Abe Mirza **Topics Review** Statistics

Part III

Point and Interval Estimation

Learning Objectives

What do we estimate? **Population Mean** ($\mu = ?$) or **Population Proportion** ($P = ?$)
Why do we estimate?
Due to our limited resources (Time, Money, manpower, destruction of tested subjects, widely scattered data, hardly accessible subjects).
Know all the new **terminologies** and related **notations** (Point estimate $\overline{x}$, $\hat{p}$, Margin of error)
Know all the new **formulas** on **formula sheet** and their related **TI commands**.
Know in estimating **population mean** ($\mu = ?$) when to use **normal distribution** versus **t- distribution**.
Know how to use TI (**option 7 or 8**) or (**formula** $\mu = \overline{x} \pm E$ **+ table**) to estimate **population mean** ($\mu = ?$).
Know how to use TI (**option A**) or (**formula** $P = \hat{p} \pm E$ **+ table**) to estimate **population proportion** ($P = ?$).
Know how to use TI (**option 9**) or (**formula** $\mu_1 - \mu_2 = (\overline{x}_1 - \overline{x}_2) \pm E$ **+ table**) to estimate **difference between two population Means** ($\mu_1 - \mu_2$).
Know how to use TI (**option B**) or (**formula + table**) to estimate **difference between two population proportions** ($P_1 - P_2$).
Know how to use (**formula + table**) to **determine sample size** for **Population Mean** ($\mu = ?$) or **Population Proportion** ($P = ?$)

Important Note 1: Start part 3 with quick start along with topics review

Important Note 2: As you study each page of **topics Review,** do all the problems listed at the bottom of the page from practice problem before going to the next page.

Important Note 3: For all practice problems the answers and complete solutions are given on later pages.

Important Note 4: Doing all related practice problems.

Quizzes for Part 3

Be sure you have formula sheet and the 2 pages of table (Normal and T-Distribution).

Quiz # 8: This quiz covers pages **8 through 11** of topics review.

Quiz # 9: This quiz covers pages **8 through 13** of topics review.

Quiz # 10: This quiz covers pages **8 through 18** of topics review.

Quiz # 11: This quiz covers pages **8 through 20** of topics review.

Overview

One major application of inferential statistics involves the use of sample data to estimate the value of a population parameter such as means (μ) and proportions (P).

Objectives:

a) To introduce methods for estimating values of two important population parameters: mean (μ) and proportions (P). This can be done by using a point estimate $(\bar{X} \text{ or } \hat{P})$ that is the value of a statistic to estimates the value of a parameter $(\mu \text{ or } P)$.
b) To present methods for determining sample sizes necessary to estimate those parameters.

Example:

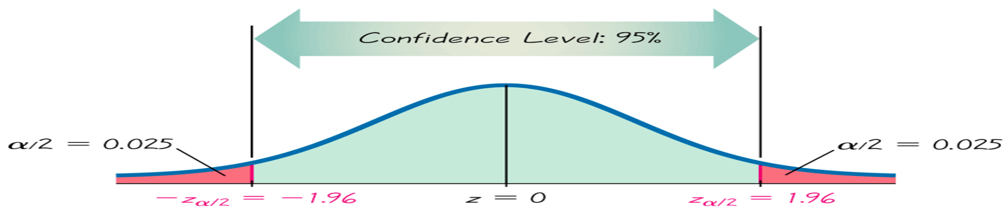
1. Estimate the **average** life of Diehard batteries? $\mu = ?$
2. Estimate the **average** waiting time at a supermarket register? $\mu = ?$
3. Estimate the **average** clarity (in depth) of water at Lake Tahoe? $\mu = ?$
4. Estimate the **percentage** of residents in North America that only speak English at home? $P = ?$
5. Estimate the **percentage** of drivers text while driving? $P = ?$
6. Estimate the **percentage** of registered voters will vote for next democratic candidates? $P = ?$
7. Estimate the **average difference** in battery life between Diehard and Everlast brand? $\mu_D - \mu_E = ?$
8. Estimate the **percentage difference** between female and male who pass stat class? $P_f - P_m = ?$

Definitions:

Point estimate: Sample statistics such as $(\bar{X} \text{ or } \hat{P})$

Confidence Interval: A confidence interval (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.

A confidence level*: a* confidences level is the probability $(1-\alpha)$ (often expressed as the equivalent percentage value) usually 90%, 95%, or 99%.that is the proportion of times that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. Percentage outside confidence level is called **critical area** (α). So for example with $95\% = (1-\alpha)$ confidence level then the critical **area** will be $(\alpha = 5\%)$.



Critical Value(s): *The* $z_{\alpha/2}$ value *that can be found from the table based on different confidence level.*
Margin of error: (also called error, error bound or maximum error) is the maximum likely difference observed between point estimates $(\bar{X} \text{ or } \hat{P})$ and population parameter $(\mu \text{ or } P)$.

Estimating One Population **Mean** $\mu = \bar{x} \pm E$

$\bar{X}$ = **Point** estimate (Sample Mean)		E = Margin of error(error bound)
Decision making process based on σ (population standard deviation)		
Margin of Error	$E = z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ σ (known or given) **(For $z_{\alpha/2}$, use Table page 3)** $E = z_{\alpha/2}\dfrac{s}{\sqrt{n}}$ σ (unknown or not given) and $n > 30$ **(For $z_{\alpha/2}$, use Table page 3)** $E = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ σ (unknown or not given) and $n \le 30$ **(For $t_{\alpha/2}$, use Table page 4)** Population is normally distributed	
Interval Estimate	$\mu = \bar{x} \pm E$	$\mu = \bar{x} \pm E$
TI-83/84	*stat* → *tests* → ***Option 7(Z-interval)***	*stat* → *tests* → ***Option 8(t-interval)***

