INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

February 2025 Examination

Subject CS1A – Actuarial Statistics (Paper A)

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

Total Marks: 100

INDICATIVE SOLUTION

Introduction:

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions are only indicative. It is realized that there could be other points as valid answers and examiners have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1: Correct Answer is **Option B.**

Getting a sum of six on both dice is considered to be "success"

As X_i be the number of throws required in chance i to get success.

So, Xi has a Geometric Distribution (Type I) with parameter p (where p represents the probability of success).

Variance (X) = $(1 – p) / p^2$. **[1]**

Solution 2: Correct Answer is **Option E.**

$P(X_i >= x)$
$= P(X_i > (x-1))$
= probability that the number of throws required in a chance i is greater than (x-1)
= probability of failure in (x-1) throws
$= (1 – p)^{(x-1)}$. **[2]**

Solution 3: Correct Answer is **Option D.**

Value of probability p
= P (Getting six on both dice)
= Probability of getting the outcomes – {(5,1), (1,5), (2,4), (4,2), (3,3)}
= 1/36 + 1/36 + 1/36 + 1/36 + 1/36
= 5/36. **[1]**

Solution 4: Correct Answer is **Option A.**

P (Y >= y)
= P (Y > (y-1))
$= P (Min (X_1, X_2, ……., X_n) > (y-1))$
$= P (X_1 > (y-1)) * P(X_2 > (y-1)) * …………….. * P(X_n > (y-1))$
…… as X_is are independent random variables
$= (1 – p)^{(y-1)} * (1 – p)^{(y-1)} * ……………….. * (1 – p)^{(y-1)}$
$= ((1 – p)^{(y-1)})^n$ …….. as the same expression is multiplied n times
$= ((1 – p)^n)^{(y – 1)}$. **[2]**

Solution 5: Correct Answer is **Option C.**

A player will lose his money when the reward earned < charges paid to play the game.

Reward = (10 – Y) * 100 where $Y = Min (X_1, X_2, …….., X_n)$.

In numerical terms, player will lose his money when –
(10 – Y) * 100 < 700
i.e. (1000 – 100Y) < 700
i.e. 300 < 100Y
i.e. 100Y > 300
i.e. Y > 3
i.e. Y >= 4

P(Player losing his money)
= P(Y >=4)
= $((1 – (5/36))^{10})^{(4 – 1)}$
= 0.011266
= 1.1266%. **[3]**

Solution 6: Correct Answer is **Option A.**

S_{xx} = 3355 – 8 * 19.875^2 = 194.875
S_{yy} = 437 – 8 * 6.875^2 = 58.875
S_{xy} = 1128 – 8 * 19.875 * 6.875 = 34.875

$r_{PEARSON}$
= S_{xy} / sqrt (S_{xx} * S_{yy})
= 34.875 / sqrt (194.875 * 58.875)
= 0.32559. **[2]**

Solution 7: Correct Answer is **Option E.**

Z_r
= ½ * $\log_e$ ((1+r) / (1-r))
= $\tanh^{-1}$ (r)
= 0.337887. **[1]**

Solution 8: Correct Answer is **Option B.**

95% confidence interval for Z_r is [-0.45577, 0.684747].

Z_r = ½ * $\log_e$ ((1+r) / (1-r))
2 Z_r = $\log_e$ ((1+r) / (1-r))
$e^{2 Z_r}$ = (1+r) / (1-r)
r = ($e^{2 Z_r}$ - 1) / ($e^{2 Z_r}$ + 1)

Lower limit of 95% confidence interval for population correlation coefficient
= ($e^{2 * -0.45577}$- 1) / ($e^{2 * -0.45577}$+ 1)
= -0.49195

Upper limit of 95% confidence interval for population correlation coefficient
= ($e^{2 * 0.684747}$ - 1) / ($e^{2 * 0.684747}$+ 1)
= 0.837997.

So, 95% confidence interval for population correlation coefficient
= [-0.49195, 0.837997]. **[3]**

Solution 9: Correct Answer is **Option C.**

Let us calculate Spearman Rank's Correlation Coefficient.

