VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009748**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **09/05/2024**	Slot: **A1+TA1**	Session: **FN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: Yes
2. Reference tables permitted: YES (Statistical Tables: Z, t, Chi-square, F with seal and signature of CoE are allowed)

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	Calculate the median of a flowing distribution.	**10**	**1**	**1**

Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Number of students	7	10	10	20	20	15	8

Q2.	A factory production line is manufacturing bolts using three machines, A, B and C. Of the total output, machine A is responsible for 25%, machine B for 35%, and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C. A bolt is chosen at random from the production line and found to be defective. What is the probability that it came from (a) machine A, (b) machine B.	**10**	**2**	**1**
Q3.	(a) A continuous random variable X can assume non-zero value only between $x = 2$ and $x = 5$ and has the probability density function (p.d.f.) $f(x) = k(1 + x)$. Find the value of k and $P(X) < 4$. (b) A test engineer discovered that the cumulative distribution function of the lifetime of an equipment (in years) is given by $F_X(x) = 1 - e^{-\frac{x}{5}}$, $x \geq 0$. What is the expected lifetime of the equipment?	**10**	**3**	**2**
Q4.	The top 5% of applicants (as measured by GRE scores) will receive scholarships. If GRE score is normally distributed with mean 500 and standard deviation 100, what is the cut-off for GRE to be qualified for a scholarship?	**10**	**3**	**2**
Q5.	From the following data, compute the rank correlation coefficient.	**10**	**5**	**2**

X	82	68	75	61	68	73	85	68
Y	81	71	71	68	62	69	80	70

Q6. After investigation, the demand for automobiles in a city depends mainly, if not entirely, upon the number of families residing in that city. Below are the given figures for the sales of automobiles in the five cities for the year 2023 and the number of families residing in those cities. **10 5 3**

City	No. of Families (in lakhs)	Sale of automobiles (in 1000)
Belagavi	70	25.2
Bangalore	75	28.6
Hubli	80	30.2
Kalaburagi	60	22.3
Mangalore	90	35.4

Fit a linear regression equation by the least square method and estimate the sales for the year 2024 for the city Belagavi which is estimated to have 100 lakh families assuming that the same relationship holds.

Q7. North Carolina University looked at factors that affected the success of students in a required chemical engineering course. Students must receive a C or better in the course to continue as chemical engineering majors, so we consider a grade of C or better as a success. Is there a difference in the proportions of male and female students who succeeded in the course at 5% level of significance? The data showed that 23 of the 34 women and 60 of the 89 men succeeded. **10 4 3**

Q8. Suppose the life expectancy of Seattleites has a population that is normally distributed with a standard deviation of 1. You go out and sample 45 Seattleites from this population and obtain a mean life expectancy of 88.51 and a standard deviation of 1.0815. Using 5% level of significance, is this observed mean significantly different than a life expectancy of 89? **10 4 3**

Q9. A group of 5 patients treated with medicine A with a mean weight of 44 kg and a variance of 82.5 kg. The second group of 7 patients from the same hospital were treated with medicine B with a mean weight of 57 kg and a variance of 154.33 kg. Find whether there is any difference between the mean weight of the patients taking medicine A and medicine B at a 5% level of significance? (Assume population variances are equal.) **10 6 4**

Q10. Two sources of raw materials are under consideration by a bulb manufacturing company. Both sources A and B seem to have similar characteristics but the companies are not sure about their respective uniformity. A sample of 12 lots from source A yields a variance of 125 and a sample of 10 lots from source B yields a variance of 112. Is it likely that the variance of source A significantly differs from the variance of source B at 5% level of significance? **10 4 4**

Q11. A researcher collected 400 records from a company of which day of the week employees called in sick to work. Can the researcher conclude that the proportion of employees who call in sick is not the same for each day of the week? Design and conduct a Chi-square test for 1% level of significance. **10 4 4**

Day of Week	**Frequency**
Monday	95
Tuesday	65
Wednesday	60
Thursday	80
Friday	100
Total	**400**

Q12. The variance of a certain article produced by a machine is 7.2 over a long period. A random sample of 20 articles gave a variance 8. Is it justifiable to conclude that variance has increased at 5% level of significance assuming that the measurement of the article is normally distributed? **10 6 4**

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009786**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **10/05/2024**	Slot: **A2+TA2**	Session: **FN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. **"*fx* series" - non-Programmable calculator are permitted.**
2. **Students are allowed to use the statistical tables endorsed by COE.**
3. **Assume data wherever necessary.**
4. **Any assumptions made should be clearly stated.**

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks) **Marks** **CO** **BL**

Q1. In a biology class, students were asked to record the number of hours they spent studying for an upcoming exam. The data collected is as follows: **10** **1** **1**

Study Hours	Frequency
1-5	4
6-10	8
11-15	12
16-20	10
21-25	6

Find the mean, median of the number of study hours spent by the students.

