

[This question paper contains 8 printed pages.]

Sr. No. of Question Paper : 2156 GEC-3 Your Roll No.....

Unique Paper Code : 32373901

Name of the Paper : Statistical Data Analysis Using Software Packages (SEE)

Name of the Course : **B.Sc. (H) Statistics (CBCS) – Skill Enhancement Course**

Semester : III

Duration : 2 Hours

Maximum Marks : 50

Instructions for Candidates

1. Write your Roll No. on the top immediately on the receipt of this question paper.
2. Attempt questions from Any **One** of the **Two Parts I or II**.
3. All questions are compulsory.

Part – 1 (Based on SPSS)

1. Fill in the blanks :

- (i) In SPSS, Data file is saved with the extension _____ and output file is saved with the extension _____ .
- (ii) 'Insert variable' inserts a variable to the _____ of the variable containing the active cell in SPSS.
- (iii) The data editor has two views _____ and _____ .
- (iv) In SPSS the default variable type is _____ .
- (v) In order to completely delete cases and variables in SPSS _____ is used from the edit menu. (1×5)

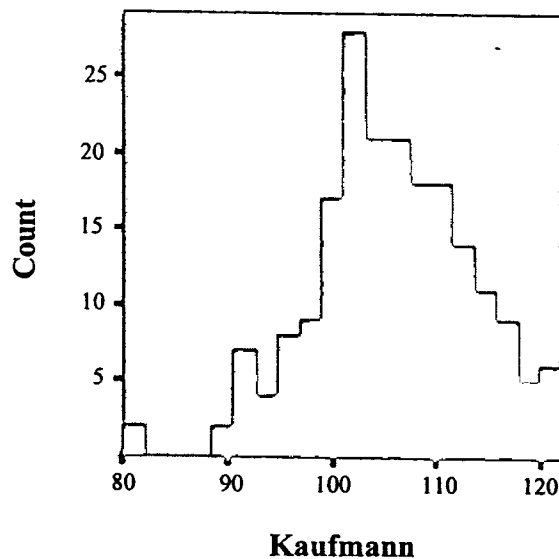
P.T.O.

2. Answer any five of the following :

- (i) Data of marks (out of 100) of 20 students in four different subjects are given. One needs to know in how many subjects each student has got marks greater than 70. Give the procedure to obtain it using SPSS.
- (ii) Differentiate between variable labels and value labels in SPSS.
- (iii) Differentiate between frequencies and descriptive menus of SPSS.
- (iv) Name five standard command pushbuttons in most dialogue boxes of SPSS.
- (v) Give the procedure for constructing bar chart in SPSS.
- (vi) While using the Count Values within Cases transformation, which combinations of values to count a user can specify. (2×5)

3. Answer the following :

- (i) Is the Kaufman score approximately normally distributed ?



- (ii) A conservation scientist in Northern California conducted an experiment to know the average speed of jackrabbits. A random sample of 20 jackrabbits was taken and the speed was recorded. After conducting the test in SPSS following result was obtained :

Test Value = 8.3					
t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
.813	19	.426	.42200	-.6641	1.5081

What is the p-value in this case ? Interpret the result on the basis of above table.

- (iii) Which statement concerning SPSS Statistics application windows is correct ?
- At least one Data Editor window must be open in each SPSS Statistics session.
 - At least one Output Viewer window must be open in each SPSS Statistics session.
 - At least one Syntax window must be open in each SPSS Statistics session.
 - Closing all open Syntax windows will result in the program automatically shutting down.
- (iv) Which statement is true for this crosstab table obtained using SPSS ?

Gender *Eye Crosstabulation

			Eye		Total
			color-blind	not color-blind	
Gender	male	Count	38	442	480
		Expected Count	21.1	458.9	480.0
	female	Count	6	514	520
		Expected Count	22.9	497.1	520.0
Total		Count	44	956	1000
		Expected Count	44.0	956.0	1000.0

P.T.O.

- (a) This table gives the chi-square test for independence of attributes.
 - (b) The expected numbers of females who are not color blind are 497.
 - (c) The observed numbers of color-blind males are 38.
 - (d) All the above.
- (v) Emma tests the following set of hypotheses with an independent samples t-test :

H_0 : Students that spend more than 15 minutes in the first round of registration give the same or a higher evaluation for the current course scheduling procedure than students that spend 15 minutes or less.

H_a : Students that spend more than 15 minutes in the first round of registration give lower evaluation for the current course scheduling procedure than students that spend 15 minutes or less.

She conducted the experiment in SPSS and gets the following output:

Group Statistics

Time_categories	N	Mean	Std. Deviation	Std. Error Mean
Evaluate <= 15 minutes	37	3.24	1.461	.240
> 15 minutes	59	3.47	1.443	.188

Independent Samples Test

	Levene's Test for Equality of variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Evaluate Equal variances assumed	.074	.786	-.761	94	.449	-.231	.304
Equal variances not assumed			-.759	75.889	.450	-.231	.305

Which of the following statements is true ?

