National University of Computer & Emerging Sciences, Karachi
Spring-2021 CS-Department
Final Examination
15th June 2021, 12:30 pm – 3:30 pm

Course Code: MT205	Course Name: Probability & Statistics
Instructor Name Names: Mr. Osama Bin Ajaz, Dr. Fahad Riaz, Miss Amber, Miss Javeria Iftikhar	
Student Roll No:	Section No:

Instructions:

- Return the question paper.
- Read each question completely before answering it. There are **8 questions and 2 pages**
- In case of any ambiguity, you may make assumptions. But your assumption should not contradict any statement in the question paper.
- All the answers must be solved according to the sequence given in the question paper.
- This paper is subjective

Time: 3 hrs. **Max Points**: 70

Q1) A study was conducted to examine the influence of the drug succinylcholine on the circulation levels of androgens in the blood. Blood samples were taken from wild, free-ranging deer immediately after they had received an intramuscular injection (X) of succinylcholine administered using darts and a capture gun. A second blood sample (Y) was obtained from each deer 30 minutes after the first sample, after which the deer was released. The levels of androgens at time of capture and 30 minutes later, measured in nanograms per milliliter (ng/mL), for 15 deer are given in the table below. Assuming that the populations of androgen levels at time of injection and 30 minutes later are normally distributed, test at the 0.05 level of significance whether the androgen concentrations are altered after 30 minutes. (critical value = ± 2.145) **[6]**

X	2.76	5.18	2.68	3.05	4.10	7.05	6.60	4.79	7.39	7.30	11.78	3.90	26.00	67.48	17.04
Y	7.02	3.10	5.44	3.99	5.21	10.26	13.91	18.53	7.91	4.85	11.10	3.74	94.03	94.03	41.70

Q2) A Zoologist conducted a research to determine if there is a significant difference in the density of organisms at two different sewage treatment plants. At 5% level of significance, can it be concluded that the average densities are equal for the two plants? Assume that the observations come from normal populations with **different variances**. The necessary data shown below give the density measurements, in number of organisms per square meter, at the two treatment plants. (CV ± 2.093) **[6]**

Plant - I	5030	13700	107300	11400	860	2200	4250	15040
	4980	11910	8130	26850	17660	22800	1130	1690
Plant - II	2800	4670	6890	7720	7030	7330	2810	1330
	3320	1230	2130	2190				

Q3) The number of grams of carbohydrates contained in 1-ounce servings of randomly selected chocolate and non chocolate candy is listed here. Is there sufficient evidence to conclude that the difference in the means is significant? Also show a 90% confidence interval. Use $\alpha = 0.10$ and assume population variances to be equal. (critical value ±1.717) **[8]**

Chocolate:	29	25	17	36	41	25	32	29
	38	34	24	27	29			
Nonchocolate:	41	41	37	29	30	38	39	10
	29	55	29					

OR

The FBI reported the mean value lost to mobile snatching was $417 in 2009. For last year, 12 randomly selected mobile-snatching offenses yielded the following values lost, to the nearest dollar: 364, 488, 314, 428, 324, 252, 521, 436, 499, 430, 320, and 472. Apply appropriate hypothesis testing procedure at the 5% level of significance, to test whether last year's mean value lost to mobile snatching has decreased from the 2009 mean. Also construct a 95% confidence interval. (critical value -1.79) **[8]**

Q4) A researcher wishes to see if the average length of the major rivers in the United States is the same as the average length of the major rivers in Europe. The data (in miles) of a sample of rivers are shown. At α=0.01, is there enough evidence to reject the claim? Assume $\sigma_1 = 450$, $\sigma_2 = 474$, $\bar{x}_1 = 662.611$, $\bar{x}_2 = 758.875$, $n_1=36$, and $n_2 = 32$ (critical value = ± 2.58)

(a) Use the appropriate technique to test the claim. **[6]**

(b) Find the 99% confidence interval for the difference between the means, and confirm your result in part (a) that either there is enough evidence to support the claim or not? **[2]**

Q5) Suppose 25 colored balls are distributed in three bags, which are identical in appearance as shown to the right. A bag is selected at random from three bags. From which a ball is drawn at random out of 25 balls. Given that the ball is yellow, what is the probability that bag 2 was selected. **[6]**

	Bags		
	1	2	3
GREEN	1	3	4
YELLOW	2	2	3
RED	4	5	1

Q6) According to the International Data Base of the U.S. Census Bureau, the population of the world grows according to the table given below.

year	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015
population (million)	2558	2782	3043	3350	3712	4089	4451	4855	5287	5700	6090	6474	6864

a) Plot a scatter diagram and find correlation between the years and population. **[6]**

b) Find the equation of the regression line to predict the World's population in 2020 and 2025. **[6]**

