SOUTH PACIFIC BOARD

FOR

EDUCATIONAL ASSESSMENT

PSSC MATHEMATICS

Statistics Project

HANDBOOK

Effective from January 2008

CONTENTS OF THIS HANDBOOK

A. BACKGROUND

During the year you are required to complete a statistics project.

To complete this project, you will use the statistics knowledge you will be learning this year.

This unit of work is internally assessed, meaning that the project you do will be marked by your teacher. The mark you get for the project will contribute to your final mark in PSSC Mathematics.

The topic of Statistics will not be assessed in the end of year examination.

The project should take you about 10 to 12 hours to complete, half of which should be in school time. Your teacher will tell you when this will be, but as a rule it should be used for doing your calculations and writing up the final report.

The project will involve you:

- Identifying a population
- Asking a question or formulating a hypothesis about this population
- Taking a random sample of **numerical** data from the population
- Calculating the appropriate statistics of your sample
- Drawing conclusions and trying to answer your question or prove/disprove your hypothesis.

The project will consist of four chapters. Each chapter will be reviewed by your teacher as you proceed to help you to know whether or not you are on the right track. Each chapter must be completed by the deadline date selected by your teacher, and before the next chapter is started. At the end of this section of the handbook your will find the marking criteria which explains how your project will be marked. *Initiative and independent thought will be rewarded.*

Here are some of the hypotheses investigated in projects that students carried out:

- The mean number community announcements made each day in the local radio station over the last 4 weeks - is more than 19.
- The mean number of words on the pages of the National Geographic magazine is more than 100.
- The mean weight of fish caught in a week by fishermen is less than 225Kg.
- The mean number of guesses by students of the number of marbles in a jar is within 10% of the actual number?
- The mean circumference of the trunks of a block of 10 year old coconut trees is greater than 0.5m?

You can choose a project which may be useful to your family, or which maybe just for fun. It does not matter, as long as you choose a project where you can collect a variety of data (but not a data string such as 0, 1, 1, 0, 0, 1, 2, 0, 0, 1 etc as there is not much you can do with these) and then work out some sensible statistics. Your data may be continuous (eg. measuring the height of students) or discrete (counting marbles), it does not matter, but how you handle the statistics may differ slightly, depending on your choice of project.

The next section describes the process you will need to follow in carrying out your investigation.

B. MANAGING THE PROJECT

Chapter 1: The Topic

In this chapter you are required to:

- Select a topic and formulate a question or hypothesis relevant to the topic
- Submit your topic to the teacher for approval
- Prepare a title page
- Write a short preamble outlining the aims and objectives of your project
- Clearly state how the sampling procedure you intend use in your sample selection.

You are advised to think carefully before choosing your topic as ***you will not be allowed to change your topic*** once you have started.

The topic must allow you to collect ***numeric data*** so that you can calculate statistical parameters such as mean, standard deviation etc. Hence a survey on people's opinion, colour of cars etc. would **not** be suitable. You must also select a population of about not less than 200, and from this population select a sample of 30 items.

The report on your work for Chapter 1 should clearly show the following:

- the aim of your project
- the question/hypothesis under investigation
- your sampling method and process, including an example of your written questionnaire if you are going to use one.
- any other information not shown above but relevant to your investigation.

This Chapter will not be marked, but your teacher will be checking your work using this guide:

- Promptness, i.e. was it handed in on time?
- Content, i.e. does it contain everything that was required? Does it go beyond the minimum requirements i.e. give extra information?
- Presentation, i.e. does the way it is set out enhance the story?
- Sampling procedure, is it mathematically sound?

Your teacher will comment on your work. You may use this advice both to improve your subsequent work and the final write up of Chapter 1.

Due date: Will be announced by the teacher

Chapter 2: Sampling and Data Collection

You are required to:

- Select a random sample of 30 from your population.
- List all your sample data. Remember it is the data you are interested in. Background information is not relevant at this stage. Information such as names, addresses or telephone numbers of respondents of subjects are excluded from the sample data list.

