②

Name: Roll number:

Q. 1 – (5 marks + 10 marks)
Let X,Y,Z be three normal random variables independently distributed. Each of X,Y,Z has mean = 10, variance = 36

Let W be a random variable which denotes the arithmetic average of X,Y and Z.

Further, let V be a random variable which is given by "W – 2X"

Find Expectation of V and Variance of V

E(V): -10 **Var(V):** 108

Q.2 (15 marks)
   Let z be a function of x and y given as follows:

$$z = (x-1)^2 + (y-1)^2 + x^2y^2$$

We are interested in finding the minimum value of the function z, using gradient descent method.

Let the starting point (x,y) be (2,1) . Learning rate (or step size) is 0.2

One iteration of the gradient descent method would move the point from (2,1) to

(0.709, 0.56) ~~r~~ ~~x=0.409~~ $(0.8, -0.6)$

1

Q.3 a.) (5 marks)
The Dean of Student Welfare (DOSW) receives a lot of emails on a daily basis. On any day, the number of emails received by DOSW can be assumed to follow a Binomial distribution with parameters (10000,0.015).

Answering an email takes some time. Some emails can be responded quickly , while some emails take longer. Assume that time required to respond to a email follows a uniform distribution with minimum possible value as 1 minute and maximum possible value as 3 minutes.

Find the expected time spent by DOSW in answering to emails.

~~6705~~   20000

Q.3b.) (5 marks)
Let T denote the random variable which represents the time required by DOSW to respond to emails.
Find the VARIANCE of random variable T.

640.5

Q.4 (2+2+2+2+2 marks)
Let X be a random variable that follows a normal distribution with mean $= 100$ and standard deviation $= 11$.

a.) Find $P(X = 100)$ 0

b.) Find $P(X<100)$ 0.5

c.) Find $P(X<105)$ 0.6752

d.) Find $P(X>110)$ 0.1816

e.) Find $P(85<X<105)$ 0.5889

Q.5a.) [10 marks]

Amul wants to launch "Kaju Katli" as a new product to give competition to local sweets vendor. The objective of Amul is to capture atleast 40% of the market.

In a pilot study, Amul surveyed 1000 customers who were regular buyers of Kaju Katli from local sweets vendor, out of which 410 said that they would switch to Amul Kaju Katli. From this data, what is the symmetric confidence interval for the true proportion of customers who would switch to Amul.

0.3795, 0.4404

Q.5b.) [10 marks]

Using the data, CEO of Amul would like to test the hypothesisthat Amul's Kaju Katli would be able to capture 40 % of the market.

Taking Amul's objective as the alternative hypothesis, what is the p-value of the hypothesis test.

0.3795, 0.4404   p-value = 0.261

Q.5c.) [5 marks]

By looking at the p-value in part b.) of the problem, should the alternative hypothesis be accepted? Answer briefly in less than 20 words.
[Assume level of significance as 0.05]

p value > 0.05

Q.6 [10 marks]

A coin having probability of heads as 0.8 is tossed 10000 times. What is the approximate probability that observed number of heads is less than 7950?

[The solution must use central limit theorem to approximate something somewhere, the answer obtained by using binomial distribution directly will not be accepted]

0.1058

Name :

Q.7 [10+5+5+5 marks]

Consider the following dataset of 10 values.

S.No	X	Y	Z
1	20	5	8
2	30	4	7
3	40	3	9
4	50	2	6
5	60	1	7
6	70	6	14
7	80	7	15
8	90	8	18
9	100	9	20
10	110	10	23

A multiple linear regression is built considering Z as the dependent variable and "X" and "Y" as the independent variables.

a.) What is the multiple regression model? [Hint: $E(Z) = alpha + beta*X + gamma*Y$]

$$Z = -0.01897 + 0.0922X + 1.22224Y$$

b.) By looking at the p-value of the F-test, should the proposed linear relationship between these variables considered or not?

