Name: **Enrolment No:**	UPES UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, May 2021

Course: Advanced Statistics **Semester: IV**
Program: BBA Analytics and Big Data **Time: 03 Hours**
Course code: DSQT 2004 **Max. Marks: 100**
Instructions: Z table is given at last.

SECTION A

	Attempt all Questions	**Marks**	**CO**
	Select the most appropriate answer.	**6 X 5=30**	**CO_1**
1.	The correlation coefficient for X and Y is known to be zero. We then can conclude that: (a) X and Y have standard distributions (b) the variances of X and Y are equal (c) there exists no relationship between X and Y (d) there exists no linear relationship between X and Y		
2.	What is the probability of picking a card that was red or black?		
3.	Which one of the following is non-random sampling method (a) Simple random sampling (b) Cluster Sampling (c) Systematic Sampling (d) Purposive Sampling		
4.	Once a week a merchandiser replenishes the stocks of a particular product brand in six stores for which she is responsible. Experience has shown that there is a one-in five chance that a given store will run out of stock before the merchandiser's weekly visit. (a) Which probability distribution is appropriate in this problem? Is this probability density function?		
5.	If X is a random variable follows standard normal distribution then its (a) Mean=0, Variance=0 (b) Mean=0, Variance=1 (c) Mean=1, Variance=0 (a) None of these		

6.	The probability that a ticketless traveler is caught during trip is 0.4. If the traveler makes 10 trips, the probability that he/she will be caught during at least one of the trips is : (a) $1-(0.9)^4$ (b) $(1-0.9)^4$ (c) $1-(1-0.9)^4$ (d) None of these		
	SECTION B		
Q	**Attempt all the questions**	**10X 5=50**	
1.	(a) Explain with an example the difference between correlation and regression. (b) What the requirement of calculating the partial correlation coefficient? Explain with an example.		**CO_2**
2.	**Name the sampling method used in each of the following situations:** (a) A woman in the airport is handing out questionnaires to travellers asking them to evaluate the airport's service. She does not ask travellers who are hurrying through the airport with their hands full of luggage, but instead asks all travellers who are sitting near gates and not taking naps while they wait. (b) A teacher wants to know if her students are doing homework, so she randomly selects rows two and five and then calls on all students in row two and all students in row five to present the solutions to homework problems to the class. (c) The marketing manager for an electronics chain store wants information about the ages of its customers. Over the next two weeks, at each store location, 100 randomly selected customers are given questionnaires to fill out asking for information about age, as well as about other variables of interest. (d) The librarian at a public library wants to determine what proportion of the library users are children. The librarian has a tally sheet on which she marks whether books are checked out by an adult or a child. She records this data for every fourth patron who checks out books. (e) A political party wants to know the reaction of voters to a debate between the candidates. The day after the debate, the party's polling staff calls 1,200 randomly selected phone numbers. If a registered voter answers the phone or is available to come to the phone that registered voter is asked whom he or she intends to vote for and whether the debate changed his or her opinion of the candidates.		**CO_2**
3.	A web-based travel agency was not working good in the market after discussion with several experts agency decided to use its website to market its travel products (holiday packages). The agency receives an average of five web-based enquiries per day for its different travel products. (a) What is the probability that, on a given day, the agency will receive only three web based enquiries for its travel products? (b) What is the probability that, on a given day, the travel agency will receive *at most* two web-based enquiries for travel packages? (c) What is the probability that the travel agency will receive *more than four* web-based enquiries for travel packages on a given day?		**CO_3**