Width (difference between upper and lower bounds) $= 2E = UB - LB$ **so** $E = (UB - LB)/2$

Point Estimate (**middle** of upper and lower bounds) $= \bar{x} = (UB + LB)/2$

Estimating Population Proportion $P = \hat{p} \pm E$

$\hat{p} = \dfrac{x}{n}$ **(Called p-hat is sample proportion and point estimate for population proportion)**	**E = Margin of error** $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Width (difference between upper and lower bounds) $= 2E = UB - LB$ **so** $E = (UB - LB)/2$

Point Estimate (middle **of upper and lower bounds**) $= \hat{p} = (UB + LB)/2$

TI-83 ***stat* → *test* → *Option A***

Estimating the *difference* between Two Populations *Means* or *Proportions*

Mean $\mu_1 - \mu_2$	Proportion $P_1 - P_2$
$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm E$	$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm E$
Point **estimate** $= (\bar{x}_1 - \bar{x}_2)$	Point **estimate** $= (\hat{p}_1 - \hat{p}_2)$
$E = z_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	$E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
TI-83/84 ***stat* → *test* → *Option 9***	TI-83/84 ***stat* → *test* → *B***

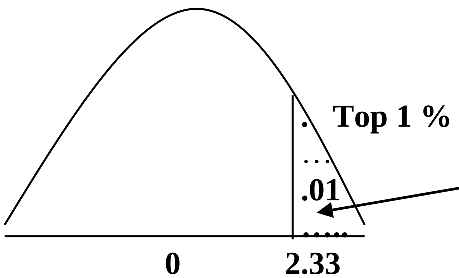
Sample Size Determination for the Estimation **of Population**

Mean = μ	**Proportion** = P
$n = (z_{\alpha/2} S / E)^2$	$n = (z_{\alpha/2} / E)^2\, \hat{p}(1-\hat{p})$
If S **is** unknown **then estimate it by** $S = Range / 4$	**If** $\hat{p}$ **is** unknown **then estimate it by** $\hat{p} = 0.5$

Important: If confidence level is not given use 95% as a default.

Based on Standard Normal Distribution $\mu = 0$ and $\sigma = 1$

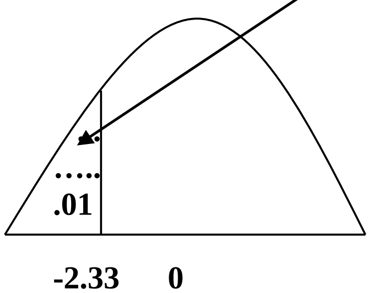
Out Side Area



OR

Out Side Area

Bottom 1 %

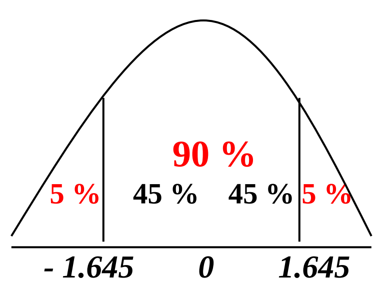


Confidence Level	Out Side Area On left or right Cut-off Point	Z - Value (±) Critical Value $= Z_{\alpha/2}$
99%	.005	± 2.5758
98%	.01	±2.3263
97%	.015	±2.1701
96%	.02	±2.0537
95%	.025	±1.9600
94%	.03	±1.8808
92%	.04	±1.7507
90%	.05	±1.6450
88%	.06	±1.5548
86%	.07	±1.4758
84%	.08	±1.4051
82%	.09	±1.3408
80%	.10	±1.2816
78%	.11	±1.2265
76%	.12	±1.1750
70%	.15	±1.0364
60%	.20	±0.8416
50%	.25	±0.6749
40%	.30	±0.5244

How to find the Z -value for confidence intervals.

Example: Find the Z - value for 90% confidence interval

1. Divide 90% = 0.90 by 2, $\Rightarrow .90/2 = 0.45$
2. Subtract 0.45 from 0.5 $\Rightarrow .5 - 0.45 = .05$
3. Look for area close to 0.05from **inside** the table (page1).
4. ***Find its corresponding Z-value (- 1.645)***

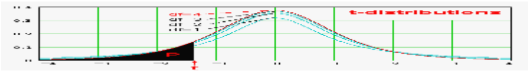


TI-83/84 *2nd* $\rightarrow$ *Distr* $\rightarrow$ *Option 3* *input* *(% ,0 ,1)*

Example: *2nd* $\rightarrow$ *Distr* $\rightarrow$ *Option 3* *input (.05 ,0 ,1) enter , then the answer will be - 1.645*

Example: *2nd* $\rightarrow$ *Distr* $\rightarrow$ *Option 3* *input (.95 ,0 ,1) enter, then the answer will be 1.645*

Hint for TI *% is the area to the left of the cut off point.*



t -Distribution for small sample $n < 30$ and σ Unknown

df = n-1	<-------------------- alpha α -------------------->							
2-Tailed	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.005**
1-Tailed	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0025**
Conf. Levl.	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**
1	1.376	1.963	3.078	6.314	12.706	31.821	63.656	127.321
2	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089
3	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453
4	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598
5	***0.920***	***1.156***	***1.476***	***2.015***	***2.571***	***3.365***	***4.032***	***4.773***
6	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833
9	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690
10	***0.879***	***1.093***	***1.372***	***1.812***	***2.228***	***2.764***	***3.169***	***3.581***
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326
15	***0.866***	***1.074***	***1.341***	***1.753***	***2.131***	***2.602***	***2.947***	***3.286***
16	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252
17	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222
18	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174
20	***0.860***	***1.064***	***1.325***	***1.725***	***2.086***	***2.528***	***2.845***	***3.153***
21	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135
22	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119
23	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104
24	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091
25	***0.856***	***1.058***	***1.316***	***1.708***	***2.060***	***2.485***	***2.787***	***3.078***
26	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067
27	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057
28	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047
29	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038
30	***0.854***	***1.055***	***1.310***	***1.697***	***2.042***	***2.457***	***2.750***	***3.030***
n>30 $\Rightarrow$ ***Z***	***0.842***	***1.036***	***1.282***	***1.645***	***1.96***	***2.326***	***2.576***	***2.807***

for $n > 30$ Use Bottom row

2-T	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.005
1-T	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0025
Conf. Levl.	60%	70%	80%	90%	95%	98%	99%	99.5%

T-Distribution vs. the Normal Distribution for Confidence Interval for Means

Main Point to Remember:
You must use the t-distribution table when working problems when the population standard deviation (σ) is not known and the sample size is small (n<30).

