Course	:	Diploma in Electronic Systems Diploma in Telematics & Media Technology Diploma in Aerospace Systems & Management Diploma in Electrical Engineering with Eco-Design Diploma in Mechatronics Engineering Diploma in Digital & Precision Engineering Diploma in Aeronautical & Aerospace Technology Diploma in Biomedical Engineering Diploma in Nanotechnology & Materials Science Diploma in Engineering with Business Diploma in Information Technology Diploma in Financial Informatics Diploma in Cybersecurity & Forensics Diploma in Infocomm & Security Diploma in Chemical & Pharmaceutical Technology Diploma in Biologics & Process Technology Diploma in Chemical & Green Technology

Module	:	Engineering Mathematics 2B / Mathematics 2B/ Computing Mathematics 2	–	EG1761/2008/2681/2916/2961 EGB/D/F/H/J/M207 IT1201/1531/1631/1761 CLB/C/G201

Topic 9 : Estimation of Parameters

Objectives :

At the end of this lesson, the student should be able to:

1. calculate the point estimators of population parameters
2. construct and interpret confidence intervals for the population mean using the appropriate distributions (standard normal or *t*-distribution)
3. explain how the confidence interval is related to sample size and confidence level

Topic 9: Estimation of Parameters

9.1 Estimation of the Population Mean

- It is common that we **do not know the population mean** for a random variable we are interested in. Using the example in Chapter 8, it would be impossible for us to determine the population mean of all the peanuts in this world.

- Hence a common approach is to take a **sample** and use the information from it to **estimate** the population mean.

Definition:
*A **point estimate** is a single value estimate for a population parameter.*
The most unbiased point estimate of the population mean μ is the sample mean $\bar{x}$

Example 9.1-1

The following set of data is the heights (in cm) of 16 children:

101	118	125	116	113	102	117	126
106	114	109	121	107	119	116	115

Find a point estimate of the mean height, μ of ***all*** the children (the population).

Solution:

9.2 Confidence Interval for the Mean (Large Samples)

- In example 9.1-1, the probability that the population mean height of the children is exactly 114.0625 cm is virtually nil. So instead of using a point estimate to estimate µ to be exactly 114.0625 cm, we can estimate that µ lies in an interval.

Definition:
*An **interval estimate** is an interval, or a range of values, used to estimate a population parameter.*

Although the point estimate in example 9.1-1 is not equal to the actual population mean, it is probably close to it. To form an interval estimate, use the point estimate as the centre of the interval, then add and subtract a margin of error.

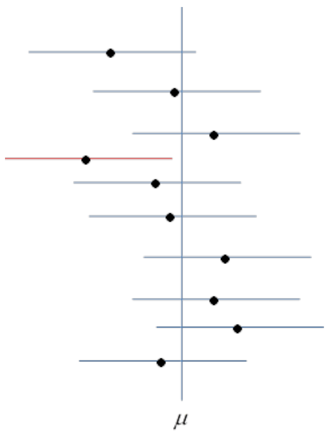
If the margin of error is 3.95, then the interval estimate in example 9.1-1 will be computed as 114.0625 ± 3.95 or $110.1 < \mu < 118.0$.

- Before finding a margin of error for an interval estimate, we must first determine how confident we need to be that our interval estimate contains the population mean µ.

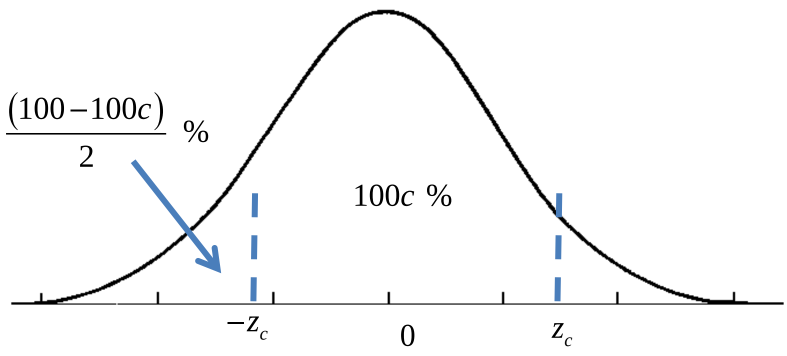
Definition:
*A confidence level **100c** % refers to the **percentage of the intervals from all possible samples** that we can expect to contain the true population mean.*

For example,

The diagram shows that there are 10 intervals obtained from 10 samples. If it is a 90% confidence interval, then it is expected that 9 out of 10 intervals contain the population mean.



