

INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

20th October 2009

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00 Hrs)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. Please read the instructions on the front page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception
2. Mark allocations are shown in brackets.
3. Attempt all questions, beginning your answer to each question on a separate sheet. However, answers to objective type questions could be written on the same sheet.
4. In addition to this paper you will be provided with graph paper, if required.

AT THE END OF THE EXAMINATION

Please return your answer book and this question paper to the supervisor separately.

- Q. 1)** A sample of 280 observations has mean 54 and standard deviation 3. It was found later that two observations which should read as 62 and 82 had been wrongly recorded as 64 and 80 respectively.

Calculate the correct mean and standard deviation.

[3]

- Q. 2)** Find the probability that a leap year selected at random will contain 53 Sundays.

[2]

- Q.3)** A random variable X has a mixed distribution with probability function

$$P(X = 0) = a;$$

$$f_X(x) = \begin{cases} bx^2(1-x), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

If mean of X is " a ", find a and b , and hence, find the standard deviation of X .

[3]

- Q.4)** If $G(t)$ is the Probability Generating Function of a random Variable X and $a > 0$, then show that

$$E\left(\frac{1}{X+a}\right) = \int_0^1 t^{(a-1)} G(t) dt$$

[2]

- Q. 5)** Let X and Y be jointly distributed with probability density function,

$$f(x, y) = \begin{cases} \frac{1}{4}(1+xy), & |x| < 1, |y| < 1 \\ 0, & \text{otherwise} \end{cases}$$

Show that X and Y are not independent but X^2 and Y^2 are independent

[5]

- Q. 6)** If the random variable X follows Poisson distribution with parameter λ , then show that $E(|X-1|) = \lambda - 1 + 2e^{-\lambda}$

[4]

- Q. 7)** A random variable X has the following probability function:

$$p(x) = \begin{cases} \frac{c}{x(x+1)}, & x = 1, 2, 3, \dots \\ 0, & \text{otherwise} \end{cases}$$

- (a) Find the value of C. (1)
- (b) Show that $E(X)$ does not exist (1)
- (c) If $M_X(t)$ (for $t < 0$) is the Moment Generating Function then, prove that
- $$(1 - e^t) \frac{1 - e^t}{e^t} = e^{(M_X(t) - 1)} \quad (4)$$

[6]

- Q.8)** An executive (HR) of a life insurance company desires to analyse the salaries of 27 actuarial staff employed by his company. The Manager (HR) advised the executive that salary should be based on number of actuarial papers cleared and years of professional experience in life insurance industry.

Suggest a suitable statistical model to quantify, defining the symbols used.

[3]

- Q.9)** Let X_1, X_2, \dots, X_n be a random sample with common *pdf*

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

If $U = \text{Min}(X_1, X_2, \dots, X_n)$ and $V = \text{Max}(X_1, X_2, \dots, X_n)$

- (a) Find the *pdf* of U (2)
- (b) Find mean and variance of U (3)
- (c) Find the *pdf* of V (1)
- (d) Show that
- $$E(V) = n E(U)$$
- $$\text{Var}(V) = \text{Var}(U) \quad (2)$$

[8]

- Q.10)** (a) Define Critical Region of a test? (1)
- (b) Let p denote the probability of getting a head when a given coin is tossed once. Suppose that the hypothesis $H_0: p = 0.5$ is rejected in favor of $H_1: p = 0.6$, if 10 trials results in 7 or more heads.

Calculate the size and power of the test (4)

[5]

Q.11) Two independent normal random variables X and Y have means μ and $\rho\mu$ respectively, and both have variance 1.

(a) Show that the distribution of random variable $Y - \rho X$ depends on ρ , but not on μ (2)

(b) A single observation is made on X and Y , resulting in values x and y respectively.

Use your result to test the hypothesis $\rho = 1$ against the alternative $\rho \neq 1$ at 5% level (2)
when $x = 0.3$ and $y = 2.7$

(c) Comment on the result of (b) (1)

[5]

Q.12) Let Y be the number of patients dying in a region among x patients tested positive H1N1 virus, is to be modelled as a Poisson random variable with mean θx , where θ is unknown.

Suppose data is available from n independent regions: region i with x_i patients tested positive H1N1 virus of which y_i have died, $i = 1, 2, 3, \dots, n$.

The least squares estimator of θ is that value of θ for which

$\sum_{i=1}^n (Y_i - E(Y_i))^2$ is minimized.

(a) Show that the least squares estimator of θ is given by:

$$\bar{\theta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} \quad (3)$$

(b) Find $\bar{\theta}$, the maximum likelihood estimator of θ . (3)

(c) Examine whether $\bar{\theta}$ and $\bar{\theta}$ provide unbiased estimators of θ . (2)

(d) The table below gives the region- wise patients affected by H1N1 virus and the number of patients died:

| Region | Number of patients affected by H1N1 virus | Number of deaths |
|--------|---|------------------|
| 1 | 41,773 | 534 |
| 2 | 77,578 | 1,025 |
| 3 | 1,205 | 2 |
| 4 | 22,422 | 74 |
| 5 | 9,965 | 97 |
| 6 | 29,223 | 67 |

Find the value of $\bar{\theta}$ and $\bar{\theta}$ from the given data.

