = 13.5$

$$\sigma_p = \sqrt{\frac{.55*.45}{30}} = .09083$$

Probability that $\bar{p}$ will lie between .5 & .6 is:

=NORMDIST(.6,.55,.09083,1) − NORMDIST(.50,.55,.09083,1)

= .418011

The Crossett Trucking Company

The Crossett Trucking Company claims that the mean weight of their delivery trucks when they are fully loaded is 6000 lbs. And the standard deviation is 150 lbs. Assume that the population follows the normal distribution. 40 trucks are randomly selected and weighed.

Within what limits will 95% of the sample means occur?

This problem is similar to chapter 8. In chapter 7, we know what the population mean, μ, is, and we can say things like "we are 95% sure that $\bar{X}$ occurs between 6047 lbs. and 5954 lbs." In chapter 8, we do not know what the population mean, μ, is and we say things like "we are 95% sure that μ occurs between two calculated values."

① $\mu = 6000$ lbs.

$\sigma = 150$ lbs.

$n = 40$

$$\left\{\begin{array}{l}\text{standard error} \\ \text{"standard deviation} \\ \text{for Distribution of} \\ \text{Sample Mean"}\end{array}\right\} = \frac{\sigma}{\sqrt{n}} = \frac{150 \text{ lbs.}}{\sqrt{40}} = 23.71708$$

② In this problem we are not given $\bar{X}$ and asked to find the probability, we are given the probability and asked to find the $\bar{X}$. ⟶

Remember we can solve for $\bar{x}$ from our Z formula →

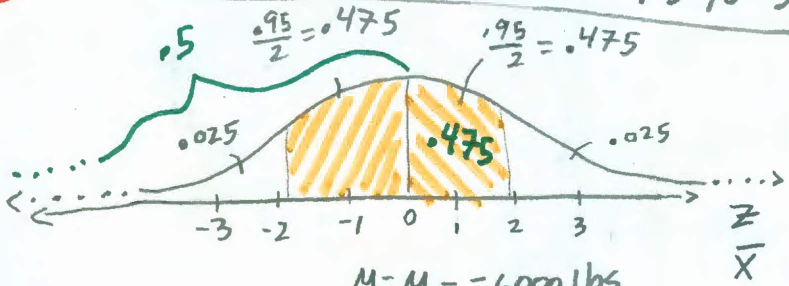
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad ①$$

$$Z\frac{\sigma}{\sqrt{n}} = \bar{x} - \mu \quad ②$$

$$\mu + Z\frac{\sigma}{\sqrt{n}} = \bar{x} \quad ③$$

$$\bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

③ Because we want to be 95% sure, we need to divide 95%/2



$\mu = \mu_{\bar{x}} = 6000$ lbs

$$\left\{\begin{array}{c}\text{SD} \\ \text{of} \\ \text{sample} \\ \text{mean}\end{array}\right\} = \frac{\sigma}{\sqrt{n}} = 23.71708$$

④ Find Z

① Book Table method, look .475 in Table

or

② =NORM.S.INV(.475+.5,1) ≈ 1.96

1.96

⑤ Find $\bar{X}$ (one on upperside and one on lower side)

$$\bar{X} = \mu \pm Z\frac{\sigma}{\sqrt{n}}$$

$6000 \pm 1.959964 * 23.71708$

$\bar{X}$ Rounded to the pound: 6046 and 5954

Answer:

It is reasonable to assume that the sample means for truck weight will occur between the limits 5954 lbs. and 6046 lbs. 95% of the time. However we do run of 5% risk that they will not occur between our limits.

Chapter 7:

Population:

All elements of interest

Sample:

A subset of the population

Why do we sample instead of look at whole population?

1) Some populations are impossible to check
    Can't count all the fish in the ocean
2) The cost of studying all the items in a population
    General Mills hires firm to test a new cereal:
    Sample test: cost ≈ $40,000
    Population test: cost ≈ $1,000,000,000
3) Contacting the whole population would often be time-consuming
    Political polls can be completed in one or two days
    Polling all the USA voters would take nearly 200 years!
4) The destructive nature of certain tests
    Test each bottle of wine?!!?
    Testing all seeds from Burpee ➔ there'd be none left
5] The sample results are usually adequate
    Consumer price index constructed from a sample is an excellent estimate for a consumer price index that could be constructed from the population

Why do we need sample data?

We select samples to collect data to answer research question about a population
    What do community college presidents think of the new federal proposal to rate colleges?
    Is the filling machine filling accurately?
    Do Highline College students think that advising should be mandatory?
    What is the mean annual accounting salary in Seattle?

Samples are only estimates:

In point estimation we use the data from the sample to compute a value of a **sample statistic** that serves as an estimate of a **population parameter**.
    We refer to **Xbar** as the point estimator of the population mean = **μ** = mew.
    We refer to **s** as the point estimator of the population standard deviation = **σ** = sigma.
    We refer to **Pbar** as the point estimator of the population mean **p**.