Your teacher will be reviewing this chapter using the following criteria;

- Promptness
- Detail of Fieldwork Procedures used in sampling and data collection
- Actual practical effort, i.e. what did you actually do to get your sample?
- List of all sample data collected
- Presentation and setting out.

In your report you should mention any changes that you had to make to your sampling procedure as you outlined it in Chapter 1 and clearly explain the reason(s) for the change(s) . When you submit your final project you may re-submit Ch 1 and Ch 2 unaltered, or you may wish to take advantage of the advice your teacher gave you and rewrite these two chapters to improve them.

Due Date: Will be announced by the teacher.

Chapter 3: Numerical Processing of Data

In this chapter you are required to:

- Draw an appropriate graph to show the frequency distribution of your data
- Draw other graphs where appropriate (eg. Cumulative Frequency for continuous data)
- Use not more than two types of graphs to display your data.
- Calculate the median, mode (if appropriate), lower and upper quartiles and semi-inter quartile range, range, mean and standard deviation of your sample
- All working must be shown. (a tabular form is best).

Chapter 4: Findings and Conclusions

Your task in this chapter is to answer/verify the question/hypothesis you asked/stated in Chapter 1. Make sure you *interpret* the results of your calculations with reference to your original question. Following this, you need to establish the reliability of your statistical results calculated in Chapter 3 and the best way to do this is to calculate the 95% Confidence Interval of the mean. This procedure will be taught in class during the Statistics unit. When you have found this Confidence Interval, write a paragraph explaining what it means in your case.

Now go on to form some general conclusions, for example;

- Point out any strengths or weaknesses in the method you used
- In what ways could your procedures be improved?
- What degree of confidence do you have in the accuracy of your results? Why?
- Would another person have come up with the same conclusion?
- What advice would you give to a student doing a similar project next year?

- List at least one way in which your project could be extended.
- What part of the project did you find most interesting?

This list is not exhaustive. Your teacher will want you to summarise your project as you see it and to make appropriate comments about what you have done.

When marking your project, your teacher will be taking special interest in these things:

- Graphs, are they appropriate, clear and correct, do they add to your "story"
- Accuracy of calculations
- Calculation of Confidence Interval
- Explanation of Confidence Interval as it applies to your topic
- How your conclusions relate to the data and your findings.
- Explanations, arguments and justifications for your methodology, findings and conclusions.
- Recommendations for follow up, improvements and replication of your project.

Due dates for Chapters 3 and 4: Will be announced by the teacher

After the completion of the project:

The interview:

Your teacher will want to interview you for two or three minutes about your project. It is important that you are able to speak about your work, and be able to explain anything the teacher may wish to ask you about it.

What if you hand any part in late?

Should you be aware of circumstances which may prevent you from handing in any chapter of the project by the due date you must see your teacher if you wish to negotiate a time extension. While there may be a possibility for this for Chapters 1 and 2, it is most unlikely that it will be granted for Chapters 3 & 4. Only under very special circumstances will your teacher recommend a time extension. Experience has shown that too many students "leave doing their project to the last minute". Don't you be one of these students this year. There is plenty of time available for you to complete all aspects of the project on time. Use your time wisely and plan and organise your work efficiently.

What happens to your project after you have completed it?

After your project has been marked and returned to you for checking, it will be collected again by the teacher. All projects will be retained by the school as some will be selected as part of the sample for external moderation.

C. ACHIEVEMENT BASED ASSESSMENT CRITERIA FOR MARKING THE STATISTICS PROJECT

The project will be marked from three perspectives,

- *Information Processing:* how well you gather, manage and analyse the information
- *Communication:* how you explain the facts, the results, and the significance of your findings,
- *Numerical Skills:* how accurately you perform the calculations.

