Name: Leung Ko Tsun SID:20516287

Table

Description automatically generated with medium confidence

The class benign is set to be 1 and malignant is set to be 0. Result for the logistic model is shown above.

A picture containing graphical user interface

Description automatically generated

The G value is 781.462, which is significantly greater than the 95% quantile of chi-square distribution, which is 16.91898. So, we can conclude that the model is significant.

b).

Graphical user interface, text, application

Description automatically generated

The 95% confidence interval for beta\_hat{CI.thickness} is [-0.8134, -0.2567]. For the hypothesis testing for beta\_{cell.shape}, we can see that the p-value of beta\_hat{cell.shape} is larger than 0.1, so we fail to reject the null hypothesis at the significant level of 0.1.

c).

Text

Description automatically generated

The summary of the model is shown above. We can see that the AIC values of the reduced model and the full model are similar. So, we can conclude that the reduced model and the full model are equally predictable regression model. For further comparison, we can make a hypothesis testing that H0 = reduced model is suitable versus H1 = full model is suitable, where p = 5, p + q = 9, and 2[L(p+q) - L(p)] = 9.68 < qchisq(0.01,4) = 13.7267. So, we fail to reject the null hypothesis, implying that we should use the reduced model instead of the full model.

d).

P(Class = “benign” | Cl.thickness=6, Cell.shape=3, Marg.adhesion=8, Bare.nuclei=2, Bl.cromatin=5) = 0.1506149Text

Description automatically generated

e). Text

Description automatically generated

Using bestglm() function in R, the model in part (c) has the smallest AIC (124.57). So the model is part (c) is the best model by the AIC method, and the following variables are used: CI.thickness, Cell.shape, Marg.adhesion, Bare.nuclei, and Bl.cromatin.