- When the sample size is large, i.e. $n \geq 30$, by Central Limit Theorem, the sampling distribution of sample means is a normal distribution. The level of confidence $100c$ % is the area under the standard normal curve between the critical values,. $-z_c$ and z_c.



Example 9.2-1

Find the critical values z_c necessary to form a confidence interval at the following given level of confidence: (i) 80%, (ii) 85%, (iii) 97%.

Solution:

- Given a level of confidence $100c$ %, the **margin of error**, ***E*** is the greatest possible distance between the point estimate and the value of the parameter it is estimating. ***E*** is also known as the maximum error of estimate or error tolerance.
- The margin of error, E can be calculated as follows:
-

$$\boxed{E = z_c\left(\frac{\sigma}{\sqrt{n}}\right) \text{ or } E = z_c\left(\frac{s}{\sqrt{n}}\right)}$$

If the population standard deviation, σ is **known** or when $n \geq 30$, the sample standard deviation, *s* is used in place of σ.

Example 9.2-2

Find the margin of error for the given values of *c, s* and *n*.

(i) $c = 0.90,\ s = 2.5,\ n = 36$;

(ii) $c = 0.95,\ s = 3.0,\ n = 60$;

(iii) $c = 0.975,\ s = 4.6,\ n = 100$

Solution:

Note: In general, the margin of error **<u>decreases</u>** as the sample size **<u>increases</u>**.

- Using a point estimate and a margin of error, an interval estimate for a population parameter such as µ can be constructed. This interval estimate is called a **<u>confidence interval</u>**.
- Hence, a $100c$ % confidence interval for the population mean μ is given as:

$$\boxed{x - E < \mu < x + E \quad \text{or} \quad \left(x - E, x + E\right)}$$

where the probability that the confidence interval contains μ is $100c$ %.

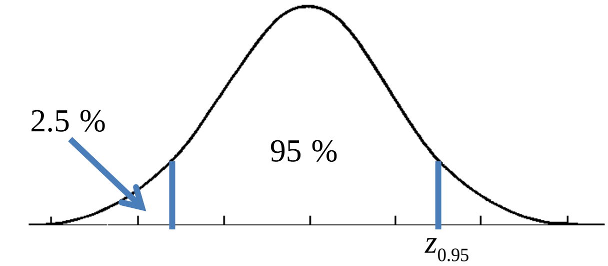
- Steps for constructing a confidence interval for a population mean ($n \geq 30$ or σ is known with a normally distributed population) are:

1. Find the sample statistics *n* and $\bar{x} = \frac{1}{n}\sum x$.
2. Specify σ if known. Otherwise, if $n \geq 30$, find the sample standard deviation $s = \sqrt{\frac{1}{n-1}\sum\left(x - \bar{x}\right)^2}$ and use it as an estimate for σ.
3. Find the critical value z_c that corresponds to the given level of confidence.
4. Find the margin of error, $E = z_c\left(\frac{\sigma}{\sqrt{n}}\right)$.
5. Form the confidence interval; $\left(\bar{x} - E, \bar{x} + E\right)$.

Example 9.2.3

After a few rainy days, numerous tadpoles appeared on a wet field. 12 tadpoles were randomly picked and their lengths measured. It is found that the sample mean is 11.1 mm. If this sample came from a normally distributed population with variance 4, calculate a 95% confidence interval for the mean length of all the tadpoles in the field.

