Tutorial Letter 201/2/2018

Statistics Education in Intermediate and Senior Mathematics

MAE202N

Semester 2

Department of Mathematics Education

This tutorial letter contains feedback of Assignment 02.

BARCODE

Define tomorrow.

UNISA | university of south africa

Question 1

1.1 1) Determine the sampling interval; 2) select a random starting point between one and k; 3) select every k^{th} element (including and starting with the element selected in step two). Note that the definition of systematic sampling includes all these steps.

1.2 A representative sample is a sample that resembles the total population. It is important to use a sampling method that produces representative samples when your goal is to understand the characteristics of a population based on the study of a sample (meaning, when you want to directly generalise from your sample to your population).

1.3

    1.3.1

**Accidents caused by drunken driving:
Nov 2015 -Feb 2016**
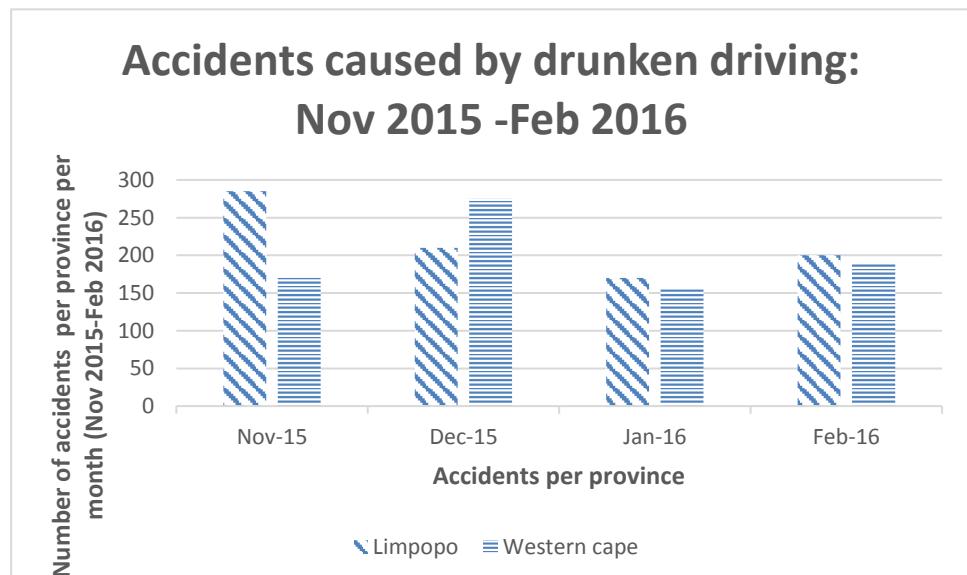
1.3.2 Limpopo province had an increase in accidents in November 2015, January 2016 and February 2016 and the Western Cape had an increase in accidents in December 2015.

Question 2

Marking rubric		Poor (0)	Good (1.5)	Very good (2)
Defining median and mode				
Planning	Plan the "before" phase.			
	Plan the "during" phase.			
	Plan the "after" phase			
Evaluation / Conclusion				

Question 3

3.1

3.1.1 The population mean is

$$\text{Population mean} = \mu = \frac{sum \ of \ all \ the \ observations}{Number \ of \ obsertions} = \frac{\sum x_i}{N}$$

$$= \frac{71+72+34+82+55+\cdots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+50+46+74+60}{64}$$

= 51.562. SHOW ALL THE STEPS, OTHERWISE 1 MARK FOR FINAL ANSWER

3.1.2 The median

The median is the middle number of an ordered set of data.

- Start by ordering the data in either ascending or descending order. The middle data value will be the median.

The median = 51.5.

3.1.3 The mode is the value that occurs most often. In this case, we have trimodal:

24, 34 and 55.

3.2 The mean is the best measure of central tendency if there are no outliers, because it includes all the values in the data set for its calculation. However, if the data contains outlier(s), the median is the best measure of central tendency, because it is not affected by outliers.