Q2. Suppose that the four inspectors at a film factory are supposed to stamp the expiration date on each package of film at the end of the assembly line. John, who stamps 20% of the packages, fails to stamp the expiration date once in every 200 packages; Tom, who stamps 60% of the packages, fails to stamp the expiration date once in every 100 packages; Jeff, who stamps 15% of the packages, fails to stamp the expiration date once in every 90 packages; and Pat, who stamps 5% of the packages, fails to stamp the expiration date once in every 200 packages. If a customer complains that her package of film does not show the expiration date, what is the probability that it was inspected by John? **10** **4** **3**

Q3. **10** **3** **2**

Consider the density function $f(x) = \begin{cases} kx^{1/2} & 0 < x < 1 \\ 0 & elsewhere. \end{cases}$

a. Evaluate k.

b. Find cumulative distribution function $f(x)$ and use it to evaluate $P(0.3<X<0.6)$.

Q4. On average, 3 traffic accidents per month occur at a certain crossing. What is the probability that in any given month at this intersection **10 2 2**

a. exactly 5 accidents will occur?
b. fewer than 3 accidents will occur?
c. at least 2 accidents will occur?

Q5. Eight students were surveyed regarding the number of hours they studied per week and their corresponding scores on a test. The data is as follows: **10 3 5**

Student	Hours Studied (x)	Test Score (y)
1	5	60
2	7	65
3	3	55
4	6	70
5	4	58
6	2	50
7	8	75
8	5	62

Calculate the Karl Pearson coefficient (Pearson correlation coefficient) for these data points and interpret the Pearson coefficient in the context of this study.

Q6. An article in the Journal of the Environmental Engineering Division ["Least Squares Estimates of BOD parameters" (1980, Vol. 106, pp.1197-1202)] took a sample from the Holston River below Kingsport. Tennessee, during August 1977. The biochemical oxygen demand (BOD) test is conducted over a period of times in days. The result data are shown below: **10 3 3**

Time (days)	1	2	4	6	8	10	12	14
BOD (mg/litre)	0.6	0.7	1.5	1.9	2.1	2.6	2.9	3.7

a. Fit the simple linear regression model using the method of least squares to find the true regression line of BOD on time.
b. What is the estimate of expected BOD level when the time is 9 days?

Q7. A new rocket-launching system is being considered for deployment of small, short-range rockets. The existing system has $p = 0.8$ as the probability of a successful launch. A sample of 40 experimental launches is made with the new system, and 34 are successful. Use 5% level of significance. Would you conclude that the new system is better? **10 6 3**

Q8. Seven students were given intensive coaching and 5 tests were conducted in a month. The score of first and 5^{th} tests are as following: **10 5 4**

Test 1: 52, 43, 52, 27, 36, 43, 61
Test 5: 63, 41, 62, 36, 31, 53, 70
Use paired t-test to determine if there is evidence that the score from first to fifth test show an improvement? Use 5% level of significance.

Q9. A telecom service provider claims that individual customers pay on an average Rs. 400 per month with standard deviation of Rs. 25. A random sample of 50 customers bills during a given month is taken with mean of Rs. 390 and standard deviation of Rs.15. What to say with respect to claim made by the service provider? Use 4% level of significance. **10 4 4**

Q10. A nutritionist wants to compare the effectiveness of three different diets (A, B, and C) in reducing some hormone levels. She randomly selected 15 participants and divided them equally into three groups. Each group was assigned one of the three diets. After 8 weeks, she recorded the hormone levels for each participant. The data is as follows: **10 6 5**

Diet A: 18, 17, 17, 18, 19
Diet B: 16, 15, 15, 16, 17
Diet C: 19, 19, 18, 20, 20
Perform a one-way ANOVA to determine if there is a significant difference in mean reduction of hormone levels between the three diets at a 5% level of significance.

Q11. A survey was conducted to study the relationship between gender and preference for smartphone operating systems among users. The data collected is shown in the table below: **10 6 5**

	iOS	Android	Windows	Symbian
Male	45	60	65	35
Female	55	40	45	30

Determine whether there is a significant association between gender and preference for smartphone operating systems by testing the independence of these two attributes at a 5% level of significance.