Solution:



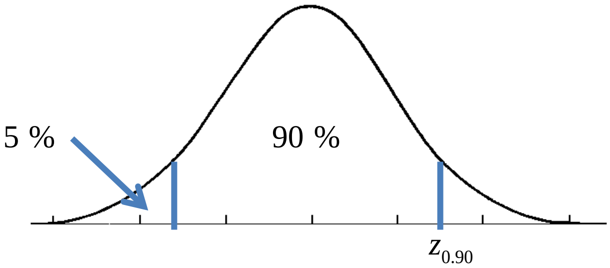
$$P\left(Z \le z_{0.95}\right) = 0.975 \Rightarrow z_{0.95} =$$

$$95\% \text{ confidence interval} = \left(\overline{x} - z_{0.95}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{0.95}\frac{\sigma}{\sqrt{n}}\right) =$$

Example 9.2-4

Fifty 2-year-old cows were injected with an antibiotic A, at a dosage of 12 mg/kg body weight. It is found that the sample mean of the blood serum concentrations ($\mu g / ml$) of the antibiotic 2 hrs after injection is 25.5 and the sample standard deviation is 3.03. Construct a 90% confidence interval for the population mean.

Solution:



$$P\left(Z \le z_{0.90}\right) = 0.95 \Rightarrow z_{0.90} =$$

$$90\% \text{ confidence interval} = \left(\overline{x} - z_{0.90}\frac{s}{\sqrt{n}},\ \overline{x} + z_{0.90}\frac{s}{\sqrt{n}}\right) =$$

Example 9.2-5

A random sample of 150 readings was taken from a population with mean μ and variance σ^2. Given that $\Sigma x = 1623$ and $\Sigma x^2 = 17814.36$,

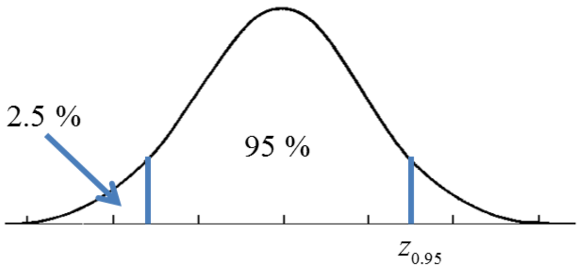
(a) calculate $\bar{x}$ and s.

(b) construct a 95 % confidence interval for the population mean.

Solution:

a) Recall $\bar{x} = \dfrac{\sum x}{n} =$, $s^2 = \dfrac{1}{n-1}\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}\right] =$

b)



$P\left(Z \le z_{0.95}\right) = 0.975 \Rightarrow z_{0.95} =$

95% confidence interval = $\left(\bar{x} - z_{0.95}\dfrac{s}{\sqrt{n}}, \bar{x} + z_{0.95}\dfrac{s}{\sqrt{n}}\right) =$

9.3 Confidence Interval for the Mean (Small Samples)

- In many real-life situations, the population standard deviation is ***unknown***. Moreover, due to constraints such as cost and time, it is often not practical to collect samples of size 30 or more. If the random variable is normally or approximately normally distributed, we can use a *t*-distribution.

- When X is a normal random variable, with the population standard deviation, σ **unknown**, the random variable T

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

follows a ***t* – distribution** with degrees of freedom, d.f. = $n - 1$.

- The value of t_c can be obtained from the t – distribution table.
- For example, if $n = 7$ and we want to construct a 95% confidence interval, we can obtain the value of $t_{0.95}$ as follows:

  d.f. $= 7 - 1 = 6$

	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	1.638	2.353	3.182	4.541	5.841
4		.741	1.533	2.132	2.776	3.747	4.604
5		.727	1.476	2.015	2.571	3.365	4.032
6		.718	1.440	1.943	2.447	3.143	3.707

$\therefore t_{0.95} = 2.447$

Example 9.3-1

Twelve packets of a particular brand of sweets are selected at random and their weights noted. The weights obtained (in grams) are

407.3, 409.6, 391.0, 402.9, 406.8, 390.0, 407.6, 402.1, 390.8, 390.6, 396.8, 400.2.