Information Processing (Total = 25 Marks)

Gathering and analysing information or data

Mark	Achievement Criteria	Assessment Schedule
5	Gathers some relevant information.	Presents an incomplete report which indicates the use of an appropriate method for gathering data.
10	Gathers some relevant information and assembles it systematically.	Presents an incomplete report which indicates the use of an appropriate method for gathering data and which displays the data in a systematic and appropriate manner.
15	Gathers relevant information, assembles it systematically and analyses it.	Presents a complete report which meets all the deadlines for the survey, indicates the use of an appropriate method for gathering and displaying data, and in which the appropriate calculations are performed.
20	Assembles and analyses relevant information and draws a valid conclusion.	Presents a complete report which meets all the deadlines for the survey, and from the results of the data analysis draws a valid conclusion in response to the question behind the survey.
25	Assembles and analyses relevant information, draws a valid conclusion or conclusions and evaluates findings.	Presents a complete report which meets all the deadlines for the survey, draws a valid conclusion in response to the question, and includes a searching, thorough, and critical evaluation of the process and its relevance to the original question.

Communication (Total = 20 Marks)

Expressing mathematics in written or oral form, using symbols, graphs, diagrams etc.
Recognition will be given to presentation which enhances the visual impression of the report.

Mark	Achievement Criteria	Assessment Schedule
4	Attempts to communicate mathematical ideas.	Presents a report with some written commentary.
8	Communicates mathematical ideas.	Presents a report which explains the process used to conduct the survey.
12	Communicates linked mathematical ideas.	Presents a report which appropriately conveys reasoned views on the outcome of the survey relative to the question asked.
16	Communicates linked mathematical ideas logically and clearly.	Presents a completed report which relates each chapter to the question asked; which supports the conclusion by the use of appropriate graphs and/or diagrams, and which links the conclusion back to the question asked.
20	Communicates complete mathematical arguments logically and in an appropriate style.	Presents a report which leaves the reader in no doubt that the conclusion is the most appropriate one for the data collected. The argument is to be cogent, reasoned, supported by clear and appropriate graphs and/or diagrams, and the whole is presented as a logical package linked to the question asked.

Numerical Skills (Total = 15 Marks)

Applying the correct mathematical formulae as required, and calculating outcomes accurately.

Mark	Achievement Criteria	Assessment Schedule
3	Attempts to calculate statistics appropriately.	Some calculations are made, but not all the appropriate formulae are used.
6	Calculates some statistics appropriately.	Some calculations are made using most of the appropriate formulae, but not all calculations are performed correctly.
9	Calculate statistics appropriately and with a measure of success.	All formulae are applied appropriately, but not all are calculated correctly.
12	Correctly calculates all statistics.	All formulae are applied appropriately, all calculations are correct, but not all results are related back correctly to the question being researched.
15	Correctly calculates all statistics, and correctly interprets the results of their calculations in relation to the question being researched.	All formulae are used appropriately, all calculations are correct, and the interpretations of the results are correctly related back to the question being researched.

Compilation of marks:

Add together:

- the Information Processing mark out of 25,
- the Communication mark out of 20,
- the Numerical Skills mark out of 15,

The Total will be out of 60.

D. A METHOD OF RANDOMLY SAMPLING A POPULATION

1. BACKGROUND

People everywhere are concerned about the many issues which influence their way of life, and the values they consider important. These issues may be political, educational, economic etc, and because people often hold strong views on such issues they want to know if other people (or what proportion of all the people) agree with their views. To find these things out, radio, television, newspapers and private companies often commission surveys and publish their findings for everyone to see. It would be too expensive and too time consuming to ask all adults in the country for their opinion on a particular issue, therefore the company doing the survey will *sample the views of a fraction of the population* and publish the findings with its opinion on how accurately these findings represent the views of the whole country.

