MT-2002: Statistical Modeling

Student Name	Roll No.	Course Section	Student Signature

DO NOT OPEN THE QUESTION BOOK OR START UNTIL INSTRUCTED.

Instructions:

1. Attempt on question paper. Attempt all of them. Read the question carefully, understand the question, and then attempt it.
2. No additional sheet will be provided for rough work. Use the back of the last page for rough work.
3. If you need more space, write on the back side of the paper and clearly mark question and part number etc.
4. After asked to commence the exam, please verify that you have **nine (9)** different printed pages including this title page. There are a total of **5** questions.
5. Calculator sharing is strictly prohibited.
6. Use permanent ink pens only. Any part done using soft pencil will not be marked and cannot be claimed for rechecking.

	Q-1	Q-2	Q-3	Q-4	Q-5	Total
Marks Obtained						
Total Marks	10	10	10	10	10	50

Imagine you are conducting a simple experiment where you have a bag filled with red and green marbles. Assume you generated such data synthetically. You want to estimate the probability of drawing a red marble from the bag. You have no prior knowledge about the composition of the bag, so you assume that all possible proportions of red to green marbles are equally likely.

i. Select a prior with proper justification.

Prior : Uniform Or Beta with 1,1 parameter

* *Justification : Above mentioned priors reflect lack of prior information about the Composition of Bag. alternatively we can say that all possible proportion of red and Green marbles are equally likely.*

ii. Select a likelihood with proper justification.

Likelyhood : Bernouli Distribution.

* *Justification : Because having no Count Or n given in Description*

iii. What goes into the Bayesian inference engine in this problem?

1. *priors*
2. *Likelyhood.*
3. *Data*

iv. What will be the outcome of the Bayesian inference engine in this problem?

Posterior

Question 2 [10 Marks]

The owner of FAST cafeteria conducted a survey and used Bayesian method to get an estimate of the students who prefer to dine out on Friday. Figure below shows the prior and posterior curves for the proportion of students who prefer to dine out on Friday. What two observations can be made based on this figure?
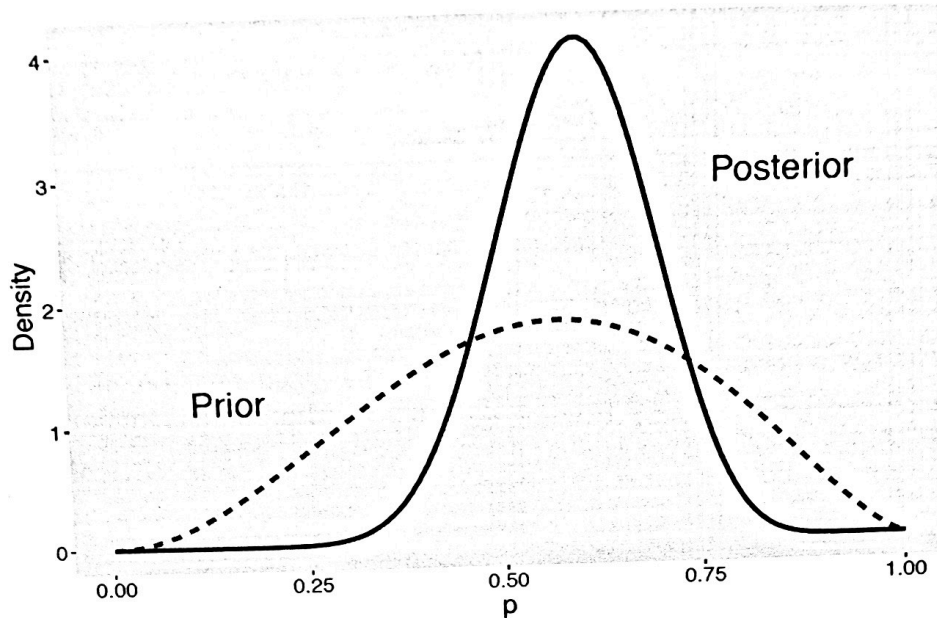
Answers:

Observations:

① The mean value of the prior belief and the posterior is more or less same.

② One see's a much wider spread of the prior then that of posterior. initially the Owner was Unsure about the proportion of students favoring Friday to dine out. After Observing the results of the Survey, The Solid posterior curve indicates that he is more certain that p is b/w .5 and .7. This sheds light on general feature of Baysian inference. The data helps, Sharpen the belief about the parameter of interest, producing a posterior distribution with a smaller spread than the prior distribution.

Question 3 [10 Marks]

Carefully read the following description and try to understand.

[*Description*]: As an analyst, assume you are working in a light bulb manufacturing company that operates every day except Sundays. Each day, the company produces a specific quantity of light bulbs, with variations in daily production volumes. You have the daily data on the number of bulbs produced. You already implemented a model and estimated the probability from which you can easily determine whether the number of faulty bulbs (denoted by 0) overtakes the number of flawless (denoted by 1) ones. Consider the following plots and next answer the following.

