INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

February 2025

Subject CS1A – Actuarial Statistics (Paper A)

Time allowed: 3 Hours 15 Minutes (09.30 – 12.45 Hours)

Total Marks: 100

Use the following information for attempting questions 1 to 5:

Regal Club is a popular club in the city known for its game called "Power of Six". In this game, the player is given a total of n chances to play.

In every chance the player,

- Has to throw two unbiased dice.
- Has to continue throwing them until the sum of numbers on the dice is equal to "six".

Let X_i be the number of throws required in chance i to get a sum of six on both dice and let p be the probability of getting a sum of six on both dice in each throw.

X_is are independent and identically distributed random variables.

The player is rewarded based on the chance with the minimum number of throws.

Let Y = min $(X_1, X_2, \ldots\ldots\ldots, X_n)$.

The player is rewarded only when Y < 10. In case if Y >= 10, then the player is not entitled to any reward. Reward to the player is defined as INR [(10 – Y) * 100].

The charges to be paid by the player for playing this game is INR 700.

Q. 1) Variance of X_i is given by which of the following expressions?

A. $p * (1 - p)$
B. $(1 - p) / p^2$
C. $(1 - p) / p$
D. p^2
E. $1/p^2$. **[1]**

Q. 2) For each chance i = 1, 2, ………., n, $P(X_i >= x)$ is given by –

A. $(1 - p)^{(1 - x)}$
B. $(1 - p)^{(x + 1)}$
C. $(1 - p)^{x}$
D. $p * (1 - p)^{(x - 1)}$
E. $(1 - p)^{(x - 1)}$. **[2]**

Q. 3) What is the value of the probability p?

A. 1 / 36
B. 2 / 36
C. 3 / 36
D. 5 / 36
E. 6 / 36. **[1]**

Q. 4) P (Y >= y) is given by –

A. $((1-p)^n)^{(y-1)}$
B. $((1-p)^n)^{(1-y)}$
C. $((1-p)^n)^{y}$
D. $((1-p)^n)^{(y+1)}$
E. $p^n *((1-p)^n)^{(y-1)}$ **[2]**

Q. 5) If a player is given 10 chances to play (i.e. n = 10), what is the probability that the player will be losing his money in this casino game?

Hint: [*Loss in the game will arise when the reward earned < charges paid to play the game*]

A. 0.2526%
B. 5.0256%
C. 1.1266%
D. 22.4177%
E. 0.0566%. **[3]**

<u>Use the following information for attempting questions 6 to 10:</u>

Pro-Fitness is a company which manufactures fitness essentials for gymnasts. Two of its products are very popular across eight cities in the country –

- Pro-Fitness Whey Protein Bar
- Pro-Fitness Peanut Butter

Sales team of the company has collected the following data relating to the average number of units sold per day for whey protein bar (X) and the average number of units sold per day for peanut butter (Y) across eight cities in the country.

The data has been given below –

Region	Panjim	Delhi	Ranchi	Cochi	Shimla	Pune	Surat	Agra	Average
X	15	25	21	29	13	18	21	17	19.875
Y	4	3	11	10	6	9	7	5	6.875

For these observations we obtain: $\sum x^2 = 3355$, $\sum xy = 1128$, $\sum y^2 = 437$.

Q. 6) What is the value of Pearson's sample correlation coefficient between X and Y i.e. r_{PEARSON}?

A. 0.32559
B. 0.57992
C. 0.29314
D. 0.50547
E. 0.0102. **[2]**

Q. 7) The company wants to develop a 95% confidence interval for the population correlation coefficient using Fisher's transformation. Under Fisher's transformation, what is the value of Z_r based on r_{PEARSON} calculated in question 6?

A. 0.953368
B. 0.650840

C. 0.285019
D. 0.641977
E. 0.337887. **[1]**

Q. 8) Analysts from the company have obtained a 95% confidence interval for Z_r calculated in question 7 as given below:

[-0.45577,0.684747]

Based on the above interval, which of the following represents a 95% confidence interval for the population correlation coefficient?

A. [- 0.322690, 1.08115]
B. [- 0.49195, 0.837997]
C. [- 0.423940, 0.79315]
D. [- 0.422688, 0.95417]
E. [0.4233092, 0.876386] **[3]**

Q. 9) Which of the following is TRUE regarding the values of the sample correlation coefficients i.e. $r_{PEARSON}$ and $r_{SPEARMAN}$ calculated based on the given dataset?