Public opinion polls are only one important use of sampling techniques. There are many others. Sampling is carried out in industry as a method of quality control. When goods are mass produced by machinery, continual monitoring is necessary to make sure that the product is the exact size, shape or weight it should be. In marketing, sampling is often used to see if a new product will sell, or in what form it should be produced. For health purposes, sampling is necessary. Purity of water supplies, quality of food production and pollution levels are continuously monitored by sampling methods. Body fluids and blood are often sampled to help doctors solve health problems in patients. Sampling is important here, as you would not want the doctor taking all your blood to see if there was a problem with it!

Unless sampling is carried out carefully, the information produced may not be accurate, and the statistics produced could lead to incorrect decisions.

Some reasons why we should *sample* and not *survey* a whole population are;

- ***practicality:-*** testing can often be destructive, eg testing the strength of wire. You would not test it all, otherwise you would not have any left to use!
- ***economic:-*** it costs a lot of money to survey a whole population, so why do it when it is not necessary.
- ***availability of the population:-*** if you are collecting information about the length of earthworms or fish, how would you know that you have caught and measured them all?
- ***time:-*** it would probably take too long to collect the data, analyse it, and publish the findings. Statistics are often required quickly so that decisions can be made.

TERMINOLOGY

- ***population:-***the complete set of statistics under consideration, eg. the weights of all the pupils at a secondary school. If we wanted the opinions of all the pupils on a particular topic, then the pupils themselves would be the population.
- ***sample:-*** a selection from, or sub-set of, the population.
- ***census:-*** a survey of the whole population.
- ***parameter:-*** one of the measures found by a census of the population, eg. mean, median, standard deviation.
- ***statistic:-*** one of the measures of a sample, eg. mean, median, standard deviation.

NOTATION

- Greek letters are used to denote the parameters of a population.
- Roman letters are used to denote the statistics of a sample.

Three of the most common measures used in statistics are mean, standard deviation and proportion. In Form 6 you will be particularly interested in the mean and standard deviation.

	Sample	Population
Mean	⊠	μ
Standard Deviation	s	σ

The population parameters are small Greek letters, while the letters ⊠ and s are used for the variables in sample statistics.

In your previous work on mean and standard deviation you used x for any measurement value. To calculate the standard deviation you used the formula

$$s = \sqrt{\frac{(x - \bar{x})^2}{n}}$$

The assumption in these calculations is that the values of measurements were derived from samples, not the population, so that μ and σ were not used.

2. STATISTICAL EXPERIMENTS

There are 6 steps that should be considered when designing and conducting a statistical experiment or survey.

Step 1*:-* Decide exactly what the experiment aims to do. An important part of the planning is to determine the population about which the question is asked - this is called the *target population.*

Step 2*:-* The question to be answered must be clear and unambiguous. Do not be too "short and to the point" if the meaning is going to be lost. Spend some time to make it clear what question you are trying to answer. When you have framed your question, predict an outcome.

eg. Target Population:	Senior students of the school
Question:	How much homework do senior student do a week?
Postulated outcome:	Is it more than 5 hours?

Step 3*:-* What information (data) is to be collected, and how. In your sample, you must aim for a random selection of the sample from the population, and avoid any bias. The whole purpose of the survey can be lost if bias is present because of the poor design of the question and the process. Bias will be explored shortly.

Step 4*:-* Select the sample and collect the data. Think carefully about both these aspects and consult your teacher at all times about the decisions you make.

Step 5*:-* Analyse the data and draw conclusions. Some estimation of the expected accuracy of the numerical results should be given. This will be linked to the size of the sample.

Step 6:- Finally the presentation of results and their interpretation is most important and takes a lot of thought and care.

TYPES OF SAMPLING

In all types of sampling, efforts must be made to overcome bias, and introduce randomness. If you are successful in these two, then you have a good chance of representing the parameters of the population by means of sampling.

In dealing with the human population, you are probably in the greatest danger of introducing some measure of bias into your survey.

Listed here are 5 ways of sampling a human population, and some of the problems associated with each.

1. ***Representative Sample***
   Try to make the sample reflect a true cross section of the target population. In your efforts to introduce people of each race, religion, social and economic background, gender etc. you may easily over-emphasise the importance of one group, and hence introduce bias.