Assuming that the sample is taken from an approximately normal population with mean mass μ, calculate

(a) the 95% confidence interval for μ,

(b) the 99% confidence interval for μ.

Solution:

Using calculator, $\bar{x} =$, $s =$, df $= n - 1 =$

a) Since sample size, $n < 30$ and population variance unknown, $t_{0.95} =$

95% confidence interval is $\left(\bar{x} - t_{0.95}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.95}\frac{s}{\sqrt{n}}\right) =$

b) Since sample size, $n < 30$ and population variance unknown, $t_{0.99}$ =

99% confidence interval is $\left(\overline{x} - t_{0.99} \frac{s}{\sqrt{n}}, \overline{x} + t_{0.99} \frac{s}{\sqrt{n}} \right) =$

9.4 Minimum Sample Size to Estimate Population Mean μ

- Sometimes we will need to determine the sample size required before we conduct an experiment. Given a pre-determined confidence level $100c$ % and margin of error,

$$\boxed{n = \left(\frac{z_c \sigma}{E} \right)^2 \quad \text{or} \quad n = \left(\frac{z_c s}{E} \right)^2 \quad \text{or} \quad n = \left(\frac{t_c s}{E} \right)^2}$$

Example 9.4-1

You want to estimate the mean number of sentences in a magazine advertisement. How many magazine advertisements must be included in the sample if you want to be 95% confident that the sample mean is within one sentence of the population mean? Assume that the population standard deviation is 5.0 and the number of sentences is normally distributed.

Solution:

Given $E = 1$, $\sigma = 5.0$, & $z_{0.95} =$

Hence the number of advertisements required in the sample is at least :

$$n = \left(\frac{z_c \sigma}{E} \right)^2 =$$

In summary,

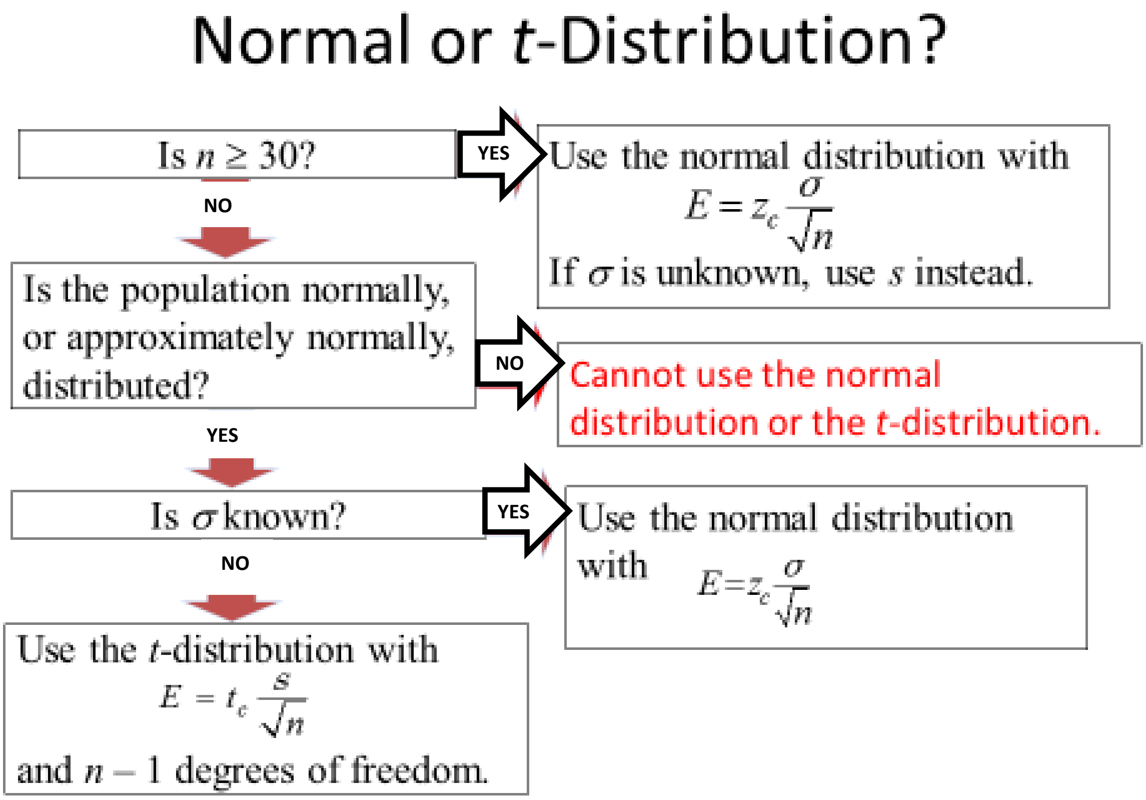
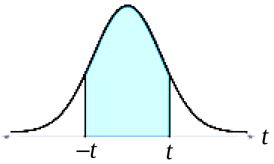
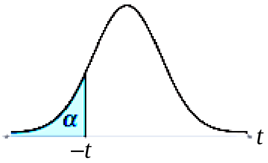

