Analysis of Body Fat Predictors

Samuel Van Gorden (995998955)

[spvangorden@ucdavis.edu](mailto:spvangorden@ucdavis.edu)

**Abstract**

The main objective of this paper is to discover and compare linear models that can be used to predict body fat percentage in individuals. Given our set of available predictors, we have found that body density has almost perfect correlation with body fat percent and can therefore be used in a simple linear model, with a slight response transformation, to predict body fat. Given that the method of obtaining body density requires specific scientific tools that may not be available to most (Katch and McArdle, 1977), we also set out to determine if some of the more easily measurable predictors could be used to generate a model that is comparable to the density-only model. We found that while there are several models that can be shown to have moderately good predictive power (R-squared > 0.6), none of them come close to the predictive power of the density-only model. We also set out to observe how each model performs with respect to mean squared estimation error (MSEE) and its components of squared bias and model variance. We performed validation of the assumptions of the normal error model on each of our models and made corrections where needed. Finally, we conducted an assessment of how our models performed in a randomized test scenario by splitting our data into training and validation subsets and running the trained models on the validation data.

**Introduction**

Our project is motivated by finding an inexpensive and accurate alternative method of measuring body fat. The dataset we are using includes numerous body metrics, such as body weight and height, thigh, abdomen, and bicep circumference, and age. It also includes body density, which is measured by taking one's underwater weight. All subjects in the data collected were men. Using the density measurement is likely not a widely available or easy to access option, so one of our goals is to assess whether the body metrics can be used by themselves to obtain an accurate and precise model for body fat percentage. We also set out to discover which single predictor (outside body density) is the strongest predictor of body fat, how our models perform when subsetting the data into training and validation sets, which models provide the best bias-variance tradeoff, whether converting some of the variables to categorical can provide a better model, and how our models hold up against the normal error model assumptions.

**Methods and Results**

We started by doing an exploratory data analysis. This included looking at histograms of each variable (see figure 1), creating a correlation matrix for all the potential predictor variables to determine if there may be multicollinearity in our models, and plotting each potential predictor variable against percent body fat to get a sense of what type of relationship they have (linear, non-linear, no relationship) (see figure 2). We did a preliminary round of data cleaning by removing a few observations that had values which seemed incorrect, such as a body fat percentage of 0.0, and then did several more rounds of removing highly influential points. This involved calculating the Cook’s distance for every observation, plotting them against a cutoff threshold calculated as 4/(n-p-2), removing points above this threshold, and then repeating the process until most of the points appeared to have similar values for the Cook’s distance (see figures 3 and 4).

Once the exploratory data analysis was complete we generated a few models based on our results. We generated a simple linear model using density as the only predictor. The adjusted R-squared value was extremely high (0.9997), and the plot of density against percent body fat shows an almost perfectly linear relationship (see figure 2). However, the residual vs fitted plot shows nonlinearity and the normal Q-Q plot shows non-normality of the errors (heavy right tail, light left tail). Because of this, we suspected that a variable transform may be necessary. We performed the Box-Cox procedure and found that using a lambda value of 0.93 seemed to correct the assumption violations and the adjusted R-squared even increased to 0.9999 (see figure 5). This was a bit surprising considering 0.93 is so close to 1 (i.e. close to no transformation at all).

Since the plots of each potential predictor (excluding density) against percent body fat showed that the relationships were either linear or non-existent, we considered only first-order models. In order to determine the Cook’s distances we had to run the full model with all variables (adjusted R-squared = 0.7312). From this we saw that age, neck, abdomen, hip, thigh, and wrist were significant at alpha <= 0.05. We began by considering the model comprised of forearm and ankle circumference since these variables seemed to be correlated with percent body fat while not being correlated with the other variables. This turned out to be a pretty bad model (ankle p-value > 0.2, very low adjusted R-squared of 0.1441). We then tried a model with all variables except hip and thigh, since they seemed to be the most highly correlated with the largest number of other potential predictor variables. This model had a much better adjusted R-squared (0.7262), however the only variables significant at < 0.1 were abdomen and wrist. We then took the significant variables (alpha <= 0.05) from the full model and created a new model, which was better than the previous one (adjusted R-squared = 0.7309) and possibly more useful since it was less complex (just over half the number of predictors). This model will be referred to as the “reduced additive model” going forward. It is worth noting that while density is clearly the best single predictor of body fat percentage, abdomen circumference appears to stand alone as a distant second best. This makes intuitive sense since body fat tends to accumulate in the abdomen in men due to visceral fat (Nauli and Matin, 2019). Because of this we also trained a simple linear model with abdomen as the only predictor. The adjusted R-squared was 0.6753, which is decent but not as high as most of the other models.