2. ***Stratified Sample***
   The population is divided into groups with a common characteristic (such as gender or age groups). Separate random samples would be taken from each of the strata. This is a type of sampling often undertaken by professional survey organisations.

3. ***Cluster Sample***
   If you wanted the opinion of third formers throughout the country on a particular topic, you could save yourself a lot of time and effort by merely choosing your sample from one class in a particular school. You can probably see many problems in this. Did the class include boys and girls? Did the class include boys and girls from various social and economic backgrounds? Was the class streamed on ability? etc. Cluster sampling can be used with reasonable accuracy for some statistics, but generally, it is less accurate than simple random sampling.

4. ***Convenience Sample***
   The first persons in the population that you meet are sampled. This can allow forceful and pushy characters to make sure that they are chosen. Real chances for bias here.

5. ***Systematic Sample***
   You may randomly choose some page of the telephone directory, and then systematically interview every 20^{th} entry until 100 replies are obtained. Or you may interview a member from every 4^{th} house in the street. The first house chosen could be picked randomly. Of course, if you only chose one street in your survey, you can hardly draw any widely held conclusions.
   In a manufacturing operation, a systematic sample is often considered a random sample eg. you may choose every 100^{th} item off an assembly line to test for quality.
   In your practical work this year, you will not be attempting any of the sampling methods mentioned above. You will use only simple random sampling methods which will be explained in detail later.

In any sampling that you attempt, you must always be aware of falling into situations where bias might occur.

BIAS

Bias is the preference for one section of the population over another. When you set out to choose a sample, you must be aware that it is almost inevitable that the result will be influenced by conscious or sub-conscious bias.

As in representative sampling above, bias may easily be introduced into a survey while making a conscious effort to avoid it.

Examples of situations where bias occurs:

1. Interviewing people "at random" on the street by an interviewer. Such samples are too small to be meaningful. Interviewers using TV are likely to choose "pretty" people, and ignore old people or others who may appear inarticulate.
2. Household surveys where the opinion of the "head of the house" or "lady of the house" is sought. The target population could then be "heads of households" rather than say, the adult population in general.
3. Newspapers and magazines inviting reader response may only hear from people holding strong views on the subject, and not from the "silent majority". Furthermore the sample is restricted to only those who read those newspapers and magazines.
4. Telephone surveys can introduce bias. eg. During the day, working mothers are missed. At night, any concerned people may be missed while they attend meetings. Shift workers are also missed.

RANDOM SAMPLING

A timely, random method of selecting from a population is required if the sample is to be unbiased and is to reflect the composition of the population.

For small populations, names can be "drawn out of a hat", but for larger populations, you need to make use of a table of random numbers or the random number generator function on your scientific calculator.

Random *numbers* might more correctly be called random *digits*. A table of random numbers consists of large numbers of random digits. The appearance of a digit in any position in the table has the same likelihood as any other digit. The presence of a digit in one position in the table is independent of the digit in the previous position; eg. if there is an 0 in one position, there are 10 digits that can fill the next place, all with the same probability; ie. 1/10. There is therefore one chance in ten that another zero would follow the first one. There is no pattern to the string of digits. The next digit in the string will be unpredictable.

The random numbers are often in groups of "four" for easy reading. They should be viewed as a continuous string of random digits. Random numbers under 100 can be chosen from the table by starting at any point in the table and taking pairs of digits in turn. eg. 72,63 ,65, 07 (= to 7)

To use random numbers:

Number in order each member of the target population.

Select your random numbers from either a table or your calculator.

(i) Using a table:

Starting at a random point in the random number table, read off sets of digits of the appropriate length. (eg. if hundreds are in the target population, read the digits off in threes). Ignore repeats and numbers outside the population total (eg. 729 in a population total of 650 people).

The population members corresponding to the chosen random numbers will be your sample.