In order to check if the data contains the outliers or not, we use this formula:

Lower limit: Q_1 - 1.5 (IQR) and upper limit Q_3 + 1.5 (IQR) (see page 72 of Tutorial Letter 501)

Lower limit: 38.5 -1.5 (65-38.5) = -1.25 and

Upper limit: 65 + 1.5 (65-38.5) = 104.75

Data outside these limits are considered outliers. Looking at our data, the data values range from 18 to 96, which fall within the limits; therefore, the data do not have the outliers.

In this situation, the mean is the best measure of central tendency.

3.3

3.3.1

$$\textbf{Variance } (\sigma^2) = \frac{\textit{Sum of squared deviations}}{(\textit{population size})} = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{(N)}$$

$$= \frac{(71-51.562)^2 + (34-51.562)^2 + (47-51.562)^2 + (39-51.562)^2 + (54-51.562)^2 + \cdots\ldots + (80-51.562)^2}{(64)}$$

= 314.058594 SHOW ALL THE STEPS, OTHERWISE 1 MARK FOR FINAL ANSWER

To calculate sample variance	
1. Calculate the population mean	μ
2. Calculate the deviation of each data value from the mean	$(x_i - \mu)$
3. Square these deviations	$(x_i - \mu)^2$
4. Sum these squared deviations	$\sum (x_i - \mu)^2$

5. Lastly, average the squared deviations by dividing by	(N)

3.3.2   To find the standard deviation, you take the square root of variance you calculated in 3.3.1

Standard deviation =

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

$=\sqrt{314.058594}$

$= 17.722$

3.3.3   The range = maximum value – minimum value = 96 -18 = 78.

3.3.4   The interquartile range = upper quartile (q3) - lower quartile (q1)

(a) To determine the lower quartile (q1):

1. Remember to start by ordering the data in ascending or descending order if the data is not ordered.
2. Use the median to divide the ordered data set into two halves.
   - Since there are an even number of data points in the original ordered data set, split this data set exactly in half.
3. The lower quartile value is the median of the lower half of the data.

Therefore, lower quartile = (38 + 39) ÷ 2 = 38.5

(b) To determine the upper quartile:

1. Use the median to divide the ordered data set into two halves.
   - Since there are an even number of data points in the original ordered data set, split this data set exactly in half.
2. The upper quartile value is the median of the upper half of the data.

Therefore, upper quartile = (64 + 66) ÷ 2 = 65.