The evaluation of students that spend > 15 minutes is not significantly different from the evaluation of students that spend ≤ 15 minutes because

- (a) The p-value of the t-test is .450.
- (b) The p-value of the t-test is .225.
- (c) The p-value of the t-test is .449.
- (d) The p-value of the t-test is .2245 (1×5)

4. Answer any six of the following :

- (i) What are the different levels of measurement specified in SPSS and briefly explain them ?
- (ii) A community activist believed that there was a relationship between membership in the police SWAT Team and prior military experience. He collected data from several police departments in an effort to support his belief. He found that there were 57 members of the SWAT team with prior military experience and 13 members with no prior military service. There were also 358 police personnel who had military experience but were not members of SWAT and another 413 with no military experience and not members of SWAT. Write the null and alternative hypotheses. Select the correct statistical method and write the procedure for conducting the test in SPSS.
- (iii) Name any three interfaces of SPSS and give their applications.
- (iv) Explain the functions of any three of the following functions used in SPSS: recode, crosstabs, descriptive, compute.
- (v) Write the procedure for constructing box-plot graph in SPSS and its interpretation.
- (vi) Briefly explain the concept of syntax in SPSS and also give its advantage ?

P.T.O.

(vii) Using 'recode function' in SPSS, write the procedure for constructing frequency distribution from the given raw data in case of

(a) equal class intervals

(b) unequal class intervals

(5×6)

Part – 2 (Based on MATLAB)

1. Attempt any ten of the following :

Consider the following to attempt the questions.

$$A = \begin{bmatrix} 2 & 8 & 0 \\ 0 & 5 & 5 \\ 7 & 9 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ -2 \\ 9 \end{bmatrix} \quad a = [-5 \quad 7 \quad -3]$$

- (i) What are the values of $A.*b$, $b.*a$, $a.*b'$ (Any Two) ?
- (ii) What is the command that extracts or isolates 2nd and 3rd rows of the matrix A.
- (iii) Delete the 2nd row of A, 2nd column of A and multiply the resultant matrices.
- (iv) What is the result of $A < b$?
- (v) What is the result of $\text{diag}(b)$ and $\text{diag}(a)$?
- (vi) Explain the commands `clc`, `mkdir`, `delete` with one example each.
- (vii) What is the purpose of `path(...)` ? Explain the command `path(path, 'c:\')`.
- (viii) What are system variables in MATLAB ? Give two examples.
- (ix) Explain in brief the 'rules' for naming variables in MATLAB. Give one example each of a valid and invalid Variable.
- (x) Fill in the blanks :
 - (a) `semilogx` is similar to `plot` function, except x-axis is in _____ scale.

- (b) A function file must have same name as the _____ name.
- (c) A script file must have an extension _____ .
- (xi) Given $x = [2,7,4,8,6,7,10]$, what will be the values of y in the following statements :
- (a) $y = x > 5$,
- (b) $y = \text{length}(\text{find}(x > 5))$ and
- (c) $x(\text{find}(x < 5)) = 9 * \text{ones}(\text{size}(\text{find}(x < 5)))$.
- (xii) For a matrix $C_{5 \times 100}$, what does (a) $C(:, 1:3: \text{end})$, (b) $C([1, 3], :) = []$ mean ? (3×10)

2. Attempt any **two** of the following :

- (i) Write a while loop that computes the sum $1 + x + x^2 + \dots + x^n$ for user input value of n and x .
- (ii) Explain the different data types available in MATLAB.
- (iii) Discuss the difference between `plot(...)` and `fplot(...)` functions available in MATLAB. (4×2)

3. Attempt any **one** of the following :

- (i) What will be the output of the following program ? Explain.

```
n=4;
y=[ ];
j=1;
for i=ones (n)
    j=j+1;
    i(j:n)=0;
    y=[y i];
end
y
```

- (ii) Write the functions, in general, you would use to compute 'pdf', 'cdf' and 'rnd' for the (a) Exponential distribution, (b) Gamma distribution and (c) Uniform distributions. (5×1)

4. Attempt any **one** of the following :

- (i) Generate two variates $Y_1 = U_1^{1/\alpha}$ and $Y_2 = U_2^{1/\beta}$, where the U_i are from the Uniform distribution. If $Y_1 + Y_2 \leq 1$, then

$$X = \frac{Y_1}{Y_1 + Y_2}$$

is from a Beta distribution with parameters α and β . Write a program to generate 100 r.v.s using this method and calculate mean, variance.

- (ii) Plot the functions $N(x; \mu, \sigma)$ and $\frac{d}{dx} N(x; \mu, \sigma)$ for $x = -3(0.1)3$ and $\mu = 10$, $\sigma = 2$ using `plotyy(...)`. (7×1)