<u>Explore the association between bone mineral density and hip fracture. Comment on the predictive ability of bmd on the probability of hip fracture (6 marks)</u>

The EDA showed that there were 169 patients in the study. The variables consist of binary variables with 0 being no fracture and 1 being a fracture. A table revealed that 50 patients had a hip fracture and 119 did not. To predict hip fracture probability the variable bone mineral density (BMD) was included as a continuous measurement. This was included because it is a predictor of orthopaedic health.
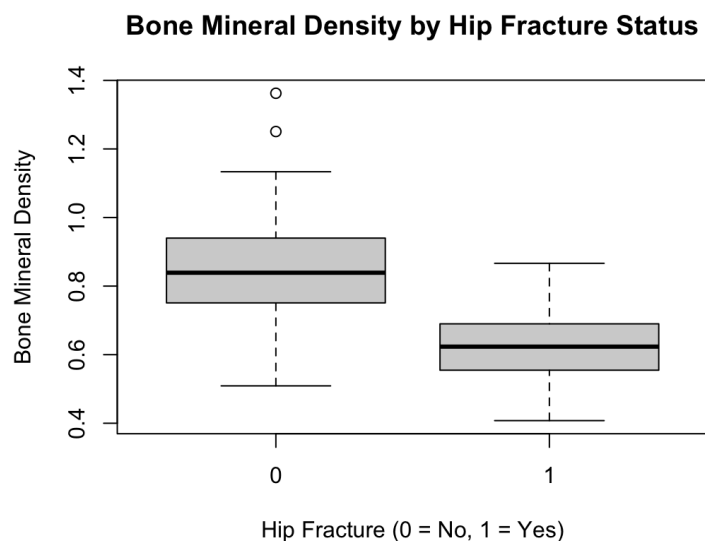

Figure 4. Boxplot of Bone Mineral Density (BMD) Vs Hip Fracture Score

This boxplot shows that the median BMD is higher in the non hip fracture group compared to the hip fracture group. Indicating there is an association between the amount of BMD and hip fracture rates. To test this association and develop a deeper understanding, more in detail statistical analysis will be completed. Due to the variables being binary a logistic regression was picked to assess the relationship between BMD and hip fractures.

The logistics regression model showed there was a significant association between BMD and hip fractures. The model produced results of a negative coefficient, supporting the idea that as BMD increases, the chance of a hip fracture decreases. In the model the predictor was BMD and the outcome was whether there was a hip fracture or not. The p-value was <0.001 which means it was a very significant result. This evidence supports that BMD is a key factor for predicting hip fractures, however does not highlight if there are any other factors that could also contribute.

<u>Subsequently, carry out a suitable regression based analysis to investigate the collective usefulness of bmd and other patient information (as collected) in predicting hip fracture. Present your best model for the prediction of hip fracture and briefly</u>

The new logistic regression model measures BMD against other variables such as age, sex, weight, height, waiting time and medication. It revealed that there was a statistically significant association between higher BMD and lower hip fracture rates. The p-value for BMD vs hip fracture was <0.001. It's important to mention that other factors influence this and those variables have been adjusted to achieve the best model. Medication has a p-value of 0.053 which is close to being significant, however cannot confidently give a clear answer. This model with multiple variables had an Akaike Information Criterion (AIC) of 120.9. A lower score is possible by making the model more simple. AIC is useful because it compares and selects a model that estimates the risk of hip fractures in individual patients.

To reduce the AIC, a stepwise model was used and included the variables BMD, height and medication, these gave the lowest AIC score of 114.7. This final model supports that BMD is a very significant predictor of hip fractures with a p-value <0.001. Height had a p-value of < 0.05 which is statistically significant. Medication had a p-value of 0.054, this is just outside of being significant and was included because overall it benefits the model.

Overall, logistic regression was selected because the predictor hip fracture variable was binary. The full model included all patient variables to understand their initial predictor values. Subsequently, a stepwise model was used to find the simplest model that contained the lowest AIC value. This process involves comparing different models against each other and adjusting the number of useful variables to produce the lowest AIC result. The final model produced the lowest AIC making it more efficient in understanding the data and being the strongest predictor of hip fractures.

Comment on the validity and predictive ability of your model. You may want to include an illustration of how the model can be used (10 marks)

To assess the validity of the model, a stepwise model was produced. This deepened our understanding of how the models compare against each other and develop a good idea of which variables are the best predictors for hip fractures. It was important to do the stepwise model because after adjustments were made it produced the final model with the lowest AIC, therefore the most simple. The final model consists of the predictor variables BMD, height and medication. Key interpretations from this model included that BMD was by far the most significant predictor of hip fracture, with a p-value of <0.001.

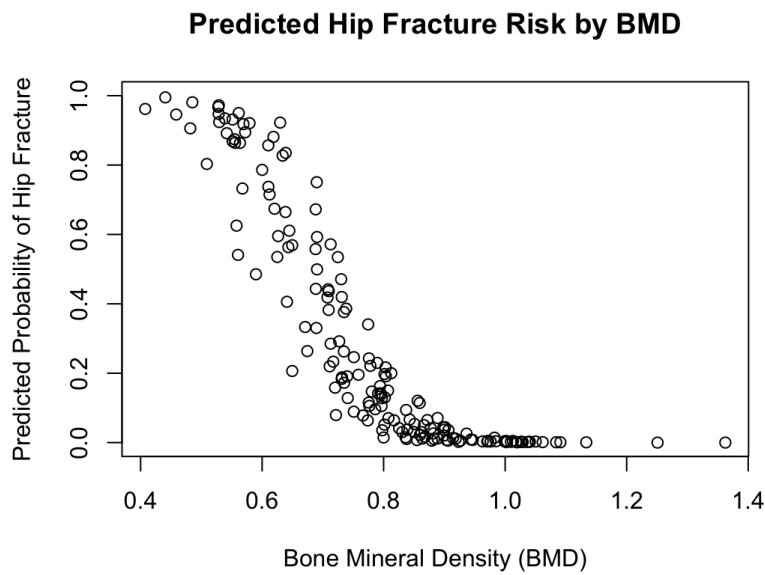**Predicted Hip Fracture Risk by BMD**

Figure 5. Scatterplot of Predicted Probability of Hip Fracture Vs BMD Based on The Logistic Regression Model

Figure 5 shows a scatterplot that supports BMD as a valid predictor of hip fracture. It shows that as BMD increased the chance of a hip fracture rapidly decreased. The final model could be used to identify patients more at risk and predict the probability of hip fractures if they know their BMD, height and medication. The final model's moderate predictive ability is supported by the negative coefficient between BMD and hip fracture rates.

In conclusion, other factors also contribute to hip fractures, however BMD is the most significant. There are certain limitations to the model such as other factors not considered such as bone health and injuries. Therefore the model is valid to predict risk probability, although is not a definitive statistical predictor such as a diagnosis. Overall, it has a clinically useful predictive ability, with BMD being the most efficient predictive variable.