Therefore,

The interquartile range = 65 - 38.5

= 26.5

3.3.5   The semi-interquartile range = (The interquartile range) ÷ 2

$$= 26.5 \div 2$$

$$= 13.25$$

3.4   Standard deviation is the best measure of spread if there are no outliers in the data, because it includes all the values in the data set and since, in 3.2, we have shown that there are no outliers. We conclude that standard deviation is the best measure of spread.
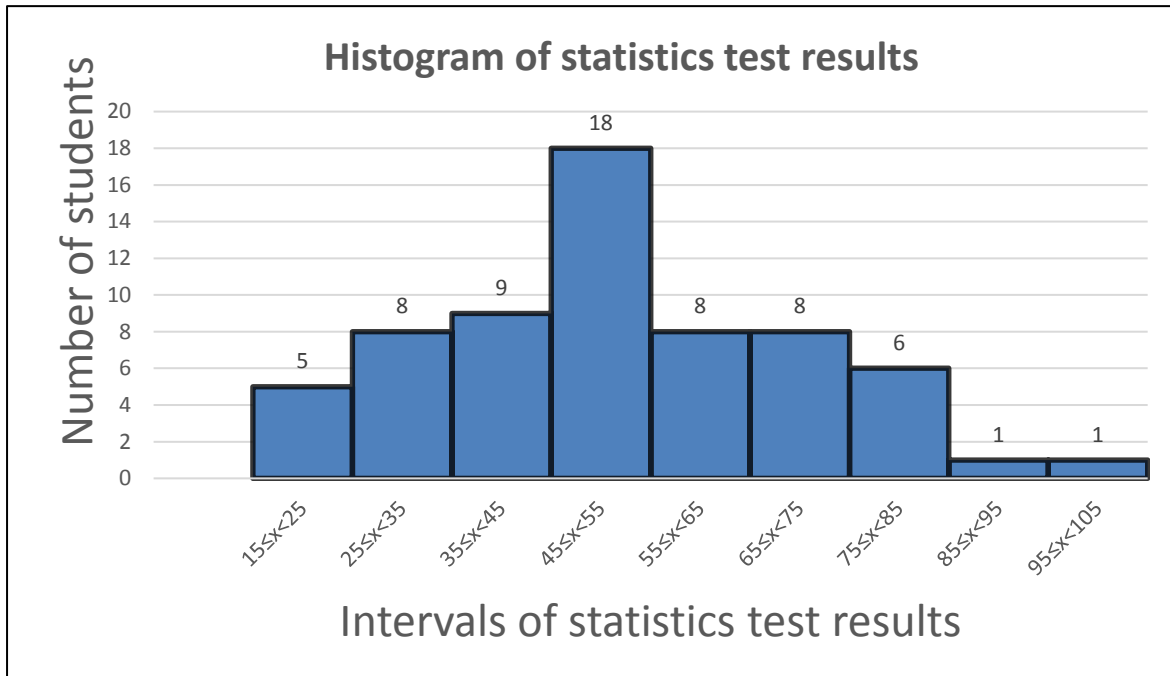
3.5
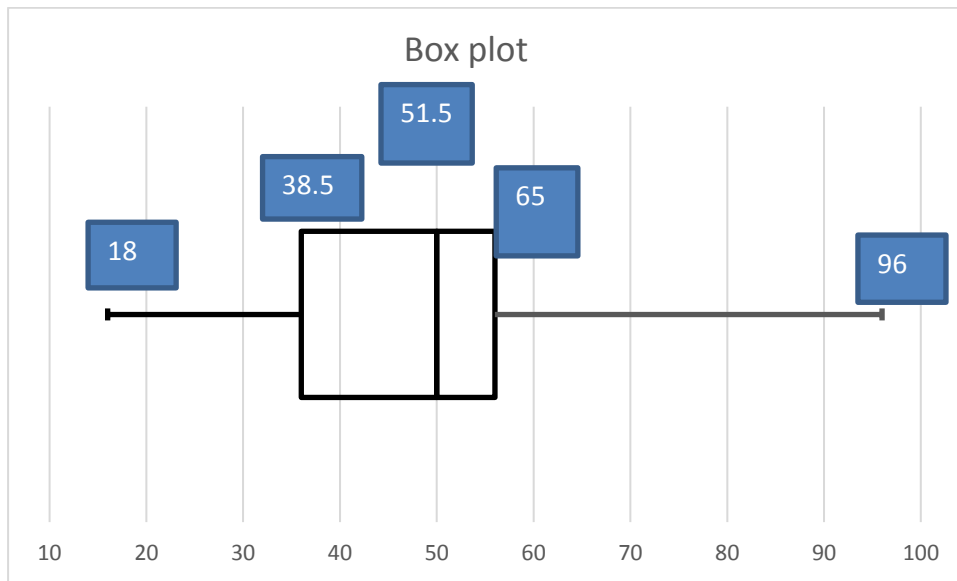
3.6

Statistics test result intervals	Frequency
15≤x<25	5
25≤x<35	8
35≤x<45	9
45≤x<55	18
55≤x<65	8
65≤x<75	8
75≤x<85	6
85≤x<95	1
95≤x<105	1
Total	64

3.7

Histogram of statistics test results

3.8 Most of the students scored between 45 and 55.

Question 4

Leaf - females	Stem	Leaf - males
8	2	1
8 0	3	3
8	4	3
5 3 3 0	5	5 6
5 2 1	6	1 7 7
7 2	7	1 1 4 5 8

