INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

22nd November 2024

Subject CS1A – Actuarial Statistics (Paper A)

Time allowed: 3 Hours 15 Minutes

Total Marks: 100

Use the following information for attempting questions 1 to 5:

The maximum number of seconds a yogic practitioner can hold his breath (Y) is being studied by research scientists of a Yogic Research Centre. It is being analysed by studying two explanatory variables viz. –

- Weight of the person in kilograms (X_1) and
- Age of the person in years (X_2).

The centre has collected data for 7 practitioners who have volunteered for this research project.

Q. 1) A multi-variate linear regression model has been fit which has the following equation:

$$Y = 8.928 + 0.993\ X_1 + 0.476\ X_2$$

Which of the following statements regarding the relationship between Y, X_1 and X_2 is necessarily TRUE based on the above equation?

A. Y and X_1 are negatively correlated.
B. Y and X_2 are positively correlated.
C. Y and X_2 are negatively correlated.
D. Y and X_1 are un-correlated. **[2]**

Q. 2) It is generally believed that as we age, we tend to put on weight. This is proposed to be statistically tested using a students' t-distribution based on the data collected.

Sample correlation coefficient between X_1 and X_2 based on the sample of 7 volunteers using Pearson's Method i.e. $r_{PEARSON}$ is determined as – 0.1599543.

Select the correct null hypothesis (H_0) and the correct value of the test statistic under the null hypothesis from the following options:

A. H_0: "X_1 and X_2 are un-correlated";
t-statistic = - 0.35318
B. H_0: "X_1 and X_2 are correlated";
t-statistic = 0.10162
C. H_0: "X_1 and X_2 are un-correlated";
t-statistic = - 0.36233
D. H_0: "X_1 and X_2 are correlated";
t-statistic = 0.32164. **[2]**

Q. 3) At 5% level of significance, determine the critical value from the students' t-distribution and the result of the t test using the null hypothesis and the test statistic calculated in question 2. (Kindly note that it is a two-sided test).

A. Critical value = 2.365;
Result: With age, we tend to put on weight
B. Critical value = 2.571;
Result: With age, we tend to lose weight
C. Critical value = 2.365;
Result: We may put on or lose weight irrespective of age
D. Critical value = 2.571;
Result: We may put on or lose weight irrespective of age. **[2]**

Q. 4) One of the research scholars Mr. Anshul has commented that if Spearman's Rank method is used, then the sample correlation coefficient between weight and age of the practitioner will be exactly equal to 0.

What is the implicit value of the sum of squares of deviations in ranks when $r_{SPEARMAN} = 0$?

A. 56
B. 336
C. 48
D. 120. **[2]**

Q. 5) Another research scholar Mrs. Eesha has suggested that Kendall's method should be used as it has better statistical properties as compared to Spearman's method. However, Mr. Anshul is claiming that this is not required as the value of the sample correlation coefficient under Kendall's method would also be equal to 0 i.e. $r_{KENDALL} = 0$.

Which of the following statements is correct in relation to the claim made by Mr. Anshul?

A. As total number of rank pairs is an even number, number of concordant pairs and discordant pairs can be equal leading to $r_{KENDALL} = 0$.
B. As total number of rank pairs is an odd number, number of concordant pairs and discordant pairs can be equal leading to $r_{KENDALL} = 0$.
C. As total number of rank pairs is an odd number, number of concordant pairs and discordant pairs cannot be equal leading to $r_{KENDALL} \neq 0$.
D. As total number of rank pairs is an even number, number of concordant pairs and discordant pairs cannot be equal leading to $r_{KENDALL} \neq 0$. **[2]**

Use the following information for attempting questions 6 to 10:

Let X represent the height of one-year old infants in inches. It is assumed to be normally distributed.

$Y = e^X$ is lognormally distributed with parameters $\mu = 25$ and $\sigma^2 = 36$.

A paediatrician believes that generally infant height lies in the interval of mean height plus / minus 4 inches.

Q. 6) The claim of the paediatrician is proposed to be quantified statistically. Which of the following expression represents this correctly?

