|  |  |  |
| --- | --- | --- |
| Cover Page | P1 | Title and member information |
| Introduction Project Background | P2 | Why conduct research on body fat + Limitations of current body fat measurement methods |
| Data Description | P3 | Describe what data is included (this page describes what variables are available and what the units are) |
| Data Description | P4 | Description of what data is included (this page explains how Density and Pcr.BF were obtained and what the relationship between the two is) |
| Data preprocessing | P5 | It shows that there is a recording problem with Density and Pct.BF data (the Density has data out of range that should be deleted) |
| Data preprocessing | P6 | It shows that there is a recording problem with Density and Pct.BF data (the latter is calculated based on Density but the image shows outliers, so recalculate Pct.BF based on Density) |
| View dependent and independent variable relationships | P7 | Using the scatterplot to examine other variables in relation to Pct.BF while identifying any outliers |
| View dependent and independent variable relationships | P8 | Using a heatmap to examine the correlation coefficients between variables and to confirm whether there is multicollinearity |
| Assumption check | P9 | Using VIF to check for collinearity (high, but planning to retain and address it) (Simultaneously address singularity) |
| Assumption check | P10 | Using residual plots to check for homoscedasticity and normality |
| Using stepwise regression & stepwise subset selection to reduce multicollinearity | P11 | Present the results and models of forward stepwise, backward stepwise regression and stepwise subset selection |
| Using stepwise regression to reduce multicollinearity | P12 | Use cross-validation to evaluate the model, indicating that the stepwise subset selection method performs better but does not eliminate heteroscedasticity. |
| Using GLS to Address Multicollinearity | P13 | Introduction to GLS (Innovative Parts) |
| Using GLS to Address Multicollinearity | P14 | Using cross-validation and resampling to demonstrate that GLS performs better. (Innovative Parts) |
| Introduction of Derived Variables | P15 | Addressing the issue of multicollinearity while maintaining practical significance. |
| Re-experimentation | P16 | Presenting experimental conclusions: all models experienced a decrease in performance, but it was noted that GLS showed a slight improvement after resampling. |
| Analysis of the Unsuitability of BMI/WHR | P17 | Highlighting the specificity of BMI and WHR, which may render them unsuitable for predicting body fat percentage. |
| Summary | P18 | Summary |
| Limitations | P19 | Limitations |
| Future Direction | P20 | Future Direction |