Question 5

5.1   Dependent variable = y = Output
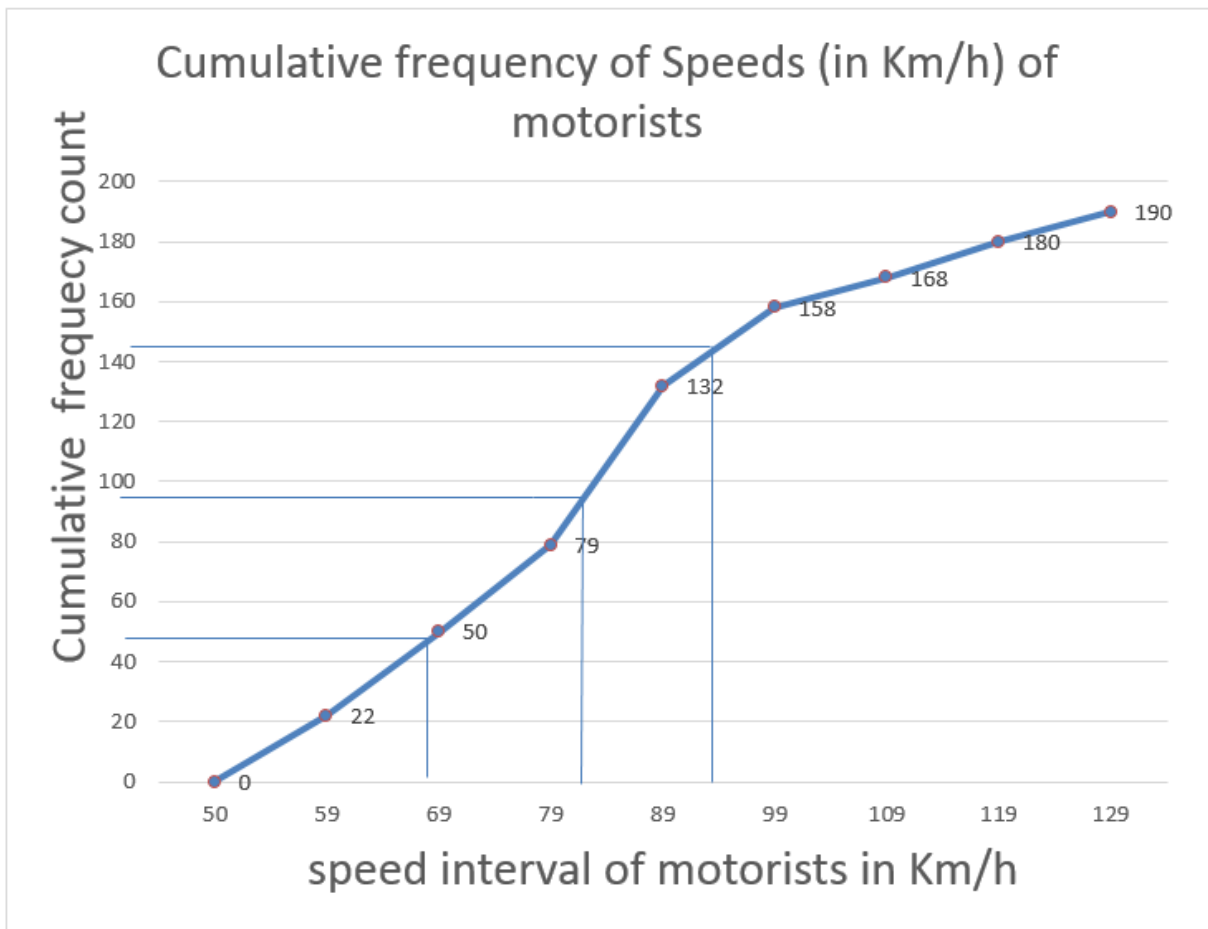
Independent variable = x = Training hours

5.2