A. $P(| X - 4 | < 25)$
B. $P(| X - 25 | < 4)$
C. $P(| Y - 4 | < 25)$
D. $P(| Y - 25 | < 4)$. **[2]**

Q. 7) How much probable is the claim made by the paediatrician?

A. 49.52%
B. 50.00%
C. 74.76%
D. 25.24%. **[2]**

Q. 8) Which of the following is necessarily TRUE about E[Y]?

A. E[Y] = E[X]
B. E[Y] < E[X]
C. $E[Y] = e^{(E[X])}$
D. $E[Y] = E[e^{X}]$ **[2]**

Q. 9) Select the correct expression which represents the moment generating function of the random variable X i.e. $M_X(t)$ –

A. $M_X(t) = \exp(\frac{1}{2} t^2)$
B. $M_X(t) = \exp(25t + 36t^2)$
C. $M_X(t) = \exp(25t + 18t^2)$
D. $M_X(t)$ does not exist in closed form. **[2]**

Q. 10) Which of the following represents simulated value from the normal distribution of X corresponding to the value 0.40 from U (0, 1)?

[**Hint:** *For simulating from a normal distribution, following steps are to be followed:*
1. *Generate random number from U (0,1).*
2. *If u > 0.5, using standard normal tables, find z such that P (Z <= z) = u. Simulated value is z in this case.*
3. *If u < 0.5, find z such that P (Z <= z) = 1 – u. Simulated value is – z in this case.*
4. *Depending on the simulated value generated in 2 or 3, generate the simulated x value as* $x = \mu + \sigma z$.]

A. 26.52
B. 6.52
C. 14.12
D. 23.48. **[2]**

Use the following information for attempting questions 11 to 15:

Green-City is known for having rainfall all throughout the year. During the year, a few days are marked by torrential rainfall exceeding 5 centimetres.

A bi-variate linear regression model ($Y = \alpha + \beta * X + e$) is being fit to understand the dependency of –

- the number of trees falling in Green-City on a day with torrential rainfall (Y) on
- the rainfall in the city on that day in centimetres (X).

Data for 6 such days has been collected from the records of the municipal authorities. Summary for that data is presented below:

$$\sum_1^6 x = 68, \quad \sum_1^6 y = 85, \quad \sum_1^6 x^2 = 842, \quad \sum_1^6 xy = 1062, \quad \sum_1^6 y^2 = 1463.$$

Q. 11) What is the value of the coefficient β based on the collected data?

A. 1.3832
B. – 1.5093
C. 0.7230
D. 0.1132. **[2]**

Q. 12) A cyclone "Pawan" is expected to pass from the coastline adjoining Green City and an average daily rainfall of 15 centimetres is expected for the next three days in Green City.

The municipal authorities want to know the number of trees which are expected to fall during the cyclone "Pawan" so that they can keep their disaster task force ready. How many trees are expected to fall based on the fitted bi-variate regression model (round off your answer to the next integer)?

A. 10
B. 20
C. 25
D. 58. **[2]**

Q. 13) What is the value of residual sum of squares for the bi-variate linear regression model?

A. 258
B. 136
C. 122
D. 71. **[2]**

Q. 14) For predicting the results more accurately, an alternative model is being fit to the data which has the following regression equation:

$$Y = \sigma * \exp(\lambda * X + e)$$

Following additional calculations are made based on the collected data:

- $\sum_1^6 \ln y_i = 15.039379$,
- $\sum_1^6 x_i \ln y_i = 178.519074$.

What is the value of the coefficient λ for the alternative model?

A. 7.9256
B. 0.1132
C. 1.1345
D. 3.4007. **[2]**

Q. 15) What is the value of the coefficient σ for the alternative model fitted in question 14?

A. 1.0697
B. 3.4007
C. 0.7191
D. 2.0526. **[2]**

Use the following information for attempting questions 16 to 20:

A start-up insurance company offering bite-size insurance products has recently launched a policy which offers "performer and entertainer liability insurance" for stand-up comedians.