(ii) Using your calculator

On your scientific calculator, generate a long string of random digits then divide them into groups of two or three depending on what you need.

Example: To choose a random sample of 10 students from a class of 35 students.

We should actually use 'names in a hat' for such a small target population.
However, if we wish to use random number tables, the steps are shown below.

Step 1:- Use the class roll as the population. Number each member of the class 00 to 34 (or 01 to 35)

Aneau	00
Bulu	01
...*...	.*..
Taufu	24
....*...	.*..
Watson	33
Wong	34

Step 2:- Starting at any point in the random table read off pairs of digits. (Cross out repeats and numbers over 34)

SOME SAMPLING EXPERIMENTS

Statistics requires an orderly process and an accurate record keeping system. For your project to be successful you need to be careful at all times, take special care to keep neat and systematic records of data, perform accurate calculations, show good presentation of results, and draw conclusions that are meaningful and valid with respect to your research question or hypothesis.

EXERCISE 1

In these *census* surveys, make a list of the names of the class, and record results against each name in each experiment. The results will be used in further work later.

In the set of experiments in this exercise, you are not taking random samples from the class population. You will do that in a later exercise.

When finding the standard deviation of the population, use the "n" button on the calculator. This is equivalent to using "n" as the divisor in the standard deviation formula.

Later, when you use the standard deviation of a sample to estimate the population standard deviation, you will use the "n - 1" button in some cases.

1. Measure the height of all members of your class to the nearest cm. Record results against each name, and with the aid of your calculator, find the mean height and standard deviation of the class heights. (to 0.1 cm).
2. The purpose of this experiment is to measure hand spans. The population is the length of hand spans of all pupils in the class. Class members are to suggest ways of measuring hand spans to the nearest half-centimetre, and the class can decide by vote which method they think would be the most accurate. Record all results and find the mean and standard deviation of the distribution using your calculators. (to 0.1 cm).

3. Using a set of scales graduated in kilograms, weigh all members of the class to the nearest kilogram. Record all results, and use calculators to find the mean and standard deviation of the population of class weights (to 0.1 cm).

4. Your teacher will place a long piece of timber (between 1 and 2 m) on a desk at the front of the room. All students will be able to approach the desk to view the object, but must not physically touch it.
   Each student will write down his or her estimate of the length to the nearest cm, and the teacher will record the class results on the blackboard. All students will then copy the results into their own books and use calculators to find the mean and standard deviation of the population of estimated lengths.

5. In this experiment, you will estimate the length of a time interval. Your teacher will give a signal at the start of the interval, and the class will wait in silence for the end of the interval to be signalled. Each member of the class will write down their estimate of the length of the time interval in seconds.
   Record all results, find the mean and standard deviation (to 0.1 cm).

3. ESTIMATION OF THE POPULATION MEAN

In the last section you learnt that it is impractical to measure some parameters of a population. To overcome this difficulty, a sample is chosen from which to measure the corresponding population parameters. When used in this way, it is said that the statistic concerned is an estimator of the corresponding population parameter.

Two of the most important parameters of a normal distribution are the ***mean*** and ***standard deviation***.

From a simple random sample you can:
- use the value of the sample mean ⊠ to estimate the population mean value, μ.
- use the value of the sample standard deviation s to estimate the population standard deviation, σ.

In this 6th Form course, you will concentrate on the mean value, and try to find with what degree of precision the sample mean can be used as an estimate of the population mean.
ie. how close to μ can you predict your sample mean value, ⊠ will be.

In Exercise 1 you carried out a number of experiments using members of your Maths class. You will continue with those experiments here in Exercise 2.

EXERCISE 2

1. Use a random number process to select a sample of 2 students from your Maths class. From your previous results from Exercise 1, record the heights of the two class members selected and find their mean height. Compare this mean, ⊠, with the population mean, μ. Your teacher will record on the board the mean values of this sample which the class found.
2. Repeat 1 with a sample size of 5.
3. Repeat 1 with a sample size of 10.
4. As the sample size increases, what can you say about the mean of your sample, ⊠, as compared to the mean of the population, μ.