A. $r_{PEARSON} = r_{SPEARMAN}$
B. $r_{PEARSON} < r_{SPEARMAN}$
C. $r_{PEARSON} > r_{SPEARMAN}$
D. $r_{SPEARMAN}$ does not exist for this dataset
E. None of the above. **[3]**

Q. 10) Which of the following is TRUE regarding the mathematical properties of the sample correlation coefficients i.e. $r_{PEARSON}$ and $r_{SPEARMAN}$?

A. Both $r_{PEARSON}$ and $r_{SPEARMAN}$ only measure the monotonic relationship between X and Y.
B. Both $r_{PEARSON}$ and $r_{SPEARMAN}$ measure the strength of the linear relationship between X and Y.
C. Whereas $r_{PEARSON}$ only measures the monotonic relationship between X and Y, $r_{SPEARMAN}$ measures the strength of the linear relationship between X and Y.
D. Whereas $r_{PEARSON}$ measures the strength of the linear relationship between X and Y, $r_{SPEARMAN}$ only measures the monotonic relationship between X and Y.
E. Both $r_{PEARSON}$ and $r_{SPEARMAN}$ measure the monotonic relationship between X and Y, but $r_{SPEARMAN}$ also measures the strength of the linear relationship between X and Y. **[1]**

Use the following information for attempting questions 11 to 15:

A supermarket store is investigating into the number of visits by individuals in the nearby vicinity during the month among different age groups. It has categorized the population among three age groups *viz*. Gen X (born before 1st January 1981), Millennials (born on or after 1st January 1981 but before 1st January 2001) and Gen Z (born on or after 1st January 2001).

It has appointed an actuary for this purpose.

The actuary has estimated that the joint distribution of the number of supermarket visits by individuals (N) and the age groups of the individuals (G) is given by the following table:

	Gen X	Millennials	Gen Z
0 visits per month	0.120	0.150	0.150
1 visit per month	0.200	0.100	0.030
> 1 visit per month	0.190	0.050	0.010

For example, the probability that a randomly selected individual belongs to the group "Gen Z" and made no visit to the supermarket store during the month is 0.15.

Q. 11) The probability that a randomly selected individual has made more than one visit during the month to the supermarket store given that he / she is 47 years old as at 22nd November 2024 is –

A. 37.25%
B. 19.00%
C. 39.00%
D. 76.47%
E. None of the above. **[2]**

Q. 12) The owner of the supermarket store wants to know the likelihood of an empty and deserted supermarket during the month. Required probability is –

A. 55.00%
B. 42.00%
C. 30.00%
D. 24.00%
E. 20.00%. **[1]**

Q. 13) The probability that a randomly selected individual does not belong to "Gen Z" given that he / she has made at least one visit to the supermarket store during the month is –

A. 60.61%
B. 67.24%
C. 90.91%
D. 85.38%
E. 93.10%. **[2]**

Q. 14) The actuary has further bifurcated the data based on nature of employment (E) into two categories – "Employed individuals" and "Self-employed individuals" and obtained the following two joint distributions:

Employed Individuals:

	Gen X	Millennials	Gen Z
0 visits per month	0.175	0.105	0.070
1 visit per month	0.225	0.135	0.090
> 1 visit per month	0.100	0.060	0.040

Self-Employed Individuals:

	Gen X	Millennials	Gen Z
0 visits per month	0.080	0.080	0.040
1 visit per month	0.200	0.200	0.100
> 1 visit 'per month	0.120	0.120	0.060

Note:
N: the number of supermarket visits by individuals
G: and the age groups of the individuals
E: nature of employment.

Considering the above information, which of the following statements are TRUE?

A. N | E = "Employed" is independent of G | E = "Employed" **[3]**
B. N | E = "Self-Employed" is independent of G | E = "Self-Employed"
C. Both (A) and (B)
D. Either (A) or (B)
E. Neither (A) nor (B).

Q. 15) Considering the original joint distribution formulated by the actuary and the employment category-wise joint distributions constructed by the actuary as given in question 14, which of the following statements is TRUE?

