LEARNER GUIDE

Unit Standard Title:	**Apply knowledge of statistics and probability to critically interrogate and effectively communicate findings on life related problems**
Unit Standard No:	**9015**
Unit Standard Credits:	**6**
NQF Level:	**4**

This outcomes-based learning material was developed by Masifunde Training Centre with funding from INSETA in July 2014. The material is generic in nature and is intended to serve as a minimum standard for the industry.

Table of Contents

Module 1

Techniques for collecting, organising and representing data

This Module deals with:

- Situations or issues are dealt with through statistical methods and identified correctly
- Appropriate methods for collecting, recording and organisation data are used to maximise efficiency
- Selection of data sources and databases in a manner that ensures the representativeness of the sample
- Activities that could result in contamination of data
- Gathering data using appropriate methods to data type
- Using data collection methods correctly
- Correct calculation and use of statistics
- Graphical representations and numerical summaries are consistent with data are clear and appropriate to the situation
- Resolutions for a situation or issue are supported with data

Introduction

Numerical information is part of our everyday life. Whether you are reading newspapers, watching television or reading magazines you will find numbers and graphs being used to report information.

This information has been **collected** by researchers, scientists, governments, organisations such as the World Health Organisation etc. They **organise** this data and **represent** it in a way that is easy to read (usually in tables or graphs). It can then be **analysed** and **interpreted** to determine trends etc. We need to be aware that data can be manipulated to mislead people and you must therefore always critically analyse information that is presented to you as a statistic.

Simply put, we are going to look at problems in our lives and work that can be investigated by posing questions.

- What is the problem?
- What information do we need to investigate the problem and how do we get it?
- How should the information be recorded?
- How should the information be analysed?
- How should the findings be communicated?
- Are the statistics valid or do they contain biased or incorrect measurements?

1.1 Identifying situations for investigation

There are many issues in South Africa that have been and continue to be investigated through statistical methods. These surveys are funded by the government, by private companies and even by overseas organisations in order to establish trends or patterns and therefore make predictions about the future, and hopefully act on them to improve the life of South Africans. A wide range of issues are investigated relating to crime, health, living standards, literacy, weather patterns etc.

Let’s look at some types of questions that could lead to investigation.
Determining trends:

- Is crime in South Africa decreasing?
- Is the average life-expectancy of South Africans increasing or decreasing?
- Are South Africans smoking less now than they did five years ago?
- Are South African children becoming more obese?
- Are women more at risk for HIV/AIDS than men?
- Is child abuse on the increase in South Africa?
- Is the number of cars on South African roads increasing?

Considering attitudes of people:

- Are South Africans satisfied with their police force?
- Are taxis the preferred form of transport to South African commuters?
- Are South African youth satisfied with the education they are receiving?

Activity 1

Turn to your Learner Work File and complete Activity 1

LEARNER WORK FILE

Once you have identified an issue for investigation and posed your question, you then need to identify variables which will affect the problem.

For example, let’s use the question: “Is crime in South Africa decreasing?”
There are a number of variables which contribute to the problem and need to be addressed

- Which type of crime? Murder, rape, theft, assault, kidnapping etc.
- Is it affected by location?
  I.e. Variations between provinces, variations between urban and rural
- Does it affect socio-economics groups differently?
- Is age of victim a factor?
- Was the criminal apprehended?
- How many crimes are committed by people out on bail?

These variables will be taken into account when data is gathered for the investigation.

Activity 2

Turn to your Learner Work File and complete Activity 2

LEARNER WORK FILE

1.2 Collecting data

Once we have identified the problem to be investigated, we can then ask the next question: What information do we need to investigate the problem and how do we get it?

Selection of samples

Terms to know:

A **population** is a collection of people, objects or events that the researcher is interested in studying.

A **sample** is a smaller group chosen from the population to take part in the study.