$r_{SPEARMAN}$
= 1 – (6 * $\sum d^2$ / (n * (n^2 – 1)))

Let us calculate $\sum d^2$

X	Y	R_X	R_Y	D	d^2
15	4	2	2	0	0
25	3	7	1	6	36
21	11	5	8	-3	9
29	10	8	7	1	1
13	6	1	4	-3	9
18	9	4	6	-2	4
21	7	5	5	0	0
17	5	3	3	0	0
				Total:	**59.00**

r_{SPEARMAN}
$= 1 - (6 * 59 / (8 * (8^2 - 1)))$
= 0.297619.

From question 6,
r_{PEARSON}
= 0. 32559

So, $r_{\text{PEARSON}} > r_{\text{SPEARMAN}}$. **[3]**

Solution 10: Correct Answer is **Option D.**

r_{PEARSON} measures the strength of the linear relationship between X and Y.

r_{SPEARMAN} only measures the monotonic relationship between X and Y. It does not measure the strength of linear relationship between X and Y. **[1]**

Solution 11: Correct Answer is **Option A.**

The individual is 47 years old in 2024. So, the individual belongs to "Gen X".

P (N > 1 | X = "Gen X")
= 0.190 / (0.120 + 0.200 + 0.190)
= 37.25%. **[2]**

Solution 12: Correct Answer is **Option B.**

Probability of empty supermarket
= P (N=0)
= 0.12+0.15+0.15= 42%. **[1]**

Solution 13: Correct Answer is **Option E.**

P (X ≠ "Gen Z" | N >= 1)
= 1 – P (X = "Gen Z" | N >= 1)
= 1 – (0.030+0.010)/ (0.200 + 0.190 + 0.100 + 0.150 + 0.030 + 0.010)
= 1 – 0.068966
= 93.10%. **[2]**

Solution 14: Correct Answer is **Option C.**

Employed Individuals: (Observed Values)

	Gen X	**Millennials**	**Gen Z**	**Total**
No visit per month	0.175	0.105	0.070	**0.350**
One visit per month	0.225	0.135	0.090	**0.450**
More than one visit per month	0.100	0.060	0.040	**0.200**
Total:	**0.500**	**0.300**	**0.200**	**1.000**

G | E = "Employed" and N | E = "Employed" will be independent if the product of column totals and row totals equal to the observed values.

Employed Individuals: (Expected Values = Row Totals * Column Totals)

	Gen X	**Millennials**	**Gen Z**	**Total**
No visit per month	0.175	0.105	0.070	**0.350**
One visit per month	0.225	0.135	0.090	**0.450**
More than one visit per month	0.100	0.060	0.040	**0.200**
Total:	**0.500**	**0.300**	**0.200**	**1.000**

So, G | E = "Employed" and N | E = "Employed" are independent.

Self-Employed Individuals: (Observed Values)

	Gen X	**Millennials**	**Gen Z**	**Total**
No visit per month	0.080	0.080	0.040	**0.200**
One visit per month	0.200	0.200	0.100	**0.500**
More than one visit per month	0.120	0.120	0.060	**0.300**
Total:	**0.400**	**0.400**	**0.200**	**1.000**

G | E = "Self-Employed" and N | E = "Self-Employed" will be independent if the product of column totals and row totals equal to the observed values.

Self-Employed Individuals: (Expected Values = Row Totals * Column Totals)

	Gen X	**Millennials**	**Gen Z**	**Total**
No visit per month	0.080	0.080	0.040	**0.200**
One visit per month	0.200	0.200	0.100	**0.500**
More than one visit per month	0.120	0.120	0.060	**0.300**
Total:	**0.400**	**0.400**	**0.200**	**1.000**

So, G | E = "Self-Employed" and N | E = "Self-Employed" are independent. **[3]**

Solution 15: Correct Answer is **Option D.**

From question 14, we know that N and G are conditionally independent of each other for both the values of E.

However, we need to check based on original tables whether they are unconditionally independent of each other.