A. Random variables N and G are unconditionally independent of each other.
B. Random variables N and G are not conditionally independent of each other for a given value of E.
C. Although random variables N and G are not conditionally independent of each other for a given value of E, they are unconditionally independent of each other.
D. Although random variables N and G are conditionally independent of each other for a given value of E, they are not unconditionally independent of each other.
E. None of the above. **[2]**

<u>Use the following information for attempting questions 16 to 20:</u>

General elections are being conducted in Actuaria. The following table represents the number of seats estimated to be won by eight national political parties in the country as per exit poll results (X) and the actual number of seats won by the respective parties as per election commission poll results data.

Party	XJP	NNC	CP	MMC	DTP	JSU	HSH	PNC
X:	340	60	6	20	16	13	10	7
Y:	240	99	37	29	16	12	9	8

A regression model is to be fit with Y as the response variable and X as the explanatory variable.

For this dataset you are given:

$$S_{xx} = 92362,\ S_{yy} = 44844,\ S_{xy} = 62350,\ \overline{X} = 59,\ \overline{Y} = 56.25.$$

Q. 16) What is the value of the coefficient β?

A. 0.675061
B. 0.719222
C. 0.485519 **[1]**

D. 0.517280
E. 0.368712.

Q. 17) What is the value of standard error of $\hat{\beta}$?

A. 0.06104
B. 0.07049
C. 0.23434
D. 0.27060
E. 0.15632. **[2]**

Q. 18) If we test the null hypothesis H_0: $\beta = 0$ against the alternate hypothesis H_1: $\beta \neq 0$, what will be the result of the test at 95% level of significance?

A. We have sufficient evidence to reject the null hypothesis as 0 lies in the 95% symmetrical confidence interval for β.
B. We do not have sufficient evidence to reject the null hypothesis as 0 lies in the 95% symmetrical confidence interval for β.
C. We have sufficient evidence to reject the null hypothesis as 0 does not lie in the 95% symmetrical confidence interval for β.
D. We do not have sufficient evidence to reject the null hypothesis as 0 does not lie in the 95% symmetrical confidence interval for β.
E. None of the above. **[2]**

Q. 19) A political party newly formed out of a nation-wide anti-corruption campaign is contesting its first general election. The exit polls have estimated that the party will draw a blank in the general election. How many seats will the party win based on the fitted regression model (round off your answer to the nearest integer)?

A. 0
B. 4
C. 9
D. 16
E. 25. **[2]**

Q. 20) Under the hypothesis as mentioned in question 18 relating to testing whether the slope coefficient is equal to 0, the expression of the test statistic under the null hypothesis is given by:

$$\hat{\beta} / se(\hat{\beta})$$

Which of the following is an equivalent expression to the above expression of the test statistic?

A. $r * \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$

B. $r * \frac{\sqrt{n-1}}{\sqrt{1-r^2}}$

C. $r * \frac{\sqrt{1-r^2}}{\sqrt{n-1}}$

D. $r * \frac{\sqrt{n-2}}{\sqrt{r^2+1}}$

E. $r * \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$. **[3]**

Use the following information for attempting questions 21 to 25:

The random variables X_1, X_2, …………., X_n denote the length of time a swimmer i (for i = 1, 2, ……., n) can swim underwater at a stretch and is modelled using exponential distribution with unknown parameter λ.

Q. 21) What is the maximum likelihood estimator for the unknown parameter λ?

A. $\sum x$ / n
B. n / $\sum x^2$
C. n / $\sum x$
D. $\sum x^2$ / n
E. $\sum x$ / n^2. **[3]**

Q. 22) What is the Cramer's Rao Lower Bound (CRLB) for the unknown parameter λ?

A. $-n / \lambda^2$
B. $-\lambda^2 / n$
C. n / λ^2
D. λ^2 / n
E. λ^2 / n^2. **[1]**

Q. 23) Instead of observing the values of X_1, X_2, …….., X_n precisely, we now decide to only observe whether or not a swimmer can swim under water and use a Poisson distribution with mean λ to model the same

So, we only observe that for m swimmers, $X_i = 0$ and for the remaining swimmers $X_i > 0$.

What is the likelihood function for this new sample?

A. $L(\lambda) = \lambda * e^{-\lambda m}$
B. $L(\lambda) = \lambda^m * e^{-\lambda m}$
C. $L(\lambda) = \lambda^m * (1 - \lambda)^{(n - m)}$
D. $L(\lambda) = \lambda * e^{-\lambda (n-m)}$
E. $L(\lambda) = e^{-\lambda m} * (1- e^{-\lambda})^{(n-m)}$. **[2]**

Q. 24) Based on the new sample as determined in question 23, determine the maximum likelihood estimator for the unknown parameter λ?