It is not usually feasible or even necessary to collect data from everyone in the population group. So the researcher will conduct his study using the **sample** which is specially chosen to represent the **population** and then generalise his results to the population group. This process of generalisations from sample to population will always be subject to some error as each sample chosen from a population will not have exactly the same characteristics as the population. To keep this error as small as possible it is best to use **random samples.**

A **random sample** is a sample chosen from a population in such a way as to ensure that every person or object in the population has an **equal chance** of being selected for the sample.

For a small population, a random sample can be obtained in the following manner:

Sampling without replacement

List all the objects in the population on separate cards and place them in a container. Decide on your sample size and draw out the required number of cards one at a time, shuffling well between each draw.

Sampling with replacement

The same process is used, but this time each card that is drawn is recorded and then put back into the container so that there is an equal chance of it being drawn again. If a card is drawn twice (or more), the information from the questionnaire is recorded twice (or however many times the card was drawn)

For a very large population, more sophisticated methods such as number tables or computer generated numbers would be used to generate a random sample.

Discussion

Your facilitator will conduct a selection of random samples exercise with your class. Compare the method of sampling with replacement to the method of sampling without replacement. Identify the more reliable method of selecting a sample, justifying your answer. (Use your knowledge of probability to decide whether each card has an equal chance of being drawn in each method)

Sample size

The error in drawing conclusions about a population from sample information can be limited by selection of random samples. It can also be affected by sample size. The bigger the sample, the smaller the error. However this is not always possible and so the following figures are given by John Roscoe as a guideline.

- No sample size should be less than 10
- In most research, sample sizes from 30 to 500 are acceptable. Within these limits, a sample size of about one-tenth the size of the population is recommended.
- Larger samples, where they can be afforded are better smaller ones. However, well-chosen small samples are better than badly chosen larger ones.

Case Study:

The Medical Research Council (MRC) wished to document and monitor the prevalence of tobacco use of adolescents aged 13-15. They also wanted to assess pupil's knowledge, beliefs and attitudes relating to tobacco-using behaviour. In 2012 they conducted a survey using a sample of 207 schools nationally. They had an 80% participation rate.

Activity 3

Turn to your Learner Work File and complete Activity 3

LEARNER WORK FILE

There are a number of methods that can be used to collect and record data.
In Activity 2, we discussed using questionnaires and interviews.
If you were collecting data on the most commonly driven car on the road, you would use an observation tally sheet, which would record the exact number of each make of car that you observed in a specific time period.
It is also possible to use information that has already been collected and recorded (called a **database**) and is made available to the public in journals, directories, the Internet etc. The newspapers frequently publish the findings of studies for public interest. One such database is that produced by Statistics South Africa, as a result of Census 2001. This database gives information about South Africans in terms of

population figures, languages spoken, unemployment rates, education levels, birth rates, death rates, access to water etc.

Measurement scales

When collecting data from a survey, there are a number of ways that information can be recorded.

In a questionnaire or interview, you could use a simple yes/no scale.

Eg.1

	YES	NO
Do you smoke		
Have you ever smoked		
If you smoke, would you like to stop?		

It might be more informative to use a 5 point scale:

1 = Never 2 = Seldom 3 = Sometimes 4 = Mostly 5 = Always

Eg. 2

	1	2	3	4	5
Do you read the newspaper					
Do you read magazines					
Do you watch television					

Terms to know:

Qualitative variables: Variables which fit research subjects into categories where one category does not have a higher or lower value than another. eg. Gender or sex of subject. Males may be assigned the variable 1 and females the variable 0. It doesn't mean that males have a higher value than females!! In example 1 above, a yes response could be assigned a 1 and a no response could be assigned a 0.

Quantitative variables: Variables which can be "measured" and for which the idea of having more or less than something else has meaning. eg. Height, weight, intelligence.

If a variable can only take on a finite set of values, it is said to be a **discrete variable**. This means it is either a qualitative variable or a quantitative variable which does not allow fractions eg. children in a classroom. If a variable can take on **any** value, including fractions, it is said to be a **continuous variable.**

Formulating Questions in Surveys

To conduct a meaningful survey, the right questions must be asked and they must be asked in the right way!

