BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF AGRICULTURE AND ENVIRONMENTAL SCIENCE

DEPARTMENT: NATURAL RESOURCES

PROGRAMME: FAES Part1

COURSE CODE NR122 (2) Introduction to Statistics

DURATION: 2hrs TOTAL MARKS: 70

INSTRUCTIONS TO CANDIDATES

Answer THREE questions out of the following five questions. You must answer question ONE from SECTION A and any TWO questions from SECTION B.

SECTION A (COMPULSORY)

1. (a) The table below shows the results of a national wide survey on the on the preference of a National Park one would like to visit in Zimbabwe in 2015 and in 2020

National Park	2015	2020
Hwange	62.15	65.9
Gonarezhou	22.23	16.6
Mana pools	6.35	-
Matobo	2.40	4.55
Victoria falls	0.67	7.42
Zambezi	6.20	5.47

(i) Summary the results of the survey in a bar chart. **[5 Marks]**
(ii) Write an R code for the barplot in (i) **[5 Marks]**
(iii) Using a single plot compare the results of the 2015 and 2020 survey Results. **[5 Marks]**

(b) A national park in Namibia determines the weight (in kg) of a sample of common eland antelopes:
450 730 700 600 620 660 850 520 490 670 700 820
910 770 760 620 550 520 590 490 620 660 940 790
Assuming that the weight in normally distributed calculate

(i) the point estimate of μ and σ **[5 marks]**
(ii) the confidence interval for μ ($\alpha = 0.05$). **[5 Marks]**

(c) A national park conducts a study on the behaviour of their leopards. A few of the park's leopards are registered and receive a GPS device which allows measuring the position of the leopard. Use this example to describe the following concepts:

(i) population, **[1 Mark]**
(ii) sample, **[1 Mark]**
(iii) observation, **[1 Mark]**
(iv) variable, and **[1 Mark]**
(v) value. **[1 Mark]**

Section B

2 A hiking enthusiast has a new app for his smartphone which summarizes his hikes by using a GPS device as shown in the table below.

Distance(Km)	12.5	29.9	14.8	18.7	7.6	16.2	16.5	27.4	12.1	17.5
Altitude (m)	342	1245	502	555	398	670	796	912	238	466

(a) Calculate the following for distance and Altitude
(i) Mean **[1 Mark]**
(ii) Median **[1 Mark]**
(iii) interquartile range **[2 Marks]**
(iv) Standard deviation **[2 Marks]**
(b) Draw and interpret the boxplot for:
(i) Distance **[4 Marks]**
(ii) Altitude **[4 Marks]**
(c) Write the R codes that reproduce results of (a) and (b) **[6 marks]**

3. The body mass index (BMI) and the systolic blood pressure of 6 people were measured to study a cardiovascular disease. The data are as follows:

Body mass index	26	23	27	28	24	25
Systolic blood pressure	170	150	160	175	155	150

(a) The research hypothesis is that a high BMI relates to a high blood pressure. Estimate the linear model where blood pressure is the outcome and BMI is the covariate. Interpret the coefficients. **[15 Marks**

(b) Calculate R^2 to judge the goodness of fit of the model. **[5 Marks]**

4. (a) For Susan Dean's spring pre-calculus class, scores for the first exam were as follows (smallest to largest):

33; 42; 49; 49; 53; 55; 55; 61; 63; 67; 68; 68; 69; 69; 72; 73; 74; 78; 80; 83; 88; 88; 88; 90; 92; 94; 94; 94; 94;96; 100

Create a sten and leaf diagram using the data. [5 Marks]

(b) A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

	High School	**Bachelors**	**Masters**	**Ph.d.**
Female	60	54	46	41
Male	40	44	53	57

Are gender and education level dependent at 5% level of significance? **[15 Marks]**

5. (a) Calculate an independent samples t-test for the following data sets:

Data set A: 1,2,2,3,3,4,4,5,5,6

Data set B: 1,2,4,5,5,5,6,6,7,9

Test the claim that set A is greater that set B. **[20 Marks]**

End of paper