Q12. In a cooking competition, 800 contestants participated. After the event, it was discovered that 320 contestants failed to impress the judges, 270 secured a third-place finish, 190 secured second place, and the rest were awarded first place. The prevailing belief is that these rankings should follow a certain ratio: 4:3:2:1 for failing, third place, second place, and first place respectively. Test the hypothesis that the general opinion about the ranks is appropriate at 5% level of significance. **10 6 5**

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009803**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **11/05/2024**	Slot: **B1+TB1**	Session: **FN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: Yes
2. Reference tables permitted: Yes. Statistical tables Z, t, chi-square and F are allowed if they have COE seal and signature

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	Calculate the standard deviation from the following frequency distribution of marks obtained by 200 students of a school in an examination.	**10**	**1**	**2**

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students	5	12	30	45	50	37	21

Q2.	Two cities A and B of different states have rains on 60% and 75% days respectively during the monsoon. For the cities A and B, find the probability that on a certain monsoon day, a) Both the cities have rains, b) At least one city has rains, c) Only one city has rains. Note: The event of rains on a day in these two cities are independent.	**10**	**1**	**3**
Q3.	There are 2 black and 2 white balls in a box. Two balls are drawn without replacement from it. Obtain probability distribution of the number of white balls in the selected balls. Find its mean and variance.	**10**	**2**	**3**
Q4.	The maximum temperature of a city during summer follows a normal distribution. On a particular day, the probability that the maximum temperature of the city is more than 31 is 0.3085, whereas the probability that during some other day, the maximum temperature is less than 27 is 0.0668. Find the mean and standard deviation of the maximum temperature of the city.	**10**	**2**	**5**
Q5.	From the following information of weekly minimum temperature (in Celsius) and the sale (in hundred units) of heaters during a week in a city of North India for five weeks, calculate the Karl Pearson correlation coefficient between minimum temperature and sale of heaters and Interpret the results.	**10**	**3**	**2**

Minimum Temperature (Celsius) x	3	4	6	7	9
Demand of heaters (hundred units) y	16	15	14	11	9

Q6.	The monthly sale of different types of laptops (in hundred units) and its profit (in lakh Rs.) for the last 6 months for a company is given below.	**10**	**3**	**4**

Month	1	2	3	4	5	6
No. of Laptops sold (hundred units) x	5	7	5	12	8	3
Profit (Lakh Rs.) y	8	9	10	15	10	6

Obtain the regression line of Y on X. Also find the error in estimating y for x = 7.

Q7. Time magazine reported the result of a telephone poll of 800 adult Americans. The question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?" The results of the survey were: **10 4 5**

Non-Smokers	Smokers
$N_1 = 605$	$N_2 = 195$
$X_1 = 351$ said "yes"	$X_2 = 41$ said "yes"

Is there sufficient evidence at the $\alpha = 0.05$, to conclude that the two populations — smokers and non-smokers — differ significantly with respect to their opinions?

Q8. An ambulance service claims that it takes on the average 8.9 minutes to reach its destination in emergency calls. To check on this claim, the agency which licenses ambulance services has then timed on 50 emergency calls, getting a mean of 9.3 minutes with a standard deviation of 1.6 minutes. What can they conclude at 5% level of significance. **10 4 2**

Q9. The following table gives the scores (out of 15) of two batches of students in an examination. **10 5 5**

Batch I	6	7	9	2	13	3	4	8	7	11
Batch II	5	6	5	7	1	7	2	7		

Sample variance of Batch 1 is 12 and of Batch 2 is 5.4. Test at 1% level of significance the average performance of the students in Batch I and Batch II are equal where population variance is equal.

Q10. A test was given to five students taken at random from statistics background of three different Universities. The individual scores are given below: **10 5 4**
Test at 5% significance if the means are same using one-way ANOVA technique.

University	S_1	S_2	S_3	S_4	S_5
A	9	7	6	5	8
B	7	4	5	4	5
C	6	5	6	7	6

Q11. A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table: **10 6 4**

	High School	Bachelors	Masters	Ph.D.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Test whether gender and education level are dependent at 5% level of significance. In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

Q12. The following table shows the distribution of digits in numbers chosen at random from a telephone directory. **10 6 5**

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the occurrence of the digits in the directory are equal at 5% level of significance.
Note: Assume that every digit has equal probability of occurrence.

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009754**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **12/05/2024**	Slot: **B2+TB2**	Session: **FN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: yes
2. Reference tables permitted: yes

Z-table, t-table, F-table and Chi-square tables with sign of CoE are allowed.