Employed + Self-Employed = All Individuals: (Observed Values)

	Gen X	Millennials	Gen Z	Total
No visit per month	0.120	0.150	0.150	**0.420**
One visit per month	0.200	0.100	0.030	**0.330**
More than one visit per month	0.190	0.050	0.010	**0.250**
Total:	**0.510**	**0.300**	**0.190**	**1.000**

Employed + Self-Employed = All Individuals: (Expected Values = Row Totals * Column Totals)

	Gen X	Millennials	Gen Z	Total
No visit per month	0.2142	0.1260	0.0798	**0.4200**
One visit per month	0.1683	0.0990	0.0627	**0.3300**
More than one visit per month	0.1275	0.0750	0.0475	**0.2500**
Total:	**0.5100**	**0.3000**	**0.1900**	**1.0000**

As the values in the observed values table and expected values table are not equal, we can conclude that N and G are not unconditionally independent of each other.

So, although N and G are conditionally independent of each other, they are not unconditionally independent of each other. **[2]**

Solution 16: Correct Answer is **Option A.**

β
$= S_{xy} / S_{xx}$
$= 62350 / 92362$
$= 0.675061.$ **[1]**

Solution 17: Correct Answer is **Option B.**

σ^2_hat
$= 1 / (n-2) * (S_{yy} – S^2_{xy} / S_{xx})$
$= 1 / 6 * (44844 – 62350^2 / 92362)$
$= 1 / 6 * 2753.936$
$= 458.9893.$

$se(\hat{\beta})$
$= (\sigma^2\text{\_hat} / S_{xx})^{1/2}$
$= (458.9893 / 92362)^{1/2}$
$= 0.07049.$ **[2]**

Solution 18: Correct Answer is **Option C.**

H_0: β = 0 against the alternate hypothesis H_1: β ≠ 0

95% symmetrical confidence interval for β
= $\hat{\beta}$ +/- {$t_{2.5\%, 6}$ * se($\hat{\beta}$)}
= 0.670651 +/- (2.447 * 0.07049)
= [0.498162, 0.84314].

As the 95% symmetrical confidence interval for β does not contain the value 0, we have sufficient evidence to reject H_0 at 5% level. **[2]**

Solution 19: Correct Answer is **Option D.**

For predicting the value of Y using the fitted regression line we need to estimate the value of parameter α.

α_hat
= y_bar – β_hat * x_bar
= 56.25 – 0.670651 * 59
= 16.4214.

So, the regression line of Y on X is:
Y = 16.4214 + 0.675061 * X

In case of the new political party X = 0.
So, Y_hat = 16.4214 = 16 (rounded off to nearest integer). **[2]**

Solution 20: Correct Answer is **Option E.**

$\hat{\beta}$ / se ($\hat{\beta}$)
= 0.675061 / 0.07049
= 9.58.

r
= S_{xy} / sqrt (S_{xx} * S_{yy})
= 62350 / sqrt (92362 * 44844)
= 0.968808

Let us evaluate each of the options in order to check equivalence with the value of the test statistic.

A. $r * \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$ = 0.968808 * sqrt (1 – 0.968808^2) / sqrt(6) = 0.098014

B. $r * \frac{\sqrt{n-1}}{\sqrt{1-r^2}}$ = 0.968808 * sqrt (7) / sqrt (1 – 0.968808^2) = 10.34337

C. $r * \frac{\sqrt{1-r^2}}{\sqrt{n-1}}$ = 0.968808 * sqrt (1 – 0.968808^2) / sqrt(6) = 0.036867

D. $r * \frac{\sqrt{n-2}}{\sqrt{r^2-1}}$ = = 0.968808 * sqrt (6) / sqrt (1+0.968808^2) = 1.704396

E. $r * \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$ = 0.968808 * sqrt (6) / sqrt (1 – 0.968808^2) = 9.576096 = 9.58

Value of Option E is equal to the value of the test statistic. So the right answer is Option E. **[3]**

Solution 21: Correct Answer is **Option C.**

$L(\lambda)$ = Product (λ * exp($-\lambda x$))

Log $L(\lambda)$
$= \sum(\log\lambda - \lambda x)$
$= n.\log\lambda - \lambda \sum x$

$d/d\lambda$ (Log$L(\lambda)$) = n/ - $\sum x$

Equating $d/d\lambda$ (Log$L(\lambda)$) to 0

$n/\lambda - \sum x = 0$

λ_hat = $n / \sum x$.