Here are some guidelines for writing survey questions:

- Questions must be short, simple and unambiguous
- Questions should not lead the respondent to a particular answer
- Questions must be necessary and relevant to the study
- The respondent must be willing to answer the questions.
- Questions must be specific – avoid asking two questions at once

It is a good idea to pilot your questionnaire. This means testing the questionnaire first, by asking a group of typical respondents to fill it in and then analysing your results.

Potential problems with data collection in a survey

- We have already seen how important it is to design the **right questions**.
- In addition, the **interviewer must not be biased**. Helping the respondent and leading them to make the right statements must be avoided as the results of the study will then be unreliable.
- **Sample bias** is common and difficult to avoid. This results from taking a sample that is not completely representative of the population you wish to study. For example, you might conduct a survey to determine the opinion of factory workers on the benefits of belonging to a union. You use a random sample of workers from a factory in your area.

Discussion

Using the example of the factory workers above, how representative will the sample be of all factory workers in South Africa. Discuss the potential for sample bias in this case.

1.3 Organising data

The data that you collect from questionnaires, interviews or observation tallies, is called raw data – it has not yet been organised. Very often, we may wish to group data for a clearer, more comprehensive display. Let's consider the following example.

A survey is conducted to determine people's opinions on cell phone service providers. The questionnaire contains 3 questions: 1 – How old are you?
2 – Which service provider do you use?
3 – Are you happy with the service?

A random sample is chosen of 25 cell phone users in a shopping centre and the following data is collected (M = MTN; V = Vodacom; C = Cell C)

Age	Provider	Contented
13	M	Yes
25	M	No
36	V	No
51	C	No
27	C	Yes
20	V	Yes
16	V	Yes
41	M	Don't Know
26	M	Yes
15	C	Yes
34	V	Yes
17	C	No
19	V	Yes

32	M	Yes
40	M	No
47	C	Yes
51	V	Yes
23	M	Don't Know
18	C	Yes
20	V	Yes
14	C	Yes
41	M	Yes
54	V	Yes
30	V	Don't Know
25	M	Yes

The ages could be grouped into the following intervals and a tally kept of the numbers.

Age	**M**	**V**	**C**
10-19	I	II	IIII
20-29	IIII	II	I
30-39	I	III	
40-49	III		I
50-59		II	I

This would result in the following frequency table:

Table 1: Preferred cell phone service provider by age

Age	**MTN**	**Vodacom**	**Cell C**	**Total**
10-19	1	2	4	7
20-29	4	2	1	7
30-39	1	3		4
40-49	3		1	4
50-59		2	1	3
Total	9	9	7	25

We could also organise the data by contentment with service provider.

	Happy	Unhappy	Don't Know
MTN	IIIII	II	II
Vodacom	IIIII II	I	I
Cell C	IIIII	II	

This would result in the following frequency table:

Table 2: Contentment with service provider

	Happy	Unhappy	Don't Know	Total
MTN	5	2	2	9
Vodacom	7	1	1	9
Cell C	5	2	0	7
Total	17	5	3	25

These tables could now be analysed and the results of the survey communicated to people who would benefit from the findings. We can also investigate issues by using data that has been collected by others. Let's look at **extracting information that has been organised and displayed in databases** in the form of tables.

Activity 4

Turn to your Learner Work File and complete Activity 4

1.4 Representing data

Once data has been collected and organised, it may be represented in the form of tables or graphs.
Let's revise the drawing of bar graphs and line graphs.