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	1. If the median of the distribution given below is 28.5. Then find the value of x and y.	**10**	**1**	**1**

Class	0-10	10-20	20-30	30-40	40-50	50-60	Total
Frequency	5	x	20	15	y	5	60

(10M)

Q2.	2. Three urns are there containing white and black balls; first urn has 3 white and 2 black balls. Second urn has 2 white and 3 black balls and the third urn has 4 white and 1 black balls. Without any biasing, one urn is chosen from that one ball is chosen randomly which was white. Then find the probability that it came from the third urn. (10M)	**10**	**2**	**1**
Q3.	3. A pharmaceutical company states that a drug causes negative side effects in 3 of every 100 patients. To confirm this affirmation, another laboratory chooses 5 people at random who have consumed the drug. Then find the probability that (a) none of five patients experience side effects (b) at least two experience side effects (c) Compute the average number of patients that the laboratory should expect to experience side effects if they choose 100 patients at random. (10M)	**10**	**2**	**1**
Q4.	4. The Edwards theatre chain has studied its movie customers to determine how much money they spend on concessions. The study revealed that the spending distribution is approximately normally distributed with a mean of 4.11 dollars and a standard deviation of 1.37 dollars. Then find the percentage of customers will spend less than 3 dollars on concessions. (10M)	**10**	**3**	**1**
Q5.	5. We have data of income and expenditure of 11 workers of an organization in the following table.	**10**	**3**	**1**

Income	65	40	35	75	63	79	35	20	80	60	50
Expenditure	60	55	50	66	30	71	40	35	80	75	80

Find the Spearman rank correlation coefficient and interpret the given data.

Q6.	6. Sam found how many hours of sun shine (x) versus how many ice creams (y) were sold at the shop from Monday to Friday.	**10**	**4**	**1**

x	2	3	5	7	9
y	4	5	7	10	15

Find the equation of regression line y on x that best fits the data. (10M)

Q7. **10 5 1**

7. Smoking rate in a town in past was 21%. 100 samples were picked and found 14 are smokers. Has smoking habit changed at 5% level of significance? (10M)

Q8. **10 5 1**

8. A random sample of 64 bags of white cheddar popcorn weighed, on average, 5.23 ounces with a standard deviation of 0.24 ounce. Test the hypotheses that $\mu = 5 \cdot 5$ ounces against the alternate hypotheses $\mu < 5 \cdot 5$ ounces, at the 0.05 level of significance. (10M)

Q9. **10 5 1**

9. To find out whether a new serum will arrest leukaemia, 9 mice, all with an advanced stage of the disease, are selected. Five mice receive the treatment but 4 do not. Survival times, in years, from the time the experiment commenced are as follows.

Treatment	2.1	5.3	1.4	4.6	0.9
No treatment	1.9	0.5	2.8	3.1	

At the 0.05 level of significance, can the serum be said to be effective. Assume the two populations to be normally distributed with equal variances. (10M)

Q10. **10 5 1**

10. An experiment was conducted to compare the alcohol content of soy sauce on two different production lines. Production was monitored five times a day. The data are shown here.

Production line 1	0.48	0.39	0.42	0.52	0.4
Production line 2	0.38	0.37	0.39	0.41	0.38

Assume the populations are normal. It is suspected that production line 1 is not producing as consistently as production line 2 in terms of alcohol content. Test the hypotheses $\sigma_1 = \sigma_2$ against the alternative that $\sigma_1 \neq \sigma_2$ at 0.05 level of significance.

Q11. **10 6 1**

11. In a shop study, a set of data was collected to determine whether the proportion of defectives produced was the same for workers on the day, evening, and the night shifts. The data collected are shown in the following table.

Shift	Day	Evening	Night
Defectives	45	55	70
Non-Defectives	905	890	870

Use a 0.025 level of significance to determine if the proportion of defectives is the same for all the three shifts.

Q12. **10 6 1**

12. A packet consists of 130 ball pens. The distribution of the number of defective ball pens in each packet is given below.

x	0	1	2	3	4	5
f	61	14	10	17	15	13

Examine whether the Poisson distribution is appropriate for the above data at 5% level of significance.

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009844**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **14/05/2024**	Slot: **C1+TC1**	Session: **AN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: Yes
2. Reference tables permitted: Statistical tables Z, t, chi-square, F are allowed if they have COE seal and signature

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks) **Marks** **CO** **BL**

Q1. The following frequency distribution shows the price per share of the 30 companies in the Dow Jones Industrial Average **10** **1** **1**

Price per share	Number of Companies
$0-10	4
$10-20	5
$20-30	7
$30-40	3
$40-50	4
$50-60	4
$60-70	0
$70-80	2
$80-90	0
$90-100	1

Calculate the median.