What about Sample Error?

Is the Xbar (Sample Mean) and μ (population mean, mew) always equal?
    Almost never!

If Xbar (Sample Mean) and μ (population mean, mew) are not equal, is this okay?
    It depends.
    We will build a model to check to see if our Xbar is a "reasonable" estimate for the population mean!! :)

Random Variable:

Numerical Description of the outcome of an experiment
If we consider the process of selecting a "Random Sample" as an experiment, the Xbar is the numerical description of the outcome of the experiment.
Thus Xbar is the random variable.

If the population is so big, how do we know where to go and get sample data?

We must be careful!!
We have to make sure that the Samples Population is the same as the Target Population.

Sampled Population:

Population from which the sample is drawn

Target Population:

Population we want to make an inference about

Sampled Population and Target Population are not always the same!

Example 1: If your goal was to gather data about students attending South Community College:
    If you took a sample from a College Registration List, Sampled Population and Target Population are the same
    If you took a sample from people eating at the Culinary Arts restaurant at South Community College you might get some people who are:
        students at South (Target Pop)
        and
        some who are not students at South (NOT Target Pop).
    The Sampled Population would be people who eat at Culinary Arts Restaurant.
    Sampled Population NOT EQUAL TO Target Population
Example 2: If you take a sample from only matinée movie-goers and you want to make inferences about all movie goers your
        Sampled Population (matinée) is different than your Target Population (all movie goers): they are not the same.
Conclusion: When a sample is used to make inferences about the population, make sure that the sampled and target population are in close agreement.
        This is not a mathematical calculation, it is a judgment call.

Chapter 7:

Frame:

List of all elements in the population

List of elements that sample will be selected from.

- Frames cannot always be created

Finite Population

A population where we can create a Frame

Examples:

- List of student names at a college
- List of Sales Invoices

“Sampling from a Finite Population”: Use “Simple Random Sampling” method to select a sample

Infinite Population

A population where we can NOT create a Frame

Examples:

- Population too big (like all the fish in the sea)
- Take a sample of cereal box weights from a cereal box filling machine
- Customers entering a store

“Sampling from an Infinite Population or Process”: Use “Random Sampling” method to select a sample

Frame that CAN be constructed:

Take sample at Highline Community College to see how many people have iPods

- Sampled Population = List of registered students
- Frame = List of registered students
- The sampled population has a finite number of elements
- This is called “Sampling from a Finite Population”. Use “Simple Random Sampling” method to select a sample

Frame that CANNOT be constructed:

Population is too big (like counting all the fish in the sea)

Take a sample of cereal box weights from a cereal box filling machine

- Sampled Population = conceptual population of all boxes that could have been filled at that particular point in time.

  In this sense, the sampled population is considered infinite.
- Frame = impossible to construct frame from infinite population because all the elements are not present
- The sampled population has a conceptually infinite number of elements

This is called “Sampling from an infinite Population or Process”. Use “Random Sampling” method to select a sample

“Random Sampling” is the same as “Simple Random Sampling”, except for two assumptions have to hold true (more later)

Processes (Sampling from an infinite Population):

Examples of processes:

- Machine fills boxes of cereal
- Machine fills bags of lettuce
- Machines make bolts and screws for airplanes
- Router makes boomerangs
- Transactions occur at bank
- Calls arrive at Highline help desk
- Customers entering store

All are viewed as coming from a process generating elements from a conceptually infinite population

How a sample can help to decide whether the process is working properly:

Processes not working properly (like machine filling too much) will produce

- sample statistics that are not close (statistically significant) to the population parameter

Processes working properly (like machine filling just right) will produce

- sample statistics that are close (statistically insignificant) to the population parameter

Chapter 7:

Simple Random Sample (for Finite Populations)

Textbook: A simple random sample of size n from a finite population of size N is a sample selected such that
each possible sample of size n has the same probability of being selected

Each sample element has the same probability of being selected, (if TRUE, then each sample has the same chance)

Think of: Uniform Distribution

You have a frame and you can randomly select from frame.

How to select a sample:

Example 1: Select any n units in a random way

Book method:

Step 1: Assign Random Number to each element in population

Step 2: Select “n” elements that have the “n” smallest (or largest) random numbers

Using Excel’s RAND or RANDBETWEEN functions (each follows a Uniform Distribution between 0 and 1)

Example 2:

RAND, SMALL and VLOOKUP functions can automat the Simple Random Sample

RAND function generates Radom numbers between 0 and 1 with 15 significant digits and generate these numbers based on a Uniform Distribution

Example 3: Names of classmates in a hat, mix up names, select until sample size, “n” is reached

Random Sample (for Infinite Population)

These must hold true:

1) Each element selected comes from same population (Sampled and Targeted Populations are the same).