Bar Graphs and Line Graphs

Activity 5

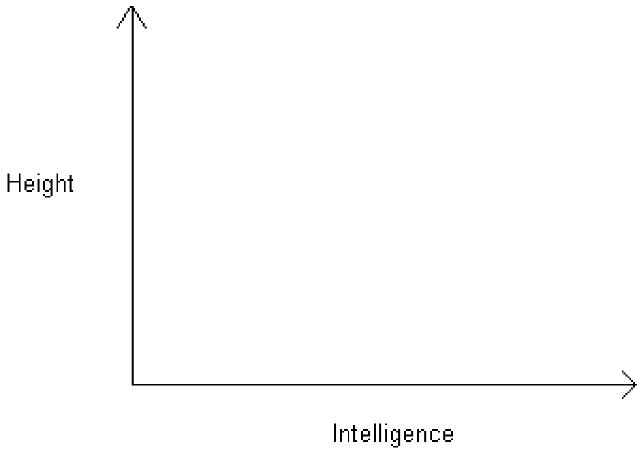
Turn to of your Learner Work File and complete Activity 5

Scatter graphs

Another very useful tool to represent and analyse data is the scatter graph. This is used when there is paired information and we want to establish whether a relationship exists between the two sets of information. For example, is there a relationship between height and shoe size? If we took a sample of people and recorded their height and shoe size, we would expect a graph of height against shoe size to look something like this:

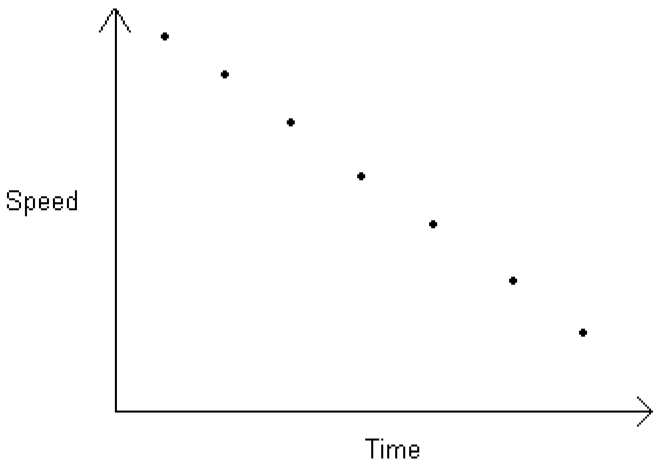
From the scatter graph we can see that there **is a relationship** between height and shoe size. As a person's height increases, the size of their shoes also increases. This is called a **positive correlation** and we could use this correlation to **make predictions** about the shoe size a person measuring a certain height would wear.

Do you think there is a relationship between a person's height and their intelligence? What do you think a scatter graph of height against intelligence would look like? Fill in some points on the axes below of your prediction.



If the points of a scatter graph are scattered all over, we say that there is **no correlation** between the two variables and we would therefore be unable to make any predictions about a person's IQ based on their height.

Suppose we plotted a scatter graph of the relationship between the speed of a car and the time taken to drive a certain distance. We would expect the following situation.



This is called a **negative correlation.** The faster the car, the less time it will take to travel the distance.

Activity 6

The Human Resources Manager of a company employing 200 workers conducts a survey to investigate whether there is a relationship between the age of workers and the number of days they are absent per year. He uses the correlation and a **line of best fit** to make predictions about absenteeism based on a worker's age.

Turn to your Learner Work File and complete Activity 6

1.5 Analysing data

Once the data has been collected and recorded, we have a number of tools which can be used to analyse the data.
We might want to analyse the data in terms of percentages.
For example, in the survey on cell phone users on pages 9-11, we could calculate the percentage of cell phone users falling within age groups.

Let's look again at table 1:

Table 1: Preferred cell phone service provider by age

Age	MTN	Vodacom	Cell C	Total
10-19	1	2	4	7
20-29	4	2	1	7
30-39	1	3		4
40-49	3		1	4
50-59		2	1	3
Total	9	9	7	25

We can now see that 7 out of 25 people interviewed were in the 10 – 19 age group. This translates to 28% of the sample.
Similarly, 3 out of 25 people interviewed were in the 50 – 59 age group ie. 12% of the sample.

Activity 7

Discussion

Could the results of this survey now be used to make inferences about the population of all cell phone users?

