This question paper contains 4 printed pages

Roll No ....

MCA 103: Statistical Techniques
Master of Computer Applications
Semester First, Nov/Dec-2017

Time: Three Hours **Maximum Marks: 70**

(Write your Roll No. on the top immediately on receipt of this Question paper)
Attempt **all** questions. All parts of a question must be answered together.
Use of Scientific Calculator is allowed.
Tabular values of various test-statistic are given at end.

Q1. Consider the following dataset.

Height (cm.)	Frequency
144.55-149.55	1
149.55-154.55	3
154.55-159.55	24
159.55-164.55	58
164.55-169.55	60
169.55-174.55	27
174.55-179.55	2
179.55-184.55	2

(a) Find Mean, Median and Mode for the given data. 3

~~(b) Find~~ the Kurtosis of the data and comment on it. 5

~~(c)~~ Find the second & fourth corrected moment after applying Sheppard's corrections. 4

Q2. (a) A certain city has three television stations. During prime time on Saturday nights, Channel 12 has 50 percent of the viewing audience, Channel 10 has the 30 percent viewing audience, and Channel 3 has the 20 percent of the viewing audience. Find the probability that among the eight television viewers in that city, randomly chosen on a Saturday night, five will be watching Channel 12, two will be watching Channel 10, and one will be watching Channel 3. 3

(b) The marks of 500 candidates in an examination are normally distributed with a mean of 45 marks and a standard deviation of 20 marks. If 20% of the candidates obtain a distinction by scoring x marks or more, estimate the approximate value of x. 3

(c) A roulette wheel in a Casino has the numbers 1 through 36, as well as 0 and 00. If anyone bet \$10 that an odd number comes up, he/she wins or loses \$10 according to whether or not that event occurs, respectively. If 100,000 games of roulette are played on a weekend, what is the average winning/losing amount for the Casino? **3**

(d) Given the joint probability density $f(x, y) = (2/3)*(x+2y)$ for $0<x<1$, $0<y<1$ and $f(x, y) = 0$ elsewhere, find the marginal densities of X and Y. **4**

Q3. (a) In 16 test runs the gasoline consumption of an experimental engine had a standard deviation of 2.2 gallons. Construct a 99% confidence interval for σ^2, which measures the true variability of the gasoline consumption of the engine. **3**

(b) Show that $Y = (1/6)*(X_1 + 2X_2 + 3X_3)$ is not a sufficient estimator of the Bernoulli parameter θ. **3**

$\frac{\bar{X}-\mu}{\sigma}$

(c) State the Central Limit Theorem. If a soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 millimeters and a standard deviation of 15 millimeters, what is the probability that the average amount dispensed in a random sample of size 36 is at least 204 millimeters? **4**

$\chi^2_{\alpha/2,\nu}$ $\frac{(n-1)S^2}{\sigma^2}$

Q4. Consider the following data table for computations.

Hours Studied (x)	Test Score (y)
4	31
9	58
10	65
14	73
4	37
7	44
12	60
22	91
1	21
17	84

(a) Compute the sample correlation coefficient. Test whether this coefficient is statistically significant at 0.05 level of significance. **5**

(b) Find the equation of the least squares line that approximate the regression of the test scores on the number of hours studied. Also, predict the average score of a person who studied 16 hours for the test. **5**

5.74×10^{-3}

Reject 17.

Q5. (a) For comparing the variability of the tensile strength of two kinds of structural steel, an experiment yielded the following results: $n_1 = 13$, $s_1^2 = 19.2$, $n_2 = 16$ and $s_2^2 = 3.5$, where the units of measurement are 1,000 pounds per square inch. Assuming that the measurements constitute independent random samples from two normal populations, test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative $\sigma_1^2 \neq \sigma_2^2$ at the 0.02 level of significance. 5

(b) For comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average 546 sq. ft. with a standard deviation of 31 sq. ft., whereas four 1-gallon cans of another brand cover on the average 492 sq. ft. with a standard deviation of 26 sq. ft. Assuming that the two populations sampled are normal and have equal variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 > 0$ at the 0.05 level of significance. 5

S=31

Q6. (a) The following data, in tons, are the amounts of sulfur oxides emitted by a large industrial plant in 40 days: 5

17	15	20	29	19	18	22	25	27	9
24	20	17	6	24	14	15	23	24	26
19	23	28	19	16	22	24	17	20	13
19	10	23	18	31	13	20	17	24	14

Use the *sign test* to test the null hypothesis $\mu = 21.5$ against the alternative hypothesis $\mu < 21.5$ at the 0.01 level of significance.

(b) The following are the final examination grades of samples from three groups of students who were taught German by three different methods: 5

First method: 94, 88, 91, 74, 87, 97

Second method: 85, 82, 79, 84, 61, 72, 80

Third method: 89, 67, 72, 76, 69

Use the *Kruskal-Wallis* test at the 0.05 level of significance to test the null hypothesis that the three methods are equally effective.

Q7. Based on the given output in R language for prediction of Number of Species of Tortoise found on an island, answer the following questions. 5

(a) What could be the possible R-Code for this output (Assuming "Tortoise" dataset has the historical data for modeling)?

(b) State with explanation, the significant predictors and relevant prediction model.

(c) Comment on the strength of the model and its statistical significance.

--------------------------X------------------------------

R-Code Output

```
Residuals:
     Min       1Q   Median       3Q      Max
-111.68   -34.90    -7.86    33.46   182.58

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.06822   19.15420    0.37   0.7154
Area         -0.02394    0.02242   -1.07   0.2963
Elevation     0.31946    0.05366    5.95  3.8e-06
Nearest       0.00914    1.05414    0.01   0.9932
Scruz        -0.24052    0.21540   -1.12   0.2752
Adjacent     -0.07480    0.01770   -4.23   0.0003

Residual standard error: 61 on 24 degrees of freedom
Multiple R-Squared: 0.766,      Adjusted R-squared: 0.717
F-statistic: 15.7 on 5 and 24 degrees of freedom,      p-value: 6.84e-07
```

Statistical Table Values

$t_{0.05,6}$	1.943	$U_{0.10}$	24	$\chi^2_{0.005,15}$	32.801
$t_{0.05,7}$	1.895	$T_{0.05,14}$	21	$\chi^2_{0.005,16}$	34.267
$t_{0.05,8}$	1.860	$f_{0.01,12,15}$	3.67	$\chi^2_{0.995,15}$	4.601
$t_{0.05,9}$	1.833	$f_{0.01,15,12}$	4.01	$\chi^2_{0.995,16}$	5.142
$t_{0.05,10}$	1.812	$f_{0.05,12,15}$	2.48	$\chi^2_{0.05,2}$	5.991
$t_{0.025,6}$	2.447	$f_{0.05,15,12}$	2.62	$\chi^2_{0.025,2}$	7.378
$t_{0.025,7}$	2.365	$-Z_{0.01}$	-2.33	$\chi^2_{0.05,3}$	7.815
$t_{0.025,8}$	2.306	$Z_{0.01}$	2.33	$\chi^2_{0.025,3}$	9.348
$t_{0.025,9}$	2.262	$Z_{0.025}$	1.96	$-Z_{0.05}$	-1.645
$t_{0.025,10}$	2.228	$Z_{0.005}$	2.575	$Z_{0.05}$	1.645

$\frac{n}{2}+1$

$\frac{n}{2}$th

Z-Table (Normal Distribution)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.6	0.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	0.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	0.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	0.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389

Z-Table (Standard Normal Distribution)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	0.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545