Performer liability has two aspects which need to be statistically modelled –

- Number of people from the audience filing a legal suit against the comedian during the year (N) and
- Total amount payable to the comedian under the policy during the year (Y).

$Y = \sum_1^N X = X_1 + X_2 + \ldots\ldots\ldots\ldots\ldots + X_N \ldots\ldots$ where X_i denotes the amount (INR in lakhs) payable in respect of the i^{th} claim for i = 1, 2, ……., N.

N follows a Binomial distribution with parameter p = 0.005 and X_i follows a Gamma Distribution with parameters $\alpha = ½$ and $\beta = 1/10$.

Mr. Prakhar, a renowned stand-up comedian in the country who is one of the policyholders of the company has participated in 10 stand-up events during the year with average audience size of 100 persons per event.

Q. 16) What is the value of $E(X_i)$ and $Var(X_i)$?

A. $E(X_i) = 5$ and $Var(X_i) = 50$
B. $E(X_i) = 10$ and $Var(X_i) = 100$
C. $E(X_i) = 20$ and $Var(X_i) = 100$
D. $E(X_i) = 10$ and $Var(X_i) = 50$. **[2]**

Q. 17) What is the expected number of claims and the standard deviation in the number of claims that Mr. Prakhar will file with the insurance company during the year?

A. E(N) = 5 and SD(N) = 4.975
B. E(N) = 5 and SD(N) = 2.230
C. E(N) = 10 and SD(N) = 9.950
D. E(N) = 10 and SD(N) = 3.154. **[2]**

Q. 18) What is the expected value of total claim amount from Mr. Prakhar during the year under the policy (INR in lakhs)?

A. 5
B. 50
C. 25
D. 100. **[2]**

Q. 19) Which of the following represents the correct expression for the unconditional variance of Y i.e. Var(Y)?

A. Var(Y) = E[Var(X|N)] + Var[E(X|N)]
B. Var(Y) = E[Var(N|Y)] + Var[E(N|Y)]
C. Var(Y) = E[Var(N|X)] + Var[E(N|X)]
D. Var(Y) = E[Var(Y|N)] + Var[E(Y|N)]. **[2]**

Q. 20) What is the standard deviation of the total claim amount from Mr. Prakhar during the year under the policy (INR in lakhs)?

A. 15.77
B. 19.35
C. 11.15
D. 26.92. **[2]**

Use the following information for attempting questions 21 to 25:

VIDEA Ltd is a leading telecommunication company in the country of Actuaria. Strategy team of the company has developed a metric "Customer Value" which measures the worth of the product or service to the customer.

Customer value (Y) is derived based on multiple explanatory variables:

- Usage (in seconds) (X_1),

- Subscription length (in months) (X_2),
- Age (in years) (X_3),
- Number of call failures (X_4).

The analysts from the strategy team have developed a multi-variate linear regression model based on data of 25 customers. Following information is presented to you:

Particulars	Coefficient	Standard Error	99% confidence interval for the coefficient
β_0	172.3588	9.6541	
β_1	0.0885	0.0036	(0.0783, 0.0987)
β_2	0.7825	0.1586	(0.3313, 1.2337)
β_3	-8.0144	0.6283	(-9.8019, -6.2269)
β_4	-0.4372	0.2537	(-1.1590, 0.2846)

The analyst has also constructed the following ANOVA table:

Source of variation	Degrees of freedom	Sum of squares	Mean sum of squares
Regression	4	?	1602.748925
Residual	?	2136.9986	?
Total	?	?	?

Q. 21) Based on the 99% confidence intervals calculated by the analysts, which explanatory variables can be considered to be significant?

A. X_1 and X_2
B. X_1 , X_2 and X_3
C. X_1 , X_2, X_3 and X_4
D. None of the variables are significant. **[2]**

Q. 22) What is the value of the Coefficient of Determination (R^2)? Answer using the ANOVA table constructed by the analyst.