Turn to your Learner Work File and complete Activity 7

1.6 Data sources and databases

Data organisation techniques enable data collectors to create a database that can be used in future should there be need to refer to the data that has already been collected. **Databases** also make it easier to make follow ups on the situation that is being dealt with, for instance, follow up can be made on recorded previous statistics of adults infected with HIV after implementation of proper measures to curb HIV infection on adults. Also, the data sources and databases used in a sample should therefore introduce no bias. This means there is need for careful and well planned choice of data sources and databases

1.7 Contamination of Data

Data can be contaminated in many different ways. The problem with the use of contaminated data is that the conclusions will be biased and this might lead to wrong decisions being made on real life problems and issues. Data can be contaminated due to errors in the collection of data, errors in entering or capturing the data, errors in choosing the right sample for the population, manipulation of data for other reasons such as political reasons.

1.8 Measures of central tendency

Most sets of data tend to cluster around a central value and so it is often useful to calculate this single number that will summarise or represent a set of numbers. Ie. it gives one a good idea of the set of data. This number is called a measure of central tendency.

There are three different average measures or measures of central tendency: the mean, median and mode.

The **mean** of a set of a data is the sum of all values divided by the number of values (also called the average).

$$\text{The mean} = \frac{\text{Sum of the total number of items}}{\text{Total number of items}}$$

E.g. We calculate the mean of 2, 10, 11, 15 and 21 as follows:

Mean $= \frac{2 + 10 + 11 + 15 + 21}{5}$

$= \frac{59}{5}$

$= 11.8$

The **median** of a set of data is the middle value when the values are arranged in size.
The median is the middle value. Line up the values in your data from smallest to biggest (or biggest to smallest) and the one in the dead centre is the median.

E.g. 1 The median of 2, 10, 11, 15, 21 is 11

If there are an even number of values, the median is the average of the 2 middle values.

E.g. 2 The median of 2, 10, 11, 15, 16 and 21 is $\frac{11 + 15}{2}$ i.e. 13.

The **mode** of a set of data is the value that appears most often (there may be more than one mode in a set of data).

The **range** is the difference between the highest number and lowest number in a set of numbers.

E.g. 21 - 2 = 19

The mean is used most often as the average of a set of scores. However the mean can be distorted by extremely high or extremely low values (called **outliers**). In this case, the median would give a better indication of the average of the scores. When we talk about the "average" factory workers salary, for example, it would be more meaningful to give the median salary. The median is the point on the score scale below which one-half of the scores fall.

Activity 8

Turn to your Learner Work File and complete Activity 8

LEARNER WORK FILE

Measures of Spread

We have examined measures of central tendency but it also useful to have an understanding of whether the measurements are clustered tightly around the mean or spread over the range. We have already discussed the range ie. the difference between the largest and the smallest scores in a collection. There are times when the range is a very good measure of spread and *should* be calculated. However the range is not always a good representation of the spread of a set of data. Why?

Another measure of variability is the difference between two percentiles – usually the 25^{th} and the 75^{th} percentile.

<u>Note:</u> The 25^{th} percentile is a number such that 25% of the measures in a set of data are less than or equal to it. It is also called the *lower quartile.*

The 75^{th} percentile is a number such that 75% of the measures in a set of data are less than or equal to it. It is also called the *upper quartile.*

If the difference between the lower and upper quartiles is small, the measures are tightly clustered around the mean. If the difference is large, the measures are widely scattered.

1.9 Manipulating Statistics to mislead

Statistics are used to "make sense" of data and to inform people about that data. However, statistics are sometimes used to **mislead people** and you must therefore **always critically analyse information** that is presented as a statistic. You need to ask who is presenting the information. For example if research is done on whether smoking lowers life expectancy and the resulting statistics show that it has little or no effect, you need to find out who paid for the research. It might turn out to be a cigarette company!

Let's investigate how statistics can be manipulated to support a particular cause.

