3. A p-value is the probability of observing a result given the hypothesis is true.

In our example, we created a confusion matrix between the iris$Species data set and sample(iris$Species). This shows that we are not applying any model/information to the data, other than random sampling. By doing so, we resulted with a high p-value of 0.9786. In other words, we have a 97.86% chance of predicting this exact confusion matrix given we use randomly sampled data as our predictive data set.

In answer to the question, the p-value is large because the confusion matrix is being affected by a no information rate. When there is no information for the prediction, the confusion matrix shows that the relationship between the predicted and truth values is not very accurate. This is not difficult to recreate, explaining why there can be a 97.86% chance you can have the same confusion matrix given you have no information for the relation between the predicted vs. true values.

5. Bayes’ theorem is a mathematical definition for conditional formatting. It can take on many different frameworks using derivation, propositional, and subjective logic, but in its simplest form, it states the probability of event A occurring given event B occurred. What might seem to be a primitive concept is actually responsible for understanding multiple health related tests such as predicting genes, cancers, and drug successes in humans. This theory can be used to predict many bi/multi variate questions. In fact, one of the most well-known statistical concepts, p-value, has a definition written in Bayes’ theorem’s conditional format. Hence, Bayes’ theorem is one of the most fundamental ideas of statistics.