2) Each element is selected independently: to prevent selection bias (prevent all from similar group, similar attitudes, or choosing to get desired result)

Used for Infinite Populations or Populations where it is not feasible to list all elements

How to select a sample:

Select any n units in a random way

Examples:

Machines filling boxes or bags, choose sample from same point in time

This is done to make sure that each element selected in selected from the same population.

People arriving at a restaurant, choose customer directly after customer who uses coupon (McDonald’s did this to simulate a random selection).

This is done to make sure that each element is selected independently (without bias)

Probability Sample

Each possible sample has a known probability of selection and a random process is used to select the elements for the sample.

Good because we have techniques to help us understand if our sample is “good”, reasonable – more later…

Examples:

1) Simple Random Sample taken (for Finite populations)

2) Random Sample (for Infinite population)

3) Stratified Random Sampling (allows smaller sample size and lower cost)

Population divided into mutually exclusive strata where elements in each strata are similar

Each element in the strata are similar (location, age, department, major)

Works best when the variation among the elements in each strata is relatively small.

4) Cluster Sampling

Elements in clusters are not a like – each cluster is like a min population

Helps to reduce cost.

5) Systematic Sampling

Like with invoices or other ordered populations.

Non-probability Sampling

Not good because we can’t calculate how reasonable the sample results are

Convenience Sampling

Judgment Sampling

Chapter 7:

Leading up to the Central Limit Theorem:

If all samples of a particular size are selected from any population, the sampling distribution of the sample mean Xbar is approximately normal distribution.

This approximation improves with larger samples

Sampling Distribution of Xbar (SDofXbar)

The sampling Distribution of Xbar is the probability distribution of all possible values of sample mean Xbar.

Sample Mean of Sampling Distribution of Xbar = E(Xbar) = μ = Pop Mean!!!! Called Unbiased point estimator

Standard Deviation of SDofXbar = σ_{Xbar} = Standard Error = (σ/SQRT(n))*SQRT((N-n)/(n-1)), Don't need SQRT((N-n)/(n-1)) if population is finite AND n/N <= 0.05

If Pop is Finite AND n/N > 0.05, then use the Finite Population Correction Factor: SQRT((N-n)/(n-1))

As sample size increases, Standard Error will decrease, and thus provide a higher probability that Xbar will fall within a specified distance of the population mean.

Book assumes all problems use s/SQRT(n), unless otherwise stated

The Sampling Distribution can be used to provide probability information about how close the sample statistic is to the population parameter

Sampling Distribution of Pbar (SDofPbar)

Sample Proportion = point estimator of Population Proportion = Pbar = x/n

x = the number of elements in the population that posses the characteristic of interest = binomial variable (bi = 2, nominal = nominal variable)

n = sample size

The sampling Distribution of Pbar is the probability distribution of all possible values of sample proportion Pbar

Sample Proportion of Sampling Distribution of Pbar = E(Pbar) =p = Pop Proportion!! Called Unbiased point estimator

Standard Deviation of SDofPbar = σ_{Pbar} = Standard Error = SQRT(p*(1-p)/n)*SQRT((N-n)/(n-1)), Don't need SQRT((N-n)/(n-1)) if pop is finite AND n/N <= 0.05

If n/N > 0.05, then use the Finite Population Correction Factor: SQRT((N-n)/(n-1))

As sample size increases, Standard Error will decrease, and thus provide a higher probability that Pbar will fall within a specified distance of the population mean.

Book assumes all problems use SQRT(p*(1-p)/n), unless otherwise stated

The sampling Distribution of Pbar (and Binomial Distribution) can be approximated by a normal distribution whenever n*p>=5 AND n*(1-p) >=5

Standard Error

Standard Error stands for the Standard Deviation of a point estimator

Examples:

Standard Error of the Mean = σ_{Xbar} = "Standard Error"

Standard Error of the Proportion = σ_{Pbar} = "Standard Error"

Central Limit Theorem:

In selecting random samples of size n from a population, the sampling distribution of the sample mean Xbar can be approximated by a normal distribution as the sample size becomes large

If population distribution is symmetrical but not normal, the distribution will converge toward normal when n > 10

Skewed or thick-tailed distributions converge toward normal when n > 30

Heavily skew distributions converge n > 50

Use of Central Limit Theorem:

We can reason about the Sampling Distribution of Xbar with absolutely no information about the shape of the original distribution from which the sample is taken

This means that:

We can take one sample and compare it to the Standard Normal Curve (NORM.S.DIST) or Normal Curve (NORM.DIST) to see if our sample result is reasonable.

If the sample mean seems reasonable, the original process or claim is reasonable

If the sample mean does not seem reasonable, the original process or claim is not reasonable

The Sampling Distribution can be used to provide probability information about how close the sample statistic is to the population parameter

Statistical Inference:

The process of using data obtained from a sample to make estimates or test hypotheses about characteristics of the population (like mean).

Draw reasonable conclusions about population from statistics

Infer:

Conclude from evidence