A. 75%
B. 43%
C. 80%
D. 25%. **[2]**

Q. 23) What is the value of Adjusted R^2?

A. 10%
B. 76%
C. 32%
D. 70%. **[2]**

Q. 24) The Chief Operating Officer of the company is of the view that none of the explanatory variables have any significant impact on the customer value. Analysts from the strategy team are statistically testing this claim using one-sided analysis of variance (ANOVA).

What is the outcome of the ANOVA test at 5% level of significance?

A. View of the Chief Operating Officer is justified. None of the four explanatory variables have any impact on the customer value. **[2]**

B. View of the Chief Operating Officer is not justified. At least one of the four explanatory variables has a significant impact on the customer value.
C. View of the Chief Operating Officer is not justified. All the explanatory variables have a significant impact on the customer value.
D. ANOVA test cannot be applied to the dataset.

Q. 25) The Chief Strategy Officer has suggested that the model should be optimized and only those explanatory variables which have a significant impact on the customer value must be retained. Analysts have decided to use the process of backward selection for selecting the optimal number of explanatory variables.

Some output of this process is shown in the below table:

Model	**R^2**	**Adjusted R^2**
$X_1 + X_2 + X_3 + X_4$	?	?
$X_1 + X_2 + X_3$	76%	73%
$X_1 + X_2$	73%	70%
X_1	69%	68%

Select the option which represents the optimal set of explanatory variables for estimating customer value –

A. X_1 only
B. X_1 and X_2
C. X_1, X_2 and X_3
D. X_1, X_2, X_3 and X_4. **[2]**

Use the following information for attempting questions 26 to 30:

A random variable X is a continuous random variable representing the age in years of a historical artefact in a country's premier museum. X is believed to have the following probability density function: $f(x) = 3\lambda^3 (\lambda + x)^{-4}$ for $x > 0$.

In order to test the null hypothesis $H_0 : \lambda = 50$ against the alternative hypothesis $H_1 : \lambda = 60$, a single value is observed. If this value is greater than 93.50, H_0 is rejected.

Q. 26) In the context of the above hypothesis, what is definition of Type I Error and Type II Error?

	Type I Error	**Type II Error**
A.	Fail to reject H_0 when it is true	Reject H_0 when it is false
B.	Fail to reject H_0 when it is true	Fail to reject H_0 when it is false
C.	Reject H_0 when it is true	Reject H_0 when it is false
D.	Reject H_0 when it is true	Fail to reject H_0 when it is false

[2]

Q. 27) What is the probability of Type I Error?

A. 4.23%
B. 5.97%
C. 0.09%
D. 12.69%. **[2]**

Q. 28) What is the probability of Type II Error?

A. 4.23% **[2]**

B. 94.03%
C. 5.97%
D. 17.92%.

Q. 29) What is the definition of power of the test and size of the test?

	Power of Test	**Size of Test**
A.	1 – Probability (Type I Error)	Probability (Type II Error)
B.	Probability (Type II Error)	1 – Probability (Type I Error)
C.	1 – Probability (Type II Error)	Probability (Type I Error)
D.	1 – Probability (Type II Error)	1 – Probability (Type I Error)

[2]

Q. 30) What is the definition of sensitivity of the test and specificity of the test?

	Sensitivity of Test	**Specificity of Test**
A.	1 – Probability (Type I Error)	1 – Probability (Type II Error)
B.	Probability (Type II Error)	Probability (Type I Error)
C.	1 – Probability (Type II Error)	1 – Probability (Type I Error)
D.	Probability (Type I Error)	Probability (Type II Error)

[2]

Use the following information for attempting questions 31 to 35:

The civic administration of a metropolis is trying to implement preventive measures to reduce road accidents due to drink and drive.

Following is a sample of results of Breath Alcohol Content (BAC) test of civilians who have breached the legal limit of 8 basis points per 100 ml of blood. The sample (in basis points) is assumed to be taken from a normal distribution with mean μ and variance of 20.

56, 32, 49, 57, 44.

The administration has made it compulsory for every lounge in the city to administer a drug "Neurontin" as a part of the drink so that the alcohol absorption in the body will be delayed and it will continue to promote better alertness and attentiveness while driving. This is believed to reduce the BAC below the legally permissible level and thereby reduce the accidents due to drink and drive.

