Exercise 1

a) **Model A**

Sensitivity = 400 / (400+112) = 0.781

Specificity = 406 / (406+82) = 0.832

Accuracy = (400+406) / (400+112+406+82) = 0.806 -> 80.6%

Model B

Sensitivity = 402 / (402+110) = 0.785

Specificity = 405 / (405+83) = 0.830

Accuracy = (402+405) / (402+110+405+83) = 0.807 -> 80.7%

Model C

Sensitivity = 393 / (393+119) = 0.768

Specificity = 398 / (398+90) = 0.816

Accuracy = (393+398) / (393+119+398+90) = 0.791 -> 79.1%

Given that the results found are so similar between the three models, there would probably be no statistical difference between them, at least in the case of model A and model B. Therefore, the decision of which model to choose would depend on other factors, such as, which is most time or space efficient because the models have similar accuracies, sensitivities, specificities and little to no bias.

Exercise 2

a) Model A is the most appropriate because it has a higher correlation coefficient and also less error than model B. This means that model A will be more accurate and its errors, on average, are smaller than model Bs.

Exercise 3

a) Assuming that the CI is 95%, the p-value of 0.858 means that the difference between the means of model A and model B is not statistically significant. Therefore, we accept the null hypothesis meaning there's no difference between the models in terms of their statistical performance. So, the choice between the models would depend on other factors like time efficiency.

b) If both models were examined using the same data set, the most appropriate t-test would be the Student's t-test because we know that the underlying variance is the same. If they had different variance, we would choose a Welch's t-test.

Exercise 4

a) This table tells us that $F_{(2,2997)} = 5.0478$, $p = 0.0065$. Assuming a CI of 95%, we can see that the p-value is below 0.05 meaning that we reject the null hypothesis. Therefore, we can say that there is a statistically significant difference between the models of Q1A.

b) From analysing the contents of the TUKEY Post Hoc table, we can see that there seems to be a statistically significant difference between the performance of model A and model C because the p-value associated to them is less than 0.05. The same can be said when we compare the performances of model B and model C. However, we can observe that there is no statistically significant difference between model A and model B as the p-value is 1. Therefore, the decision on which model to choose would depend on other factors.