From these experiments you will have learnt an important property of sampling; ie. as the sample size increases, the mean of the sample becomes a more accurate estimator of the population mean. Of course you must balance any increase in accuracy with the time and cost of taking larger samples.

4. CONFIDENCE INTERVALS

It is time now to consider the heights of all Sixth Formers in the South Pacific. Because there are so many, their heights will be normally distributed.

From your studies of normal distribution, you should know about such a distribution of heights; namely;

* 68% of the population of heights lie within 1sd, either side of the mean.
* 95% of the population of heights lie within 2sd, either side of the mean.

This means that if you randomly choose one 6th Form student, then there is a probability of 0.95 that the height of the student will be within 2 standard deviations of the population mean.

ie. $\mu - 2\sigma < x < \mu + 2\sigma$.

The logic applies in reverse too, meaning that there is a 0.95 (95%) probability that the population mean will lie within 2 standard deviations of a randomly chosen height.

ie. $x - 2\sigma < \mu < x + 2\sigma$.

If, rather than choosing one height, we select a sample and calculate a mean ⊠ of the sample, then it follows that;

⊠ $- 2s < \mu <$ ⊠ $+ 2s$ with 95% probability.

However, the experiments you have done clearly show that the sample mean, ⊠, will be much closer to the population mean, μ, than this relationship suggests. More advanced Statistics than is required for this course can show that for a sample of size *n,* the closeness of ⊠ to μ depends on $\sqrt{n}$.

The true relationship is expressed as

⊠ $- \frac{2\sigma}{\sqrt{n}} < \mu <$ ⊠ $+ \frac{2\sigma}{\sqrt{n}}$ with 95% probability.

There is however a little problem with this relationship. You can calculate ⊠ from your sample, *n* (the number in your sample) is known, but s (the population standard deviation) is not known. However, if you have chosen an appropriately sized sample, then you can assume that the sample standard deviation s will be the same as the population standard deviation σ. The formula then becomes one you can use, and it is

⊠ $- \frac{2s}{\sqrt{n}} < \mu <$ ⊠ $+ \frac{2s}{\sqrt{n}}$ with 95% probability.

This range of values is often written as ⊠ $\pm \frac{2s}{\sqrt{n}}$.

The three ranges of confidence can be summarised as follows:

68% of the time μ lies in the range ⊠ $\pm \frac{s}{\sqrt{n}}$. We can say "likely"

95% of the time μ lies in the range ⊠ $\pm \frac{2s}{\sqrt{n}}$. We can say "very likely"

99% of the time μ lies in the range ⊠ $\pm \frac{3s}{\sqrt{n}}$. We can say "almost certain"

The intervals $\frac{s}{\sqrt{n}}$, $\frac{2s}{\sqrt{n}}$, $\frac{3s}{\sqrt{n}}$, are called **Confidence Intervals (CI).**

$\frac{s}{\sqrt{n}}$	is called a 1s or 68% confidence interval,
$\frac{2s}{\sqrt{n}}$	is called a 2s or 95% confidence interval,
$\frac{3s}{\sqrt{n}}$	is called a 3s or 99% confidence interval.

Therefore for a Normal population, a sample of size n, with a mean ⊠, and standard deviation s,

you are 68% confident that μ lies in the range	⊠ $\pm \frac{s}{\sqrt{n}}$
you are 95% confident that μ lies in the range	⊠ $\pm \frac{2s}{\sqrt{n}}$
you are 99% confident that μ lies in the range	⊠ $\pm \frac{3s}{\sqrt{n}}$

The most popular confidence interval for statisticians is the 95% one. Any particular measurement will lie within the 68% C.I. only 2 times out of 3, whereas the same measurement will lie in the 95% C.I. 19 times out of 20. The 99% C.I. gives wider bounds within which the measurement may lie, but it only improves the certainty by a margin of 4%.