1. Statistics that are true but misleading

<u>Eg.</u> A study done on gender equality in top positions in a particular company results in the following conclusion:

The number of women in senior managerial posts has risen by 200%.

You would be suitably impressed. What you don't know is that there used to be 1 woman and now there are 2!

<u>Eg.</u> An advert uses the statistic that ***4 out of 5 dentists recommend a particular brand of toothpaste.*** Impressive!

What you need to ask is how big was the sample group and who was paying for the survey?

If the toothpaste manufacturer was financing it, they would **want** to get the result that 4 out of 5 dentists recommended their brand. They could decide on a sample size of 5 dentists! If in the first sample of dentists, only 2 out of the 5 recommended their toothpaste, they could pick another sample of 5 dentists and keep trying until they got the desired result.

Activity 9

A survey using data from an insurance company statistics makes the inference that women are better drivers than men. The study actually shows that on average a woman in the 20 – 65 year age group will have fewer accidents than a man of the same age.
Do you think the inference that women are better drivers than men is accurate from this study? Discuss.

Turn to your Learner Work File and complete Activity 9

2. Manipulating scales on graphs

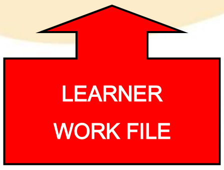
Statistics can be represented in a way that makes the results of a study **appear** better (or worse) depending on which side you are on!

Activity 10

Turn to your Learner Work File and complete Activity 10

Activity 11

Turn to your Learner Work File and complete Activity 11



3. Wrong questions being asked

Some surveys might be constructed in a way to encourage a particular answer.
For example, a survey is being conducted to determine people's opinions on providing assistance to the unemployed. The question is asked as: *Do you feel you should be taxed so that some people can stay at home and get paid for doing nothing?*
There would probably be lots of *NO* responses.
On the other hand if the question was worded: *Do you think the government should help people who are unable to find work?* – There would be lots of *YES* responses.

4. Sample size error

We have already seen the effects of sample size error. With small samples, a change in one piece of data can completely change the results. The bigger the sample size, the more accurate the results.

Conclusion

We have seen that if you want to deceive people, statistics makes it easy. When you are evaluating statistics based arguments, you should be aware of potential areas of misuse of the statistics. These are the things you should be looking for:

- **Where did the data come from?** Who ran the survey and do they have an interest in the outcome? Who is paying for the study?
- **How was the data collected?** What questions were asked? How were they asked? Who was asked?
- **Have the numbers been taken out of context?** Does the analysis only focus on data that supports the prediction and ignores everything else?

Activity 12

Turn to your Learner Work File and complete Activity 12

LEARNER WORK FILE

Statistics Project 1

In this module you have learnt how to:

- Identify situations or issues for investigation
- collect data
- organise data
- represent data
- analyse and interpret data
- critically evaluate statistics

You are now going to put this knowledge to good use in your statistics project.

Turn to your Learner Work File for instructions on your project.

LEARNER WORK FILE

Module 2

Use theoretical and experimental probability to develop models, make predictions and study problems.

This Module deals with:

- Appropriately experiments and simulations are chosen and/or designed
- Predictions based on validated experimental and theoretical probabilities
- Correctly and clearly interpreting the results of experiments and simulations
- Clearly communicating the outcomes of experiments and simulations

2.1 Probability

Probability is a measure of the likeliness that an event will occur.
If you flip a coin 20 times how many times do you think you will get heads?_________
Let's conduct **a trial** to determine whether you are correct.

Activity 13

Turn to your Learner Work File and complete Activity 13

Experimental probability

Experimental probability is the **chance of something happening, based on repeated testing and observing results**. It is the ratio of the number of times an event occurred to the number of times tested. For example, to find the experimental probability of winning a game, one must play the game many times, then divide the number of games won by the total number of games played.
The ratios you calculated in the previous activity are the probabilities of getting heads or tails every time you flip a coin. You probably found that the number of heads and tails were pretty much equal – especially when you combined results with the other groups. The more tosses, the more chance there is that heads and tails will be tossed equally.