Q2. A company produces 1,000 refrigerators a week at three plants. Plant A produces 350 refrigerators a week, plant B produces 250 refrigerators a week, and plant C produces 400 refrigerators a week. Production records indicate that 5% of the refrigerators produced at plant A will be defective, 3% of those produced at plant B will be defective, and 7% of those produced at plant C will be defective. All the refrigerators are shipped to a central warehouse. If a refrigerator at the warehouse is found to be defective, what is the probability that it was produced at plant C? **10** **1** **2**

Q3. Stella has 5 hens. The number of eggs she gets in a particular day is a random variable and its probability distribution function is given as **10** **2** **1**

X	0	1	2	3	4	5
p(x)	1/8	1/8	1/8	1/4	1/4	1/8

Find the Cumulative Distribution Function, mean and variance of this probability distribution.

Q4. Suppose there are 15 washing machine came in a service center for repairing. The probability of a washing machine gets repaired is 3/4 . Find the probability of **10** **3** **3**

(1) At most two washing machine gets repaired
(2) At least 2 washing machine gets repaired
(3) Exactly 5 washing machine gets repaired

Q5. The data below gives the marks obtained by 8 students in Mathematics (out of 30) and Physics (out of 40). Find the Karl Pearson's correlation coefficient of the given set of data. Interpret the value with the given data. **10 4 4**

Marks in Mathematics	20	23	8	29	14	11	11	20
Marks in Physics	30	35	21	33	33	26	22	31

Q6. A textile company, wanting to know the effect of temperature on the tearing strength of a fiber, obtained the data shown in the following table. **10 5 4**

Temperature (°C) x	20	22	25	35	18	29	31	16
Tearing Strength (g) y	1600	1700	2100	2500	1550	2600	2550	1100

Fit a straight line to the given data by the method of least squares and use it to predict the tearing strength one can expect when the temperature is 30°C.

Q7. A supplier of imported vernier calipers claims that 90% of their instruments have a precision of 0.999. Testing the null hypothesis p=0.90 against the alternative hypothesis p ≠ □0.90, what can we conclude at the level of significance α =0.10, if there were 665 calipers out of 700 with a precision of 0.999? **10 6 3**

Q8. An effective way to tap rubber is to cut a panel in the rubber tree's bark in vertical spirals. In a pilot process, an engineer measures the output of latex from such cuts. Eight cuts on different trees produced latex (in liters) in a week are 26.8, 32.5, 29.7, 24.6, 31.5, 39.8, 26.5, 19.9 with a mean 28.913 and standard deviation 5.962. Conduct a test of hypotheses with the intent of showing that the mean production is less than 36.2. Take $\alpha = 0.01$ and assume a normal population. **10 6 4**

Q9. An investigation of two types of bulldozers showed that 50 failures of one type of bulldozer took on an average 6.8 hours to repair with a standard deviation of 0.85 hours, while 50 failures of the other type of bulldozer took on an average 7.3 hours to repair with a standard deviation of 1.2 hours. Test the null hypothesis $\mu_1 - \mu_2 = 0$ (namely, the hypothesis that on an average, it takes an equal amount of time to repair either kind of bulldozer) against the alternative hypothesis $\mu_1 - \mu_2 \neq 0$ at the level of significance, $\alpha = 0.10$. **10 6 4**

Q10. The following are the numbers of mistakes made in 5 successive days for 3 technicians working for a photographic laboratory: **10 6 4**

Technician 1	Technician 2	Technician 3
5	17	9
12	12	11
9	15	6
8	14	14
11	17	10

Test at the level of significance $\alpha = 0.01$ whether the differences among the 3 sample means can be attributed to chance.

Q11. A Fortune study found that the variance in the number of vehicles owned or leased by subscribers to Fortune magazine is 0.94. Assume a sample of 12 subscribers to another magazine provided the following data on the number of vehicles owned or leased: 2, 1, 2, 0, 3, 2, 2, 1, 2, 1, 0, and 1. Test the hypothesis $H_0: \sigma^2 = 0.94$ to determine whether the variance in the number of vehicles owned or leased by subscribers of the other magazine differs from $\sigma^2 = 0.94$ for Fortune. At a 0.05 level of significance, what is your conclusion? **10 6 4**

Q12. The following is the distribution of the hourly number of trucks arriving at a company's warehouse. Fit a Poisson distribution to the data and test the goodness of fit at the 0.05 level of significance. **10 6 4**

Trucks arriving per hour	Frequency
0	52
1	151
2	130
3	102
4	45
5	12
6	5
7	1
8	2

The mean of this distribution is 2.02 and the recurrence relation of the expected frequency is given as, Expected Frequency at (x+1)=($\frac{\lambda}{x+1}$) Expected Frequency at x, where λ is the parameter of the Poisson distribution. Kindly consider the number of parameter estimated here as 1.