General Correct Rule:
If σ is not known, then using t-distribution is correct. If σ is known, then using the normal distribution is correct.

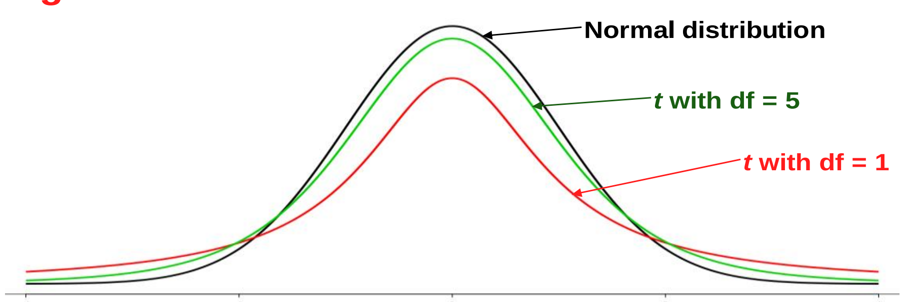
What is Most Common Practice:
Since people often prefer to use the normal, and since the t-distribution becomes equivalent to the normal when the number of cases becomes large, common practice often is:

- If σ known, then use normal.
- If σ not known:
  - If n is large, then use normal.
  - If n is small, then use t-distribution.

What is Another Common Way Textbooks Teach This:
Textbooks often simplify this to "large-sample" vs. "small-sample" methods; use normal distribution with large samples and t-distribution with small samples. This is right almost all the time, because in real sampling problems we seldom have a basis for knowing σ. However, there can be some situations when we do have a basis for assuming a value for σ, such as using a σ based on past data, and in those situations even if sample size is small the correct procedure would be to use the normal distribution, so the simplified "large-sample" vs. "small sample" approach would lead to an error.

t distribution

- *t* distribution looks like a normal distribution, but has "thicker" tails. The tail thickness is controlled by the **degrees of freedom**



- The smaller the degrees of freedom, the thicker the tails of the *t* distribution
- If the degrees of freedom is large (if we have a large sample size), then the t distribution is pretty much identical to the normal distribution

Estimating one population Mean $\mu = \overline{x} \pm E$

a) What do we estimate? Population mean (μ) or sample mean ($\bar{x}$) or both?

b) Why do we need to estimate? Cite some reasons?

c) What is the point estimate?

d) What is the confidence level?

e) What is the criteria of t-distribution?

f) Under what condition we use t-distribution?

g) What is the formula for degree of freedom $df = ?$

h) What is the margin of error and what are three different possible formulas for it?

i) Where you can find the z table and under what condition you will be using this table?

j) Where you can find the t table and under what condition you will be using this table?

k) What is the width of a confidence interval?

l) How we can use the width of a confidence interval to find point estimate?

m) How we can use the width of a confidence interval to find margin of error?

n) How to use **TI calculator** to find the boundaries of a confidence interval when we use **normal distribution**?

o) How to use **TI calculator** to find the boundaries of a confidence interval when we use **t-distribution**?

A) For the following problems decide to z or t value or neither?

1) Sample size $n = 20$, $s = 4$ and the population is normally distributed?
2) Sample size $n = 19$, $\sigma = 4$ and the population is normally distributed?
3) Sample size $n = 18$, $s = 4$ and the population is normally distributed?
4) Sample size $n = 10$, $\sigma = 4$ and the population is normally distributed?
5) Sample size $n = 100$, $s = 4$ and the population is normally distributed?

Find the margin of error for the following problems?

B) $E = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ **B)** $E = Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ **C)** $E = t_{\alpha/2, df}\left(\frac{s}{\sqrt{n}}\right)$

1) Sample size $n = 36$, $\sigma = 4$ and 95% confidence level? **Answer: 1.31**
2) Sample size $n = 16$, $s = 8$ and 99% confidence level? **Answer: 5.894**
3) Sample size $n = 18$, $\sigma = 30$ and 90% confidence level? **Answer: 11.6**
4) Sample size $n = 100$, $s = 4$ and 97% confidence level? **Answer: 0.868**
5) Sample size $n = 14$, $s = 10$ and the 95% confidence level? **Answer: 5.773**

C) Fill in the blanks with one of the following: *increases, decreases,* or *stays the same* where $E = z\frac{\sigma}{\sqrt{n}}$.

a) As the sample size (n) decreases, the margin of error (E) ______________.

b) As the confidence level (C) decreases, the margin of error (E) _________________.