Training hours and output
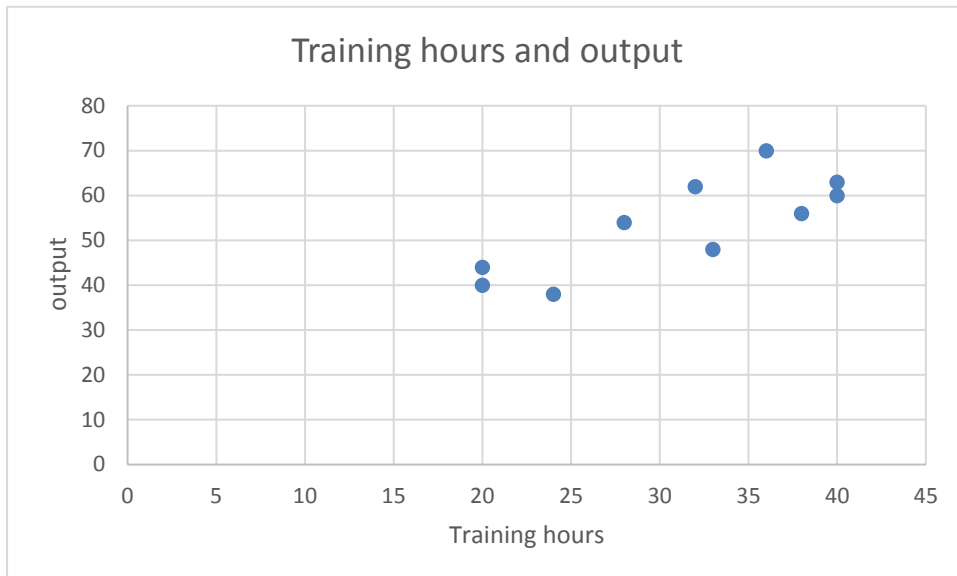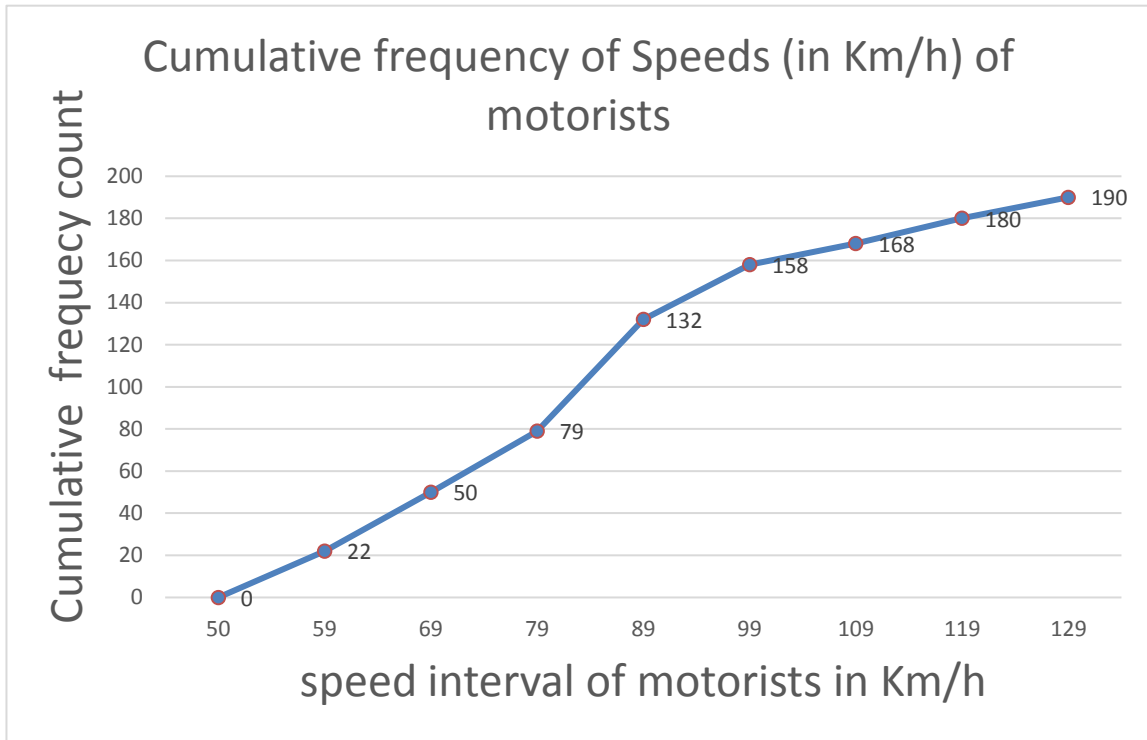
5.3 positive, medium (strong), linear relationship