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **MAT1011**	Course Title: **Applied Statistics**	
Question Paper ID: **009710**	Exam Type: **CLOSED BOOK**	School: **SAS**
Date: **15/05/2024**	Slot: **C2+TC2**	Session: **FN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: No
2. Reference tables permitted: No

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	The frequency distribution of number of grains per ear-head on 50 wheat ear-heads is given below: Classes: 4-10 | 10-16 | 16-22 | 22-28 | 28-34 | 34-40 | 40-46 Frequency: 2 | 4 | 10 | 18 | 8 | 5 | 3 Calculate the mean and median values of this distribution.	**10**	**1**	**1**
Q2.	A gene can be either type X or type Y, and it can be either dominant or recessive. If the gene is type Y, then there is a probability of 0.31 that it is dominant. There is also a probability of 0.22 that a gene is type Y and it is dominant. What is the probability that a gene is of type X?	**10**	**2**	**1**
Q3.	A continuous random variable X has the cumulative distribution function $F(x) = \begin{cases} 0, & if \ \ x \leq 1 \\ k(x-1)^4, & if \ 1 < x \leq 3 \\ 1, & if \ x > 3 \end{cases}$ Find the a) probability density function b) value of k c) mean.	**10**	**2**	**1**
Q4.	(a) Find the probability of getting at least 5 times head-on tossing an unbiased coin for 6 times by using the binomial distribution. (b) At a certain manufacturing company, 5% of the tools produced turns out to be defective. If a sample of 10 tools chosen at random, find the probability that at least two will be defective by using Poisson distribution.	**10**	**3**	**1**
Q5.	Two referees in a flower beauty competition rank the 10 type of flowers as follows: A: 1 | 6 | 5 | 10 | 3 | 2 | 4 | 9 | 7 | 8 B: 6 | 4 | 9 | 8 | 1 | 2 | 3 | 10 | 5 | 7 Find the correlation coefficient between the referees.	**10**	**3**	**1**
Q6.	The grades of a class of 9 students on a midterm report (x) an on final examination (y) are as follows: x: 77 | 50 | 71 | 72 | 81 | 94 | 96 | 99 | 67 y: 82 | 66 | 78 | 34 | 47 | 85 | 99 | 99 | 68 (a) Estimate the linear regression line y on x. (b) Estimate the final examination grade of a student who received a grade of 85 on the midterm report.	**10**	**4**	**1**

Q7. In a winter of an epidemic flu, the parents of 2000 babies were surveyed by researchers at a well-known pharmaceutical company to determine if the company's new medicine was effective after two days. Among 120 babies who had the flu and were given the medicine, 29 were cured within two days. Among 280 babies who had flu but were not given medicine, 56 recovered within two days. Is there 5% significant indication that supports the company's claim of the effectiveness of the medicine? **10 4 1**

Q8. It is claimed by a racer bike manufacturing company that the racer bikes are driven on an average of more than 12,000 Kilometers on yearly basis. To test the company's claim 100 randomly drawn racer bike owners are asked to maintain a record of the total distance (in Kilometers) they travel. Will you favor this company's claim if the randomly selected sample exhibit an average of 13,200 Kilometers and a standard deviation of 3900 Kilometers? Find the P-value/critical value to conclude your answer with 5% level of significance. **10 4 1**

Q9. The proportions of blood types O, A, B, and AB in the general population of a particular country are known to be in the ratio 49:38:9:4, respectively. A research team, investigating a small isolated community in the country, obtained the following frequencies of blood type. **10 5 1**

Blood Type	O	A	B	AB
Frequency	87	59	20	4

Test the hypothesis that the proportions in this community do not differ significantly from those in the general population using Chi-square test with 5% level of significance.

Q10. A botanist wants to know whether corn harvest depends upon watering frequency. She plants 100 corn seeds out of which 60 are watered daily and 40 are watered weekly. Then she categorizes the plants according to the harvest amount as low, medium, and high (see the table below). Conduct a Chi-square test at 5% level of significance whether the corn harvest amounts are independent of the watering frequency. **10 5 1**

.

Q11. With reference to the experiment which was conducted to compare the heat producing capacity of coal from two different mines. The production was monitored five times a day. The data are shown here. **10 5 1**

Mine 1: 0.48 0.39 0.42 0.52 0.4
Mine 2: 0.38 0.37 0.39 0.41 0.38
Assume both the populations are normal. It is suspected that variance of Mine 1 is not producing as consistently as Mine 2 in terms of coal content.
Test the hypothesis that against with 5% level of significance:

Q12. The following table shows the performance scores of five different models of a popular mobile phone company manufactured at three different countries after passing through a quality control test; **10 6 1**

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Country 1	9	7	6	5	8
Country 2	7	4	5	4	5
Country 3	6	5	6	7	6

Test at 5% significance if the means are same.