EXAMPLE

A machine fills bags of sugar to a nominal amount of 1kg. Temperature, humidity, and other factors, causes variations in the filling process, causing random variations in the weights of the bags. The manufacturer wants to make sure that the bags contain, on the average, not less than 1kg of sugar. A worker takes random sample of 10 bags off the production line and finds that the net weights (in gm) are 975, 1015, 960, 990, 1005, 960, 980, 965, 1040, and 985.

Solution;

⊠ $= \frac{9875}{10} = 987.5gm$

s = 24.5gm, n = 10

$$\frac{2s}{\sqrt{n}} = \frac{2x24.5}{\sqrt{10}} = 16.0\text{gm}$$

It is very likely that the confidence interval 987.5 ± 16.0gm contains the population mean μ.

ie. between 971.5 and 1003.5gm. (ie. the 95% C.I.)

The estimate of the population mean, (⊠ = 987.5gm), suggests that μ is less than 1kg, meaning that the bags are generally underfilled and that the machine may have to be adjusted to add more sugar. However, before the operator does the adjustment, he should note that μ = 1kg is within the “very likely” bounds. Since the sample was a small one, and being a random sample, they could have struck a run of mainly underfilled bags. So before adjusting the machine, they could take another, possibly larger, sample just to confirm the first set of results.

Notes:

1. For large samples (>30) you can expect s to be a good estimator of σ . For smaller samples, it becomes less reliable. In fact, for small samples, the formula for s has an in-built bias, producing values of *s* generally smaller than the true population value, σ. You can compensate for this by using the divisor "n-1" in the formula (or on your calculator) when finding *s. This is only done when you plan to use the value of* s *to estimate* σ. In all other cases, use "n" in the divisor (or the n button on your calculator), no matter what the size of your sample or the population.
2. Though "n-1" may be used in the denominator to find the value of s for small samples, do not change "n" in the formula to establish the Confidence Intervals. For 95% confidence, μ lies in the range $\boxtimes \pm \frac{2s}{\sqrt{n}}$, and the value for "n" stays at "n", no matter what the size of the sample.

EXERCISE 3

1. To find the mean amount of money spent by Australian tourists on holiday in Vanuatu, a random sample of 50 holiday-makers was interviewed. The amounts spent, excluding all travel and lodging costs, had a mean of $780, and a standard deviation of $212. State the 95% C.I. of the mean amount spent by this year's holiday-makers in Vanuatu.
2. A New Zealand business manufactures car batteries. A sample of 16 batteries produced a mean voltage of 12 volt, with a standard deviation of 0.2volt. What is the 2s confidence interval for the mean of the population voltages?
3. A sample of 100 tyres randomly chosen from a production line shows, under continuous testing, a mean life of 20,000 km, with a standard deviation of 2,000 km. Find the 95% C.I. for the mean distance of travel for tyres from this manufacturer.
4. The STATE Insurance Company conducts a survey to estimate the average cost to insurance companies for panel beating and replacement parts for each car involved in an accident for which a claim was made. The mean of a sample of 40 cars chosen at random from all around Suva resulted in an average cost of $1266 with a standard deviation of $594. What is the 95% C.I. for the population mean?
5. A random sample of car owners in a taxi business in Suva showed a mean distance driven each year of 55,000 km with a standard deviation of 3,700 km. Find the 95% C.I. for the mean distance driven by all the taxi drivers in Suva.
6. A random sample of heights of 100 female students at a University provided a mean of 1.58 m with a standard deviation of 0.12 m. Find the 95% C.I. for the population mean of heights of female students at the University.
7. Students at a large secondary school sit an English exam. To estimate the mean of the population of scores, a random sample of 15 students produced scores of 58, 61, 48, 43, 62, 62, 51, 43, 54, 43, 70, 65, 42, 59, and 48 percent. Find the 95% C.I. for the mean, μ, of the English scores.