$p = 6.0174 \times 10^{-7} < 0.05$ (Reject H_0) There is linear relation-ship

c.) Give the interpretation of coefficient of "Y" in less than two lines.

for a unit increase in Y, Z increases by 1.22224 on an average, keeping other variable constant
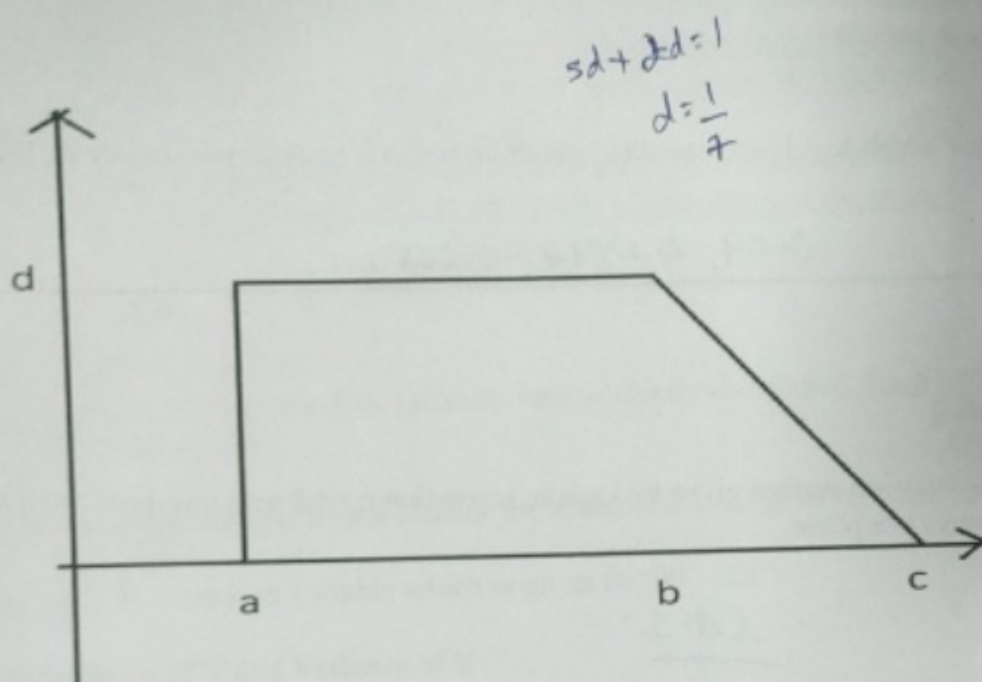
d.) (True/False) ✓

The VIF(Variance Inflation Factor) of X and Y will come out to be the same.

Q.8 [10 marks]

Consider the below graph as the pdf (probability density function) of a random variable.

$$5d + \frac{2d}{} = 1$$
$$d = \frac{1}{7}$$

The value of pdf for a value less than "a" or more than "c" is zero. Between "a" to "b", it is a straight line parallel to X-axis and value of pdf is "d". Between 'b" to "c", the value of pdf decreases linearly from "d" to 0.

If a= 5, b = 10, c = 14, find the value of d ? $1/7$

Q.9 [2+5+2+5 marks]

In a logistic regression model with a single feature (X), the model parameters (alpha and beta) come out to be alpha = 5, beta = 6. If X increases by 1, what can be said about the expected increase or decrease in probability/odds ratio. Answer the following 4 questions precisely. [Hint: You can also write "Cannot be Determined", this may be the correct answer]

a.) The probability will increase (True/False) ___True___

b.) The probability will change(increase/decrease) by ___cannot be determined___

c.) Odds ratio will increase (True/False) ___True True___

d.) Odds ratio will change by a factor of ___e^6___

Q.10 [10 marks]

We train a "softmax" model to classify something into 3 categories A,B,C using a single feature(X). Following are the parameters that come out after the optimization is done.

Alpha_A =1.5, beta_A = 2
Alpha_B = 1.5, beta_B = 3
Alpha_C = 1.5, beta_C = 4

For a point for which X = 1, what are the probabilities that it belongs to category A, B & C ?

0.09, 0.2447, 0.6652

Q.11 [10 marks]

Prove that the decision surface given by logistic regression model with threshold of "0.5" in case of 2 features is a plane.

Set-1