Next we leveraged the power of R and greedy algorithms by utilizing the stepAIC function to determine the best models based on the AIC and BIC criterions. We used backward and forward stepwise and selection procedures for both AIC and BIC. We found that the forward stepwise/selection BIC methods produced the same model (abdomen, weight, and wrist) while the backward stepwise/selection BIC methods produced the same model (age, abdomen, hip, thigh, and wrist). The forward stepwise, backward stepwise, and backward selection AIC procedures produced the same model (age, neck, abdomen, hip, thigh, forearm, and wrist) while the forward selection AIC procedure produced a model with the same predictors plus weight. The forward selection AIC model has a slightly better adjusted R-squared (0.7347) than the rest of the AIC models, and the backward selection/stepwise BIC models have better adjusted R-squared (0.7262) so we will use these going forward.

Finally, we converted the age and height variables into categorical variables and fit/refit several models. For age, we used three categories: “young” (under 36), “middle-aged” (36-54), and “old” (over 54). For height we used two categories: “short” (70.25 inches and under) and “tall” (over 70.25 inches). We determined these thresholds by using quartiles for each variable (first and third for age, second for height). We fit a model with all variables, just the categorical variables, all of the significant variables from the anova results of the full model, and a refitting of the reduced additive model and the forward selection AIC model (age, weight, neck, abdomen, hip, thigh, forearm, and wrist). The adjusted R-squared value of the refitting of the forward selection AIC model was 0.732, which was better than any (non-full) model (without density) that we had tested so far.

We decided to move forward with the backward stepwise BIC model (referred to as “BIC model” from here on), the reduced additive model, the refitting of the forward selection AIC model with categorical age variable (referred to as “categorical model” from here on), and the model with abdomen only. The residuals vs fitted plots and normal Q-Q plots for these four models indicate that the model assumptions do not seem to be violated in any major way (see figure 6), so no transformations are needed.

Next, we calculated the squared bias and model variance components of MSEE for each model we considered. For the density-only model, the squared bias was 3959 and the model variance was 0.003909. The squared bias was 4726, 3918, 3770, and 3836 and the model variance was 39.718, 100.472, 163.914, and 115.231 for the abdomen-only, BIC, categorical, and reduced additive models, respectfully. Surprisingly, there were actually two models with better overall MSEE than the density model (MSEE = 3959). These were the categorical model (MSEE = 3934) and the reduced additive model (MSEE = 3951).

Finally, we split the data sets we had used before into two random subsets of equal size - one for training and another for validation. We then refit the same five models on the training data and plugged the validation set data into each newly generated model to obtain fitted results. Next, we generated the mean squared prediction error (MSPE) for each model by taking the sum of the squared difference between the fitted results and the actual test data and dividing by the number of data points in the training data. In doing so, we observed that the density-only model had the lowest MSPE (15.076) and the categorical model had the lowest MSPE of all the non-density models (17.084). The remaining three models all had similar values of MSPE (around 20). Based on this information, the density-only model performs the best with respect to prediction, but the categorical model is not far behind.

**Conclusions and Discussion**

Our main goal was to discover whether or not there was a model for determining body fat percentage that could perform as well as any model containing density. Specifically we were comparing the simple linear model containing only density with other models *not* containing density. We came up with four such models - one which contained all of the variables found to be individually significant at the 0.05 level upon fitting the full additive model (the reduced additive model), one which was determined by using the stepAIC function to determine which model (locally) minimizes the BIC criterion, one which takes the (locally) minimized AIC criterion model and replaces the age quantitative variable with a categorical representation of age, and one which is a simple linear model using abdomen, which appeared to be the best single predictor for body fat percentage outside of density.

We utilized various measures to determine which model provided the best fit of the data. The first metric we used in assessing models, which we used to determine the final five, was the adjusted R-squared. We used adjusted R-squared instead of just R-squared because adjusted R-squared accounts for the number of parameters used in the models by calculating mean squared regression error and mean squared total error instead of just using the sum of squares. By doing this we chose models that fit the data well (high correlation with response variable) but did not overfit the data (too many parameters).

