INSTITUTE OF ACTUARIAL
& QUANTITATIVE STUDIES

Subject: Probability and Statistics II

Chapter: Unit 1 & 2

Category: Assignment Questions

1. Unit 2

Claim sizes on a fixed benefit health insurance policy are normally distributed about a mean of INR 900 and with a standard deviation of INR 100. Claim sizes on an indemnity-based health insurance policy are normally distributed about a mean INR 1,400 and with a standard deviation of INR 300. All claim sizes are assumed to be independent and in units of INR 100.

To date, there have already been fixed benefit health insurance claims amounting to INR 900 and no indemnity-based health insurance claims. Assuming that there will be further 4 fixed benefit health insurance claims and 3 indemnity-based health insurance claims in next year, calculate the probability that the total claim amount under the indemnity-based health insurance claims exceeds the total claim amount under fixed benefit health insurance claims.

2. Unit 1

Suppose $x_i: i = 1,2,\ldots n$ are n independent and identically distributed random variables, each with mean μ and variance σ^2. Let $\bar{x} = \sum_1^n \frac{x_i}{n}$ be sample mean and $S^2 = \frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$ be the sample variance.

i) Show that the mean and variance of $\bar{x}$ *is* μ *and* $\frac{(\sigma^2)}{n}$ respectively.

ii) Show that $E(S^2) = \sigma^2$.

x_i 's are from a normal population. Using the distribution of $\frac{(n-1)(S^2)}{\sigma^2}$ as chi square

iii) Show that variance of $S^2 is \frac{2(\sigma^4)}{n-1}$.

iv) Find the probability that S^2will fall between plus or minus 50% of its expected value when n=10 and $\sigma^2 = 100$.

3. Unit 2

An insurer believes that the distribution of the number of claims on a particular type of policy is binomial with parameters n = 4 and p. A random sample of 180 policies revealed the following information.

No. of claims	0	1	2	3	4
No. of policies	86	75	16	2	1

i) Obtain the maximum likelihood estimate of p.

ii) Carry out a goodness of fit test for the number of claims on each policy conforms to the binomial model.

UNIT 1 & 2

ASSIGNMENT QUESTIONS

4. Unit 1

The probability density function of a random variable X, is given by

$$\mathrm{f(x)} = \frac{\mathrm{c}\beta^2}{(\mathrm{x}+\beta)^4} : \mathrm{x} > 0, \beta > 0$$

where c is a constant and β is a parameter.

i) Determine the value of c and calculate the mean and variance of X as a function of β by using Formulae and Tables for Actuarial Examinations or otherwise.

It is required to estimate β based on a random sample $X_1, X_2, \dots, X_n$
ii) Show that the method of moments estimator $\bar{\beta}$ is $2X_n$ and verify the unbiasedness and consistency of this estimator.

iii) Consider the set of estimators of the form bX_n, where b is a constant. Show that the value of β that minimizes the MSE of $\mathrm{X_n}$ is $2/(1+3/\mathrm{n})$.

iv) Compare the unbiasedness and consistency of the estimator in (iii) with minimum b using the corresponding properties of the estimator in (ii).

5. Unit 1
i) Show that the method of moments estimate for " a ' of a continuous uniform distribution $U(a, b)$ is $\hat{a} = \bar{x} - \sqrt{3}s$ where $\bar{x}$ is sample mean and s is sample standard deviation.

ii) State a formula for $\hat{b}$ (method of moments estimate for 'b') in terms of sample mean and sample standard deviation.

iii) Create a sample of five observations and use the sample to demonstrate potential weakness of the method of moment's estimation of ' a ' and ' b ' for $U(a, b)$.

iv) Find the maximum likelihood estimate of b for $U(a, b)$ based on a sample $x_1, x_2, \dots, x_n$, when $a = 0$.

6. Unit 1

ABC Space agency, responsible to protect a planet from asteroid collision, developed new space-to-space missiles to be loaded in satellites.

The agency planned missile trial in two steps for testing $H_0: p = 0.1$ versus $H_1: p > 0.1$, where p is the proportion of hits of missiles, each missile targeted at similar asteroid.