Normal or t-Distribution?
Is $n \geq 30$?
YES
Use the normal distribution with $E = z_c \frac{\sigma}{\sqrt{n}}$
If σ is unknown, use s instead.
NO
Is the population normally, or approximately normally, distributed?
NO
Cannot use the normal distribution or the t-distribution.
YES
Is σ known?
YES
Use the normal distribution with $E = z_c \frac{\sigma}{\sqrt{n}}$
NO
Use the t-distribution with $E = t_c \frac{s}{\sqrt{n}}$ and $n - 1$ degrees of freedom.

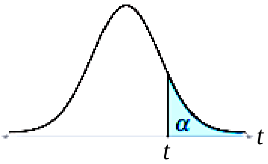
Table 2: *t* – Distribution



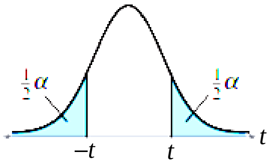
c-confidence interval



Left-tailed test



Right-tailed test

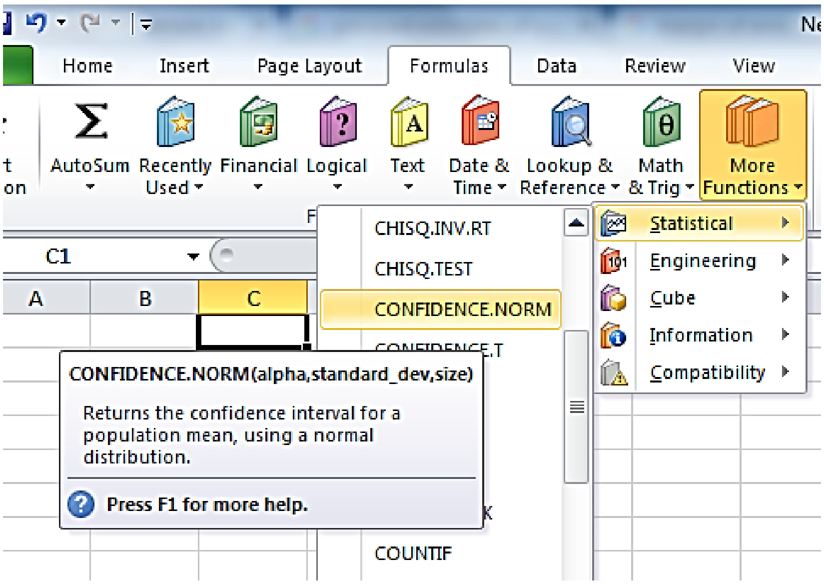


Two-tailed test

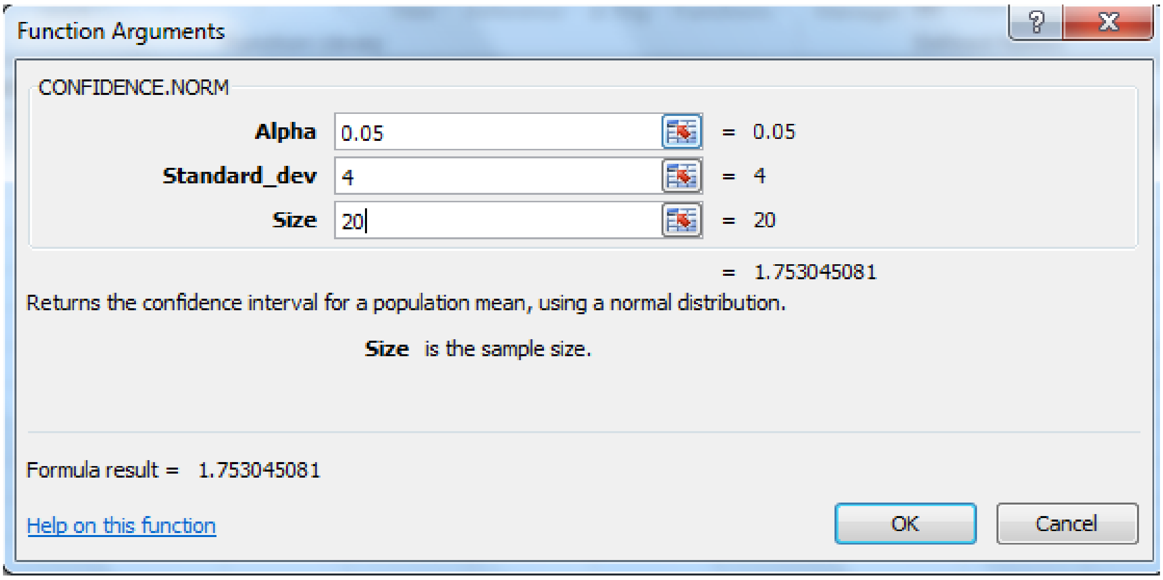
	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		0.816	1.886	2.920	4.303	6.965	9.925
3		0.765	1.638	2.353	3.182	4.541	5.841
4		0.741	1.533	2.132	2.776	3.747	4.604
5		0.727	1.476	2.015	2.571	3.365	4.032
6		0.718	1.440	1.943	2.447	3.143	3.707
7		0.711	1.415	1.895	2.365	2.998	3.499
8		0.706	1.397	1.860	2.306	2.896	3.355
9		0.703	1.383	1.833	2.262	2.821	3.250
