SECOND YEAR OF SURVEYING DEPARTMENT

THEORY OF ERROR 2017

LECTURE 4

CONFIDENCE INTERVALS

DR. ENG. HASSAN MOHAMED

LECTURE ELEMENTS

1. "INTRODUCTION"
2. "DISTRIBUTIONS USED IN SAMPLING THEORY"
3. "CONFIDENCE INTERVAL FOR THE MEAN: T STATISTIC"
4. "SELECTING A SAMPLE SIZE"
5. "CONFIDENCE INTERVAL FOR A POPULATION VARIANCE"

DR. ENG. HASSAN MOHAMED

INTRODUCTION

- SINCE THE MEAN OF A SAMPLE SET Y AND ITS VARIANCE S^2 ARE COMPUTED FROM RANDOM VARIABLES, THEY ARE ALSO RANDOM VARIABLES. THIS MEANS THAT EVEN IF THE SIZE OF THE SAMPLE IS KEPT CONSTANT, VARYING VALUES FOR THE MEAN AND VARIANCE CAN BE EXPECTED FROM THE SAMPLES, WITH GREATER CONFIDENCE GIVEN TO LARGER SAMPLES.

Population of 100 Values

18.2	26.4	20.1	29.9	29.8	26.6	26.2
25.7	25.2	26.3	26.7	30.6	22.6	22.3
30.0	26.5	28.1	25.6	20.3	35.5	22.9
30.7	32.2	22.2	29.2	26.1	26.8	25.3
24.3	24.4	29.0	25.0	29.9	25.2	20.8
29.0	21.9	25.4	27.3	23.4	38.2	22.6
28.0	24.0	19.4	27.0	32.0	27.3	15.3
26.5	31.5	28.0	22.4	23.4	21.2	27.7
27.1	27.0	25.2	24.0	24.5	23.8	28.2
26.8	27.7	39.8	19.8	29.3	28.5	24.7
22.0	18.4	26.4	24.2	29.9	21.8	36.0
21.3	28.8	22.8	28.5	30.9	19.1	28.1
30.3	26.5	26.9	26.6	28.2	24.2	25.5
30.2	18.9	28.9	27.6	19.6	27.9	24.9
21.3	26.7					

INTRODUCTION

- FLUCTUATIONS IN THE MEANS AND VARIANCES COMPUTED FROM *VARYING SAMPLE SETS* RAISES QUESTIONS ABOUT THE ABILITY OF THESE VALUES TO ESTIMATE THE *POPULATION VALUES RELIABLY*.
- IN STATISTICS, THIS RELATIONSHIP BETWEEN THE SAMPLE SETS, THE *NUMBER OF SAMPLES*, AND THE VALUES COMPUTED FOR THE *MEANS AND VARIANCES* IS PART OF *SAMPLING DISTRIBUTION THEORY*.