A) $\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ **B)** $\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ **C)** $\bar{x} \pm t_{\alpha/2, df}\left(\frac{s}{\sqrt{n}}\right)$

1) A random sample of 25 life insurance policy holders showed that the average premiums paid on their life insurance policies was $340 per year with a standard deviation of $62. Construct a 90% confidence interval for the population mean. $n =$ $\bar{x} =$ $\sigma =$ **or** $s =$

$E =$ $\mu =$ $318.79 < \mu < 361.21$

2) A random sample of 49 life insurance policy holders showed that the average premiums paid on their life insurance policies was $340 per year with a standard deviation of $62. Construct a 90% confidence interval for the population mean. $n =$ $\bar{x} =$ $\sigma =$ **or** $s =$

$E =$ $\mu =$ $325.43 < \mu < 354.57$

3) A company that produces white bread is concerned about the distribution of the amount of sodium in its bread. The company takes a simple random sample of 25 slices of bread and computes the sample mean to be 100 milligrams of sodium per slice. Construct a 90% confidence interval for the unknown mean sodium level assuming that the sample standard deviation is 10 milligrams.
$n =$ $\bar{x} =$ $\sigma =$ **or** $s =$

$E =$ $\mu =$ $96.58 < \mu < 103.42$

4) The football coach randomly selected eight players and timed how long it took to perform a certain drill. The times in minutes were: 12, 9, 15, 7, 6, 15, 7, 10. Assuming that the times follow a normal distribution, find a 90% confidence interval for the population mean. $n =$ $\bar{x} =$ $\sigma =$ **or** $s =$
$E =$ $\mu =$ $7.73 < \mu < 12.51$

5) A company that produces white bread is concerned about the distribution of the amount of sodium in its bread. The company takes a simple random sample of 25 slices of bread and computes the sample mean to be 100 milligrams of sodium per slice. Construct a 90% confidence interval for the unknown mean sodium level assuming that the population standard deviation is 10 milligrams
$n =$ $\bar{x} =$ $\sigma =$ **or** $s =$

$E =$ $\mu =$ $96.71 < \mu < 103.29$

6) The actual time it takes to cook a ten-pound turkey is a normally distributed. Suppose that a random sample of 9 ten pound turkeys is taken. Given that an average of 2.9 hours and a standard deviation of .24 hours was found for a sample of 9 turkeys, calculate a 95% confidence interval for the average cooking time of a ten-pound turkey. $n =$ $\bar{x} =$ $\sigma =$ **or** $s =$

$E =$ $\mu =$ $2.72 < \mu < 3.08$

Estimating the μ = average life of Diehard batteries by using **95%** confidence Level. A sample of **64** batteries has $\bar{\mathbf{x}} = 50$ **months** .From prior study, we know $\sigma = \mathbf{10}$ **months**

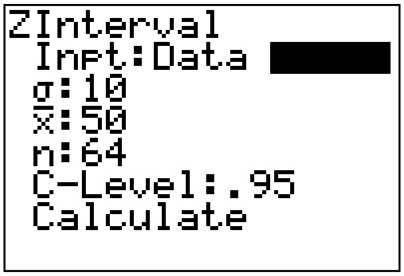
Solution by Formula

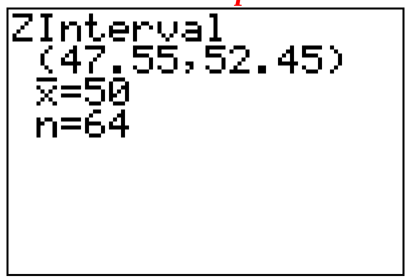
σ **known,** $\rightarrow E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{10}{\sqrt{64}}$ $E = 2.45$ $\mu = 50 \pm 2.45$ $47.55 < \mu < 52.45$

By 95% confidence, the average life of Diehard batteries is between 47.55 to 52.45 months

Solution by TI 83/84 Calculator

σ **known** $\rightarrow$ **TI-83/84** ***stat*** $\rightarrow$ ***tests*** $\rightarrow$ ***Option 7***





$E = (UB - LB)/2 = (52.45 - 47.55)/2 = 2.45$

Estimating the μ = average life of Diehard batteries by using **95%** confidence Level when a sample of **6** batteries provides these data 48,54,57,45, 56,52

Solution by Formula

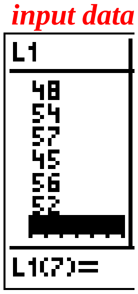
Hint: to use the formula, you need to calculate $\bar{\mathbf{x}} = ?$. $S = ?$ $\bar{\mathbf{x}} = 52$ **months** . $S = 4.69$ **months**

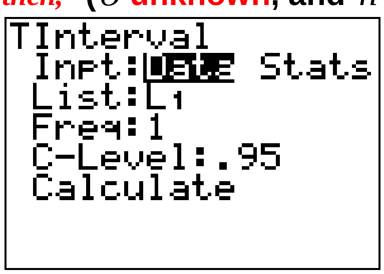
(σ **unknown, and** $n \le 30$) **(for t- value use table page 4)** $df = 6 - 1$ $t_{\alpha/2} = 2.571$ $E = 2.571\frac{4.69}{\sqrt{6}}$

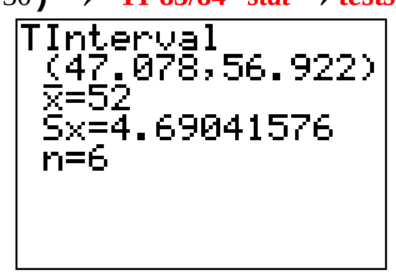
$E = 4.92$ $\mu = 52 \pm 4.92$ $47.08 < \mu < 56.92$

Solution by TI 83/84 Calculator

input data in L1 ***then,*** **(**σ **unknown, and** $n \le 30$**)** $\rightarrow$ **TI-83/84** ***stat*** $\rightarrow$ ***tests*** $\rightarrow$ ***Option 8***







$E = (UB - LB)/2 = (56.92 - 47.08)/2 = 4.92$

Summary as to decide to use Normal or t-Distribution in estimating One Population Mean $\mu = \bar{x} \pm E$