Question 6

6.1 (a) 168, (b) 180, (c) 190

6.2

Cumulative frequency of Speeds (in Km/h) of motorists

Cumulative frequency of Speeds (in Km/h) of motorists

USING THE CUMULATIVE FREQUENCY GRAPH

6.3.1 To find the median, we know that median is the middle value of the data. So, we will multiply the total sample by $\left(\frac{1}{2}\right)$ in order to find the median position.

That is,

Median position $= \frac{1}{2}$ (n) $= \frac{1}{2}$ X 190 = 95th position

Remember n is the total frequency of the data or the last cumulative frequency, which is 190.

Then, to find an estimate of the median:

You go to the graph to draw a line from 95, at the vertical axis, across curve and down; then, read the horizontal axis. We get a median of 81.

6.3.2 To calculate the interquartile range, we need to first estimate the upper quartile (third quartile) and the lower quartile (first quartile).

(a)   To find the upper quartile (third quartile), we know that the upper quartile value is the median of the upper half of the data (third quarter of the data). So, we will multiply the total sample by $\left(\frac{3}{4}\right)$ in order to find the upper quartile position.

That is,

Upper quartile position = ¾ X 190= 142.5th position

Then, to find an estimate of the upper quartile:

You go to the graph to draw a line from 142.5, at the vertical axis, across curve and down; then, read the horizontal axis. We get an upper quartile of 93.

(b)   To find the lower quartile (first quartile), we know that the lower quartile value is the median of the lower half of the data (first quarter of the data). So, we will multiply the total sample by $\left(\frac{1}{4}\right)$ in order to find the lower quartile position.

That is,

Lower quartile position = ¼ X 190= 47.5th position

Then, to find an estimate of the lower quartile:

You go to the graph to draw a line from 47.5, at the vertical axis, across curve and down; then, read the horizontal axis. We get a lower quartile of 67.

 Therefore,

Interquartile range = upper quartile - lower quartile = 93 – 67 = 26

6.3.3   The semi-interquartile range = (Upper quartile - Lower quartile) = (26) ÷ 2 = 13

Question 7

7.1

		Bag one			
		2	4	6	8
Bag two	1	3	5	7	9
	3	5	7	9	11
	5	7	9	11	13
	7	9	11	13	15

7.2 P (the sum of two values is an even number) = $\dfrac{0}{16}$

7.3 P (the sum of two values is less than 11) = $\dfrac{10}{16}$

7.4 P (the sum of two values will be 11) = $\dfrac{3}{16}$

7.5 P (the sum of two values will be the multiple of 3) = $\dfrac{6}{16}$

Question 8

Let Pi be the pineapple-flavoured

M be the melon-flavoured

L be the lemon-flavoured

B be the banana-flavoured

S be the strawberry-flavoured

8.1 P (M or L) = P (M U L) = P(M) + P(L)

$$= \frac{5}{20} + \frac{2}{20}$$

$$= \frac{7}{20}$$

8.2 P (disliking B or M) = P (disliking B) + P (disliking M)

$$= \frac{3}{20} + \frac{5}{20}$$

$$= \frac{8}{20}$$

Then P (liking other flavours) = $\overline{P \text{ (dıslıkıng B or M)}}$ = 1 – P (disliking B or M)

$$= 1 - \frac{8}{20}$$

$$= \frac{12}{20}$$

$$= \frac{3}{5}$$

OR

P (liking other flavours) = P (Pi or L or S) = P (Pi or L or S)

= P (Pi U L U S)

