Instruction: The project will entail a through data analysis of the data provided using appropriate regression models. You may use SAS software for your data analysis and estimation.

Suggested title:

Predictors of Tumor Status among Breast Cancer Patients:  
Can the Fine Needle Aspiration (FNA) Technique be used as a Substitute for Biopsy?

The project has the following outcome (Y) and predictors (Xi) variables:

Outcome: Y = tumor status

Predictors: X1 – X9

Research question

(1) Do the cell features allow us to predict tumor status? That is, can we use FNA as an alternative to the biopsy procedure for future patients?

(2) What are the sensitivity and specificity of the FNA based on the model?

(3) What features are the predictors of tumor status?

Develop your hypotheses based on the predictor variables you are interested to test and investigate.

Make a decision which variables to retain in the final model based on the results of a detailed analysis.

Submit a 10 page well written paper based on the methods applied and your findings. Attach summary results in the form of tables. Also attach your codes and final results as Appendix.

The paper has to be double-spaced using Times New Roman 12pt. Sound analysis and correct interpretation of the key results is expected.

The written report should include:  
(1) Introduction; (2) purpose of the study and research questions to be addressed; (3) hypotheses to be tested; (4) brief description of the data; (5) statistical methods applied including their full specifications; (5) results and interpretations; (6) summary and conclusions; and (7) limitations if any and suggestions how to improve the project for future analysis.

Dataset

The data is from the Wisconsin Breast Cancer which consist of 683 cases of potentially cancerous tumors. Traditionally whether a tumor is malignant or benign is determined with an invasive surgical biopsy procedure. An alternative less invasive technique called “fine needle aspiration” allows examination of small amount of tissue from the tumor. (FNA). For the Wisconsin data, FNA provided nine cell features for each case; a biopsy was then used to determine the tumor status as malignant or benign.

Don’t forget to start with simple analysis tools and build your final models step by step.

You may use the 5% level of significance for your hypothesis testing.

Name of dataset: wisbcdata.xlsx

Description of variables

|  |  |
| --- | --- |
| Variable | Definition |
| Y | Tumor status (0= benign, 1= malignant) |
|  | Clump thickness |
|  | Cell size uniformity |
|  | Cell shape uniformity |
|  | Marginal adhesion |
|  | Single epithelial cell size |
|  | Bare nuclei |
|  | Bland chromatin |
|  | Normal nucleoli |
|  | Mitoses |