$d^2/d\lambda^2$ (Log$L(\lambda)$) = $- n / \lambda^2 < 0$ ……….. so this is indicative of maxima **[3]**

Solution 22: Correct Answer is **Option D.**

Cramer's Rao Lower Round for λ
$= - 1 / d^2/d\lambda^2$ (Log$L(\lambda)$)
$= \lambda^2 / n$. **[1]**

Solution 23: Correct Answer is **Option E.**

$P(X=0) = \exp(-\lambda)$
$P(X>0) = 1 - \exp(-\lambda)$

$L(\lambda)$
$= P(X=0)^m * P(X>0)^{(n-m)}$
= exp($-\lambda$m) * (1 – exp($-\lambda$))^(n-m). **[2]**

Solution 24: Correct Answer is **Option B.**

$L(\lambda)$ = exp($-\lambda$m) * (1 – exp($-\lambda$))^(n-m).
Log$L(\lambda)$ = $-\lambda m$ + (n-m) * log(1-exp($-\lambda$))

$d/d\lambda$ (Log$L(\lambda)$) = - m + (n – m) * exp($-\lambda$) / (1-exp($-\lambda$))

Equating $d/d\lambda$ (Log$L(\lambda)$) to 0

-m + n * exp($-\lambda$) = 0
n * exp($-\lambda$) = m
λ_hat = log(n/m).

$d^2/d\lambda^2$ (Log$L(\lambda)$) < 0 ………….. maxima. **[3]**

Solution 25: Correct Answer is **Option A.**

λ_hat = log(n/m)

If m = 0, λ_hat = log(n/0) = log(∞) = ∞

So, minimum value of m (swimmers who cannot swim underwater) needs to be equal to 1 in order to ensure finite value of λ_hat. **[1]**

Solution 26: Correct Answer is **Option B.**

Average default size for the entire population based on 5-year data i.e. E[m(θ)]
= [(1250 * 85000) + (2450 * 72000) + (3410 * 90000)] / (1250 + 2450 + 3410)
= INR 82,918.425 **[2]**

Solution 27: Correct Answer is **Option C.**

Expected average default size for Education Loans for the year 2024
= Expected Aggregate Defaults on Education Loans / Target for Education Loans in 2023
= 2.52 * 10^7 / 300
= INR 84,000

Credibility Estimate per unit for Education Loans
= $Z_E * \overline{X_E} + (1 - Z_E) * E[m(\theta)]$

84,000 = Z_E * 85,000 + (1 – Z_E) * 82918.425
(84,000 – 82918.425) = (85,000 – 82918.425) * Z_E
Z_E = 1081.575 / 2081.575
Z_E = 0.5196. **[2]**

Solution 28: Correct Answer is **Option D.**

$Z_E = P_E / (P_E + E[s^2(\theta)] / Var[m(\theta)])$

Let $E[s^2(\theta)] / Var[m(\theta)]$ = a
0.5196 = 1250 / (1250 + a)
a = (-1250*0.5196 + 1250)/0.5196
a = 1155.697

Credibility Factor for Gold Loans
= Z_G
= $P_G / (P_G + E[s^2(\theta)] / Var[m(\theta)])$
= 3410 / (3410 + a)
= 3410 / (3410 + 1155.697)
= 0.7469. **[3]**

Solution 29: Correct Answer is **Option A.**

Credibility Estimate per unit for Gold Loans
= Z_G * $\overline{X_G}$ + (1 – Z_G) * E[m(θ)]
= 0.7469 * 90,000 + (1 – 0.7469) * 82,918.425
= INR 88,207.6534

Expected total defaults for the year 2024 under Gold Loans
= 88,207.6534 * 1,100
= INR 9.70 crore. **[2]**

Solution 30: Correct Answer is **Option E.**

EBCT Model 1 and EBCT Model 2 give exactly similar results only if equal weight is applied to each risk in each year. Hence, Option A gives the correct answer. **[1]**

Solution 31: Correct Answer is **Option C.**

Using the central limit theorem,
$\overline{X_N} \sim N(\mu_N, \sigma^2 / n)$
$\overline{X_S} \sim N(\mu_S, \sigma^2 / n)$ **[1]**

Solution 32: Correct Answer is **Option E.**

It is given that the two samples N and S are independent random samples from a normal distribution.