Having chosen our models for comparison to the density-only model, we then calculated the squared bias, model variance, and MSEE for each model. The squared bias told us which model fit the data best, the model variance told us which model would be best for making future predictions, and the MSEE was the sum of these two components which gave us a metric for determining which model performed the best with respect to both fit of the current data *and* prediction of future data. We saw that the density-only model had very low model variance compared to the other models, but there were actually three non-density models which had lower squared bias than the density-only model. Only the abdomen-only model had higher bias, though it was quite a bit higher. There were two non-density models (categorical and reduced additive) with lower MSEE than the density-only model which was surprising considering the very high adjusted R-squared for the density-only model compared to the non-density models.

Finally, after splitting our data into training and validation sets and refitting and testing each of the five models, we found that the density-only model did perform better than the non-density models with respect to MSPE, but not by much. This tells us that even though our initial model for density-only had a much lower model variance than the non-density models, the training model with the validation data fitted on it did not perform much better for prediction than did the non-density models.

Our results show that while the density-only model appears to have an almost perfectly linear association with percent body fat, several other models without density appear to perform at least as well if not better than the density-only model. Specifically, the model which leveraged the AIC criterion for minimizing MSEE, then converted the age variable to a categorical variable performed better with respect to squared bias, MSEE, and external validation. The categorical variable model also performed better than the other non-density models with respect to adjusted R-squared. The categorical model included measurements from different parts of the body, so it seemed like it did a good job of covering any parts where fat could be stored without overcounting body parts that are close to one another, thereby causing multicollinearity. This seems to confirm that it was worth looking at the possible models with quantitative variables converted to categorical. Perhaps the associations between the remaining qualitative variables provided a better model for body fat percent when isolated to a particular age group (i.e. different age groups have different body fat storage patterns). Due to the success of converting the age and height variables to categorical, it may also be useful to consider models in which other variables (perhaps weight) are converted to categorical.

Although the abdomen-only model appeared to perform the worst out of all five that we considered, it was not off by a huge margin when considering metrics of model success. For this reason, it could be useful as it provides a decent method of predicting body fat with minimal measurements required (only one).

Overall, our findings seem to suggest that an accurate and precise model for determining body fat percentage can be provided given the variables from the data set we looked at without having to utilize the difficult to access methods of obtaining body density. This will be helpful for normal people who wish to get a rough estimate of their body fat percentage for health purposes without having to go through expensive, invasive, or time consuming procedures. One of the limitations of our data is that it includes only men, and women tend to have different fat-accumulation patterns than men (Blaak, 2001). In the future it would be useful and interesting to create a model specifically for women and see how it differs from the ones we generated for men, and also to see what a model generated with a categorical gender variable would look like. It could also be useful to look at more advanced types of regression modeling such as ridge regression or principal component regression.

**Appendix**

**Plots/Figures**

Figure 1: Variable Distributions

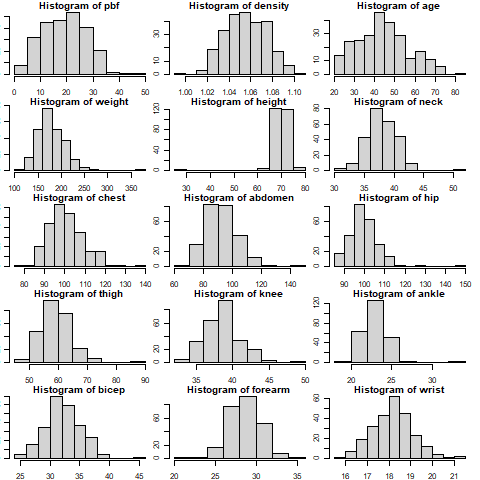


Figure 2: Predictor vs Response

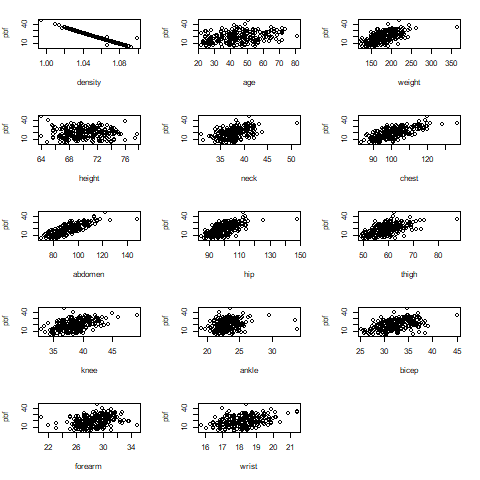


Figure 3: Cook’s Distance (Density-Only Model)