σ *(population st. dev.)* *is* ***known***	$n > 30$ σ is **unknown**	σ is **unknown**, and $n < 30$	**Raw Data Only** $n < 30$
$\sigma = 10$ ***Sample*** $n = 9$ $\bar{x} = 50$ **Use 95 %** **Confidence Level** $\mu = \bar{x} \pm E = 50 \pm E$ *Because* σ ***is known****, use Normal distribution* $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ **(for** $z_{\alpha/2}$ **- value use table on page 4)** **For 95% confidence** $z_{\alpha/2} = 1.96$ $E = 1.96 \frac{10}{\sqrt{9}} = 6.53$ $\mu = 50 \pm 6.53$ $43.47 < \mu < 56.53$ **Or** (43.47 , 56.53) ***By 95% confidence, the population average is between*** 43.47 ***and*** 56.53	***Sample*** $n = 64$ $\bar{x} = 50$ $s = 10$ **Use 95 %** **Confidence Level** $\mu = \bar{x} \pm E = 50 \pm E$ Because $n > 30$, use *Normal distribution* $E = z_{\alpha/2} \frac{s}{\sqrt{n}}$ **(for** $z_{\alpha/2}$ **- value use table on page 4)** **For 95% confidence** $z_{\alpha/2} = 1.96$ $E = 1.96 \frac{10}{\sqrt{64}} = 2.45$ $\mu = 50 \pm 2.45$ $47.55 < \mu < 52.45$ **Or** (47.55 , 52.45) ***By 95% confidence, the population average is between*** 47.55 ***and*** 52.45	***Sample*** $n = 9$ $\bar{x} = 50$ $s = 10$ **Use 95 %** **Confidence Level** $\mu = \bar{x} \pm E = 50 \pm E$ Because σ is unknown, and $n < 30$ use ***t- distribution*** $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$ **(for** $t_{\alpha/2}$ **- value use table on page 6)** **Use 95% column with df=9-1=8** $t_{\alpha/2} = 2.306$ $E = 2.306 \frac{10}{\sqrt{9}} = 7.69$ $\mu = 50 \pm 7.69$ $42.31 < \mu < 67.69$ **Or** (42.31 , 67.69) ***By 95% confidence, the population average is between*** 42.31 ***and*** 67.69	***Sample*** 42,48,52,58 $n = 4$ $\bar{x} = 50$ $s = 6.73$ **Use 95 %** **Confidence Level** $\mu = \bar{x} \pm E = 50 \pm E$ Because σ **is unknown,** and $n < 30$ use ***t- distribution*** $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$ **(for** $t_{\alpha/2}$ **- value use table on page 6)** **Use 95% column with df=4-1=3** $t_{\alpha/2} = 3.182$ $E = 3.182 \frac{6.73}{\sqrt{4}} = 10.71$ $\mu = 50 \pm 10.71$ $39.29 < \mu < 60.71$ **Or** (39.29 , 60.71) ***By 95% confidence, the population average is between*** 39.29 ***and*** 60.71
TI Instruction	**TI** Instruction	**TI** Instruction	**TI** Instruction
stat → tests → Option 7	***stat → tests → Option 7***	***stat → tests → Option 8***	***stat → tests → Option 8*** ***Hint****: You can put data in L1, and after selection option 8 select data*

Estimating population proportion

$$P = \hat{p} \pm E$$

Assumptions

1. The sample is a simple random sample.
2. All conditions for the binomial distribution are satisfied.
3. The normal distribution can be used to approximate the distribution of sample proportions because $n\,p > 5$ and $n\,q > 5$ are both satisfied.

a) What do we estimate? Population percentage ($\boldsymbol{P}$) or sample mean ($\hat{\boldsymbol{P}}$) or both?
b) Why do we need to estimate? Cite some reasons?
c) What is the point estimate?
d) What is the confidence level?
e) What is the margin of error formula for estimation population proportion?
f) What is the width of a confidence interval?
g) How we can use the width of a confidence interval to find point estimate?
h) How we can use the width of a confidence interval to find margin of error?
i) How to use **TI calculator** to find the boundaries of a confidence interval when we use **normal distribution**?

YouTube TI Calculator: https://www.youtube.com/watch?v=OVc5BCa0UvQ **General introduction**
YouTube TI Calculator: https://www.youtube.com/watch?v=e3HZ6Xv-plk **General introduction**

A) Find the margin of error for the following problems?

$$E = Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

A-1) Sample size $\boldsymbol{n = 50}$, $\boldsymbol{x = 24}$ and 90% confidence level? $E = .116$
A-2) Sample size $\boldsymbol{n = 50}$, $\boldsymbol{x = 24}$ and 95% confidence level? $E = .138$
A-3) Sample size $\boldsymbol{n = 80}$, $\boldsymbol{x = 32}$ and 95% confidence level? $E = .107$
A-4) Sample size $\boldsymbol{n = 100}$, $\boldsymbol{\hat{p} = .6}$ and 97% confidence level? $E = .106$
A-5) Sample size $\boldsymbol{n = 320}$, $\boldsymbol{\hat{p} = .45}$ and the 90% confidence level? $E = .046$

Fill in the blanks with one of the following: *increases, decreases,* or *stays the same* where $E = Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$.

1) As the sample size (n) decreases, the margin of error (E) ________________.
2) As the confidence level (C) decreases, the margin of error (E) __________________.

$$P = \hat{p} \pm E$$

1. If 64% of a sample of 550 people leaving a shopping mall claims to have spent over $25, determine a 99% confidence interval estimate for the proportion of shopping mall customers who spend over $25. Interpret your interval.
$E =$ $P =$ $58.73\% < P < 69.27\%$

2. In a random sample of machine parts, 18 out of 225 were found to have been damaged in shipment. Establish a 95% confidence interval estimate for the proportion of machine parts that are damaged in shipment. Interpret your interval.
b. $E =$ $P =$ $4.5\% < P < 11.5\%$

3. A telephone survey of 1000 adults was taken shortly after the U.S. began bombing Iraq. If 832 voiced their support for this action. Create a 99% confidence interval and interpret the interval.