For the same participants among the sample, the BAC test is taken again after the administration of the drug and the results were found to be as follows:

8, 4, 7, 6, 5.

Q. 31) Which of the following represents a symmetrical 95% confidence interval for μ based on the original sample (in basis points)?

A. (43.68, 51.52)
B. (8.40, 8.68)
C. (42.95, 52.25)
D. (1.10, 94.10). **[2]**

Q. 32) Which of the following represents a symmetrical 95% prediction interval for a single value from this distribution based on the original sample (in basis points)?

A. (34.98, 60.22)
B. (36.20, 59.00) **[2]**

C. (38.00, 57.20)
D. (39.54, 55.66).

Q. 33) We want to reduce the width of each of these intervals. What should be done to reduce their width?

Choose the correct option from those given below which will result in reduction in the width of both these intervals:

	Reduce the Width of Symmetrical 95% Confidence Interval	**Reduce the Width of Symmetrical 95% Prediction Interval**
A.	Reduce the sample size n	Reduce the sample size n
B.	Increase the sample size n	Reduce the sample size n
C.	Reduce the sample size n	Increase the sample size n
D.	Increase the sample size n	Increase the sample size n

[2]

Q. 34) Based on the two samples – before and after the administration of the drug "Neurontin", the mean value of reduction in BAC levels and the standard deviation of reduction in BAC levels is –

A. Mean = 60 basis points, Standard Deviation = 1.41 basis points
B. Mean = 41.6 basis points, Standard Deviation = 8.01 basis points
C. Mean = 60 basis points, Standard Deviation = 1.58 basis points
D. Mean = 41.6 basis points, Standard Deviation = 8.96 basis points. **[2]**

Q. 35) Which of the following represents a 90% symmetrical confidence interval for the reduction in BAC levels (in basis points)?

A. (4.73, 72.70)
B. (34.38, 48.82)
C. (35.01, 48.19)
D. (33.06, 50.14). **[2]**

Use the following information for attempting questions 36 to 40:

In the country Economa, general elections are being conducted for electing new government for a period of next five years.

Ruling Alliance and Opposition Alliance are equally strong contenders and have an equal chance of winning the poll.

The increase in the country's stock market index in the next one year is assumed to follow a Pareto distribution (two parameter version with $\alpha = 1$) with the following probability density function:

$$f(x|\lambda) = \frac{\lambda}{(\lambda+x)^2} \qquad 0 < x < \infty, \quad \lambda > 0.$$

If the Ruling Alliance is re-elected to power, λ is expected to be equal to 100 and if the Opposite Alliance is elected, λ is expected to be equal to 300.

Q. 36) Which of the following represents the correct prior distribution of λ?

A. $f(\lambda) = \frac{\lambda}{(x+\lambda)^2}$
B. $P(\lambda = 100) = 0.5$, $P(\lambda = 300) = 0.5$ **[2]**

C. $f(\lambda) = \frac{x}{(x+\lambda)^2}$

D. $P(\lambda = 100) = 0.75$, $P(\lambda = 300) = 0.25$.

Q. 37) Which of the following correctly represents the distribution function of the two parameter Pareto distribution? Answer based on Actuarial Formulae and Tables.

A. $F(x|\lambda) = \frac{\lambda}{(\lambda+x)^2}$

B. $F(x|\lambda) = 1 - \frac{\lambda}{(\lambda+x)^2}$

C. $F(x|\lambda) = 1 - \frac{\lambda}{(\lambda+x)}$

D. $F(x|\lambda) = \frac{\lambda}{(\lambda+x)}$. **[2]**

Q. 38) Technical analysts believe that the stock market index of the country will grow by at least 500 points by the end of the year. Which alliance has a better chance of winning the general elections based on this information?

Hint: [*Answer based on the posterior distribution of λ i.e. f (λ | x > 500)*]

A. Ruling Alliance
B. Opposition Alliance
C. Both Ruling Alliance and Opposition Alliance are equally likely to win
D. None of them will win, elections will have to be re-conducted. **[2]**

Q. 39) What should be the minimum level of expected growth in the stock market index during the year, which will make change in the government virtually certain?