	(d) What is the probability that the travel agency will receive *more than four* web-based enquiries for travel packages in any *two-day* period? **OR** A recent survey by a local municipality established that daily water usage by its households is normally distributed with a mean of 220 litres and a standard deviation of 45 litres. **Use standard normal *z*-tables to answer the following questions:** (a) What percentage of households is likely to use more than 300 liters of water per day? (b) What is the probability of finding a household that uses less than 100 liters of water per day? (c) What is the most amount of water used per day by the lowest-consuming 15% of households? (d) The municipality plans to implement a differential tariff policy to charge households that use more than a certain volume of water per day a higher rate per litre. If the municipality wants no more than 20% of households to pay this higher rate per litre, how much water per day.		
4.	Joey manipulates a die to increase his chances of winning a board game against his friends. In each round, a die is rolled and larger numbers are generally an advantage. Consider the random variable X denoting the outcome of the rolled die and the respective probabilities $P(X = 1 = 2 = 3 = 5) = 1/9$, $P(X = 4) = 2/9$, and $P(X = 6) = 3/9$. (a) Is distribution of random variable X a probability distribution? (b) Calculate and interpret the expectation and variance of X.		**CO_3**
5.	Telkom offers a range of telecommunication services to small businesses. A small printing business has used the services of *TalkPlus* (a value-added telephone service), *SmartAccess* (an advertising service) and *ISDN* (an internet connection) for the past three years. Their annual usage and the unit price of each service are given in the following table: (a) Calculate the *price relatives* for 2010 and 2011 for the ISDN service. Use 2009 as the base period. Interpret the meaning of each of these indexes. (b) Use Laspeyres weighted aggregates method (with 2009 as the base period) to calculate the *composite price indexes* for 2010 and 2011. By what percentage, on average, has the cost of telecommunications services for this printing company changed in 2010 and 2011 relative to 2009?		**CO_2**

Telkom services	2009		2010		2011	
	Unit price (cents/call)	Quantity (100s calls)	Unit price (cents/call)	Quantity (100s calls)	Unit price (cents/call)	Quantity (100s calls)
TalkPlus	65	14	70	18	55	17
SmartAccess	35	27	40	29	45	24
ISDN	50	16	45	22	40	32

(c) Calculate the Paasche composite quantity indexes for 2010 and 2011, using 2009 as the base period.

(d) Has the printing company's overall usage of the different telecommunication services changed significantly in 2010 and 2011 relative to 2009?

SECTION-C

Q	**Attempt the question :**	**20 X 1 =20**	**CO_4**

1. A plastics moulding manufacturer is evaluating factors that could affect the level of % *wastage* from an injection moulding process. He identified three possible 'drivers', namely *operator dexterity* (0 = min score, 30 = max score), *machine speed* (rpm), and *plastic viscosity* (PA.s). A *regression model* was built to examine the possible influence of these independent variables (drivers) on the % *wastage* per shift measure. A random sample of 31 shifts was selected and the above data recorded. The production manager wishes to identify which of these factors are significant 'drivers' of plastic wastage. The output of the regression analysis at the 5% level of significance are also given.

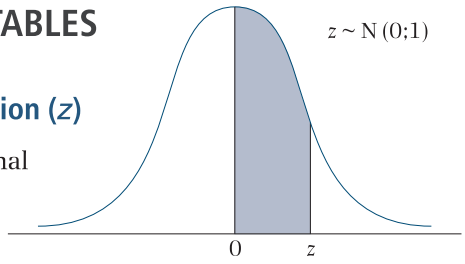
SUMMARY OUTPUT

Regression Statistics						
Multiple R	0.8061					
R Square	0.6498					
Adjusted R Square	0.6109					
Standard Error	0.5160					
Observations	31					
ANOVA						
	df	*SS*	*MS*	*F-stat*	*p-value*	
Regression	3	13.3384	4.4461	16.6992	2.466E-06	
Residual	27	7.1887	0.2662			
Total	30	20.5271				
	Coefficients	*Standard Error*	*t-stat*	*p-value*	*Lower 95%*	*Upper 95%*
Intercept	1.8179	1.2189	1.4914	0.1474	-0.6830	4.3188
Dexterity	-0.1112	0.0286	-3.8816	0.0006	-0.1700	-0.0524
Speed	0.0173	0.0047	3.6770	0.0010	0.0077	0.0270
Viscosity	1.9189	1.2581	1.5252	0.1388	-0.6625	4.5004

Explain the meaning of each cell of above given table with respect to the problem and variables under consideration. (for example, what do you understand by the value of Multiple R =0.8061 with respect to variables namely *operator dexterity* (0 = min score, 30 = max score), *machine speed* (rpm), and *plastic viscosity* (PA.s))

APPENDIX 1: LIST OF STATISTICAL TABLES

TABLE 1 The standard normal distribution (z)



This table gives the area under the standard normal curve between 0 and z

$P[0 < Z < z]$

In *Excel* (2013): use NORM.S.DIST(z, cumulative = True) to find $P(-\infty < Z < z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2793	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3557	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49897	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49991	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998