$E =$ $P =$ $80.16\% < P < 86.25\%$

4. An assembly line does a quality check by sampling 50 of its products. It finds that 16% of the parts are defective.

a. Create a 95% confidence interval for the percent of defective parts for the company and interpret this interval.
b. If we decreased the confidence level to 90% what would happen to:
   i. the critical value?
   ii. the margin of error?
   iii. the confidence interval?
c. If the sample size were increased to 200, the same sample proportion were found, and we did a 95% confidence interval; what would happen to:
   i. the critical value?
   ii. the margin of error?
   iii. the confidence interval?

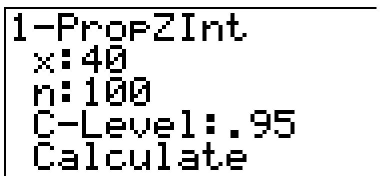
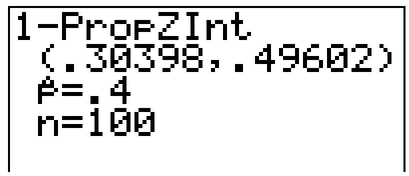
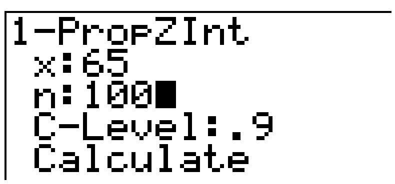
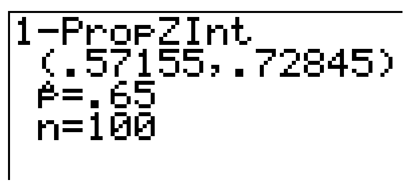
5. A nationwide poll was taken of 1432 teenagers (ages 13-18). 630 of them said they have a TV in their room.

a. Create a 90% confidence interval for the proportion of all teenagers who have a TV in their room and interpret it.
   What does "90% confidence" mean in this context?
   If we increased the confidence level to 99% what would happen to:
   iv. the critical value?
   v. the margin of error?
   vi. the confidence interval?

6. If the sample size were changed to 950, the same sample proportion were found, and we did a 90% confidence interval; what would happen to:
   ii. the critical value?
   iii. the margin of error?
   iv. the confidence interval?

7. Suppose a 90% confidence interval is stated as (0.3011, 0.4189).

a. What is the sample proportion from this sample?
b. What is the margin of error?

		TI-83 *stat* $\rightarrow$ *test* $\rightarrow$ *Option A*
6	Use **95%** confidence Level to estimate the percentage of drivers texting while driving when in a sample of 100 drivers 40 text and drive. ***Solution by Formula*** $\hat{p} = \frac{x}{n} = \frac{40}{100} = 0.4$, $E = 1.96\sqrt{\frac{0.4(1-0.4)}{100}} = .096$ $P = 0.4 \pm 0.096$ $0.304 < P < 0.496$ *or* $30.4\% < P < 49.6\%$ By 95% confidence, the percentage of drivers that text while driving is between 30% to 49.6 % ***Solution by TI 83/84 Calculator***   $E = (UB - LB)/2 = (0.49602 - 0.30398)/2 = 0.09602$	
7	Use **90%** confidence Level to estimate the percentage of drivers texting while driving when in a sample of 100 drivers 65 text and drive . ***Solution by Formula*** $\hat{p} = \frac{x}{n} = \frac{65}{100} = 0.65$, $E = 1.645\sqrt{\frac{0.65(1-0.65)}{100}} = .0785$ $P = 0.65 \pm 0.0785$ $0.5715 < P < 0.7285$   $E = (UB - LB)/2 = (0.72845 - 0.57155)/2 = 0.07695$	

Estimating Difference between Two Populations Means $(\mu_1 - \mu_2)$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm E \qquad E = z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(for z- value use table page 3)

j) What is the point estimate for the **difference** between two population **means** $\mu_1 - \mu_2$?
k) What is the error formula for the **difference** between two population **means**?
l) How to use TI to estimate for the **difference** between two population **means**?
m) What conclusion can we draw if the lower bound of the estimate happened to be zero or negative?
n) What conclusion can we draw if the both bounds of the estimate happened to be negative?

3 different ways the final answer may look like.

1) Both end signs **positive** $+ < \mu_1 - \mu_2 < +$ Indicating **group 1** has higher average than **group 2.**

2) **Different** end signs $- < \mu_1 - \mu_2 < +$ **Inconclusive** as which group has higher average.

3) Both end signs **negative** $- < \mu_1 - \mu_2 < -$ Indicating **group 2** has higher average than **group 1.**

1) Use **99%** confidence level to estimate the difference in average life (in months) of "Diehard" and "Everlast" batteries

Diehard	Everlast
$n_1 = 44$	$n_2 = 36$
$\bar{x}_1 = 51.8$	$\bar{x}_2 = 47.4$
$s_1 = 2.5$	$s_2 = 3.7$

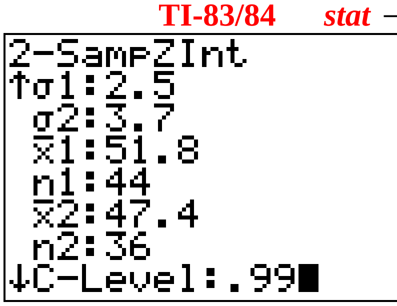
$$\mu_1 - \mu_2 = (51.8 - 47.4) \pm E \qquad E = 2.58\sqrt{\frac{2.5^2}{44} + \frac{3.7^2}{36}} = 1.86$$

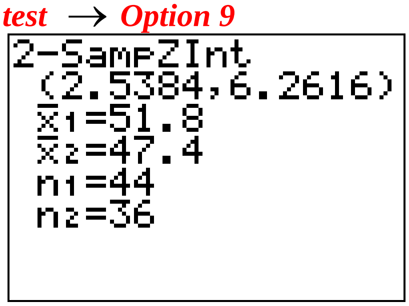
$$\mu_1 - \mu_2 = 4.4 \pm 1.86 \qquad 2.54 < \mu_1 - \mu_2 < 6.26$$

Conclusion: Both **signs are positive** indicating **group** 1 **(**Diehard**)** has higher average than **group** 2 **(**Everlast**).** By 99% confidence, Diehard batteries on the average last between 2.54 to 6.26 months longer than Everlast.