Hence,
$E(\overline{X_N} - \overline{X_S}) = E(\overline{X_N}) - E(\overline{X_S}) = \mu_N - \mu_S$
$Var(\overline{X_N} - \overline{X_S}) = var(\overline{X_N}) + var(\overline{X_S}) = 2\sigma^2 / n$

So, $(\overline{X_N} - \overline{X_S}) \sim N(\mu_N - \mu_S, 2\sigma^2 / n)$ **[1]**

Solution 33: Correct Answer is **Option B.**

Under the given rules, we reject H_0 when the confidence intervals do not overlap.

Students from North have apparently scored better than students from South. Hence, Confidence Interval N would generally be numerically higher than Confidence Interval S.

95% Confidence Interval N:
= $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}, \overline{X_N} + 1.96 * \sigma / \sqrt{n})$

95% Confidence Interval S:
= $(\overline{X_S} - 1.96 * \sigma / \sqrt{n}, \overline{X_S} + 1.96 * \sigma / \sqrt{n})$

The two will never overlap if the upper limit of Confidence Interval S is lower than the lower limit of Confidence Interval N.

In numerical terms, the two confidence intervals never overlap when:
$(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) > (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$.

[3]

Solution 34: Correct Answer is **Option A.**

Under the given hypotheses, we accept (do not reject) H_0 when the confidence intervals do overlap.

Students from North have apparently scored better than students from South. Hence, Confidence Interval N would generally be numerically higher than Confidence Interval S.

95% Confidence Interval N:
$= (\overline{X_N} - 1.96 * \sigma / \sqrt{n}, \overline{X_N} + 1.96 * \sigma / \sqrt{n})$

95% Confidence Interval S:
$= (\overline{X_S} - 1.96 * \sigma / \sqrt{n}, \overline{X_S} + 1.96 * \sigma / \sqrt{n})$

The two will overlap if the upper limit of Confidence Interval S lies in the Confidence Interval N.

In numerical terms, the two confidence intervals overlap when:
$(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) < (\overline{X_S} + 1.96 * \sigma / \sqrt{n}) < (\overline{X_N} + 1.96 * \sigma / \sqrt{n})$. **[3]**

Solution 35: Correct Answer is **Option D.**

$\overline{X_N} = 561.40$
$\overline{X_S} = 547.20$

The hypothesis will be rejected when the two confidence intervals do not overlap.

In other words, we need to check that Option which ensures that:
$(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) > (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$

Let us check each option one by one:

A. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) = 561.40 - 1.96 * 30 / \text{sqrt}(10) = 542.8058$
$(\overline{X_S} + 1.96 * \sigma / \sqrt{n}) = 547.20 + 1.96 * 30 / \text{sqrt}(10) = 565.7942$
So, $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) < (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$

B. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) = 561.40 - 1.96 * 25 / \text{sqrt}(10) = 545.9048$
$(\overline{X_S} + 1.96 * \sigma / \sqrt{n}) = 547.20 + 1.96 * 25 / \text{sqrt}(10) = 562.6952$
So, $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) < (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$

C. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) = 561.40 - 1.96 * 20 / \text{sqrt}(10) = 549.0039$
$(\overline{X_S} + 1.96 * \sigma / \sqrt{n}) = 547.20 + 1.96 * 20 / \text{sqrt}(10) = 559.5961$
So, $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) < (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$

D. $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) = 561.40 - 1.96 * 10 / \text{sqrt}(10) = 555.2019$
$(\overline{X_S} + 1.96 * \sigma / \sqrt{n}) = 547.20 + 1.96 * 10 / \text{sqrt}(10) = 553.3981$
So, $(\overline{X_N} - 1.96 * \sigma / \sqrt{n}) > (\overline{X_S} + 1.96 * \sigma / \sqrt{n})$.