A. 200 points
B. 300 points
C. 500 points
D. None of the above. **[2]**

Q. 40) In the context of Bayesian statistics, loss function $L(g(\underline{x}), \theta)$ is a measure of the loss incurred when $g(\underline{x})$ is used as an estimator of θ. Bayesian estimator $g(\underline{x})$ is that which minimises the expected loss with respect to the posterior distribution.

Which of the following is the correct expression for "All-or-nothing" loss function?

A. $L(g(\underline{x}), \theta) = 0$ if $g(\underline{x}) = \theta$, $L(g(\underline{x}), \theta) = 1$ if $g(\underline{x}) \neq \theta$
B. $L(g(\underline{x}), \theta) = [g(\underline{x}) - \theta]^2$
C. $L(g(\underline{x}), \theta) = | g(\underline{x}) - \theta |$
D. $L(g(\underline{x}), \theta) = g(\underline{x}) - \theta$. **[2]**

Use the following information for attempting questions 41 to 45:

Let θ denote the proportion of wrong answers given by an AI Chatbot. Prior beliefs about θ are described by a beta distribution with parameters α and β. Let μ and σ^2 denote the prior mean and prior variance of θ.

A random sample of 'n' questions is taken and it is observed that wrong answers have been given in respect of 'w' questions.

Q. 41) Which of the following is TRUE about μ and σ^2?

[2]

A. $\mu = \alpha / (\alpha + \beta)$ and $\sigma^2 = \mu\beta / (\alpha + \beta)^2$
B. $\mu = \alpha / (\alpha + \beta)$ and $\sigma^2 = \mu\beta / [(\alpha + \beta)^2 + (\alpha + \beta)]$
C. $\mu = \alpha / (\alpha + \beta)$ and $\sigma^2 = \mu\beta / (\alpha + \beta)$
D. $\mu = \alpha / (\alpha + \beta)$ and $\sigma^2 = \mu\beta / (\alpha - \beta)(\alpha + \beta)$.

Q. 42) Experts in the AI company have estimated μ and σ^2 to be 10% and (9 / 1100) %% respectively.

The method of moments estimates for α and β are –

A. $\hat{\alpha}_{MOM} = 3$ and $\hat{\beta}_{MOM} = 3$
B. $\hat{\alpha}_{MOM} = 9$ and $\hat{\beta}_{MOM} = 1$
C. $\hat{\alpha}_{MOM} = 1$ and $\hat{\beta}_{MOM} = 9$
D. $\hat{\alpha}_{MOM} = 2$ and $\hat{\beta}_{MOM} = 4$. **[2]**

Q. 43) What is the posterior distribution of θ based on the random sample of n questions?

A. $\theta \sim$ Beta $(\alpha + n, \beta + n - w)$
B. $\theta \sim$ Beta $(\alpha + n - w, \beta + w)$
C. $\theta \sim$ Beta $(\alpha + n - w, \beta + n + w)$
D. $\theta \sim$ Beta $(\alpha + w, \beta + n - w)$. **[2]**

Q. 44) What is the maximum likelihood estimate of θ based on the random sample of n questions?

A. $\hat{\theta}_{MLE} = w / n$
B. $\hat{\theta}_{MLE} = (n - w) / n$
C. $\hat{\theta}_{MLE} = n / w$
D. $\hat{\theta}_{MLE} = w / (n + w)$. **[2]**

Q. 45) Credibility factor Z in this case is derived as: $n / (\alpha + \beta + n)$. Which of the following is a correct expression for the credibility estimate?

A. $Z * \mu + (1 - Z) * \hat{\theta}_{MLE}$
B. $Z * \hat{\theta}_{MLE} + (1 - Z) * \mu$
C. $Z * \mu + (1 - Z) * (1 / \hat{\theta}_{MLE})$
D. $Z * \hat{\theta}_{MLE} + (1 - Z) * (1 / \mu)$. **[2]**

Use the following information for attempting questions 46 to 50:

A matrimonial website calculates a compatibility score between profiles to assist individuals in better shortlisting.