10		0.700	1.372	1.812	2.228	2.764	3.169
11		0.697	1.363	1.796	2.201	2.718	3.106
12		0.695	1.356	1.782	2.179	2.681	3.055
13		0.694	1.350	1.771	2.160	2.650	3.012
14		0.692	1.345	1.761	2.145	2.624	2.977
15		0.691	1.341	1.753	2.131	2.602	2.947
16		0.690	1.337	1.746	2.120	2.583	2.921
17		0.689	1.333	1.740	2.110	2.567	2.898
18		0.688	1.330	1.734	2.101	2.552	2.878
19		0.688	1.328	1.729	2.093	2.539	2.861
20		0.687	1.325	1.725	2.086	2.528	2.845
21		0.686	1.323	1.721	2.080	2.518	2.831
22		0.686	1.321	1.717	2.074	2.508	2.819
23		0.685	1.319	1.714	2.069	2.500	2.807
24		0.685	1.318	1.711	2.064	2.492	2.797
25		0.684	1.316	1.708	2.060	2.485	2.787
26		0.684	1.315	1.706	2.056	2.479	2.779
27		0.684	1.314	1.703	2.052	2.473	2.771
28		0.683	1.313	1.701	2.048	2.467	2.763
29		0.683	1.311	1.699	2.045	2.462	2.756
∞		0.674	1.282	1.645	1.960	2.326	2.576

Appendix A Confidence intervals using EXCEL

- In EXCEL, select the tab “Formulas” → “More Functions” → “Statistical”.
  You can calculate the margin of error using the functions:
  CONFIDENCE.NORM (z table) or CONFIDENCE.T (t table).