= P(Pi) + P(L) + P(S)

$$= \frac{4}{20} + \frac{2}{20} + \frac{6}{20}$$

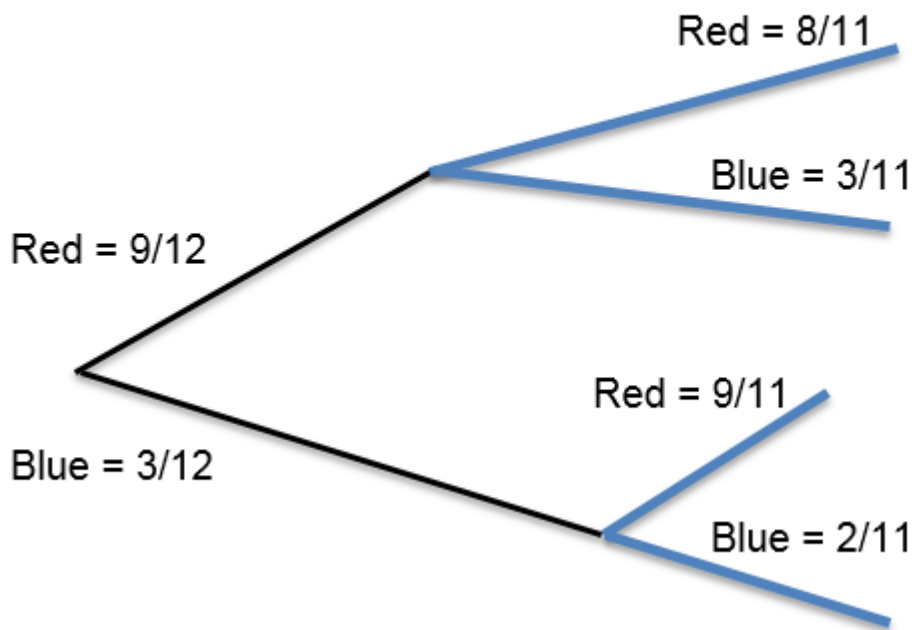$$= \frac{12}{20}$$

$$= \frac{3}{5}$$

Question 9

First draw   Second draw

```
                                              Red = 8/11
                                        _____
                                       /
                          Red = 9/12  /      Blue = 3/11
                        _____
                       /
                      /
                     /              Red = 9/11
                     \         _____
            Blue = 3/12 \     /
                         \____/      Blue = 2/11
                              \
                               _____
```

9.1 P (1ˢᵗ Red) = $\dfrac{9}{12}$

9.2 P (Blue Blue) = $\dfrac{3}{12}$ X $\dfrac{2}{11}$ = $\dfrac{6}{132}$ = $\dfrac{1}{22}$

9.3 P (Red Blue or Blue Red) = P (Red Blue U Blue Red) = $\left(\dfrac{9}{12} \text{ X } \dfrac{3}{11}\right) + \left(\dfrac{3}{12} \text{ X} \dfrac{9}{11}\right)$

$$= \dfrac{27}{132} + \dfrac{27}{132}$$

$$= \dfrac{54}{132}$$

$$= \dfrac{9}{22}$$

Question 10

10.1 P (Limpopo) = $\frac{991}{8948}$

10.2 P (Electricity user │ Western Cape) = $\dfrac{P \text{ (electricity user} \cap \text{ western Cape}}{P \text{ (western Cape)}}$ =

$$= \frac{\frac{750}{8948}}{\frac{981}{8948}} = \frac{750}{981}$$

10.3 P (Western Cape │ Gas user) = $\dfrac{P \text{ (western Cape} \cap \text{ gas user}}{P \text{ (gas user)}} = \dfrac{\frac{7}{8948}}{\frac{351}{8948}} = \dfrac{7}{351}$

Marking rubric

Semester 1: MAE202N - Assignment 02	
Marked questions	Mark per question
Question 3	44
Question 4	8
Question 5	10
Question 6	23
Question 8	7
Question 9	8
TOTAL	100