LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

FIRST YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE (STATISTICS)

STAT 123: INTRODUCTION TO COMPUTER INTERACTIVE STATISTICS

STREAM: R ***TIME: 2 HRS***

DAY: THURSDAY[8.30A.M-10.30A.M] ***DATE:18/04/2024***

THIS QUESTION PAPER CONSISTS OF FIVE (5) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO

Instructions
Answer question ONE and any other TWO questions

QUESTION ONE [30 MARKS]

a) Define briefly the following terms as used in statistics **[3 Marks]**
   i) Outlier
   ii) Descriptive statistics
   iii) Variable
b) Give a diagrammatic representation that distinguishes between a population and sample **[2 Marks]**
c) What does a model describe? **[1 Mark]**
d) What steps does one normally goes through when writing a statistical report **[5 Marks]**
e) Distinguish between nominal and ordinal scales of measurements. Cite an example for each. **[4 Marks]**
f) State any **three** high level plots that are commonly constructed in R **[3 Marks]**
g) Explain the rationale behind conducting destriptive statistics **[2 Marks]**
h) Describe any two techniques that are commonly used to handle missing values. **[4 Marks]**
i) Write down valuation results of the following R expressions
   i) *4:9-1:3* **[2 Marks]**
   ii) *c(2,4,6,8)+c(3,6,9,12* **[2 Marks]**
   iii) *rep(seq(2, 10, by = 2), 2)* **[2 Marks]**

QUESTION TWO [20 MARKS]

Suppose in an experiment, the measurements of lengths and diameters of five cylinders respectively are given as; (2.0, 1.2), (3.9, 0.7), (2.8, 0.8), (3.7, 1.5), (2.6, 0.7)
Give and explain briefly R codes that you would use to;

a) Read these data into two vectors (give the vectors appropriate names) **[4 Marks]**
b) Calculate;
   i) Mean and standard deviation of each variable **[4 Marks]**
   ii) Correlation between lengths and diameters **[3 Marks]**
   iii) Volume of each cylinder (Vol = Length * pi * (Diameter / 2)^2) **[5 Marks]**

c) Display a table with the three variables in b) above **[4 Marks]**

QUESTION THREE [20 MARKS]

a) State any three examples of continuous variables. **[3 Marks]**

b) Draw a normal distribution curve and state its properties. **[5 Marks]**

c) Assume a sample of students had a mean age of 35 years with a standard deviation of 5 years. A student was randomly picked from a group of 200 students. Find the probability that the age of the student turned out to be;

i) between 35 and 40 years **[4 Marks]**

ii) beyond 45 years **[4 Marks]**

d) Give and explain R code that can be used to generate twenty random variates whose mean and standard deviation are five and two respectively from normal probability distribution **[4 Marks]**

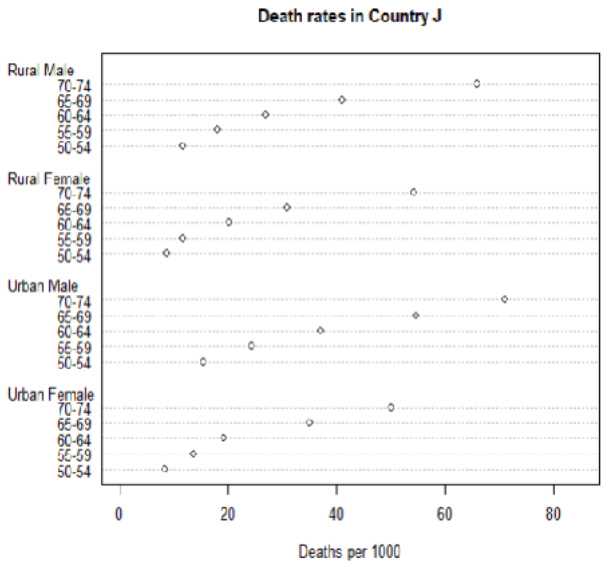
QUESTION FOUR [20 MARKS]

a) State any two R packages that are used for high level plots **[2 Marks]**

b) Describe how a box-plot is constructed. Give its diagrammatic representation. **[6 Marks]**

c) Suppose death rates for various age groups in sub-populations within counties of country J gave the output on the right in R after executing the code on the left.

```
dotchart(Deaths,
         xlim = c(0, 85),
         xlab = "Deaths per 1000",
         main = "Death rates in Country J",
         cex = 0.9)
```



i) Explain the above code **[4 Marks]**

ii) Make relevant interpretation for above dot chart **[4 Marks]**

d) Suppose in a study all women with at least one birth are classified as either cases or controls and with age at first birth as either ≤ 30 years or >30 years and gave the following cross-tabulation in R.

```
        ≤ 30 years >30 years
Case           683      2547
Control       1498      8747
```

Give and explain codes that would create the above table in R. **[4 Marks]**

QUESTION FIVE [20 MARKS]

a) What is meant by the term simple linear regression? **[3 Marks]**

b) Discuss any four assumptions that your data must satisfy before regression analysis is performed **[8 Marks]**

c) Suppose in a clinical study that was conducted to find whether there exists any relationship between the weight and blood pressure of patients gave the following output after analysis.

```
Call:
lm(formula = bp ~ wgt)

Residuals:
     Min       1Q   Median       3Q      Max
-16.5142  -8.2463  -0.5857   4.7179  24.4143

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -41.5547    45.4858  -0.914  0.38764
wgt           2.4643     0.5693   4.328  0.00252 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.27 on 8 degrees of freedom
Multiple R-squared:  0.7007,    Adjusted R-squared:  0.6633
F-statistic: 18.73 on 1 and 8 DF,  p-value: 0.002518

> summary(bp_lm)$r.squared
[1] 0.7007488
```

Make appropriate interpretation **[10 Marks]**

Table A.3 Standardized normal distribution function: a table of

$$F_U(u) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{u} e^{-x^2/2} dx,$$

for $u = 0.0$ to 3.69

u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5733
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9482	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9874	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999