MT2002 Statistical Modeling

Assignment No: 02	CLO: 02
	Semester: Spring 2023
Due date: 24-Feb-2023	Marks: 150

Instructions:

1. **Plagiarized work will result in zero marks.**

2. **No retake or late submission will be accepted.**

3. Attach complete code, results, and screenshot for questions that require programming solution. Programs/codes should not be handwritten.

4. Questions that show the icon ⟨/⟩ require partial or complete solution using the approved programming tool.

5. The assignment is to be submitted in softcopy as well as in hardcopy.

6. Jupiter notebook file is recommended.

7. The softcopy should be a single PDF file of your complete assignment including programming and non-programming questions.

8. The PDF file should be according to the following **format: id_section_A1** e.g. i22123456_A_A1. A1 in the end denotes Assignment 1.

9. The images of the by-hand solution should be properly scanned. You can use any mobile application such as Cam Scanner or Adobe Scan for scanning. Each of these applications allows you to export pdf or image files which you can use to combine with your programming solutions. Do not attach direct images from the camera application of your mobile phone, or screenshots.

Questions:

Question 1 (50 + 50 + 50 = 150 marks):

a) You are provided with Iris.csv file. For description of the mentioned dataset please visit the link. Use PyMC3 library to define a model (prior and likelihood) for the

following situation. In case your inference is not robust in comparison to real data, you can go with alternate likelihood (distribution).

      i.   Estimate the prior for the column `sepal_width` using Normal and Students- T

      ii.   Compare the prior you estimated in above likelihood i.e., Normal and Students-T. And conclude which one is robust.

b) Create group, for more understanding please check the following violin plots i.e. virginica, setosa, versicolor, you can create more groups using sepal_width etc on y axis. You need to create model and generate samples then compare each group i.e virginica, setosa, versicolor using the comparison metrics we did in class i.e Cohens'd, probability of superiority and reference values.
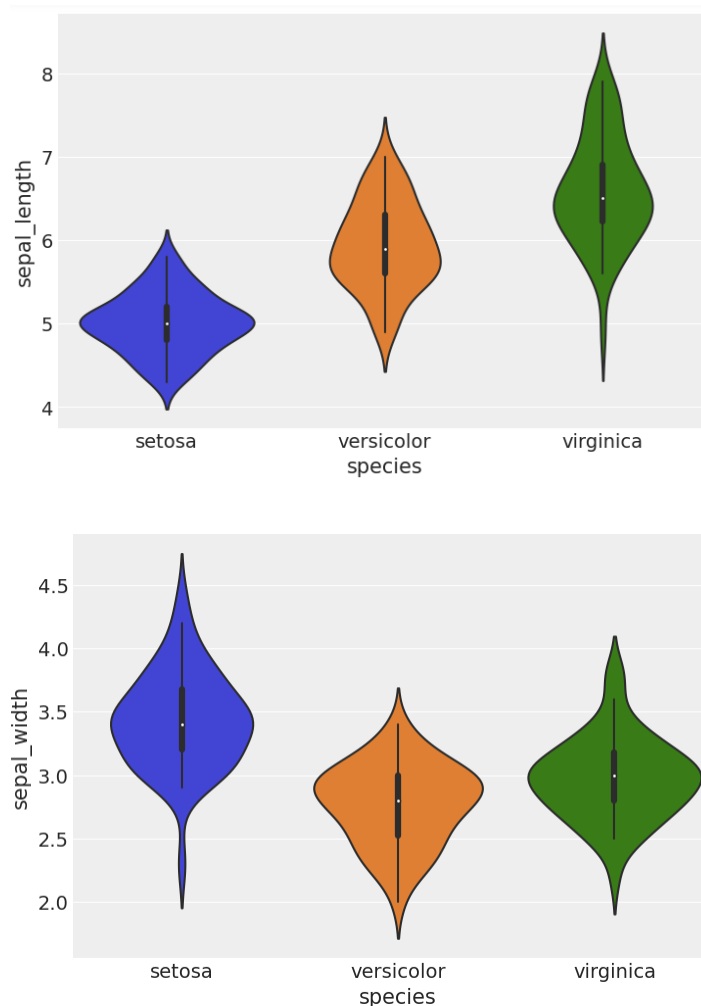
c) Interpret your results from b & apply Groups comparison techniques on your designed model such as Cohen's d and probability of superiority.

Note: You should also use other libraries such as ArviZ, Seaborn, pandas, Numpy, SciPy, and matplotlib.pyplot to properly analyze and plot your model outcomes.