Theoretical probability

Theoretical probability is the chance of events happening, found by calculating results that would occur under ideal circumstances. Ie. It is a **prediction of the outcome of an event** found by thinking about it logically.
We say that there is an **equal chance** of getting heads or tails. So when we flip a coin there are 2 **equally likely outcomes**. Of course it is possible to get 20 heads in 20 throws, but this is not a likely outcome. The **probability** is that you will get 10 heads and 10 tails.

There are 2 possible **outcomes** when a coin is flipped: heads or tails.
The probability of getting heads is 1 out of 2 outcomes or $\frac{1}{2}$
The probability of getting tails is 1 out of 2 outcomes or $\frac{1}{2}$
We can also write probability or chance as a percentage: There is a 50% chance of heads and a 50% chance of tails. (We call this a fifty-fifty outcome)

Activity 14

Turn to your Learner Work File and complete Activity 14

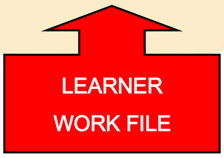
You should have seen that when you throw a die, the probability of getting a particular number is $\frac{1}{6}$.
This is calculated as:

$$\text{The probability of an event} = \frac{\text{total no. of ways in which a specific event can occur}}{\text{total number of equally likely outcomes}}$$

Let's calculate the probability of throwing a 5?
With the throw of 1 dice, how many ways can you get a 5? **1**
How many **equally likely outcomes** are there? **6** (you could get 1, 2, 3, 4, 5 or 6)
So, the probability of throwing a 5 = $\frac{1}{6}$

Activity 15

Turn to your Learner Work File and complete Activity 15



Note:

- Probability or chance can be written as a ratio $\frac{1}{6}$ or 1:6
- It can be written as a percentage : 17%
- We also talk about probability as " a 1 in 6 chance" or "1 out of 6 chance"
- Probability can also be expressed as a number between 0 and 1.
  If something is going to happen with 100% certainty the probability is 1
  If something is never going to happen the probability is 0

Statistics Project 2

Turn to your Learner Work File and complete Statistics Project 2

LEARNER WORK FILE

2.2 Laws of probability

1. Law of addition:
When two events are **mutually exclusive**, then if one occurs the other cannot occur. For example: When we flip a coin, we can get heads or tails – we cannot get both! When we toss a die, the outcome is one of 1, 2, 3, 4, 5 or 6. We cannot get a 1 **and** a 6 at the same time. Ie. they are mutually exclusive.
However, we can calculate the probability that we will throw a 1 **or** a 6 by **adding the separate probabilities** of getting a 1 or a 6 : $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
The probability of throwing a 1, 2 or 3 will be $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

2. Law of multiplication:
When two events are **independent**, the occurrence or non-occurrence of the one has no effect on the other. For example, when you throw a die you may get heads on the first throw. This does not affect the chance of getting heads on the second throw. You have a 50/50 chance of getting heads on every throw.
We could work out the probability of getting heads three times in a row by **multiplying the separate probabilities** of getting heads on each throw. $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

Example: *Tata ma chance – tata ma millions ?*
What is the probability of winning the lotto with one row of 6 numbers? (Remember that there are 49 numbers in the lotto machine each with an equal chance of being drawn)
Solution:

What is the probability of having the first number drawn? $\frac{1}{49}$

What is the probability of having the second number drawn? $\frac{1}{48}$

What is the probability of having the third number drawn? $\frac{1}{47}$

What is the probability of having the fourth number drawn? $\frac{1}{46}$

What is the probability of having the fifth number drawn? $\frac{1}{45}$

What is the probability of having the sixth number drawn? 1

44

We can now calculate the probability of getting all 6 correct numbers drawn as:

$$\frac{1}{49} \times \frac{1}{48} \times \frac{1}{47} \times \frac{1}{46} \times \frac{1}{45} \times \frac{1}{44} = \frac{1}{10\ 068\ 347\ 520} = 0,\ 000\ 000\ 000\ 1$$

Now you know why you never seem lucky!!!