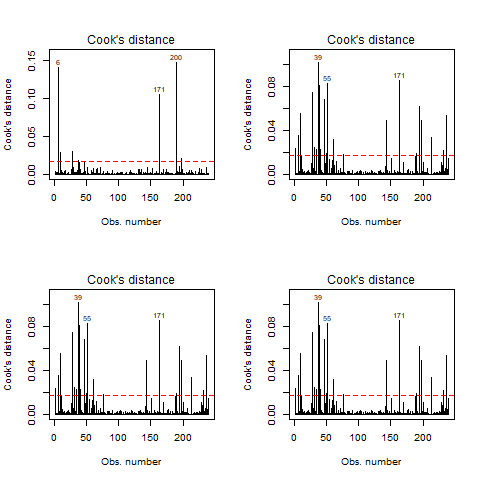


Figure 4: Cook’s Distance (Full Model w/o Density)

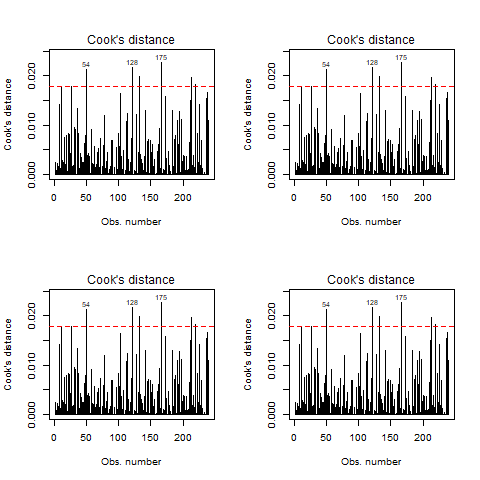


Figure 5: Density Model Assumptions

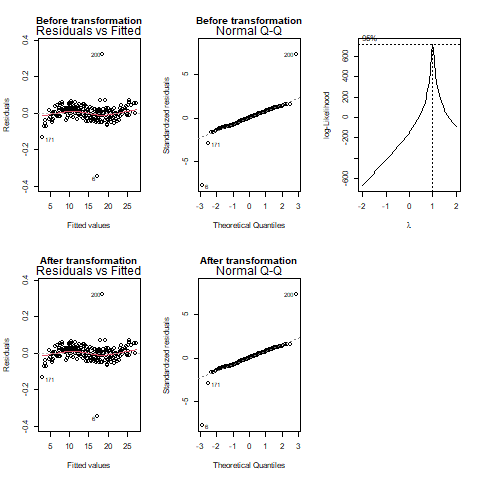
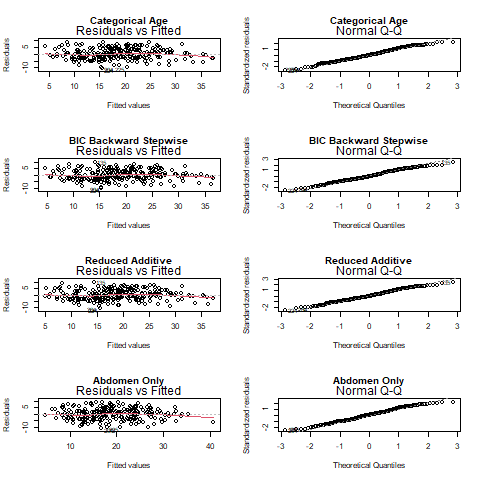