A. 1 / (n – m)
B. log (n / m)
C. m / n
D. n / m
E. 1 / m. **[3]**

Q. 25) How many minimum swimmers who cannot swim underwater need to be observed in the new sample to ensure that the maximum likelihood estimator of λ as determined in question 24 is finite?

A. 1
B. 0
C. n **[1]**

D. n – m
E. m.

Use the following information for attempting questions 26 to 30:

A Non-Banking Financial Company is analysing defaults on three different types of one-year loans viz. Education Loans, Personal Loans and Gold Loans. It has collected data for the past five years. The number of defaults in the j^{th} year for the i^{th} type of loan is denoted by P_{ij} for i = E, P, G and j = 2019, 2020, 2021, 2022, 2023. The average size of default (outstanding balance of loan unpaid by the borrower) per loan over all five years for the i^{th} type of loan is denoted by $\overline{X}_1$.

The values of P_{ij} and $\overline{X}_1$ (in INR) have been tabulated below:

Loan Type i	Number of loan-defaults					P_i	$\overline{X}_1$
	2019	2020	2021	2022	2023		
Education	170	230	210	290	350	1,250	85,000
Personal	420	510	600	550	370	2,450	72,000
Gold	430	310	620	980	1,070	3,410	90,000

The financial plan for the year 2024 has been approved by the company's board and the targets for the year 2024 are 300 education loans, 400 personal loans and 1,100 gold loans.

You have been recently employed in the Financial Risk Team as an actuary. Your predecessor, who has left the company recently, had calculated the expected aggregate defaults on education loans for the year 2024 to be INR 2.52 crore using the assumptions of Empirical Bayes Credibility Theory Model 2 (EBCT Model 2).

Q. 26) What is the average default size for the NBFC based on the historical 5-year data i.e. $E[m(\theta)]$?

A. INR 76,391.89
B. INR 82,918.45
C. INR 82,474.40
D. INR 88,658.80
E. INR 90,000. **[2]**

Q. 27) Based on the assessment of expected aggregate defaults for education loans done by your predecessor actuary for the year 2024, the credibility factor for education loans (Z_E) is –

A. 0.3205
B. 0.4804
C. 0.5196
D. 0.6795
E. 0.7434. **[2]**

Q. 28) The credibility factor for gold loans (Z_G) is –

Hint: [*Assume $E(s^2(\theta)) / Var(m(\theta))$ as 'a' and find the value of a based on Z_E calculated in question 27*]

A. 0.2531
B. 0.4324
C. 0.5436
D. 0.7469 **[3]**

E. 0.8128.

Q. 29) Based on the credibility factor for gold loans (Z_G) calculated in question 28, which of the following represents the expected total defaults under gold loans for the year 2024?

A. INR 9.70 crore
B. INR 9.54 crore
C. INR 9.46 crore
D. INR 9.32 crore
E. INR 9.15 crore. **[2]**

Q. 30) The Chief Risk Officer of the company is checking whether EBCT Model 1 can be used instead of EBCT Model 2 in the given case and in which scenario will it yield exactly similar results. Which one of the following represents that scenario?

A. EBCT Model 1 and EBCT Model 2 will always yield exactly similar results.
B. When number of loans for each year are equal but mix based on loan type is different.
C. When mean default size for each loan type is equal.
D. When mean default size for each year is equal.
E. When number of loans for each loan type are equal in each year. **[1]**

<u>Use the following information for attempting questions 31 to 35:</u>

There has been a huge uproar in the country regarding the conduct of medical and para-medical entrance examination MPMET during this year in which students from the North have apparently scored higher than students from the South. A possible paper leak is being probed by investigation agencies which led to such a difference.

Meanwhile, an actuary is trying to test whether there is such a big difference in marks by taking two independent random samples of marks obtained, of size n each, from North as well as South. Samples are from normal populations $N(\mu_N, \sigma^2)$ and $N(\mu_S, \sigma^2)$ respectively, where parameters μ_N and μ_S are unknown and σ^2 is known.

He has defined the following hypothesis for conducting a statistical test:

$$H_0: \mu_N = \mu_S \text{ against } H_1: \mu_N \neq \mu_S.$$

He wants to conduct the hypothesis test at 5% level of significance using the following two-step procedure:

Step 1: Compute the 95% confidence intervals for μ_N and μ_S.
Step 2: If the two confidence intervals do not overlap, then reject H_0.