Compatibility score (Y) is calculated based on following two important factors:

- **Education:**

Criteria	Education Score (X_1)
Both of them have professional qualifications	1
Both do not have professional qualifications	1
Other cases	0

- **Mode of Earning:**

Criteria	Mode of Earning Score (X_2)
Both of them are into public / private employment	1

Both of them are self-employed	0
One is employed and the other is self-employed	1
Other cases	0

An actuary has been employed by the website owner to predict the compatibility score. The actuary has fit a generalised linear model to predict the compatibility score based on the values of education score and mode of earning score.

Following information is available about the model:

- Poisson distribution has been used.
- Linear predictor is given by: $g(\mu) = \beta_0 + \beta_1 + \beta_2$
- "1" is the base value for Education Score (X_1).
- "0" is the base value for Mode of Earning Score (X_2).

Following is a snapshot of the model output achieved by the actuary:

Particulars	Coefficient
β_0	0.7732
$\beta_{1\ SCORE\ =\ 0}$	- 0.5725
$\beta_{1\ SCORE\ =\ 1}$	0
$\beta_{2\ SCORE\ =\ 0}$	0
$\beta_{2\ SCORE\ =\ 1}$	0.6931

In respect of profiles who have already met, both the individuals are asked to submit a feedback form and rate their compatibility with the other individual. The rating should be between 0% to 100%.

Q. 46) What is the canonical link function in case of a Poisson model?

A. $g(\mu) = \log \mu$
B. $g(\mu) = \mu$
C. $g(\mu) = \log (\mu / (1 - \mu))$
D. $g(\mu) = 1 / \mu$. **[2]**

Q. 47) Why are the values of the coefficients of $\beta_{1\ SCORE\ =\ 1}$ and $\beta_{2\ SCORE\ =\ 0}$ equal to 0?

A. Based on the current profiles of individuals registered on the website none of the pairs have got an education score of 1 or a mode of earning score of 0. Hence, these coefficients are equal to 0.
B. As the Poisson model is used for fitting the GLM, so these coefficients are equal to 0. If a normal model was used instead of Poisson, these coefficients wouldn't have been equal to 0.
C. Education score of 1 and mode of earning score of 0 are considered to be base values and hence the true value of their coefficients is implicitly covered in the intercept coefficient β_0.
D. There is an error in the model fit by the actuary and the same needs to be rectified. **[2]**

Q. 48) For Mr. Fast and Ms. Steady, following information is presented based on their profile details as per the website –

- Mr. Fast is a qualified Chartered Accountant whereas Ms. Steady is an Architect.
- Mr. Fast has his own CA practice whereas Ms. Steady works as a freelancer.

Select the correct value of compatibility score for Mr. Fast and Ms. Steady based on the fitted model. Use the link function $g(\mu) = \mu$.

[2]

A. 89.38%
B. 20.07%
C. 77.32%
D. 46.63%.

Q. 49) The log likelihood of the Poisson model is given by which of the following expressions?

A. $\log L = \sum_1^n y_i \log y_i + \sum_1^n y_i + \sum_1^n \log y_i !$
B. $\log L = \sum_1^n y_i \log y_i + \sum_1^n y_i - \sum_1^n \log y_i !$
C. $\log L = \sum_1^n y_i \log y_i - \sum_1^n y_i + \sum_1^n \log y_i !$
D. $\log L = \sum_1^n y_i \log y_i - \sum_1^n y_i - \sum_1^n \log y_i !$. **[2]**

Q. 50) Based on the fitted values of the compatibility score ($\hat{y}$) and the actual value of compatibility score (y), following information has been calculated:

$$\sum_1^n y_i \log(y_i / \hat{y}_i) = 28.6132, \qquad \sum_1^n (y_i - \hat{y}_i) = 16.2375.$$

Scaled deviance for the fitted model is –

A. 20.7672
B. 24.7514
C. 24.8475
D. 26.4799. **[2]**

****