Q7) You are given the data for the calibration constants experimentally for samples of compounds in each of four compound groups to the right. At the 5% significance level, do the data provide sufficient evidence to conclude that a difference exists in mean calibration constant among the four compound groups?(critical value 3.027) **[10]**

Esters	Alcohols	Aliphatic hydrocarbons	Aromatic hydrocarbons
0.185	0.185	0.23	0.166
0.155	0.16	0.184	0.144
0.131	0.142	0.16	0.117
0.103	0.122	0.132	0.072
0.064	0.117	0.1	
	0.115	0.064	
	0.11		
	0.095		
	0.085		
	0.075		

Q8) The average monthly electric power consumption (*Y*) at a certain manufacturing plant is considered to be linearly dependent on the average ambient temperature (x_1) and the number of working days in a month (x_2). Consider the one-year monthly data given in the table below. Determine the least-square estimates of the associated linear regression coefficients and calculate multiple correlation**. [14]**

y	210	206	260	244	271	285	270	265	234	241	258	230
$\mathbf{X_1}$	20	21	24	25	24	26	25	25	24	25	25	23
$\mathbf{X_2}$	23	21	24	25	24	26	25	25	24	25	25	23

Solution Manual

Q1)

***Solution*:** Let μ_1 and μ_2 be the average androgen concentration at the time of injection and 30 minutes later, respectively. We proceed as follows:

1. H_0: $\mu_1 = \mu_2$ or $\mu_D = \mu_1 - \mu_2 = 0$.
2. H_1: $\mu_1 \neq \mu_2$ or $\mu_D = \mu_1 - \mu_2 \neq 0$.
3. $\alpha = 0.05$.
4. Critical region: $t < -2.145$ and $t > 2.145$, where $t = \frac{\overline{d} - d_0}{s_D/\sqrt{n}}$ with $v = 14$ degrees of freedom.
5. Computations: The sample mean and standard deviation for the d_i are

$$\overline{d} = 9.848 \qquad \text{and} \qquad s_d = 18.474.$$

Therefore,

$$t = \frac{9.848 - 0}{18.474/\sqrt{15}} = 2.06.$$

6. Though the t-statistic is not significant at the 0.05 level, from Table A.4,

$$P = P(|T| > 2.06) \approx 0.06.$$

As a result, there is some evidence that there is a difference in mean circulating levels of androgen.

Q2)

10.41 The hypotheses are

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

$\alpha = 0.05$.
Degrees of freedom is calculated as

$$v = \frac{(7874.329^2/16 + 2479/503^2/12)^2}{(7874.329^2/16)^2/15 + (2479.503^2/12)^2/11} = 19 \text{ degrees of freedom.}$$

Critical regions $t < -2.093$ or $t > 2.093$.
Computation: $t = \frac{9897.500 - 4120.833}{\sqrt{7874.329^2/16 + 2479.503^2/12}} = 2.76$.
Decision: Reject H_0 and conclude that $\mu_1 \neq \mu_2$.

Q3)

t-Test: Two-Sample Assuming Equal Variances

	chocolate	*nonchocolate*
Mean	29.69230769	34.36363636
Variance	42.23076923	125.4545455
Observations	13	11
Pooled Variance	80.05975842	
Hypothesized Mean Difference	0	
df	22	
t Stat	-1.274370429	
P(T<=t) one-tail	0.107914675	
t Critical one-tail	1.321236742	
P(T<=t) two-tail	0.215829351	
t Critical two-tail	1.717144374	

OR

T-statistic = - 0.51865, t-critical = -1.79, decision = Do not reject Ho

Q4) **z = - 0.86,** p-value = 2(0.1949) = 0.3898 Do not reject Ho

Q5) P(Y) = 0.2869, P(B2|Y) = 0.232

Q6)

$\bar{x} = 1985$, $\sum x = 25805$, $\sum x^2 = 5122745$, $n = 13$, $\bar{y} = 4558.077$, $\sum y = 59255$, $\sum y^2 = 295204169$

$$\sum xy = 117958425$$

Regression Equation: $\hat{y} = -142571.8681 + 74.12088x$

Correlation: 0.9976

2020: $\hat{y}(2020) = 7152.3095 \approx 7152$

2025: $\hat{y}(2025) = 7522.9139 \approx 7523$

Q7)

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.002318	3	0.00077263	0.389408	0.761817	3.072467
Within Groups	0.041666	21	0.001984112			
Total	0.043984	24				

Q8)

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.854242448
R Square	0.72973016
	Coefficients
Intercept	-77.38
X1	8.013333333
X2	5.526666667