Q. 31) What are the sampling distributions of $\overline{X_N}$ and $\overline{X_S}$?

A. $\overline{X_N} \sim N(\mu_N, \sigma^2)$ and $\overline{X_S} \sim N(\mu_S, \sigma^2)$
B. $\overline{X_N} \sim N(\mu_N, \sigma^2 / n_N)$ and $\overline{X_S} \sim N(\mu_S, \sigma^2 / n_S)$
C. $\overline{X_N} \sim N(\mu_N, \sigma^2 / n)$ and $\overline{X_S} \sim N(\mu_S, \sigma^2 / n)$
D. $\overline{X_N} \sim N(\mu_N, \sigma / \sqrt{n})$ and $\overline{X_S} \sim N(\mu_S, \sigma / \sqrt{n})$
E. $\overline{X_N} \sim N(\mu_N, 2\sigma / \sqrt{n})$ and $\overline{X_S} \sim N(\mu_S, 2\sigma / \sqrt{n})$. **[1]**

Q. 32) What is the sampling distribution of $(\overline{X_N} - \overline{X_S})$? **[1]**

A. $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N - \mu_S, \sqrt{2}\,\sigma / \sqrt{n})$
B. $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N + \mu_S, 2\sigma^2 / n)$
C. $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N + \mu_S, \sigma^2 / n)$
D. $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N - \mu_S, 0)$
E. $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N - \mu_S, 2\sigma^2 / n)$.

Q. 33) In which of the following cases, will you reject the null hypothesis?

A. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) < (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$
B. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) > (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$
C. $(\overline{X_N} + 1.96 * \sigma / \sqrt{n}) > (\overline{X_S} - 1.96 * \sigma / \sqrt{n})$
D. Both (A) and (B)
E. None of the above. **[3]**

Q. 34) In which of the following cases, will you accept the null hypothesis?

A. $(\overline{X_N}-1.96\sigma/\sqrt{n}) < (\overline{X_S}+1.96\sigma/\sqrt{n}) < (\overline{X_N}+1.96\sigma/\sqrt{n})$
B. $(\overline{X_N}+1.96\sigma/\sqrt{n}) < (\overline{X_S}+1.96\sigma/\sqrt{n}) < (\overline{X_N}-1.96\sigma/\sqrt{n})$
C. $(\overline{X_S}+1.96\sigma/\sqrt{n}) < (\overline{X_N} -1.96\sigma/\sqrt{n}) < (\overline{X_N} +1.96\sigma/\sqrt{n})$
D. Both (A) and (B)
E. None of the above. **[3]**

Q. 35) The sample values taken by the actuary are presented below:

Sample N: 518, 720, 644, 718, 467, 356, 651, 681, 559, 300.

Sample S: 200, 613, 542, 314, 700, 711, 627, 659, 491, 615.

Which value of σ^2 from among the following options will make it imperative for the actuary to reject the null hypothesis at the 5% level of significance?

A. 900
B. 625
C. 400
D. 100
E. 81. **[3]**

Use the following information for attempting questions 36 to 40:

An actuary is fitting the following linear regression model through the origin for establishing a relationship between the speed of an aircraft in kilometres per hour (Y) and the quantity of aviation fuel required in litres (x):

$Y_i = \beta * x_i + e_i$ $\quad e_i \sim N(0, \sigma^2)$ $\quad i = 1, 2, \ldots\ldots\ldots\ldots.., n$

Q. 36) What is the distribution of $Y_i \mid x_i$?

A. $Y_i \mid x_i \sim N(\beta * x_i, \beta * var(x_i) + \sigma^2)$
B. $Y_i \mid x_i \sim N(\beta * x_i + e_i, \sigma^2)$ **[1]**

C. $Y_i \mid x_i \sim N(\beta * x_i, \beta^2 * var(x_i) + \sigma^2)$
D. $Y_i \mid x_i \sim N(\beta * x_i, \sigma^2)$
E. $Y_i \mid x_i \sim N(\beta * x_i, 0)$.