Monday

Tuesday

Thursday

Friday

Wednesday
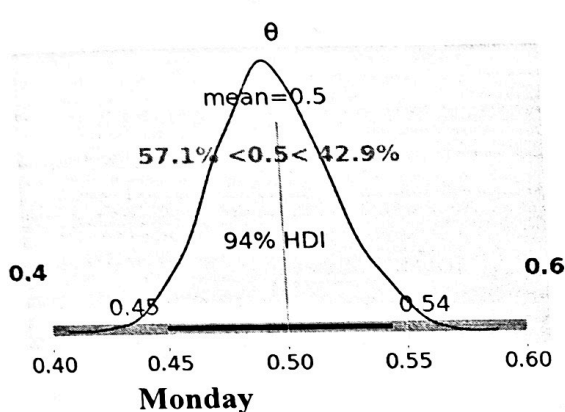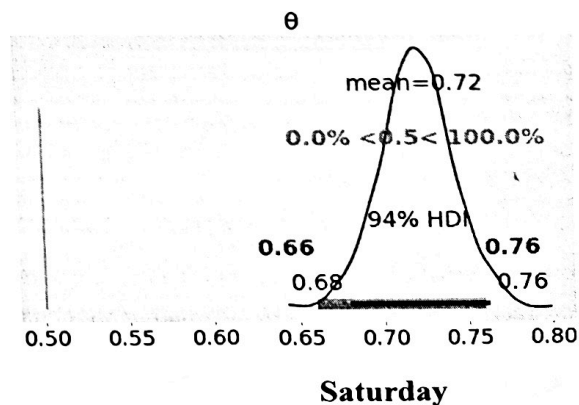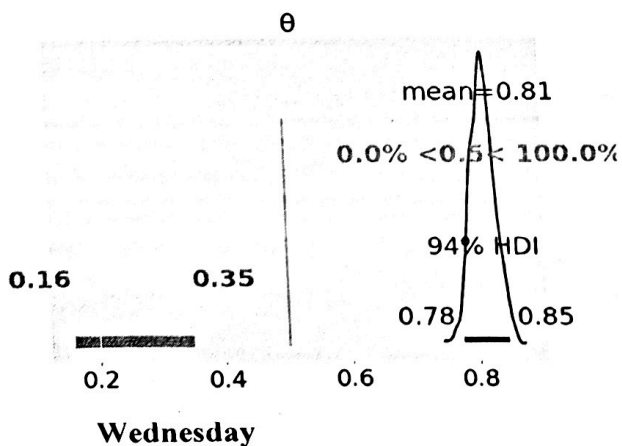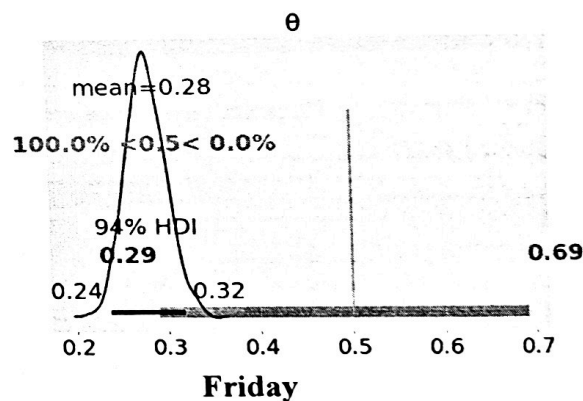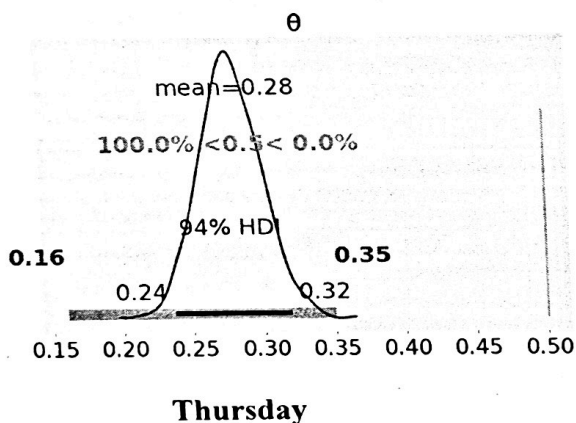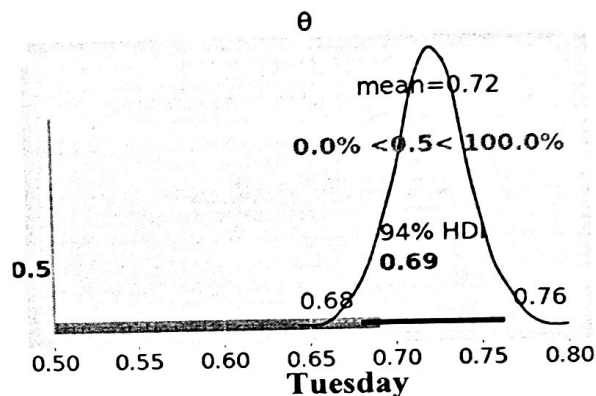
Saturday

