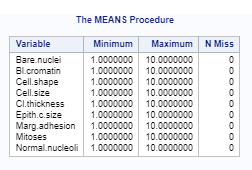
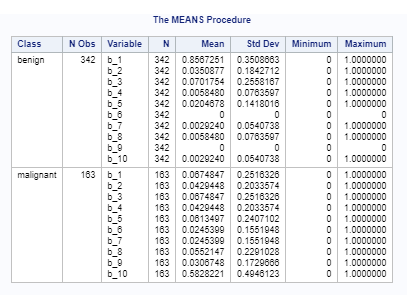
After exploring the data and understanding where boundaries can be found for the malignant and benign groups for each of the explanatory variable. The model selection process (stepwise, forward, and backward) for logistic regression was used to determine what variables are considered the best in determining cancer. However, when estimating the logistic regression, an issue arises where the convergence criterion is not satisfied because of quasi-complete separation of data points. This is an indication that the model cannot be mathematically calculated since there are explanatory variables that perfectly split the group between benign and malignant.

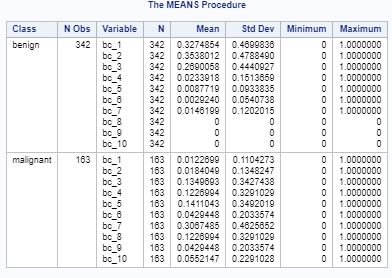
Therefore, grouping can be done on the explanatory variables to mitigate this quasi-complete separation. Since all the variables that are used in the logistic regression are dummy variables, we need to see where there are no ones between benign and malignant. The screenshot below, shows the range of values that we have for each of the explanatory variables that we have of interest. As seen below, the range of values are between 1 to 10 as explained earlier in the explanatory data analysis.



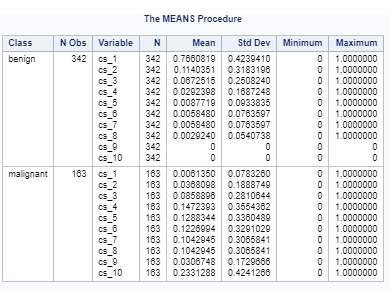
To find at which levels the factors are separating, the data must be grouped by benign and malignant by each of the 10 levels in the explanatory variables. The results of the analysis are shown.



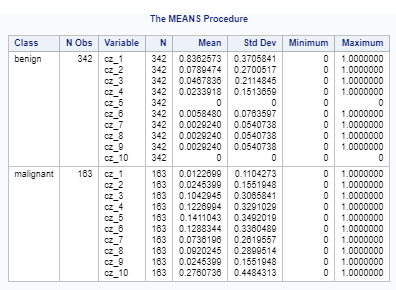
For Bare.nuclei, we can see that there are no indicators at benign for levels 6 and level 9.



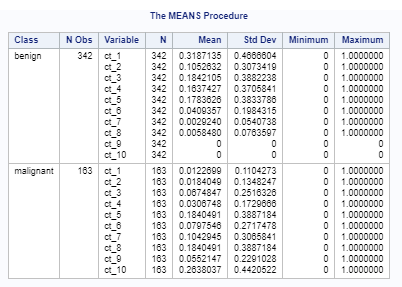
For Bl.Cromatin, we can see that there are no indicators at benign for levels 8 to 10.



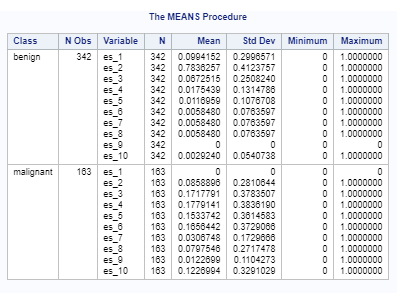
For Cell.shape, we can see that there are no indicators at benign in levels 9 and 10.



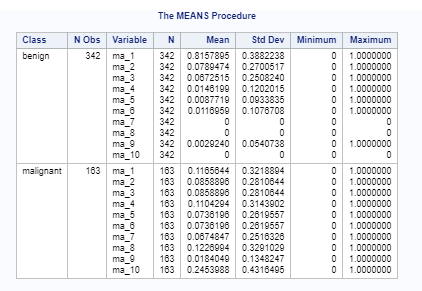
For Cell.size, we can see that there are no indicators at benign for levels 5 and 10.



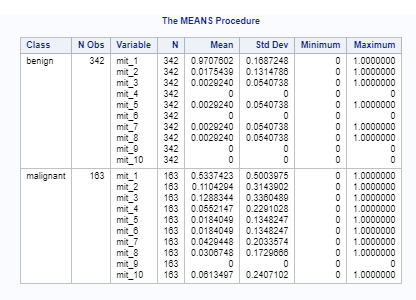
For Cell.thickness, we can see that there are no indicators at benign for levels 9 and 10.



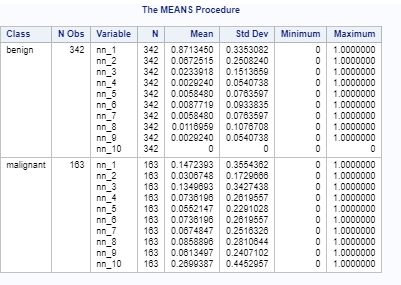
For Cell.thickness, we can see that there are no indicators at benign for level 9 and no indicators at malignant for level 1.



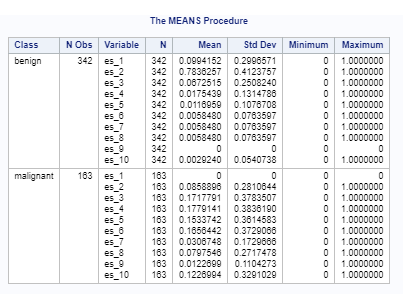
For Marg.adhesion, we can see that there are no indicators at benign for levels 7, 8, and 10.



For Mitoses, we can see that there are no indicators at benign for levels 4, 6, 9, and 10 and malignant at level 9.



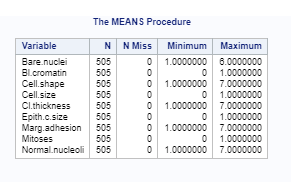
For Normal.nucleoli, we can that there are no indicators at benign for level 10.



For Epith.c.size, we can that there are no indicators at benign for level 9 and malignant for level 1.

Since we have determined at what levels that have complete separation, we have done the following criteria for grouping:

* All Bare.nuclei greater than or equal to 6 are grouped together as 6
* Bl.cromatin is split as 0 for levels 1 to 5, and 1 for levels 6 to 10
* All Cell.shape greater than or equal to 7 are grouped together as 7
* Cell.size is split as 0 for levels 1 to 5, and 1 for levels 6 to 10
* All Cl.thickness greater than or equal to 7 are grouped together as 7
* Epith.c.size is split as 0 for levels 1 to 5, and 1 for levels 6 to 10
* All Marg.adhesion greater than or equal to 7 are grouped together as 7
* Mitoses is split as 0 for levels 1 to 5, and 1 for levels 6 to 10
* All Normal.nucleoli greater than or equal to 7 are grouped together as 7



\*\*\*\*\*\*ADD EXPLANATION HERE