- Suppose we will construct a 95 % confidence interval from the z table with population deviation 4 and sample size 20:
  “Alpha” = 1 – 0.95 = 0.05



Hence the margin or error = 1.7530

Tutorial 9: Estimation of parameters

A Self Practice Questions

1 Determine the z_c value for the following:

(a) 90 % confidence interval for μ.

(b) 95 % confidence interval for μ.

(c) 98 % confidence interval for μ.

(d) 99 % confidence interval for μ.

2 Determine the t_c value for the following:

(a) $n=10$, 90 % confidence interval for μ.

(b) $n=22$, 95 % confidence interval for μ.

(c) $n=25$, 98 % confidence interval for μ.

(d) $n=18$, 99 % confidence interval for μ.

3 Given that $\bar{x}=10$, calculate the 98 % confidence interval for μ when

(a) X is a normal random variable, $n=20$, $\sigma=3$.

(b) X is **not** a normal random variable, $n=50$, $\sigma=3$.

(c) X is a normal random variable, $n=10$, σ is unknown and $s=2$.

B Discussion Questions

1 In a particular factory, the quantity of mineral water dispensed by automated machines into plastic bottles is approximately normally distributed with standard deviation of 24 millilitres. A random sample of 25 such bottles was found to have a mean quantity of 503 millilitres.

(a) Find the standard error of the mean.

(b) Find a 90 % confidence interval for the mean quantity of mineral water dispensed by the machines.

(c) Find a 98 % confidence interval for the mean quantity of mineral water dispensed by the machines.

2 The heights of a random sample of 40 NYP students yield a mean of 173.8 cm and a standard deviation of 6.8 cm. Assume population is normally distributed.

(a) Construct a 95 % confidence interval for mean height of all NYP students.

(b) With reference to the 95 % confidence interval, what is the maximum possible error of using the sample mean as an estimate of the population mean?

3 One of the objectives of a large medical study was to estimate the mean physician fee for cataract removal. For n randomly selected cases the mean fee was found to be \$1550 with a standard deviation of \$125.

(a) Find a 99 % confidence interval on μ, the mean fee for all physicians when $n = 35$.

(b) Find a 99 % confidence interval on μ, the mean fee for all physicians when $n = 25$ and the distribution of the fees is normally distributed.

4 A researcher selected a random sample of 8 chick embryos to study the development of thymus gland. He weighed the glands of these 8 chick embryos after 12 days of incubation. The thymus weights (in mg) were as follows:

28.4 20.8 27.6 33.0 40.8 36.5 29.1 31.8

(a) Using your calculator, find the sample mean and the sample standard deviation.

(b) Construct a 90% confidence interval for the population mean.

(c) State whether any assumption is required on the distribution of the embryos.

C Conceptual Questions

Determine whether the following statements are true or false. Explain your reasoning.

1 For a given standard error, lower confidence levels produce wider confidence intervals.

2 If you increase sample size, the width of the confidence interval will increase.

3 To reduce the width of a confidence interval by half, we have to increase the sample size by four times.

<u>Answers</u>

A1	a	1.645	b	1.96	c	2.33	d	2.575
A2	a	1.833	b	2.080	c	2.492	d	2.898
A3	a	(8.44, 11.6)	b	(9.01, 11.0)	c	(8.22, 11.8)		
B1	a	4.8	b	(495, 511)	c	(492, 514)		
B2	a	(172,176)	b	2.11				
B3	a	(1495.59,1604.41)	b	(14980.08,1619.93)				
B4	a	31, 6.06	b	(26.9,35.1)				
C1	F	**C2**	F	**C3**	T			