So, Option D is the right answer. **[3]**

Solution 36: Correct Answer is **Option D.**

We know that:
$Y_i = \beta * x_i + e_i$ where $e_i \sim N(0, \sigma^2)$.

$E(Y_i | x_i) = E(\beta * x_i + e_i) = \beta * x_i + 0 = \beta * x_i$
$Var(Y_i | x_i) = Var(\beta * x_i + e_i) = 0 + var(e_i) = \sigma^2$

So, $Y_i | x_i \sim N(\beta * x_i, \sigma^2)$. **[1]**

Solution 37: Correct Answer is **Option B.**

We want to find that value of β which minimizes the error squares.

$e^2 = Sum (Y – Y\_hat)^2 = Sum (Y – \beta x)^2$

Differentiating this with respect to β,

$de^2 / d\beta = d/d\beta (Sum (Y – \beta x)^2)$

$de^2 / d\beta$
$= d/d\beta (Sum (Y – \beta x)^2)$
$= Sum (2 * (Y – \beta x) * (-x))$
$= -2 * Sum(x(Y- \beta x))$

Equating $de^2 / d\beta$ to 0,

$-2 * Sum(x(Y- \beta x)) = 0$
$Sum(x(Y- \beta x)) = 0$

$Sum (xY) = Sum (\beta x^2)$
$B = Sum (xY) / Sum (x^2)$

Let us take a second order derivative to check minima.

$d(e^2)^2 / d\beta^2$
$= d/d\beta (-2 * Sum(x(Y- \beta x)))$
$= d/d\beta (-2 * Sum (\beta x^2))$
$= -2 * Sum (x^2) < 0$ …………….. this is indicative of minima.

Hence, the least square estimate of β is $Sum (xY) / Sum (x^2)$. **[3]**

Solution 38: Correct Answer is **Option A.**

$E(\beta\_hat\_LSE)$
$= E(Sum (xY) / Sum (x^2))$
$= Sum (x * E(Y)) / Sum(x^2)$
$= Sum (x * \beta x) / Sum(x^2)$
$= \beta * Sum(x^2) / Sum(x^2)$
$= \beta$. **[2]**

Solution 39: Correct Answer is **Option C.**

Var(β_hat_LSE)
= Var(Sum (xY) / Sum (x^2))
= Sum(x^2 * Var(Y)) / $(\text{Sum}(x^2))^2$
= Sum (x^2 * σ^2) / $(\text{Sum}(x^2))^2$
= σ^2 * Sum (x^2) / $(\text{Sum}(x^2))^2$
= σ^2 / Sum (x^2). **[2]**

Solution 40: Correct Answer is **Option E.**

Bias(β_hat_LSE)
= E(β_hat_LSE) – β
= β – β
= 0

MSE(β_hat_LSE)
= Var(β_hat_LSE) + Bias^2(β_hat_LSE)
= σ^2 / Sum (x^2) + 0
= σ^2 / Sum (x^2). **[2]**

Solution 41: Correct Answer is **Option B.**

X represents the number of trading days required for the trigger of the first circuit breaker during the financial year. If trigger of the first circuit breaker is termed as success, X is the number of trials which are needed to get the first success.

So, X follows a Geometric distribution. **[1]**

Solution 42: Correct Answer is **Option D.**

$L(p) = (1 – p)^{n1-1} * p$

$\text{LogL}(p) = (n_1-1) * \log(1-p) + \log p$

$d/dp\ (\text{LogL}(p)) = (-1)(n_1 – 1) / (1 – p) + 1/p$

Equating d/dp (LogL(p)) to 0,
$(-1)(n_1 – 1) / (1 – p) + 1/p = 0$
$(-p)\ (n_1 – 1) + (1-p) = 0$
$-n_1p + p + 1 – p = 0$
$1 = n_1p$
$p = 1/n_1$

Taking the second order derivative to check maxima,
d^2/dp^2 (LogL(p))
$= -1/p^2 + (-1)(n_1-1)/(1-p)^2 < 0$ ………………. this is indicative of maxima.