Figure 6: Non-Density Model Assumptions



**Code**

| ## ----setup, include=FALSE----------------------------------------------------- knitr::knit\_hooks$set(purl = knitr::hook\_purl) knitr::opts\_chunk$set(echo = TRUE)  ## ---- echo=FALSE, results='hide'---------------------------------------------- library(MASS)  # Read data and draw histograms data <- read.table("bodyfat.txt") names(data) <- c("density","pbf","age","weight","height","neck","chest","abdomen","hip","thigh","knee","ankle","bicep","forearm","wrist") data <- data[,c("pbf","density",names(data)[3:length(names(data))])] sapply(data, class) apply(sapply(data, is.na), 2, sum)  png("VarDistributions.png") par(mfrow = c(5,3), mar = c(2,1,1,1))  for (i in 1:length(names(data))) {  hist(data[,i], main = paste("Histogram of", names(data)[i])) } dev.off()  ## ---- echo=FALSE, results='hide'---------------------------------------------- # Remove obvious bad data data <- data[!(row.names(data) %in% c(42,172,182)),] orig.data <- data  # Observe correlations between potential predictor variables cor.mat <- cor(data[,2:ncol(data)]) cor.mat cor.mat < .7  # Plot response variable against each predictor variable png("VarVsResponse.png") par(mfrow = c(5,3), mar = c(5,4,2,2)) for (i in 2:length(names(data))) {  plot(data[,i], data$pbf, xlab = names(data)[i], ylab = "pbf") } dev.off()  y.cor <- cor(data$pbf, data[,2:ncol(data)]) y.cor  ## ----------------------------------------------------------------------------- density.data <- data[,c("pbf","density")]  # Fit pbf on density only model.density <- lm(pbf~density, data = density.data) summary(model.density)  png("DensityCooks.png") par(mfrow = c(2,2))  cutoff <- 4/(nrow(data)-length(model.density$coefficients)-2) plot(model.density, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  # Remove points and check again for influential points density.data <- density.data[!(row.names(density.data) %in% c(48,96,216)),] model.density <- lm(pbf~density, data = density.data) cutoff <- 4/(nrow(density.data)-length(model.density$coefficients)-2) plot(model.density, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  # Remove points and check again for influential points density.data <- density.data[!(row.names(density.data) %in% c(36,76,169)),] model.density <- lm(pbf~density, data = density.data) cutoff <- 4/(nrow(density.data)-length(model.density$coefficients)-2) plot(model.density, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  # Remove points and check again for influential points density.data <- density.data[!(row.names(density.data) %in% c(26,192,242)),] model.density <- lm(pbf~density, data = density.data) cutoff <- 4/(nrow(density.data)-length(model.density$coefficients)-2) plot(model.density, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  dev.off()  summary(model.density)  ## ----------------------------------------------------------------------------- png("DensityAssumptions.png") par(mfrow = c(2,3)) plot(model.density, which = c(1,2), main = "Before transformation") boxcox(model.density) model.density <- lm(pbf^0.93~density, data = density.data) #summary(model.density) plot(model.density, which = c(1,2), main = "After transformation") dev.off()  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- nodensity.data <- data[,!(names(data) %in% c("density"))]  # Fitting full additive model model.full.add <- lm(pbf~., data = nodensity.data) summary(model.full.add) anova(model.full.add)  png("NoDensityCooks.png") par(mfrow = c(2,2))  # Check for influential points cutoff <- 4/(nrow(nodensity.data)-length(model.full.add$coefficients)-2) plot(model.full.add, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  # Remove points and check again for influential points nodensity.data <- nodensity.data[!(row.names(nodensity.data) %in% c(39,86,221)),] model.full.add <- lm(pbf~., data = nodensity.data) cutoff <- 4/(nrow(nodensity.data)-length(model.full.add$coefficients)-2) plot(model.full.add, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  nodensity.data <- nodensity.data[!