Register No:
Date: 5-1-21

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE
I SEMESTER EXAMINATION - JANUARY 2021
M.SC IN BIG DATA ANALYTICS
BDA 1120: BASIC STATISTICAL METHODS

TIME: 2 ½ HRS **MAX MARKS 70**

THIS QUESTION PAPER CONTAINS FOUR PRINTED PAGES AND ONE PART
LAST TWO PAGES INCLUDES CHI-SQUARE TABLES.
SCIENTIFIC CALCULATORS ARE ALLOWED
STUDENTS ARE ALLOWED TO USE SCIENTIFIC CALCULATORS

ANSWER ANY SEVEN QUESTIONS **7 X10 = 70**

1. The ages of seven policyholders in a portfolio of insurance policies are as follows:
   39 34 26 41 70 34 28
   (i) Find the median age of the policyholders in this portfolio.
   (ii) Another policyholder aged 41 years is added to the portfolio. Find the median age of policyholders in the portfolio.
   (iii) Why would the mean be a poor measure of central tendency for these data?

2. i. Given the following set of data, what is twice the interquartile range?
   25,32,49,21,37,43,27,45,31

   ii. Prepare the boxplot.

   iii. Given that for the water leakage data:
   $n = 100$, $(xi\text{-}\ xbar)2 = 856{,}934.91$, $(xi\text{-}\ xbar)3 = 11{,}949{,}848.3946$

   Calculate the:

   (a) skewness (b) coefficient of skewness.

3. The table below shows the numbers of births during one month at a particular hospital classified according to whether a particular medical characteristic was or wasn't present during childbirth. Determine whether the presence of this characteristic is dependent on the age of the mother.

Age of mother	20	21-25	26-30	31-35	36+	Total
Characteristic Present	10	12	9	4	3	38
Characteristic Absent	5	51	38	25	5	124
Total	15	63	47	29	8	162

Critical value χ^2_4 at upper 0.1% is 18.47.

4. A sample of ten claims and corresponding payments on settlement for household policies is taken from the business of an insurance company.
The amounts, in units of £100, are as follows:

   Claim x: 2.10 2.40 2.50 3.20 3.60 3.80 4.10 4.20 4.50 5.00

   Payment y: 2.18 2.06 2.54 2.61 3.67 3.25 4.02 3.71 4.38 4.45

   i. If you had to fit regression line by considering y as dependent variable and x as independent variable, estimate the regression equation.
   ii. Calculate R^2.

5. Discuss usefulness of probability distribution by giving an example.

6. If X is a random variable, prove $V(X) = E(X^2) - (E(X))^2$.

7. If $X \sim N(\mu, \sigma^2)$ what is the distribution of $\bar{X}$?

8. Two random variables X and Y have the following discrete joint distribution:

		Y		
X		10	20	30
	1	0.2	0.2	0.1
	2	0.2	0.3	0

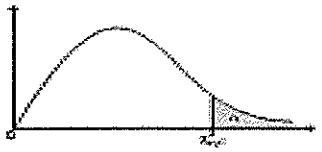
   Calculate $E(Y|X=1)$.

9. Let X and Y have joint density function given by:

   $f(X, Y) = 3/5 \times (x + y)$ when $0 < x < 1$, $0 < y < 2$

   Determine the conditional expectation $E[Y|X=x]$.

Table 7a Upper Critical Values of Chi-Square Distribution with ν Degrees of Freedom

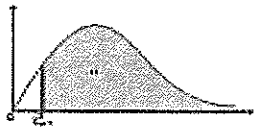


For selected probabilities α, the table shows the values $\chi^2_{\alpha,\nu}$ such that $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$, where χ^2 is a chi-square random variable with ν degrees of freedom. For example, the probability is 0.05 that a chi-square random variable with 10 degrees of freedom is greater than 18.307.

	Probability of Exceeding the Critical Value				
ν	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315
21	29.615	32.671	35.479	38.932	46.797
22	30.813	33.924	36.781	40.289	48.268
23	32.007	35.172	38.076	41.638	49.728
24	33.196	36.415	39.364	42.980	51.179
25	34.382	37.652	40.646	44.314	52.620
26	35.563	38.885	41.923	45.642	54.052
27	36.741	40.113	43.195	46.963	55.476
28	37.916	41.337	44.461	48.278	56.892
29	39.087	42.557	45.722	49.588	58.301
30	40.256	43.773	46.979	50.892	59.703
40	51.805	55.758	59.342	63.691	73.402
50	63.167	67.505	71.420	76.154	86.661
60	74.397	79.082	83.298	88.379	99.607
70	85.527	90.531	95.023	101.425	112.317
80	96.578	101.879	106.629	113.329	124.839
90	107.565	113.145	118.136	124.116	137.208
100	118.498	124.342	129.561	135.807	149.449

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, September 2011.

Table 7D Lower Critical Values of Chi-Square Distribution with ν Degrees of Freedom



For selected probabilities q (or α), these entries are the values χ^2_q such that $P(\chi^2 \le \chi^2_q) = q$, where χ^2 is a chi-square random variable with ν degrees of freedom. For example, the probability is 0.90 that a chi-square variable with 10 degrees of freedom is greater than 4.865.

	Probability Less Than the Critical Value				
ν	0.01	0.05	0.075	0.99	0.999
1	.016	.004	.001	.000	.000
2	.211	.103	.051	.020	.002
3	.584	.352	.216	.115	.024
4	1.064	.711	.484	.297	.091
5	*1.610	1.145	.831	.554	.210
6	2.204	1.635	1.237	.872	.381
7	2.833	2.167	1.690	1.239	.598
8	3.490	2.733	2.180	1.646	.857
9	4.168	3.325	2.700	2.088	1.152
10	4.865	3.940	3.247	2.558	1.479
11	5.578	4.575	3.816	3.053	1.834
12	6.304	5.226	4.404	3.571	2.214
13	7.042	5.892	5.009	4.107	2.617
14	7.790	6.571	5.629	4.660	3.041
15	8.547	7.261	6.262	5.229	3.483
16	9.312	7.962	6.908	5.812	3.942
17	10.085	8.672	7.564	6.408	4.416
18	10.865	9.390	8.231	7.015	4.905
19	11.651	10.117	8.907	7.633	5.407
20	12.443	10.851	9.591	8.260	5.921
21	13.240	11.591	10.283	8.897	6.447
22	14.041	12.338	10.982	9.542	6.983
23	14.848	13.091	11.689	10.196	7.529
24	15.659	13.848	12.401	10.856	8.085
25	16.473	14.611	13.120	11.524	8.649
26	17.292	15.379	13.844	12.198	9.222
27	18.114	16.151	14.573	12.879	9.803
28	18.939	16.928	15.308	13.565	10.391
29	19.768	17.708	16.047	14.256	10.986
30	20.599	18.493	16.791	14.953	11.588
40	29.051	26.509	24.433	22.164	17.916
50	37.689	34.764	32.357	29.707	24.674
60	46.459	43.188	40.482	37.485	31.738
70	55.329	51.739	48.758	45.442	39.036
80	64.278	60.391	57.153	53.540	46.521
90	73.291	69.126	65.647	61.754	54.155
100	82.358	77.929	74.222	70.065	61.918

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, September 2011.

BDA-1120-A-20