(3)

[11]

- Q.13)** Two catalysts are being analysed to determine how they affect the mean yield of a chemical process. Specifically, catalyst 1 is currently in use, but catalyst 2 is acceptable. Since catalyst 2 is cheaper, it should be adopted, provided it does not change the process yield.

A test is run in the pilot plant and results are shown in the following table:

| Observation number | Catalyst 1 (x) | Catalyst 2 (y) |
|--------------------|----------------|----------------|
| 1 | 91.50 | 89.19 |
| 2 | 94.18 | 90.95 |
| 3 | 92.18 | 90.46 |
| 4 | 95.39 | 93.21 |
| 5 | 91.79 | 97.19 |
| 6 | 89.07 | 97.04 |
| 7 | 94.72 | 91.07 |
| 8 | 89.21 | 92.75 |

- (a) Test at 5% level with clearly mentioning the hypothesis, is there any difference between the mean yields, if:
- (i) Variance of x (σ_x^2) = 8.04 and Variance of y (σ_y^2) = 14.84 (4)
- (ii) Equal variance and unknown (5)
- (b) Test whether the variances are equal at 5% level (2)
- (c) Find 95% confidence interval for the difference of mean yields using result in (b). (2)
- (d) Comment on the results as obtained in (c) (1)

[14]

- Q.14)** For a particular week, the following table gives minimum temperature (X), maximum temperature (Y) on the weekdays and maximum temperature (Z) on the week-end recorded in 11 Indian cities. (The temperatures are in Degree Celsius).

| City | Minimum Temperature (x) | Maximum Temperature on a Weekday (y) | Maximum Temperature on a Weekend (z) |
|------|-------------------------|--------------------------------------|--------------------------------------|
| 1 | 10.00 | 22.40 | 20.70 |
| 2 | 8.00 | 21.84 | 20.43 |
| 3 | 13.00 | 21.35 | 25.31 |
| 4 | 9.00 | 22.15 | 20.56 |
| 5 | 11.00 | 22.27 | 20.80 |
| 6 | 14.00 | 20.43 | 21.19 |
| 7 | 6.00 | 20.25 | 20.21 |
| 8 | 4.00 | 17.66 | 19.90 |
| 9 | 12.00 | 21.90 | 20.95 |
| 10 | 7.00 | 21.15 | 20.31 |
| 11 | 5.00 | 19.05 | 20.09 |

$$\sum x = 99.00; \quad \sum x^2 = 1,001.00; \quad \sum y = 230.45; \quad \sum y^2 = 4,850.30; \quad \sum xy = 2,104.85$$

- (a) Draw a scatter plot of y against x, and comment on whether a line of regression is appropriate to model the relationship between minimum temperature and maximum temperature on weekday (2)
- (b) Draw the scatter plot of z against x, and comment on your plot (2)
- (c) Fit a least squares regression line of y on x and calculate its parameters. (3)
- (d) Show that the least square fitted regression equation of z on x is same regression line as obtained in above (c) (3)
- (e) Calculate the correlation coefficient of
 - (i) x and y,
 - (ii) x and z,

Comment on result (2)
- (f) It was later found that the noting of the values for city 3 were wrong, so decided to omit those values.
 - (i) Calculate the regression line of z on x (5)
 - (ii) Calculate the correlation coefficient of x and z (1)
- (g) Comment briefly on your result as obtained in above (f)(ii) (1)

[19]

- Q.15)** (a) State the mathematical model of one-way ANOVA defining all notations and assumptions (2)
- (b) Random sample of claim amounts (in units of Rs.1,000 over a total 3-year period) under hospitalization reimbursement policies were taken from 5 different general insurance companies are shown below.

| Company | | | | |
|---------|-------|-------|-------|-------|
| y_1 | y_2 | y_3 | y_4 | y_5 |
| 60 | 40 | 33 | 55 | 88 |
| 52 | 56 | 0 | 78 | 89 |
| 11 | 123 | 0 | 99 | 44 |
| 33 | 0 | 12 | 234 | 78 |
| 1 | 12 | 19 | 45 | 85 |
| 87 | 54 | 23 | 67 | |
| 110 | 77 | | 67 | |
| | 24 | | | |

$$\sum y_1 = 354 ; \quad \sum y_2 = 386 ; \quad \sum y_3 = 87 ; \quad \sum y_4 = 645 ; \quad \sum y_5 = 384$$

$$\sum y_1^2 = 27,184 ; \quad \sum y_2^2 = 29,430 ; \quad \sum y_3^2 = 2,123 ; \quad \sum y_4^2 = 84,669 ; \quad \sum y_5^2 = 30,910$$

Test for the equality of mean claim amounts of the five companies, using ANOVA. (4)

(c) A marketing analyst claims that salaries of actuarial students do not depend on number of actuarial papers they have cleared. To test his claim, he collects data of 158 actuarial students distributed according to their annual salaries and number of papers cleared as shown below

| Papers cleared | Salary per annum (in Rs. lacs) | | | |
|----------------|---------------------------------|-------|--------|---------|
| | 3 - 5 | 5 - 8 | 8 - 10 | 10 - 12 |
| 0 - 3 | 45 | 20 | 6 | 5 |
| 4 - 6 | 7 | 20 | 9 | 6 |
| 7 - 9 | 5 | 8 | 15 | 12 |

Conclude your view statistically based on the above data using a χ^2 test (4)

[10]