Q. 37) Which of the following represents the least square estimator of β?

A. $\hat{\beta}_{LSE} = \sum x_i y_i / \sum x_i$
B. $\hat{\beta}_{LSE} = \sum x_i y_i / \sum x_i^2$
C. $\hat{\beta}_{LSE} = \sum x_i y_i / \sum y_i^2$
D. $\hat{\beta}_{LSE} = \sum x_i y_i / \sum y_i$.
E. $\hat{\beta}_{LSE} = \sum x_i y_i / \sum x_i^2 y_i^2$. **[3]**

Q. 38) What is the expectation of $\hat{\beta}_{LSE}$?

A. $E(\hat{\beta}_{LSE}) = \beta$
B. $E(\hat{\beta}_{LSE}) = n * \beta$
C. $E(\hat{\beta}_{LSE}) = \beta / \sum x_i^2$
D. $E(\hat{\beta}_{LSE}) = \beta / \sum x_i$.
E. $E(\hat{\beta}_{LSE}) = 1 / \beta$. **[2]**

Q. 39) What is the variance of $\hat{\beta}_{LSE}$?

A. $Var(\hat{\beta}_{LSE}) = \sigma^2$
B. $Var(\hat{\beta}_{LSE}) = n * \sigma^2$
C. $Var(\hat{\beta}_{LSE}) = \sigma^2 / \sum x_i^2$
D. $Var(\hat{\beta}_{LSE}) = \sigma^2 / \sum x_i$.
E. $Var(\hat{\beta}_{LSE}) = 1 / \sigma^2$. **[2]**

Q. 40) What is the mean square error of $\hat{\beta}_{LSE}$?

A. $MSE(\hat{\beta}_{LSE}) = \sigma^2$
B. $MSE(\hat{\beta}_{LSE}) = \beta * (n - 1) + n * \sigma^2$
C. $MSE(\hat{\beta}_{LSE}) = \beta^2 * (n - 1)^2 + n * \sigma^2$
D. $MSE(\hat{\beta}_{LSE}) = n\beta + \sigma^2$
E. $MSE(\hat{\beta}_{LSE}) = \sigma^2 / \sum x_i^2$. **[2]**

Use the following information for attempting questions 41 to 45:

In a stock exchange, a circuit breaker is triggered when there is a serial fall in the value of the stock market index and the fall breaches the limit of 10% of the last trading day's closing index.

Circuit breaker is designed in order to prevent a possible market crash. When the circuit breaker is triggered, trading is suspended for the rest of the day and trading resumes on the next trading day.

From the start of the financial year, it turns out that the first circuit breaker is triggered on the n_1^{th} trading day.

The probability of the first trigger is given by "p". It can be assumed that "p" is uniformly distributed over the interval [0, 1].

Q. 41) If we define X as the number of trading days required for the trigger of the first circuit breaker in the financial year, what is the distribution of X?

A. Binomial distribution
B. Geometric distribution
C. Hypergeometric distribution
D. Poisson distribution.
E. Bernoulli distribution. **[1]**

Q. 42) What is the maximum likelihood estimate of p?

A. $n_1 – 1$
B. $1 / (n_1 – 1)$
C. n_1
D. $1 / n_1$.
E. $(n_1 + 1) / (n_1 – 1)$. **[3]**

Q. 43) What is the posterior distribution of p?

A. Beta distribution with parameters $\alpha = 2$ and $\beta = n_1$
B. Beta distribution with parameters $\alpha = 2$ and $\beta = n_1 – 1$
C. Beta distribution with parameters $\alpha = 1$ and $\beta = n_1$
D. Beta distribution with parameters $\alpha = 1$ and $\beta = n_1 – 1$.
E. Beta distribution with parameters $\alpha = 2$ and $\beta = n_1 + 1$. **[2]**

Q. 44) What is the prior mean and the posterior mean of p?

A. Prior mean = 1 and Posterior mean = $2 / n_1$
B. Prior mean = 1/2 and Posterior mean = $2 / n_1$
C. Prior mean = 1/2 and Posterior mean = $2 / (n_1 + 2)$
D. Prior mean = 1 and Posterior mean = $2 / (n_1 + 2)$.
E. Prior mean = 2 and Posterior mean = $2 / (n_1 + 2)$. **[1]**

Q. 45) Which of the following represents the value of Credibility Factor for p?

A. $2 / (n_1+ 2)$
B. $1 / (n_1 + 1)$
C. $1 / (n_1 + 2)$
D. $n_1 / (n_1 + 1)$
E. $n_1 / (n_1 + 2)$.. **[3]**

Use the following information for attempting questions 46 to 50:

A newly setup health insurance company is modelling size of claims under its flagship mediclaim policy using generalised linear models. Gender – Male / Female (X_1) and Type of Hospital – Public / Private (X_2) are used as explanatory variables for predicting Claim Size (Y).