(row.names(nodensity.data) %in% c(82,207,216)),] model.full.add <- lm(pbf~., data = nodensity.data) cutoff <- 4/(nrow(nodensity.data)-length(model.full.add$coefficients)-2) plot(model.full.add, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  nodensity.data <- nodensity.data[!(row.names(nodensity.data) %in% c(3,36,81)),] model.full.add <- lm(pbf~., data = nodensity.data) cutoff <- 4/(nrow(nodensity.data)-length(model.full.add$coefficients)-2) plot(model.full.add, which = 4, cook.levels = cutoff) abline(h = cutoff, lty = 2, col = "red")  dev.off()  summary(model.full.add)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- model.forearmamkle <- lm(pbf~forearm+ankle, data = nodensity.data) summary(model.forearmamkle)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- model.nohipthigh <- lm(pbf~., data = nodensity.data[,!(names(nodensity.data) %in% c("hip","thigh"))]) summary(model.nohipthigh)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- model.red.add <- lm(pbf~age+neck+abdomen+hip+thigh+wrist, data = nodensity.data) summary(model.red.add)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- summary(lm(pbf~age+neck+abdomen+hip+thigh+wrist, data = orig.data))  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- # Fit model with abdomen (highest correlation with pbf outside of density) only model.abdomen <- lm(pbf~abdomen, data = nodensity.data) summary(model.abdomen)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- nodensity.model0 <- lm(pbf~1, data = nodensity.data)  # AIC/forward stepwise model.aic.fwstp <- stepAIC(nodensity.model0, scope=list(upper=model.full.add, lower = ~1), direction = "both", k = 2, trace = FALSE) summary(model.aic.fwstp)  # AIC/backward stepwise model.aic.bwstp <- stepAIC(model.full.add, scope=list(upper=model.full.add, lower = ~1), direction = "both", k = 2, trace = FALSE) summary(model.aic.bwstp)  # AIC/forward selection model.aic.fwsel <- stepAIC(nodensity.model0, scope=list(upper=model.full.add, lower = ~1), direction = "forward", k = 2, trace = FALSE) summary(model.aic.fwsel)  # AIC/backward selection model.aic.bwsel <- stepAIC(model.full.add, scope=list(upper=model.full.add, lower = ~1), direction = "backward", k = 2, trace = FALSE) summary(model.aic.bwsel)  # BIC/forward stepwise model.bic.fwstp <- stepAIC(nodensity.model0, scope=list(upper=model.full.add, lower = ~1), direction = "both", k = log(nrow(nodensity.data)), trace = FALSE) summary(model.bic.fwstp)  # BIC/backward stepwise model.bic.bwstp <- stepAIC(model.full.add, scope=list(upper=model.full.add, lower = ~1), direction = "both", k = log(nrow(nodensity.data)), trace = FALSE) summary(model.bic.bwstp)  # BIC/forward selection model.bic.fwsel <- stepAIC(nodensity.model0, scope=list(upper=model.full.add, lower = ~1), direction = "forward", k = log(nrow(nodensity.data)), trace = FALSE) summary(model.bic.fwsel)  # BIC/backward selection model.bic.bwsel <- stepAIC(model.full.add, scope=list(upper=model.full.add, lower = ~1), direction = "backward", k = log(nrow(nodensity.data)), trace = FALSE) summary(model.bic.bwsel)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- cat.data <- nodensity.data summary(cat.data$age) cat.age <- c() cat.age[cat.data$age <= 35] <- "young" cat.age[cat.data$age > 35 & cat.data$age <= 54] <- "mid" cat.age[cat.data$age > 54] <- "old" cat.data$age <- as.factor(cat.age)  summary(cat.data$height) cat.height <- c() cat.height[cat.data$height <= 70.25] <- "short" cat.height[cat.data$height > 70.25] <- "tall" cat.data$height <- as.factor(cat.height)  # All predictors model.cat <- lm(pbf~., data = cat.data) summary(model.cat) anova(model.cat)  # Only categorical predictors model.cat2 <- lm(pbf~height+age, data = cat.data) summary(model.cat2) anova(model.cat2)  # All significant (< 0.05) predictors from anova model.cat3 <- lm(pbf~age+weight+height+neck+chest+abdomen+thigh+wrist, data = cat.data) summary(model.cat3) anova(model.cat3)  # Redo reduced additive model with categorical age model.cat4 <- lm(pbf~age+neck+abdomen+hip+thigh+wrist, data = cat.data) summary(model.cat4) anova(model.cat4)  # Redo selected AIC model model.cat5 <- lm(pbf~age+weight+neck+abdomen+hip+thigh+forearm+wrist, data = cat.data) summary(model.cat5) anova(model.cat5)  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- png("NoDensityAssumptions.png") par(mfrow = c(4,2)) plot(model.cat5, which = c(1,2), main = "Categorical Age") plot(model.bic.bwstp, which = c(1,2), main = "BIC Backward Stepwise") plot(model.red.add, which = c(1,2), main = "Reduced Additive") plot(model.abdomen, which = c(1,2), main = "Abdomen Only") dev.off()  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- # Calculate squared bias and model variance for various models sq.bias.density <- t(density.data$pbf - model.density$fitted.values) %\*% (density.data$pbf - model.density$fitted.values) mse <- sum(model.density$residuals^2) / model.density$df.residual mod.var.density <- length(model.density$coefficients) \* mse msee.density <- sq.bias.density + mod.var.density  sq.bias.cat5 <- t(cat.data$pbf - model.cat5$fitted.values) %\*% (cat.data$pbf - model.cat5$fitted.values) mse <- sum(model.cat5$residuals^2) / model.cat5$df.residual mod.var.cat5 <- length(model.cat5$coefficients) \* mse msee.cat5 <- sq.bias.cat5 + mod.var.cat5  sq.bias.bic.bwstp <- t(nodensity.data$pbf - model.bic.bwstp$fitted.values) %\*% (nodensity.data$pbf - model.bic.bwstp$fitted.values) mse <- sum(model.bic.bwstp$residuals^2) / model.bic.bwstp$df.residual mod.var.bic.bwstp <- length(model.bic.bwstp$coefficients) \* mse msee.bic.bwstp <- sq.bias.bic.bwstp + mod.var.bic.bwstp  sq.bias.red.add <- t(nodensity.data$pbf - model.red.add$fitted.values) %\*% (nodensity.data$pbf - model.red.add$fitted.values) mse <- sum(model.red.add$residuals^2) / model.red.add$df.residual mod.var.red.add <- length(model.red.add$coefficients) \* mse msee.red.add <- sq.bias.red.add + mod.var.red.add  sq.bias.abdomen <- t(nodensity.data$pbf - model.abdomen$fitted.values) %\*% (nodensity.data$pbf - model.abdomen$fitted.values) mse <- sum(model.abdomen$residuals^2) / model.abdomen$df.residual mod.var.abdomen <- length(model.abdomen$coefficients) \* mse msee.abdomen <- sq.bias.abdomen + mod.var.abdomen  ## ---- echo=FALSE, results='hide', fig.show='hide'----------------------------- set.seed(1234567) density.indices.train <- sample(1:nrow(density.data), nrow(density.data)/2) density.data.train <- density.data[density.indices.train,] density.data.test <- density.data[-density.indices.train,] nodensity.indices.train <- sample(1:nrow(nodensity.data), nrow(nodensity.data)/2) nodensity.data.train <- nodensity.data[nodensity.indices.train,] nodensity.data.test <- nodensity.data[-nodensity.indices.train,] cat.indices.train <- sample(1:nrow(cat.data), nrow(cat.data)/2) cat.data.train <- cat.data[cat.indices.train,] cat.data.test <- cat.data[-cat.indices.train,]  model.density.train <- lm(model.density$model, data = density.data.train) model.cat5.train <- lm(model.cat5$model, data = cat.data.train) model.bic.bwstp.train <- lm(model.bic.bwstp, data = nodensity.data.train) model.red.add.train <- lm(model.red.add, data = nodensity.data.train) model.abdomen.train <- lm(model.abdomen, data = nodensity.data.train)  fit.vals.density <- predict(model.density.train, density.data.test) fit.vals.cat5 <- predict(model.cat5.train, cat.data.test) fit.vals.bic.bwstp <- predict(model.bic.bwstp.train, nodensity.data.test) fit.vals.red.add <- predict(model.red.add.train, nodensity.data.test) fit.vals.abdomen <- predict(model.abdomen.train, nodensity.data.test)  mspe.density <- sum((density.data.test$pbf - fit.vals.density)^2) / nrow(density.data.train) mspe.cat5 <- sum((cat.data.test$pbf - fit.vals.cat5)^2) / nrow(cat.data.train) mspe.bic.bwstp <- sum((nodensity.data.test$pbf - fit.vals.bic.bwstp)^2) / nrow(nodensity.data.train) mspe.red.add <- sum((nodensity.data.test$pbf - fit.vals.red.add)^2) / nrow(nodensity.data.train) mspe.abdomen <- sum((nodensity.data.test$pbf - fit.vals.abdomen)^2) / nrow(nodensity.data.train) |
| --- |

**References**

1. Katch, Frank and McArdle, William (1977). Nutrition, Weight Control, and Exercise, Houghton Mifflin Co., Boston.
2. Nauli, Andromeda M. and Matin, Sahar. “Why Do Men Accumulate Abdominal Visceral Fat?” *Frontiers in Physiology*. December 5, 2019. <https://www.frontiersin.org/articles/10.3389/fphys.2019.01486/full>
3. Blaak, E. “Gender differences in fat metabolism.” *National Library of Medicine*. November 4, 2001.