Best of Luck!

Problem 1 True or False Decide whether each of the following statements is true or false. If false or only true under some conditions, replace by the full correct statement.

a) The likelihood of a hypothesis H is the probability that H is true given the observed data.

b) The maximum likelihood estimate of a parameter θ is always our best guess for the value of that parameter.

c) In a goodness-of-fit test, if the the test statistic lies inside the acceptance region, this implies that the hypothesis we are testing is true.

d) The p-value of the data is the probability that the null hypothesis H_0 is false given the observed data.

Problem 2 Deal or no deal. Consider a game-show called "deal or no deal", where you have to choose between three suitcases labelled (A, B, C), one of which contains Rs. 1 crore and the others are empty. On this show contestants pick a suitcase (say suitcase A) and keep it temporarily without opening it. The host (who already knows what's in the suitcases) then opens up one of the other two cases (say suitcase B) which is empty. The contestant is now offered the option to keep the suitcase (A) or switch it with the remaining unopened suitcase (C). Suppose that you are on this show, would you switch or keep your suitcase to get the prize? First make a guess.

Now let us analyze this using likelihoods. Let H_i denote the hypothesis that the prize is in suitcase i where $i = A, B, C$.

a) What are the a-priori probabilities of H_i before you pick a case?

b) Calculate the likelihood of each hypothesis H_i given that the host opened suitcase B after you picked suitcase A.

c) Compute the a-posteriori probabilities of each H_i after the host has opened case B and shown that it was empty.

d) To get the prize should you (i) keep A, (ii) switch to C, or (iii) does it not make a difference? Justify.

Problem 3 Your uncle enjoys betting against you using coin tosses, but you suspect that the coin that he has is biased. You don't want to start a war by accusing him without evidence! You decide to collect data so that you can confront him.

a) Clearly state the null hypothesis and the alternate hypothesis.

b) Let us define our test statistic (q) as the number of heads minus the number of tails for N tosses. What is the exact distribution of the test statistic assuming the null hypothesis is true? Argue that for a large number of tosses, the test statistic is Gaussian distributed.

c) What is the mean and what is the standard deviation of the test statistic distribution assuming the null hypothesis is true?

d) Assuming that you need to be 95% confident that the coin is biased to catch him, sketch the probability distribution function of q and define a critical region for this test statistic.

e) Assuming that you toss the coin 100 times and you get 63 heads, is this significant evidence that your uncle is cheating?

Problem 4 You will be required to write an executable code for this problem. **You are free to use any programming language though Stata, R and SAS are preferred in that order.** Quarterly deposit and credit data is publicly available from the Reserve bank of India.
Download the quarterly district-wise data on deposits for public and private sector banks between 2008-09:Q4 to 2009-10:Q4. Use `https://dbie.rbi.org.in/` and download the data from the Time-Series Publications > Quarterly Statistics on Deposits and Credit of Scheduled Commercial Banks. Public sector banks are the bank groups "SBI AND ITS ASSOCIATES"/"NATIONALISED BANKS". For private sector banks the bank groups are "PRIVATE SECTOR BANKS". In the 2008 September, private sector banks witnessed deposit outflows. Our goal is to see what happened to the public sector bank deposits in the districts that witnessed these panic outflows.
Please provide the output from (c) to (g) in your write-up.

a) Calculate the growth in deposits at the district level between 2008-09:Q4 to 2009-10:Q4 separately for private and public sector banks (PSBs).

b) Designate the bottom tercile of growth of private sector banks as districts having had panic outflows and the rest as having had no panic outflows.

c) What was the corresponding mean deposit growth for public sector banks in districts which had private sector banks panic outflows.

d) Plot the deposit growth of public sector banks against private sector banks separately for

- Districts that had panic outflows. Districts that had no panic outflows.

e) Using either Stata or R: show 3 maps (heat maps) with the following:

- Deposit growth for PSBs.
- Deposit growth for Private sector banks
- Deposit growth for all banks.

f) Find the t-test of the difference in deposit growth rate for private and public sector banks for:

- PSB and private sector banks that had panic outflows.
- PSB and private sector banks that had no panic outflows.
- The t-test of the difference between a. and b. Do you know what empirical strategy this corresponds to?

g) Credit Growth

- Write an appropriate regression model, including the possible control variables, to check what variables affect credit growth. Please explain, giving concrete reasons, why your regression model is appropriate.
- Repeat d) and e) for credit growth. Note, panics still correspond to districts which had lowest tercile of deposit growth.