E1

a.

The tabulate below shows that

1. In the sample, the number of patients with malignant tumor is almost twice as large as that of patients with benign tumor.
2. The mean proportion uniform shape of cells for patients with malignant tumor is less than 50% but slightly higher than 30%, which is much smaller than the proportion for patients with benign tumor, which is higher than 90%.
3. Uniformity of the cell’s shape has a higher standard deviation for patients with cancer, which is 2.56, compared with those who do not have cancer (s.d. =2.56).



b. Interpretation:

1. As shown in the tables below, the models fit much better when covariate “ushape” is included, because the AIC drops from 902.53 (intercept only) to 288.25 (intercept and covariate).
2. Both intercept and ushape are significant (p-values <.0001).
3. When ushape is evaluated at 0, the log-odds of this patient having cancer is -5.06. However, as a matter of fact, evaluating ushape at zero is out of the range of plausible ushape scores (1-10).
4. More specifically, with 1-unit increase in ushape, the log-odds ratio of having cancer is expected to increase by 1.4068, so the odds ratio of having cancer (malignant tumor) is expected to increase by e^1.4068=4.08.



c.

1. The plot below shows that as the proportion of cells uniformity increases, the estimated probability of having cancer (malignant tumor) increases. However, the speed of probability increase is relatively slow when ushape is below 2 or is above 6. When ushape is rising from approximately 3 to 4, the speed is the largest of increase of the probability of having cancer (the slope is the largest).
2. According to the plot, when ushape is approximately 3.8, the probability of having malignant tumor exceeds 0.5.



E2

a.

1. The procedure that is used here for model selection is stepwise selection. As is shown below, the model includes 10 out of the 11 predictor, with only the variable “ fixed\_acidity” being left out.
2. The model tells us that quality does not predict whether the wine is red or white (p-values >.05). Given other variables in the model are held constant, with 1-unit increase in volatile\_acidity, citri\_acid, residual\_sugar, chlorides, free\_sulfur\_dioxide, total\_sulfur\_dioxide, density, sulphates, and alcohol, respectively, the odds ratio that the wine is red is expected to increase by more than 999.999, 0.10, 0.42, more than 999.999, 1.07, 0.95, more than 999.999, 19.51, and 5.15. 