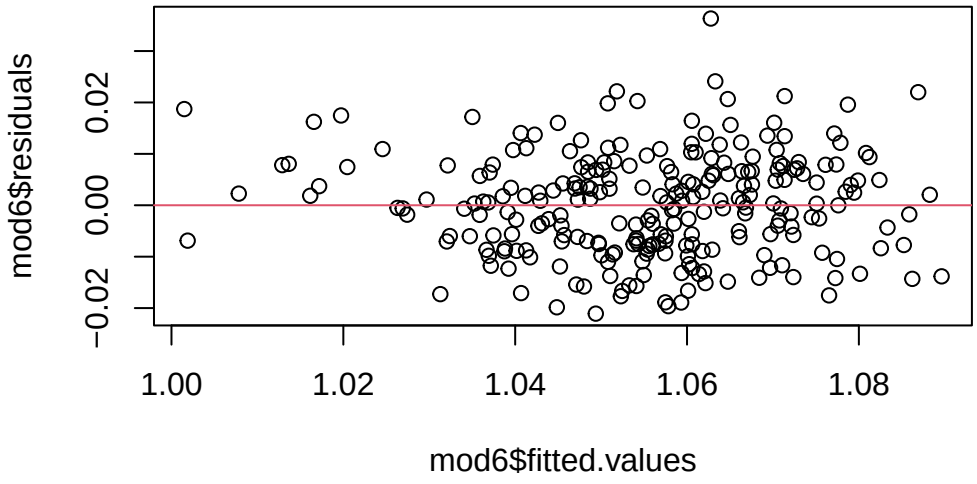
Lab_6

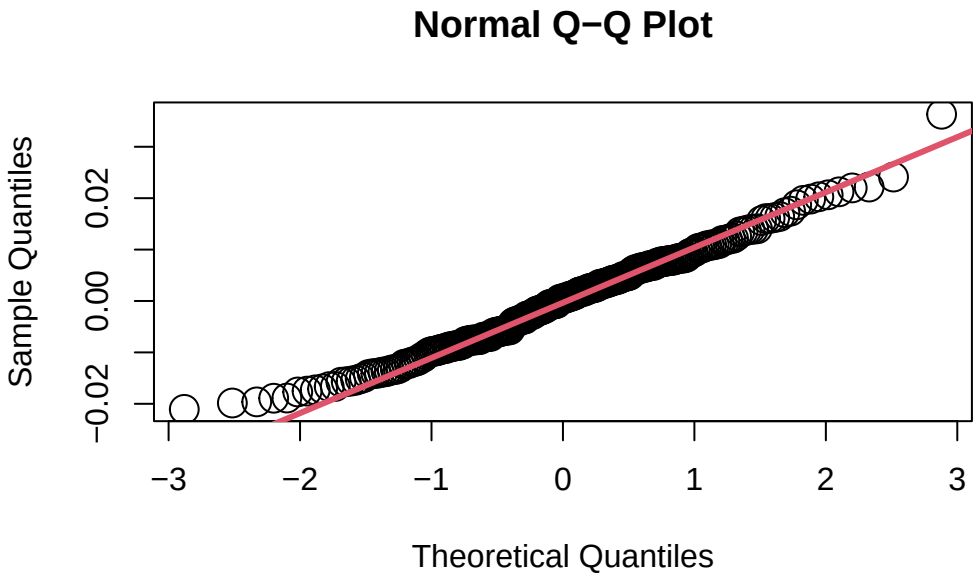
Brad Staples

1i

I would recommend the six term model created by regsubsets, which chose the predictors of intercept, weight, abdomen, biceps, neck forearm and wrists,as the best for estimating body density. These were the same terms picked by both forward and backward selection, which supports them as the best predictor variables for the model. The model we made in step g, which eliminated terms with high VIF, had a AIC of –1473.451 which was nearly 120 points higher than the six-term model, meaning the elimination model is of lesser statistical value than the other models. The VIF values in the regsubsets model are mostly acceptable, with body weight being 8.8 which is quite high but makes sense given the circumstances of the data. While this high VIF is not ideal, I feel the significant difference in AIC between the elimination model and the regsubsets model we chose makes the six-term model the better option of the two.

1j





I'd say homoscedasticity is a bit questionable. There's a slight fanning pattern from left to right and a big cluster of points near the middle, which suggests the variance isn't perfectly

equal. The QQ plot looks pretty good, with no major deviations from the normal line except near the tails. I don't see any obvious independence issues, so I'd mostly focus on transforming some of our predictors to make the linearity and homoscedasticity a little stronger.

2g

All three of the models for stepwise, backwards and forwards created the same model for the best quality wine, with the predictors of alcohol, volatile acidity, sulphates, total sulfur dioxide, chlorides, and pH. Since the models are identical, it's straightforward to recommend all three of them equally for predicting wine quality. When looking at the predictors to see if anything could be added, I noticed citric acid wasn't included. Citric acid can give a fruitier, more acidic, and robust flavor to wine, and when I created a model including it, the AIC went up just slightly. Perhaps I am biased since I enjoy the taste of citrus but I think adding citric acid could still be a real winner for improving the model's predictions.