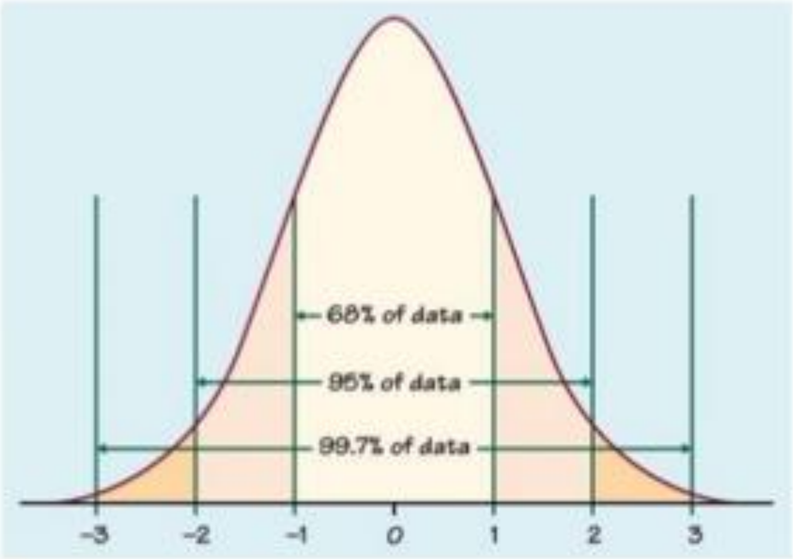
Increasing Sample Sizes

No.	$\bar{y}$	S^2
10	26.9	28.1
20	25.9	21.9
30	25.9	20.0
40	26.5	18.6
50	26.6	20.0
60	26.4	17.6
70	26.3	17.1
80	26.3	18.4
90	26.3	17.8
100	26.1	17.5

INTRODUCTION

- *BY APPLYING THESE DISTRIBUTIONS*, STATEMENTS CAN BE WRITTEN FOR THE RELIABILITY AT ANY *GIVEN LEVEL OF CONFIDENCE OF THE ESTIMATES* COMPUTED. IN OTHER WORDS, A RANGE CALLED *THE CONFIDENCE INTERVAL* CAN BE DETERMINED WITHIN WHICH THE POPULATION MEAN AND POPULATION VARIANCE CAN BE EXPECTED TO FALL FOR VARYING LEVELS OF PROBABILITY.