So, the maximum likelihood estimate of p is $1/n_1$. **[3]**

Solution 43: Correct Answer is **Option A.**

$f_{prior}(p) = 1$ …… $0 <= p <= 1$.

$L(p) = (1 – p)^{n1-1} * p$

fposterior(p)
$= f_{prior}(p) * L(p)$
$= (1 – p)^{n1-1} * p$.

This corresponds to a Beta distribution with parameters $\alpha = 2$ and $\beta = n_1$.

So, posterior distribution of p is Beta ($\alpha = 2$ and $\beta = n_1$). **[2]**

Solution 44: Correct Answer is **Option C.**

Prior mean of p = (1+0) / 2 = ½

Posterior mean of p = $\alpha / (\alpha+\beta) = 2 / (2+ n_1)$. **[1]**

Solution 45: Correct Answer is **Option E.**

The quadratic loss estimate of p

= Mean of posterior distribution
$= 2 / (2+ n_1)$
$= 1 / (2+n_1) + 1 / (2+n_1)$
$= n_1/(2+n_1) * 1/n_1 + 2/(2+n_1) * ½$
$= n_1 / (2+n_1) * 1/n_1 + (1- n_1/(2+n_1)) * ½$

This can be compared with the relationship:

Posterior Mean = Z * p_{MLE} + (1 – Z) * Prior Mean

……………..where Z is the credibility factor.

So, $Z = n_1 / (2+n_1)$. **[3]**

Solution 46: Correct Answer is **Option C.**

If the Gamma distribution with parameters α and λ is written in the form of a density function for a member of exponential family,

$\phi = \alpha$

$a(\phi) = 1/ \phi$

$a(\phi) = 1 / \alpha$

$\theta = -1/\mu = -1/(\alpha/\lambda) = -\lambda/\alpha$

$b(\theta) = - \log(-\theta) = - \log(\lambda/\alpha)$. **[2]**

Solution 47: Correct Answer is **Option A.**

Using Model 1,

$g(\mu)$
$= \beta_0 + \beta_1 + \beta_2$
$= 3.25 + 0 + (-0.72)$
$= 2.53$

Log link function has been used.
So,
$g(\mu) = \log(\mu)$
$\log(\mu) = 2.53$
$\mu = \exp(2.53) = 12.5535.$ **[2]**

Solution 48: Correct Answer is **Option E.**

So, for Ms. Merlin, the predicted value of the claim using Model 2 was 24.5537 lakhs.

$\mu = 24.5537$

Using the log link function i.e. $h(\mu) = \log(\mu)$,

$h(\mu) = \log_e(24.5537) = 3.200863.$

Using Model 2,

$h(\mu) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$

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.

Since, the given case is of a female admitted in a private hospital, only the intercept coefficient α_0 (which also captures the interaction term for the base values of female and private hospital) is relevant. All other coefficients are not relevant to the given case.

$3.200863 = \alpha_0$
$\alpha_0 = 3.20086.$ **[2]**

Solution 49: Correct Answer is **Option B.**

So, for Mrs. Ambedkar, the predicted value of the claim using Model 2 was 15.0813 lakhs.

$\mu = 15.0813$

Using the log link function i.e. $h(\mu) = \log(\mu)$,

$h(\mu) = \log_e(15.0813) = 2.713456$.

Using Model 2,

$h(\mu) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$

From the previous question, we know intercept coefficient $\alpha_0 = 3.20086$.

The given example is of a female admitted in a government hospital. So, along with the intercept coefficient α_0, the interaction term between female and public hospital i.e. α_1 will be relevant. Other coefficients are not relevant in this case.

$2.713456 = \alpha_0 + \alpha_1$
$2.713456 = 3.20086 + \alpha_1$
$\alpha_1 = - 0.4874$. **[3]**

Solution 50: Correct Answer is **Option D.**

The lower the AIC of the model, the better is the fit of the model.

Hence, Model 2 will be a better fit as compared to Model 1 if –

AIC (Model 2) < AIC (Model 1). **[1]**

****