A Suitable Sample

Assessment Standard: 12.4.1(c)
A Suitable Sample

Demonstrate an ability to draw a suitable sample from a population and understand the importance of sample size in predicting the mean and standard deviation of a population.

When you were in Grade 8, 9 or 10 you most likely had to do a survey as part of your statistics learning. For that survey you needed to capture data, and then you most likely summarised the data using:

a) a tally table
b) a stem-and-leaf diagram
c) a five-number summary
d) a box-and-whisker diagram which needed the maximum and minimum values as well as the median and upper and lower quartiles.

Collect your Paper 3 Lessons every week!!

Guys, both NSC and IEB examinations candidates have the option of writing Paper 3 at the end of the year! Paper 3 covers additional mathematics material and is out of 100 marks. Maths Paper 3 will really set you apart in the job market, and make studying technical subjects at tertiary level easier. We have hooked you up with these lessons - written by IEB Maths Paper 3 examiner Heather Frankiskos. Though the lessons apply to both IEB and NSC candidates, where there are differences, we will point them out! The lesson this week applies to candidates from both examining bodies. **Give it a go!**

You most probably also learnt how to represent your data in a:

a) pie-chart
b) bar graph
c) broken-line graph
d) frequency polygon and
e) histogram

... and in Grade 11 you further learnt to represent data graphically in an ogive and a scatter plot.

You have also learnt how to summarise data by looking at their Measures of Central Tendency: mean, median, mode and Measures of Dispersion (spread): Range, quartiles, variance and standard deviation.

However, what we need to focus on in this module is the importance of the size of the sample you used in your data collection, as well as the randomness of your sample.

A Suitable Sample

Let us look at the following example:

Jason, in Grade 11 wants to work out the average BMI (Body Mass Index) of all the learners in his school.

The standard weight status categories associated with BMI ranges for adults (both male and female) are shown in the following table:

BMI	Weight Status
Below 18.5	Underweight
18.5 - 24.9	Normal
25.0 - 29.9	Overweight
30.0 and Above	Obese

To do this he will use the formula:

$$\text{BMI} = \frac{\text{weight (kg)}}{[\text{height (m)}]^2}$$

There are 853 learners in school.

- The population is therefore all 853 learners in the school and all subjects he would like to include in his survey. However, as he is going to have to gather this data, how many members of the population should he select to ensure that his population is properly represented? If his sample is too large he will waste time and resources. If it is too small it may lead to inaccurate results.
- A sample is a selection from the population which will give us measures which are similar to those taken from the whole population.

- If you study Advanced Programme Mathematics you learn formulae which work out the ideal sample size to use, so that the difference between your sample mean and population mean is minimized. We do not cover that here.
- All you need to be aware of is that sample size is important. As a rule of thumb use a 10% sample size as a minimum.

Jason asked 10 people in each of Grade 11P, 11Q, 11R, 11S and 11T to work out their BMI. The results are tabled:

grade	mean	BMI
11P	23,88	(Normal)
11Q	27,12	(Overweight)
11R	25,35	(Overweight)
11S	25,47	(Overweight)
11T	22,98	(Normal)

If he were using the results from 11Q only he would claim that the learners in the school were overweight. However, the mean BMI of all 5 classes is 24,96 which is at the top end of normal.

The other problem with Jason's sampling (can you see it?) is that the he has chosen all his data from Grade 11 learners. This is a biased sample.

To be unbiased the subjects in your sample must be random.

By definition, a sample of size n is random if the probability of selecting the sample is the same as the probability of selecting any other sample of size n.

So let us agree that Jason samples 10% of the learners. He must find a reliable way of randomly choosing 85 learners. He can do this in one of 3 ways:

1) Simple random sampling

He can get a list of all learners from the school computer assistant. He can assign each learner a unique number from 1 to 853. He can put these numbers in a hat and get someone to draw 85 pieces of paper or he can use his random number generator on his calculator or in excel. This way every learner has an equal chance of being selected and the sampling will be random.

2) Stratified random sampling

He can get the register for each grade (8-12) and then choose 17 learners randomly from each grade. This way at least allows for all grades to be represented. In this method data is first classified into at least 2 groups and then sampled.

3) Systematic random sampling

Jason could get an alphabetical list of all learners in the school and choose every 10th person on the list.

Note:

- If you select a reasonably large sample it will have measures representative of the population - it is not necessary to include everybody in a survey to make it valid.
- In general, as our sample size increases, the error between our mean and standard deviation worked out for our sample , compared to that of our actual whole population, will decrease. They become more stable and usually the distribution becomes more normal. (See next week's module).

So, how do we test this Assessment Standard?

Here goes...

Question 1

Zulu wants to find out what percentage of children in his school from Grade 1 to Matric are able to ride a bicycle.

1. Give two things that he must consider about his sample in order to get a representative sample.
2. Describe how he could go about getting a systematic random sample.

Solution

1. His sample must be large enough - the larger the sample the more likely its statistics will correspond to those of the entire population. His sample must be unbiased. He must survey children from all grades; and of both genders.

2. He could number all the learners in the school and then select every 8th / 9th or 10th person on the list.

Question 2

A South African television station carried out a survey and asked viewers to respond to the question: "Can you ride a bicycle?"

Respondents were asked to SMS their answer of 'Yes' or 'No'.

The results indicated that 52% of the respondents had answered 'Yes'.

Can you conclude that 52% of all South Africans can ride a bicycle? Discuss by making reference to the validity of the results of the survey.

Solution

No- this is a result from only those who have a TV; only those watching this particular programme; maybe it is aired at a time when only old people are watching; possibly this station is watched by one predominant ethnic group or gender. Also respondents needed a cell phone and had to be able to SMS. Many old people don't know how to SMS - while younger people may not have had the inclination. The results are not valid for all South Africans - only for the sample that responded.

Did you know?

Although the BMI number is calculated the same way for children and adults, the criteria used to interpret the meaning of the BMI number for children and teens are different from those used for adults.
For children and teens, BMI age and sex-specific percentiles are used for two reasons:

- The amount of body fat changes with age
- The amount of body fat differs between boys and girls

If you want to know more go to www.cdc.gov/healthyweight/assessing/bmi/adult_BMI/index. html