At the first step, 12 missiles will be fired. If three or more independent hits are observed among the (first) 12 missiles, H_0 is rejected, the study is terminated, and no more missiles are fired.

Otherwise, another 12 missiles will be fired in the second step. If a total of five or more independent hits are observed among the 24 missiles fired in the two steps, then H_0 is rejected.

i) Calculate the probability of Type I error for the two-step testing procedure.

ii) Calculate the probability of rejecting the null hypothesis H_0 when $p = 0.3$.

iii) Calculate the probability of Type II error when $p = 0.3$.

Suppose that 10 Space agencies have developed missiles similar to ABC Space agency. They have performed only the first step trials and observed an aggregate of 40 hits out of 120 independent missiles fired.

iv) Calculate an approximate lower bound of 90% right-tailed confidence interval for ' p '. (show up to four decimal places)

v) Test the hypothesis that $H_0: p = 0.278$ against $H_1: p > 0.278$ at 10% level of significance.

vi) Comment on your results of confidence interval obtained in (iv) and hypothesis testing in (v).

Aptitude tests were conducted by ABC Space agency at their two Institutes (1 and 2) that provide special training on missile technology. The test scores are shown in the table below:

	Institute 1	Institute 2
No. of trainees	12	15
Mean scores	65	70
Standard deviation	54	70

vii) Test the equality in mean scores of the populations associated with the two Institutes. State any assumptions made.

7. Unit 2

A study into the average claim (in Rs. '000) per health insurance policy was performed for the claims incurred in public and private hospitals. Data for some cities is given below:

	City 1	City 2	City 3	City 4	City 5	City 6	City 7	City 8	City 9
Public	24	45	29	33	20	40	26.5	25	27.5
Private	30	54.5	30	40	28.5	36	30.5	30.5	35.5

i) Determine the sample mean and sample variance of average claim size in both the type of hospitals.

ii) State the primary condition that needs to be true for testing equal mean and verify whether that condition is satisfied in the above example (You may assume that the samples come from a normal population).

iii) Test whether the treatments in private hospitals result in higher claim size at 95% level.

8. Unit 1

If $\hat{\theta}$ is an estimator of parameter θ, answer the following:
i) Define unbiased estimator

ii) Define "bias".

iii) Define Mean Square Error (MSE) of this estimator $\hat{\theta}$.

There exists another estimator $\bar{\theta}$ of the same parameter θ, such that $\hat{\theta}$ has no bias but higher MSE than $\tilde{\theta}$ while $\tilde{\theta}$ has a positive bias.
iv) State giving reason, which estimator is 'efficient'?

v) When would either of the two estimators be termed as consistent?

vi) Outline (in one sentence each) any two methods of estimating θ

9. Unit 1

i) Define the variable t_k used in the t-test for sampling distribution of sample mean describing all the symbols used.

ii) State the mean and variance of t_k for $k > 2$.

iii) A sample of 10 numbers from normal population has sample mean and sample variance as 50 and 48.667 respectively.

Determine the confidence interval for the population mean at 99% confidence level-
a) Using the t-test tables
b) Assuming a Normal distribution with parameters as the results of part (ii) above

10. Unit 2

Number of claims in a year on an insurance policy is believed to follow a Poisson distribution. Claims on portfolio of 1000 such policies were observed for one year. It was suggested that the value of Poisson parameter is 3. If the observed number of claims in that one year is less than 3100 then the suggested value of 3 for the Poisson parameter is accepted else rejected.

You may use the result that probability distribution of summation of ' n ' Poisson variables with parameter μ is Poi $(n\mu)$.

i) Define Type I error and estimate it for the above case.

ii) Define Type II error.

iii) Define power of a test and determine the power of test in terms of μ in above case.

iv) If the actual observed number of claims is 2900, determine the confidence interval for the Poisson parameter at 99% confidence level.