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Long Summer (2023-24) - July 2024	
	Maximum Marks: 100	Duration: 3 hrs
Course Code: MAT1011	Course Title: Applied Statistics	
Set No: 2	Exam Type : **Closed Book**	School: SAS
Date: 22-07-2024	Slot: E	Session: AN
Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice		
General Instructions if any Open Book/Open Notebook/Closed Book: 1. **"fx series" - non Programmable calculator are permitted : Yes** 2. **Reference tables permitted : Yes statistical tables from COE are allowed**		

Answer any 10 questions only

Q1. The students in a class state how many siblings they have in their family. The numbers they state are given below.

$$1, 2, 1, 3, 6, 2, 8, 2, 3, 5, 4, 2, 4, 1, 5, 5$$

Find the mean, median and mode for this data. **[10M]**

Q2. Suppose there are two bags in a box, which contain the following marbles:
Bag 1: 7 red marbles and 3 green marbles
Bag 2: 2 red marbles and 8 green marbles
If we randomly select one of the bags and then randomly select one marble from that bag, what is the probability that it's a green marble? Use law of total probabiliy. **[10M]**

Q3. When an unbiased coin is tossed eight times what is the probablity of obtaining more than five heads? **[10M]**

Q4. The Edwards's Theater chain has studied its movie customers to determine how much money they spend on concessions. The study revealed that the spending distribution is approximately normally distributed with a mean of Rs 4.1 and a standard deviation of Rs 1.3. What percentage of customers will spend less than Rs 3.00 on concessions? **[10M]**

Q5. A factory produces nails and packs them in 200 boxes. If the probability that the nail is substandard is 0.006, find the probability that the box selected at random contains at most two nails that are substandard. **[10M]**

Q6. The data below show the sugar content of a fruit (x) for different numbers of days (y) after picking.

Days	Sugar
0	7.9
1	12.0
3	9.5
4	11.3
5	11.8
6	11.3
7	4.2
8	0.4

a. Draw the scatter plot for the given data.

b. Obtain the best fitted regression line to predict sugar content when the number of days after picking is 50. [2+8=10M]

Q7. Find the Pearson's coefficient of correlation between price and demand from the following data. Also interpret your result. [10M]

Price	11	13	15	17	18	19	20
Demand	30	29	24	24	21	18	15

Q8. In a poem recitation competition, ten participants were given following marks by two different judges X and Y.

X	15	17	14	13	11	12	16	18	10	9
Y	15	12	4	6	7	9	3	10	2	5

Calculate the Spearman's rank correlation for the following data and interpret your result. [10M]

Q9. A company claims to sell a coin that lands heads 60% of the time. To test this claim, you obtain one of these coins and toss it 100 times. The coin lands heads 48 times. Test the company's claim at significance level $\alpha = 0.05$. [10M]

Q10. In a study, investigators created mock identical resumés, which were sent to job placement ads in Chicago and Boston. Each resumé was randomly assigned either a commonly-white or commonly-black name. In total, 246 out of 2445 commonly-white named resumés received a callback and 164 out of 2445 commonly-black named resumés received a callback. Is there compelling evidence to conclude that callback rates are higher for common white names vs. common black names at $\alpha = 0.05$ significance level? [10M]

Q11. A researcher wants to know if there is a significant difference in the weight of newborn babies between two hospitals in a city. The researcher randomly selects 20 newborns from Hospital A and 20 newborns from Hospital B and records their weights in pounds. The mean weight for the Hospital A group is 7.5, with a standard deviation of 0.8. The mean weight for the Hospital B group is 7.1, with a standard deviation of 1.2. Is there a significant difference between the two hospitals? Use $\alpha = 0.01$ significance level. [10M]

Q12. The times required by three workers to perform an assembly-line task were recorded on five randomly selected occasions. Here are the times, to the nearest minute. Use one-way ANOVA approach to test whether the mean time is significantly different at $\alpha = 0.05$ significance level. [10M]

Hank	Joseph	Susan
8	8	10
10	9	9
9	9	10
11	8	11
10	10	9

VIT-AP UNIVERSITY	Regular Arrear Examinations (2023-24) - July 2024	
	Maximum Marks: 100	Duration: 3 Hours
Course Code: MAT1011	Course Title: Applied Statistics	
Set No: 1	Exam Type : **Closed Book**	School: SAS
Date: 05/08/2024	Slot: B	Session: FN

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice

General Instructions if any:

1. **"*fx* series" - non Programmable calculator are permitted: YES**
2. **Reference tables permitted : YES (Distribution tables for z test, t test, F test & Chi square test)**

Answer any TEN Questions, Each Question Carries 10 Marks (10×10=100 Marks)

1. The following data represents the survey regarding the heights of 51 buildings near the peripheral area of Vijayawada airport:

Height (in cm)	135-140	140-145	145-150	150-155	155-160	160-165
No. of buildings	4	7	18	11	6	5

Verify that the relation, Mode = 3 Median – 2 Mean is valid for the above grouped data.**(10 M)**

2. A manufacturer produces light-bulbs that are packed into boxes of 100. If quality control studies indicate that 0.5% of the light-bulbs produced are defective, what percentage of the boxes will contain:
(a) no defective?
(b) 2 or more defectives? **(10 M)**

3. A sample of water has a 10% chance of containing an organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that the next 18 samples
(i) exactly 3 contain the pollutant.
(ii) at least two samples contain the pollutant. **(10 M)**

4. The breaking strengths of 47 bundles of wool fibers have a sample mean 100.85 and a sample standard deviation 25. In addition, the breaking strengths of 62 bundles of synthetic fibers have a sample mean 89.32 and a sample standard deviation 20. At 5% level of significance, test the hypothesis that the wool fiber bundles have an average breaking strength more than the synthetic fiber bundles by 3? **(10 M)**

5. Suppose that the error in the reaction temperature, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$f(x) = \begin{cases} \frac{ax}{2}, & 0 < x < 2 \\ 0, & elsewhere. \end{cases}$ (a) Find the value of a. (b) Find the cumulative distribution function and use it to calculate $P(0 \leq X \leq 1)$. **(10 M)**

6. An article in the Journal of Environmental Engineering reported the results of a study on the occurrence of sodium and chloride in surface streams in central Rhode Island. The following data are chloride concentration y (in milligrams per liter) and roadway area in the watershed x (in percentage). Obtain both regression line equation.

y	4.4	6.6	9.7	10.6	11.8	12.1	14.3
x	0.19	0.15	0.57	0.70	0.47	0.70	0.60

Find the predicted chloride concentration for $x = 0.50$ **(10 M)**

7. **(a)** The heights of adult men in a certain population are normally distributed with a mean of 70 inches and a standard deviation of 3 inches. What is the probability that a randomly selected man from this population is taller than 73 inches? **(5 M)**

(b) The weights of apples in a farm are normally distributed with a mean of 150 grams and a standard deviation of 20 grams. What is the probability that a randomly selected apple weighs less than 140 grams? **(5 M)**

8. A sample of 22 wires was tested, and their resistances had a sample average of 193.7 and a sample standard deviation of 11.2. It is claimed that the average resistance of wires of this type is 200. What is your decision if α =10%? What is your decision $\alpha = 1\%$? **(10 M)**

9. An engineer who is studying the tensile strength of a steel alloy intended for use in golf club shafts knows that tensile strength is approximately normally distributed with standard deviation 60 psi. A random sample of 12 specimens has a mean tensile strength of 3250 psi. Test the hypothesis that mean strength is 3500 psi with level of significance 0.01. **(10 M)**

10. Two chemical companies can supply a raw material. The concentration of a particular element in this material is important. The mean concentration for both suppliers is the same, but we suspect that the variability in concentration may differ between the two companies. The standard deviation of concentration in a random sample of 10 batches produced by company 1 is 4.7 grams per litre, while for company 2, a random sample of 16 batches yields 5:8 grams per litre. Is there sufficient evidence to conclude that the two population variances differ? Use α= 0.05. **(10 M)**

11. (Chi-square test) Five fair coins are tossed 100 times and results are obtained as follows:

Number of heads (x)	0	1	2	3	4	5
Frequency	19	99	197	198	105	22
Probability $p(x)$	0.4066	0.3659	0.1647	0.0494	0.0111	0.0002

Test whether Poisson distribution is appropriate for the given data at 5% level of significance. **(10 M)**

12. In an air-pollution experiment, researchers wish to determine whether the three types of instruments yield the measurements of polluting percentage of sulfur monoxide in the atmosphere. The readings in the following table were recorded for the three instruments.

Instrument A	2	3	5	0	8		
Instrument B	4	6	8	4	9	0	2
Instrument C	5	2	3	2	3	3	

Test the hypothesis by using one-way ANOVA that there is a significance difference in the average reading of these instruments at 1% level of significance.

(10 M)

06/08/24

Rooms :- 301, 302, 320, 323, 324

K. Ancient History of India

Ancient