Solution by TI 83/84 Calculator

TI-83/84 ***stat*** $\rightarrow$ ***test*** $\rightarrow$ ***Option 9***





$$E = (UB - LB)/2 = (6.26 - 2.54)/2 = 1.86$$

2) Use **90%** confidence level to estimate the difference in average life (in months) of "Diehard" and "Everlast" batteries

	Diehard μ_1	**Everlast** μ_2
n	$n_1 = 40$	$n_2 = 49$
$\bar{x}$	$\bar{x}_1 = 52$	$\bar{x}_2 = 50.5$
S	$s_1 = 5.5$	$s_2 = 4.5$

Answer: $-0.279 < \mu_1 - \mu_2 < 3.279$

Conclusion:

3) Use 99% confidence level to estimate the difference in weights sugar between Regular Coke and Regular Pepsi.

	Regular Coke μ_1	**Regular Pepsi** μ_2
n	$n_1 = 36$	$n_2 = 36$
$\bar{x}$	$\bar{x}_1 = 0.82410$	$\bar{x}_2 = 0.81682$
S	$s_1 = 0.007507$	$s_2 = 0.005701$

Answer: $0.00324 < \mu_1 - \mu_2 < 0.01133$

Conclusion:

Estimating Difference between Two Populations Means $(P_1 - P_2)$

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm E \qquad E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

(for z- value use table page 3)

- **a)** What is the point estimate for the **difference** between two population **proportions** $P_1 - P_2$?
- **b)** What is the error formula for the **difference** between two population **proportions**?
- **c)** How to use TI to estimate for the **difference** between two population **proportions**?
- **d)** What conclusion can we draw if the lower bound of the estimate happened to be zero or negative?
- **e)** What conclusion can we draw if the both bounds of the estimate happened to be negative?

3 different ways the **final** answer may look like.

1) Both end signs **positive** $+ < P_1 - P_2 < +$ Indicating **group 1** has higher percentage than **group 2**
2) **Different** end signs $- < P_1 - P_2 < +$ **Inconclusive** as which group has higher percentage.
3) Both end signs **negative** $- < P_1 - P_2 < -$ Indicating **group 2** has higher percentage than **group 1.**

YouTube TI Calculator: https://www.youtube.com/watch?v=-YO0_3VqZ1g **Difference of two Proportions**

1) Use **95%** confidence level to estimate the difference in percentage between Female and Male that are passing stat class.

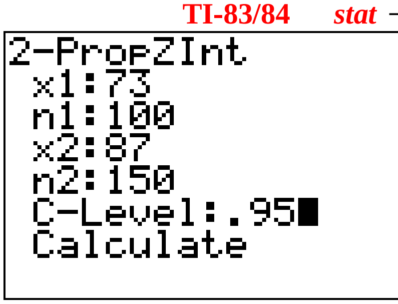
	Female	Male
Number of students passed	$x_1 =$ **73**	$x_2 =$ **87**
Number of students took stat	$n_1 =$ **100**	$n_2 =$ **150**
	$\hat{p}_1 = 73/100 = .73$	$\hat{p}_2 = 87/150 = .58$

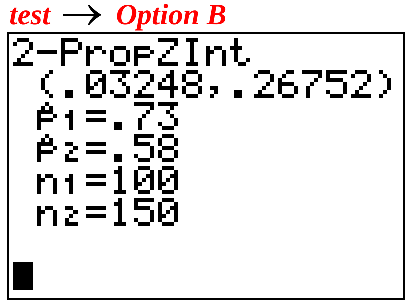
$$Point\ Estimate = \hat{p}_1 - \hat{p}_2 = 0.73 - 0.58 = 0.15 \qquad E = 1.96\sqrt{\frac{0.73(1-0.73)}{100} + \frac{0.58(1-0.58)}{150}} = 0.118 = 11.8\%$$

$P_1 - P_2 = 0.15 \pm 0.118 = 15\% \pm 11.8\%$ **Answer** $3.2\% < P_1 - P_2 < 26.8\%$

Conclusion: Both **signs are positive** indicating **group** 1 **(female)** has higher percentage in passing stat class than **group** 2**(male).**

TI-83/84 *stat* $\rightarrow$ *test* $\rightarrow$ *Option B*





$0.032 < P_1 - P_2 < 0.268 \qquad 3.2\% < P_1 - P_2 < 26.8\%$

Both **signs are positive** indicating **group** 1 **(female)** has higher percentage in passing stat class than **group** 2**(male).** $E = (UB - LB)/2 = (0.268 - 0.032)/2 = 0.118$

2) Use **90%** confidence level to estimate the difference in percentage between Female and Male that are passing DMV test.

	Female	**Male**
Number of students passed	$x_1 =$ **30**	$x_2 =$ **65**
Number of students took driving test	$n_1 =$ **78**	$n_2 =$ **80**
	$\hat{p}_1 =$	$\hat{p}_2 =$

Answer $-54.4\% < P_1 - P_2 < -031.2\%$

Conclusion:

Comparing Success rate in passing DMV driving test between Female and Male Applicants

TI-83/84 ***stat*** $\rightarrow$ ***test*** $\rightarrow$ ***Option B***

```
2-PropZInt
 x1:30
 n1:78
 x2:65
 n2:80
 C-Level:.9
 Calculate
```

```
2-PropZInt
 (-.5435,-.3123)
 p̂1=.3846153846
 p̂2=.8125
 n1=78
 n2=80
```

$-0.544 < P_1 - P_2 < -0.312$ $\quad$ $3.123\% < P_2 - P_1 < 54.35\%$

Both signs are negative indicating **group 2 (male)** has higher percentage in passing DMV test than **group 1(female).**

$$E = (UB - LB) / 2 = (-0.3123 - (-0.5435)) / 2 = 0.115$$

$$E = (UB - LB) / 2 = (0.5435 - 0.3123) / 2 = 0.115$$

3) Use **90%** confidence level to estimate the difference in percentage between Female and Male that are passing DMV test.

