

Question 1

What is the unit of observation in this data frame?

Question 2

We will be fitting models to output a diagnosis (“benign” or “malignant”). This is a categorical outcome. Which level will be considered the reference level by default in R and why?

Question 3

If you were to deploy your method in a clinical setting to help diagnose cancer, would it be worse to misclassify a benign case or to misclassify a malignant case? Explain your rationale in at least two sentences.

Question 4

Based on the glimpse, use a plot to compare the `radius_mean` for benign vs. malignant biopsies, *side-by-side*. Make sure to give your label your axes and give your plot a title. Give a shape which matches **your** expectation of the phenomenon and explain your choice in at least one sentence.

Question 5

Based on your previous sketch, what biopsies are you prepared to classify as malignant versus benign? Fill in the blanks below to make a decision rule.

```
If radius_mean > _____ : predict _____  
Otherwise predict _____
```

Question 6

Modify the side-by-side plot you made earlier to visually represent the decision rule.

Question 7

Based on the glimpse, sketch a plot that examines the association between two predictors, `radius_mean` and `area_mean`. Make sure to give your label your axes and give your plot a title. Give a shape which matches **your** expectation of the phenomenon and explain your choice in at least one sentence.

Question 8

In many realms of medicine, classification algorithms can be more accurate than the most well-trained medical doctors. What is gained and what is lost by shifting to algorithmic diagnoses? Although a book could be written about this topic, please answer in one paragraph.