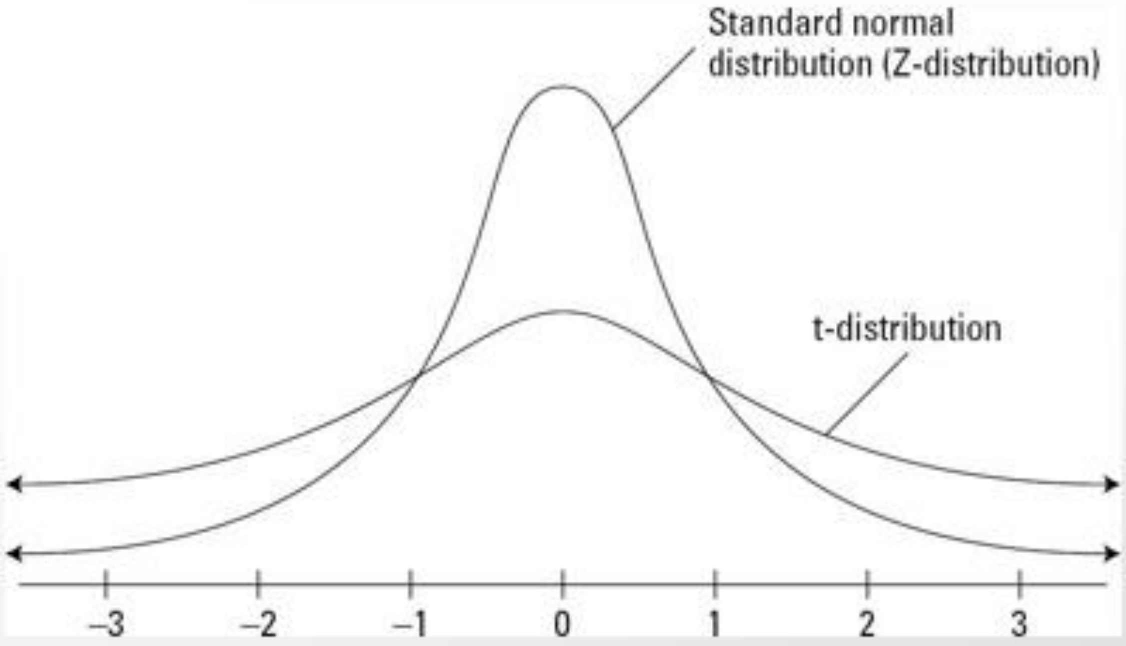
NORMAL DISTRIBUTION CURVE



DR. ENG. HASSAN MOHAMED

DISTRIBUTIONS USED IN SAMPLING THEORY

THE T DISTRIBUTION IS USED TO COMPARE A POPULATION **MEAN** WITH THE **MEAN OF A SAMPLE SET** BASED ON THE NUMBER OF **REDUNDANCIES** (U) IN THE SAMPLE SET. IT IS SIMILAR TO THE NORMAL DISTRIBUTION (DISCUSSED IN CHAPTER 3) EXCEPT THAT THE NORMAL DISTRIBUTION APPLIES TO AN ENTIRE POPULATION, WHEREAS THE T DISTRIBUTION APPLIES TO **A SAMPLING OF THE POPULATION**. ***THE T DISTRIBUTION*** IS **PREFERRED OVER** THE NORMAL DISTRIBUTION WHEN THE SAMPLE **CONTAINS FEWER THAN 30 VALUES**. THUS, IT IS AN IMPORTANT DISTRIBUTION IN ANALYZING ***SURVEYING DATA***.



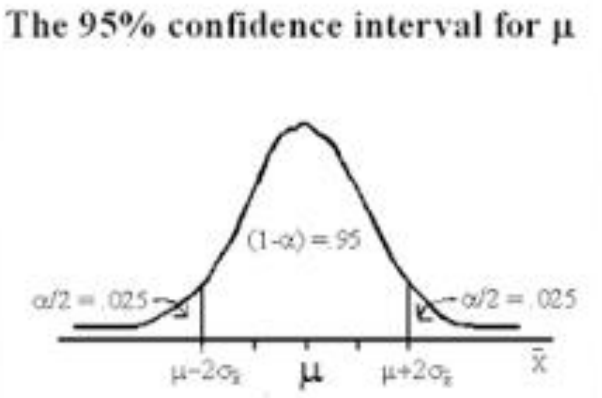
DR. ENG. HASSAN MOHAMED

CONFIDENCE INTERVAL FOR THE MEAN: T STATISTIC

- THUS, GIVEN $\bar{y}$, $t_{\alpha/2,\nu}$, n, AND S, IT IS SEEN FROM EQUATION (4.6) THAT A $1-\alpha$ PROBABLE ERROR INTERVAL FOR THE POPULATION MEAN μ IS COMPUTED AS

$$\bar{y} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{y} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

- WHERE $t_{\alpha/2}$ IS THE T VALUE FROM THE T DISTRIBUTION BASED ON υ DEGREES OF FREEDOM AND $\alpha/2$ PERCENTAGE POINTS.



DR. ENG. HASSAN MOHAMED

CONFIDENCE INTERVAL FOR THE MEAN: T STATISTIC

IN CARRYING OUT A CONTROL SURVEY, 16 DIRECTIONAL READINGS WERE MEASURED FOR A SINGLE LINE. THE MEAN (SECONDS' PORTION ONLY) OF THE READINGS WAS 25.4", WITH A STANDARD DEVIATION OF ±1.3". DETERMINE THE 95% CONFIDENCE INTERVAL FOR THE POPULATION MEAN. COMPARE THIS WITH THE INTERVAL DETERMINED BY USING A *T* VALUE DETERMINED FROM THE STANDARD NORMAL DISTRIBUTION TABLES (TABLE 3.2).

$$24.7 = 25.4 - 2.131\left(\frac{1.3}{\sqrt{16}}\right) = \bar{y} - t_{0.025}\frac{S}{\sqrt{n}}$$

$$< \mu < \bar{y} + t_{0.025}\frac{S}{\sqrt{n}} = 25.4 + 2.131\left(\frac{1.3}{\sqrt{16}}\right) = 26.1$$

$$\bar{y} \pm t_{0.025}\frac{S}{\sqrt{n}} \quad \text{or} \quad 25.4 \pm 2.131\left(\frac{1.3}{\sqrt{16}}\right) = 25.4 \pm 0.7$$