	Female	**Male**
Number of students passed	$x_1 =$ **30**	$x_2 =$ **65**
Number of students took driving test	$n_1 =$ **78**	$n_2 =$ **80**
	$\hat{p}_1 =$	$\hat{p}_2 =$

Answer $-54.4\% < P_1 - P_2 < -031.2\%$

Conclusion:

Sample Size determination needed to Estimate population Mean (μ)

In practice, before estimating population mean, we need a sample then the question is how large that sample must be? That depends on 3 factors, confidence level (z-value), margin of error (E), and prior information on standard deviation(s).

Use $n = \left(\dfrac{z_{\alpha/2} s}{E} \right)^2$ when **s** is known from prior information **(for z- value use table page 3)**

Example: How large should the sample size be if a researcher wants to estimate average life of Diehard batteries? For each case use the given confidence level, standard deviation and margin of error.

Write your **observation** for change in error and confidence level and their effects on sample size.

Case 1	Case 2	Case 3	Case 4
95% confidence s = 7.2 months E = 2 months $n = \left(\dfrac{1.96 \bullet 7.2}{2} \right)^2$ $= 49.79 = 50$ 50 batteries need to be sampled.	**95% confidence** s = 7.2 months E = 1 months $n = \left(\dfrac{1.96 \bullet 7.2}{1} \right)^2$ $= 199.15 = 200$ 200 batteries need to be sampled **Observation**: By reducing the error in half the sample size was multiplied by four.	**99% confidence** s = 7.2 months E =1 months $n = \left(\dfrac{2.5758 \bullet 7.2}{1} \right)^2$ $= 343.94 = 344$ 344 batteries need to be sampled. **Observation**: By increasing the confidence level the sample size got larger.	**90% confidence** s = 7.2 months E = 1 months $n = \left(\dfrac{1.645 \bullet 7.2}{1} \right)^2$ $= 140.3 = 141$ 141 batteries need to be sampled. **Observation**: By reducing the confidence level the sample size got smaller.

Practice: How many locations at Lake Tahoe must be tested for the purpose of estimating the average clarity of water? For each case use the given confidence level, standard deviation and margin of error.

Write your **observation** for change in error and confidence level and their effects on sample size.

Case 1	Case 2	Case 3	Case 4
95% confidence s = 2.2 ft. E = 1 ft. **Ans:** $n = 19$	**95% confidence** s = 2.2 ft. E = .5 ft. **Ans:** $n = 75$ **Observation:**	**99% confidence** s = 2.2 ft. E = .5 ft. **Ans:** $n = 129$ **Observation:**	**90% confidence** s = 2.2 ft. E = .5 ft. **Ans:** $n = 53$ **Observation:**

Extra Practice: Please do **p**roblems **25,26,27,28** from **p**ractice **p**roblems of **part III**

Sample Size determination needed to estimate population proportion (P)

In practice, before estimating population proportion, we need a sample, the question is how large of a sample? That depends on 3 factors, confidence level(z-value), margin of error(E), and prior information on population proportion $\hat{p}$. Use $n=\left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p})$ **(for z- value use table page 3)**

Use $\hat{p}$ an estimate of *p* based on a pilot study or an earlier study. If $\hat{p}$ is ***unknown*** then use $\hat{p}=0.5$ which gives the largest possible value of *n* for a given level of confidence and a given a margin of error.

Example: How large should the sample size be if a researcher wants to estimate the percentage of people who will vote for the next democratic candidate? For each case use the given confidence level, standard deviation and margin of error.

Write your **observation** for change in error and confidence level and their effects on sample size.

Case 1	Case 2	Case 3	Case 4
95% confidence $\hat{p}=0.56$ E = .06 $n=\left(\frac{1.96}{.06}\right)^2 .56(1-.56)$ $=263$ 263 voters must be sampled.	**95% confidence** $\hat{p}=0.56$ E = .02 $n=\left(\frac{1.96}{.02}\right)^2 .56(1-.56)$ $=2366$ 2366 voters must be sampled. **Observation**: By reducing the error by one third the sample size was multiplied by nine.	**99% confidence** $\hat{p}=0.56$ E = .02 $n=\left(\frac{2.5758}{.02}\right)^2 .56(1-.56)$ $=4088$ 4088 voters must be sampled. **Observation**: By increasing the confidence level the sample size got larger.	**90% confidence** $\hat{p}=0.56$ E = .02 $n=\left(\frac{1.645}{.02}\right)^2 .56(1-.56)$ $=1667$ 1667 voters must be sampled. **Observation**: By reducing the confidence level the sample size got smaller.

Practice: How large should the sample size be if a researcher wants to estimate the percentage of people who have defective vision? For each case use the given confidence level, standard deviation and margin of error.

Write your **observation** for change in error and confidence level and their effects on sample size.

Case 1	Case 2	Case 3	Case 4
95% confidence $\hat{p}=.21$ E = .02 **Ans:** $n=1594$	**95% confidence** $\hat{p}=.21$ E = .04 **Ans:** $n=399$ **Observation:**	**99% confidence** $\hat{p}=.21$ E = .04 **Ans:** $n=688$ **Observation:**	**90% confidence** $\hat{p}=$ unknown E = .04 **Ans:** $n=423$ **Observation:**

Extra Practice: Please do problems 22,23,2