



STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)
MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
END OF SEMESTER EXAMINATION
DSA 8104: FUNDAMENTAL STATISTICAL ANALYSIS

DATE: 6th October , 2022

TIME: 3 Hours

INSTRUCTIONS

1. This examination consists of **FIVE** questions.
 2. Answer Question **ONE (COMPULSORY)** and any other **TWO** questions.
 3. You may use a **SIMPLE CALCULATOR**. No **MOBILE PHONES** in the exams room.
-
-

Question One (30 Marks)

- (i) Differentiate between parametric and non-parametric tests of hypothesis. (2 marks)
- (ii) The price of a commodity increased by 5% from 1989 to 1990, 8% from 1990 to 1991 and by 77% from 1991 to 1992. Find the average price increase. (3 marks)
- (iii) Explain with the help of well labeled diagrams the difference between Two-tailed test of hypothesis and one-tailed test of hypothesis. (4 marks)
- (iv) A random variable X has the probability distribution shown below

x	1	2	3	4	5
$P(X=x)$	$7c$	$5c$	$4c$	$3c$	c

- Determine the value of the constant c and hence determine the mean and variance of $Y = 10X + 25$. (4 marks)
- (v) It is said that secondary data should be used with utmost care. Explain three characteristics of secondary data that a statistical investigate must observe before using them. (3 marks)
 - (vi) The mean weight of 800 male students at a certain college is 140kg and the standard deviation is 10kg assuming that the weights are normally distributed find how many students weigh
 - (a) between 130 and 148kg (3 marks)

(b) more than 152kg (3 marks)

(vii) A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} 0.25, & 0 < x < 2 \\ 0.5x + c, & 2 < x < 3, \\ 0, & \text{elsewhere} \end{cases}$$

Find c hence compute $P(1 \leq X < 2.5)$. (4 marks)

(viii) On experience, it is found that an executive is late for office on four days out of 30 working days. Let X denote the number of times that the executive will be late to the office in the next 60 working days. Determine $P(5 \leq X \leq 10)$ (3 marks)

(ix) What is regression analysis? (1 mark)

Question Two (15 Marks)

(i) After taking a refresher course, a salesman found that his sales (in dollars) on 9 random days were 1280, 1250, 990, 1100, 880, 1300, 1100, 950 and 1050. Does the sample indicate that the refresher course had the desired effect, in that his mean sale is now more than 1000 dollars? Assume $\sigma = 100$, and the probability of erroneously saying that the refresher course is beneficial should not exceed 0.01. Also assume that the sales are normally distributed. (3 marks)

(ii) The head of the Statistics department of a certain university is interested in the difference in writing scores between freshman Statistics students who are taught by different teachers. The incoming freshmen are randomly assigned to one of two Statistics teachers and are given a standardized writing test after the first semester. We take a sample of eight students from one class and nine from the other class. Is there a difference in achievement on the writing test between these two classes? (3 marks)

Class1	35	51	66	42	37	46	60	55	53
Class2	52	87	76	62	81	71	55	67	

(iii) A university has found over the years that out of all the students who are offered admission, the proportion who accept is 0.70. After a new director of admissions is hired, the university wants to check if the proportion of students accepting has changed significantly. Suppose they offer admission to 1200 students and 888 accept. Is this evidence at the $\alpha = .05$ level that there has been a real change from the status quo? How about at the 0.02 level? (3 marks)

(iv) In an experiment in breeding mice, a geneticist has obtained 120 brown mice with pink eyes, 48 brown mice with brown eyes, 36 white mice with pink eyes and 13 white mice with brown eyes. Theory predicts that these types of mice should be obtained in the ratios 9 : 3 : 3 : 1. Test the compatibility of the data with theory, using a 5% critical value. (6 marks)

Question Three (15 Marks)

- (i) Make a frequency distribution table for the data on mileage ratings using 5 intervals (five classes) of equal length. Include the left end point of each interval and omit the right end point. (3 marks)

36.3	41.0	36.9	37.1	44.9	40.5	36.5	37.6	33.9	40.2
38.5	39.0	35.5	34.8	38.6	41.0	31.8	37.3	33.1	37.0
37.1	40.3	36.7	37.0	33.9					

- (ii) Use your table in (i) to answer the following questions
- (a) Construct a well labeled histogram and estimate the mode of the distribution. (4 marks)
- (b) Construct a well labeled cumulative frequency curves (ogive) and estimate the median and the quartiles of the distribution. (5 marks)
- (iii) Suppose the mean monthly return on a T-Bill is 0.5% with a standard deviation of 0.58%. Suppose we have another investment Y with a 1.5% mean monthly return and standard deviation of 6%. Which of the two investments offers less risk in terms of investment. (3 marks)

Question Four (15 Marks)

- (i) In a study of the relationship between X = mean daily temperature for the month and Y = monthly charges on electrical bill, the following data were gathered:

X	20	30	50	60	80	90
Y	125	110	95	90	110	130

- (a) Find the correlation coefficient and interpret. (3 marks)
- (b) Fit regression model to the data and interpret the slope of your model. (6 marks)
- (c) Compute coefficient of determination and interpret the result. (3 marks)
- (ii) The following results are obtained by regressing personal savings on some key indicators.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	28.566087	7.354516	3.88	0.00033
pop15	-0.461193	0.144642	-3.19	0.00260
pop75	-1.691498	1.083599	-1.56	0.12553
dpi	-0.000337	0.000931	-0.36	0.71917
ddpi	0.409695	0.196197	2.09	0.04247

Residual standard error: 3.8 on 45 degrees of freedom

Multiple R-Squared: 0.338, Adjusted R-squared: 0.28

F-statistic: 5.76 on 4 and 45 degrees of freedom, p-value: 0.00079

Are the predictors having significance in the model? Explain. (3 marks)

Question Five (15 Marks)

(i) Urn A has 3 black balls and 6 white balls. Urn B has 400 black balls and 400 white balls. Urn C has 6 black balls and 3 white balls. A person first randomly chooses one of the urns and then grabs a ball randomly from the chosen urn.

(a) What is the probability that the ball be black if a person grabbed a black ball? (4 marks)

(b) What is the probability that the ball came from urn B? (4 marks)

(ii) A car salesman takes inventory and finds that he has a total of 125 cars to sell. Of these 97 are the 2001 model, 11 are the 2000 model, 12 are the 1999 model and 5 are the 1998 model.

(a) Which two types of charts are most appropriate to display the data? (2 marks)

(b) Construct the charts. (5 marks)

END