SELECTING A SAMPLE SIZE

- *A COMMON PROBLEM ENCOUNTERED IN SURVEYING PRACTICE IS TO DETERMINE THE NUMBER OF REPEATED OBSERVATIONS NECESSARY TO MEET A SPECIFIC PRECISION. IN PRACTICE, THE SIZE OF S CANNOT BE CONTROLLED ABSOLUTELY. RATHER, AS SEEN IN EQUATION (4.7), THE CONFIDENCE INTERVAL CAN BE CONTROLLED ONLY BY VARYING THE NUMBER OF REPEATED OBSERVATIONS.*

$$\bar{y} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

- FROM EQUATION (4.7), THE RANGE IN WHICH THE POPULATION MEAN (μ) RESIDES AT A SELECTED LEVEL OF CONFIDENCE (α) IS

$$I = t_{\alpha/2} \frac{S}{\sqrt{n}}$$

- REARRANGING EQUATION (4.9) YIELDS

$$n = \left(\frac{t_{\alpha/2} S}{I}\right)^2$$

DR. ENG. HASSAN MOHAMED

SELECTING A SAMPLE SIZE

- **EXAMPLE 4.2** FROM THE PRE ANALYSIS OF A HORIZONTAL CONTROL NETWORK, IT IS KNOWN THAT ALL ANGLES MUST BE MEASURED TO WITHIN ±2" AT THE 95% CONFIDENCE LEVEL. HOW MANY REPETITIONS WILL BE NEEDED IF THE STANDARD DEVIATION FOR A SINGLE ANGLE MEASUREMENT HAS BEEN DETERMINED TO BE ±2.6"?

- SOLUTION IN THIS PROBLEM, A FINAL 95% CONFIDENCE INTERVAL OF ±2" IS DESIRED.

$$n = \left(\frac{1.960 \times 2.6}{2}\right)^2 = 6.49$$

- THUS, EIGHT REPETITIONS ARE SELECTED, SINCE THIS IS THE CLOSEST EVEN NUMBER ABOVE 6.49.

CONFIDENCE INTERVAL FOR A POPULATION VARIANCE

THUS, THE $1 - \alpha$ CONFIDENCE INTERVAL FOR THE POPULATION VARIANCE ($\boldsymbol{\sigma}^2$) IS

$$\frac{\nu S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{\nu S^2}{\chi^2_{1-\alpha/2}}$$

EXAMPLE 4.3 AN OBSERVER'S POINTING AND READING ERROR WITH A 1" THEODOLITE IS ESTIMATED BY COLLECTING 20 READINGS WHILE POINTING AT A WELL-DEFINED DISTANT TARGET. THE SAMPLE STANDARD DEVIATION IS DETERMINED TO BE ± 1.8". WHAT IS THE 95% CONFIDENCE INTERVAL FOR $\boldsymbol{\sigma}^2$?

DR. ENG. HASSAN MOHAMED

CONFIDENCE INTERVAL FOR A POPULATION VARIANCE

SOLUTION FOR THIS EXAMPLE THE DESIRED AREA ENCLOSED BY THE CONFIDENCE INTERVAL $1 - \alpha$ IS 0.95. THUS, α IS 0.05 AND α /2 IS 0.025. THE VALUES OF X^2 0.025 AND X^2 0.975 WITH υ EQUAL TO 19 DEGREES OF FREEDOM ARE NEEDED. THEY ARE FOUND IN THE X^2 TABLE (TABLE D.2) AS FOLLOWS:

$$\frac{(20 - 1)1.8^2}{32.85} < \sigma^2 < \frac{(20 - 1)1.8^2}{8.91}$$

$$1.87 < \sigma^2 < 6.91$$

THUS, 95% OF THE TIME, THE POPULATION'S VARIANCE SHOULD LIE BETWEEN 1.87 AND 6.91.

DR. ENG. HASSAN MOHAMED

THANKS

PLEASE DON'T USE THIS PRESENTATION WITHOUT GETTING A PERMEATION FROM ITS ORIGINAL OWNER

DR. ENG. HASSAN MOHAMED