1. Select and mention here all those days which show the number of flawless bulbs production is more than that of the faulty bulb by the company. Justify your selection of days from the above plots by comparing HDI and ROPE. (3)

Days:

$$\underline{\underline{Saturday}}$$

Justification of your selection:

	lower	Upper
ROPE =	.66	0.76
94%HDI =	.68	0.76

As 94% HDI falls within ROPE defined with greater than 0.5. Correct option is saturday.

2. Select and mention here all those days which show an equal number of flawless & faulty bulbs production by the company. Justify your selection of days or day from the above plots by comparing HDI and ROPE. (3)

Days:

$$\underline{\underline{Monday}}$$

Justifications of your selection

	lower	Upper
ROPE region =	.4	.60
94% HDI =	0.45	0.54

Based on the ROPE and HDI Comparison we can say Monday is Correct Option. Furthermore Our 94% HDI falls within Defined region of .4 to .6 which includes .5.

Question 4 [10 Marks]

You have collected data on the daily earnings of a freelancer over four specific days and implemented a model to estimate parameters from it. Below is the table presenting the estimated mean earnings (μ) and estimated standard deviations (σ) for these days:

Day	Estimated Mean Earning (μ)	Estimated Standard Deviation (σ)
Sunday D_1	$1000	$50
Monday D_2	$1050	$55
Tuesday D_3	$980	$45
Wednesday D_4	$1100	$60

Your task is to perform statistical comparisons between these days by calculating three Cohen's d values and three probabilities of superiority.

Hint: Cohen's d: $\delta = \dfrac{\mu_2 - \mu_1}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{2}}}$

	Cohen d
D_{12}	0.95
D_{13}	0.42
D_{24}	0.86

Based on your calculations, determine the following

1. Which of the days appears to have the highest effect size by looking into Cohen d value?

Day 1 & day 2.

2. Which of the days appears to have the smallest effect size by looking into Cohen d value?

Day 1 & Day 3.

Cohen d Calculation.

Given Formula & Data.

* Cohen d b/w day 1 and day 2:

$$d_{12} = \frac{1000 - 1050}{\sqrt{\frac{50^2 + 55^2}{2}}} \approx \frac{-50}{52.7} \approx 0.95$$

* Cohen d b/w day 2 & day 3

$$d_{23} = \frac{1000 - 980}{\sqrt{\frac{50^2 + 45^2}{2}}} \approx \frac{20}{47.43} = 0.42$$

Cohen d b/w day 2 & day 4.

$$d_{24} = \frac{1050 - 1100}{57.55} \approx 0.86$$

$$\sqrt{\frac{55^2 + 60^2}{2}} = 57.55$$

Q5 [10 marks] Multiple choices

1. From the following expressions, which one corresponds to the sentence, The probability of being sunny given that it is July 9th of 1816?

   A. p(sunny)
   B. p(sunny | July)
   C. p(sunny | July 9th of 1816) ✓
   D. p(July 9th of 1816 | sunny)
   E. p(sunny, July 9th of 1816)/p(July 9th of 1816)

2. What is the definition of probability in a statistical context?

   A. The likelihood of an event occurring.
   B. The ratio of successful outcomes to total outcomes.
   C. A measure of the certainty or uncertainty of an event. ✓

3. Which of the following best describes a uniform distribution?

   A. All outcomes are equally likely. ✓
   B. The distribution is bell-shaped.
   C. Outcomes are not dependent on each other.

4. What does Bayes' Theorem allow us to do?

   A. Update beliefs based on evidence. ✓
   B. Calculate the probability of an event.
   C. Simulate random events.

5. What is the shape of the graph of a normal distribution?

   A. rectangular
   B. triangular
   C. bell-shaped ✓
   D. bimodal

6. Approximately what percentage of values lie within two standard deviations of the mean of a normal distribution?

   A. 34%
   B. 68%
   C. 95% ✓
   D. 99%

7. A normal distribution has a mean of 10 and a standard deviation of 2. What is the probability that a value chosen at random is more than 10?

7. A normal distribution has a mean of 10 and a standard deviation of 2. What is the probability that a value chosen at random is more than 10?

    A. 0.34
    B. 0.475
    C. 0.495
    D. 0.5 ✓

8. The central limit theorem states that as the sample size increases, the sampling distribution of the sample means approaches which type of distribution?

    A. Uniform distribution
    B. Exponential distribution
    C. Normal distribution ✓
    D. Poisson distribution

9. The mean, median, and mode of a normal distribution are always equal.

    A. True ✓
    B. False