A statistician working for the company has fitted two models:

Model 1:

- Claim size follows a Gamma distribution.
- Log function is the selected link function.
- Female is the base level for Gender (X_1).
- Private Hospital is the base level for Type of Hospital (X_2).
- Linear predictor is given by: $g(\mu) = \beta_0 + \beta_1 + \beta_2$.

The following model output has been achieved for Model 1:

Particulars	**Coefficient**
β_0	3.25
β_1 GENDER = FEMALE	0.00
β_1 GENDER = MALE	0.89
β_2 HOSPITAL = PUBLIC	-0.72
β_2 HOSPITAL = PRIVATE	0.00

Model 2:

- Claim size follows a Gamma distribution.
- Log function is the selected link function.
- Female is the base level for Gender (X_1).
- Private Hospital is the base level for Type of Hospital (X_2).
- This model has only interaction terms between X_1 and X_2.
- Linear predictor is given by: $h(\mu) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$.
- In the linear predictor,
- α_0 corresponds to the intercept coefficient (which includes the interaction term for base values i.e. between female and private hospital).
- α_1 corresponds to the interaction term between female and public hospital.
- α_2 corresponds to the interaction term between male and private hospital.
- α_3 corresponds to the interaction term between male and public hospital.

Q. 46) If the Gamma distribution with parameters α and λ is written in the form of a density function for a member of exponential family with the following expansion:

$$f_Y(y; \theta, \phi) = \exp [(-y/\mu - \log\mu) * \alpha + (\alpha - 1) * \log y + \alpha * \log\alpha - \log \Gamma(\alpha)]$$

This is compared with the standard probability density function of a random variable from an exponential family of distributions.

$$f_Y(y; \theta, \phi) = \exp [(y\theta - b(\theta)) / a(\phi) + c(y, \phi)]$$

Which of the following options correctly represents the values of the functions $a(\phi)$ and $b(\theta)$?

A. $a(\phi) = 1 / \lambda$ and $b(\theta) = - \log (\alpha / \lambda)$
B. $a(\phi) = 1 / \alpha$ and $b(\theta) = - \log (\alpha / \lambda)$
C. $a(\phi) = 1 / \alpha$ and $b(\theta) = - \log (\lambda / \alpha)$
D. $a(\phi) = 1 / \lambda$ and $b(\theta) = - \log (\lambda / \alpha)$
E. $a(\phi) = 1 / \alpha$ and $b(\theta) = \log (\lambda / \alpha)$. **[2]**

Q. 47) Mrs. Wadia, a policyholder under the mediclaim policy, has lodged a claim with the company in relation to her appendicitis surgery which took place in a reputed Government **[2]**

Hospital.

The predicted claim size for this claim as per Model 1 will be –

A. INR 12.5535 lakhs
B. INR 25.7903 lakhs
C. INR 30.5694 lakhs
D. INR 62.8082 lakhs
E. INR 75.6354 lakhs.

Q. 48) Using Model 2, the predicted claim size of another policyholder, Ms. Merlin who was admitted in a local private hospital for treatment of kidney stones was INR 24.5537 lakhs.

What is the estimated value of coefficient α_0?

A. + 2.34519
B. + 4.60825
C. + 1.28401
D. + 0.04531
E. + 3.20086. **[2]**

Q. 49) Using Model 2, the predicted claim size of another policyholder, Mrs. Ambedkar who was admitted in a public hospital for treatment of haemorrhoids was INR 15.0813 lakhs.

What is the estimated value of coefficient α_1?

Hint: [*Use the value of α_0 calculated in question 48*]

A. – 1.8948
B. – 0.4874
C. + 1.4295
D. + 2.6682
E. + 0.6752. **[3]**

Q. 50) Model 2 is a better fit as compared to Model 1 if –

A. AIC (Model 2) > AIC (Model 1)
B. AIC (Model 2) = 0
C. AIC (Model 2) = AIC (Model 1)
D. AIC (Model 2) < AIC (Model 1)
E. None of the above.

Note: *AIC: Akaike's Information Criteria.* **[1]**

****