11. Unit 1

i) For the following joint distribution between X and Y, Calculate E[Y|X=2]

		Y		
		1	**2**	**3**
	1	0.1	0.1	0.3
X	**2**	0.3	0.2	0

ii) If X and Y are independent standard normal variables then calculate the mean, variance and the distribution of 5X - 4Y.

iii) If a random variable X follows Gamma (10,0.1) then P(X>50) is equivalent to finding the probability using the certain chi square distribution. Identify the equivalent probability calculation formula involving chi square distribution and calculate the required probability using chi square tabulated value. You are given that chi square tabulated value is 0.0318.

iv) Calculate the required probability as mentioned in Part (iii) above using normal approximation.

v) Comment on the results obtained in part (iii) and (iv).

12. Unit 1

Let $X_1, X_2 \ldots X_n$ be a random sample from Uniform distribution over $(0, \theta)$, where θ is an unknown parameter (> 0).

[i] Outline why the Cramer-Rao lower bound for the variance of unbiased estimators of θ does not apply in this case.

Consider an estimator of θ: $\hat{\theta}(c) = cY$ for some constant c where $Y = \max(X_i)$

[ii] Show that the probability density function of Y is given as:

$$g_Y(y) = \frac{n}{\theta^n} \cdot y^{n-1} \text{ for } 0 < y < \theta$$

Hence, show that:

$$E[Y^k] = \frac{n\theta^k}{n+k} \text{ for any non negative real number k}$$

[iii] Show that the bias and mean square error (MSE) of the estimator $\hat{\theta}$(c) are given as follows:

$$\text{Bias } [\hat{\theta}(\text{c})] = \left(\frac{cn}{n+1} - 1\right) \cdot \theta$$

$$\text{MSE } [\hat{\theta}(\text{c})] = c^2 \cdot \frac{n\theta^2}{n+2} - c \cdot \frac{2n\theta^2}{n+1} + \theta^2$$

[iv] Find the value of $c = c_u$ for which $\hat{\theta}(c)$ becomes an unbiased estimator of θ.

[v] Find the value of $c = c_m$ for which the mean square error of $\hat{\theta}(c)$ is minimized.

[vi] Which of the two estimators $\hat{\theta}(c_u)$ or $\hat{\theta}(c_m)$ will you prefer for estimating θ ? Give reasons. What happens when n is large?

13. Unit 2

It is known that for a shop of a particular category, the mean invoice amount towards sakes is INR 2000 and standard deviation is INR 500. Stating clearly any assumptions that you make, calculate the following:

i) Probability that mean amount of next 10 invoices is less than INR 1700.

ii) Probability that standard deviation of amount of next 10 invoices is less than INR 250.

iii) Probability that both (i) and (ii) occur together. (i.e. mean amount of next 10 invoices is less than INR 1700 and standard deviation of amount of these 10 invoices is less than INR 250)

14. Unit 2

An Insurance company is analyzing the claim data. Determine the probabilities of the following events.

i) The number of claims reported in a year by 100 policyholders is less than 6.

Assume claims reporting from each policyholder follows Poisson distribution with mean 0.03 per year independently of the other policyholder.

ii) The number of claims examined up to and including the fourth claim that exceeds £50,000 is less than 7.

Assume the above follows negative binomial distribution with probability of a claim exceeding £50,000 as 0.4 independent of any other claim.

iii) The number of deaths in the coming year amongst a group of 1000 policyholders is less than 10.

Assume each policyholder has a 0.015 probability of dying in the coming year independently of any other policyholder.

15. Unit 2

i) Wickets taken by a cricket team 'A' follows Poisson process with rate of 1 wicket per100 balls bowled. How many wickets will the team take with 95% confidence after bowling 500 balls?

ii) A cricket team 'B' has batsmen for the last wicket that can score 1 run per ball with a probability of 40% and 0 runs with probability of 60%. The team B (playing against the above team A and having only 1 wicket in hand) needs to score 26 runs in 50 balls. Team A wins if it takes the 1 wicket in these 50 balls and team B wins in case it scores the required runs in 50 balls. Determine which team has a higher probability of win. State any assumptions you make.

iii) Hence determine the probability that team B will bat for at least 30 balls.