Activity 16

Turn to your Learner Work File and complete Activity 16

Now we are ready to look at issues in society and the workplace which can be investigated using statistics

Module 3

Critically interrogate and use probability and statistical models in problem solving and decision making in real-world situations.

This Module deals with:

- Meaningfully interpreting statistics generated from data
- Appropriately defining assumptions made in the collection or generation of data and statistics
- Appropriate analysis of tables, diagrams, charts and graphs
- Predictions, conclusions and judgements made on the basis of valid arguments and supporting data, statistics and probability models
- Evaluating statistics to identify potential source of bias ,errors in measurement ,potential uses and misuses and their effects

3.1 Statistical Experiments

All **statistical experiments** have three things in common:

- The experiment can have more than one possible outcome.
- Each possible outcome can be specified in advance.
- The outcome of the experiment depends on chance.

A coin toss has all the attributes of a statistical experiment. There is more than one possible outcome. We can specify each possible outcome (i.e., heads or tails) in advance. And there is an element of chance, since the outcome is uncertain

The Sample Space

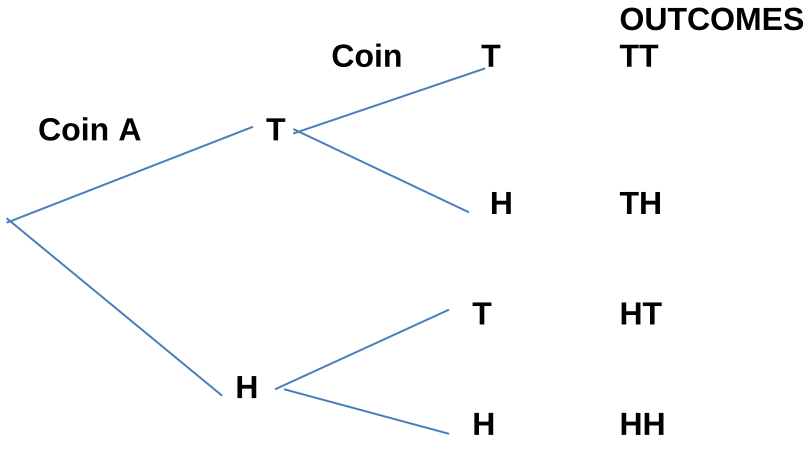
- A **sample space** is a set of elements that represents all possible outcomes of a statistical experiment.
- A **sample point** is an element of a sample space.
- An **event** is a subset of a sample space - one or more sample points

For this experiment, the sample space consists of six sample points: {1, 2, 3, 4, 5, 6}. Each sample point has equal probability. And the sum of probabilities of all the sample points must equal 1. Therefore, the probability of each sample point must be equal to 1/6.

3.2 Probability diagrams

a. Tree diagrams

Possible outcomes when 2 coins are tossed. The possibilities are represented by branches on a tree.



How many possible outcomes are there in this situation? **4 possible outcomes**

b. Probability tables

Dice 2

Dice 1

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	?
4	5	6	7	8	9	**10**
5	6	7	8	9	**10**	11
6	7	8	9	**10**	11	12

What value should go where there is question mark?

Value = 3 + 6

= 9

How many possible outcomes are there in total?

Total outcomes are = 36

How many outcomes give a score of 10?

3

Two fair dices are thrown together .Find the probability that the sum of the resulting number is odd.

Let O be the event that the sum is odd. From the probability diagram, n (O) = 18

$$P(O) = \frac{18}{36} = \frac{1}{2}$$

3.3 Bias or errors

If the coin or dice is damaged or uneven the results of the experiment will biased or have errors. The experiment must be fair to eliminate or reduce bias or errors. Analysis of a biased statistical experiment will result in wrong conclusions being made.

Activity 17

Turn to your Learner Work File and complete Activity 17

LEARNER
WORK FILE