MULTIPLE LINEAR REGRESSION ANALYSIS

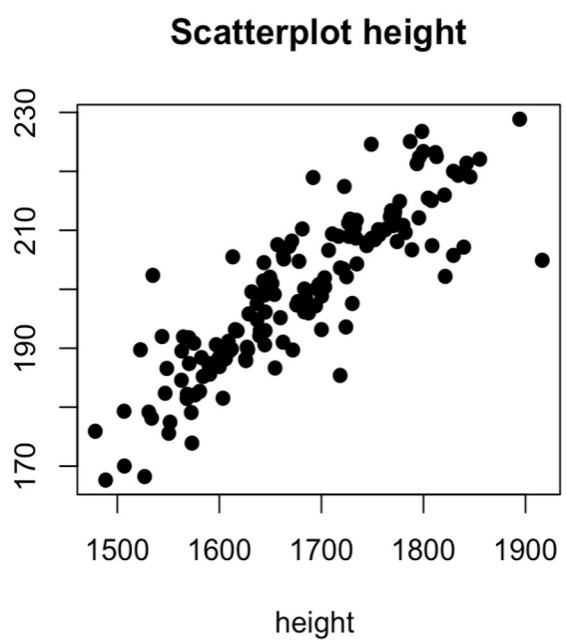
Jaiden Neff

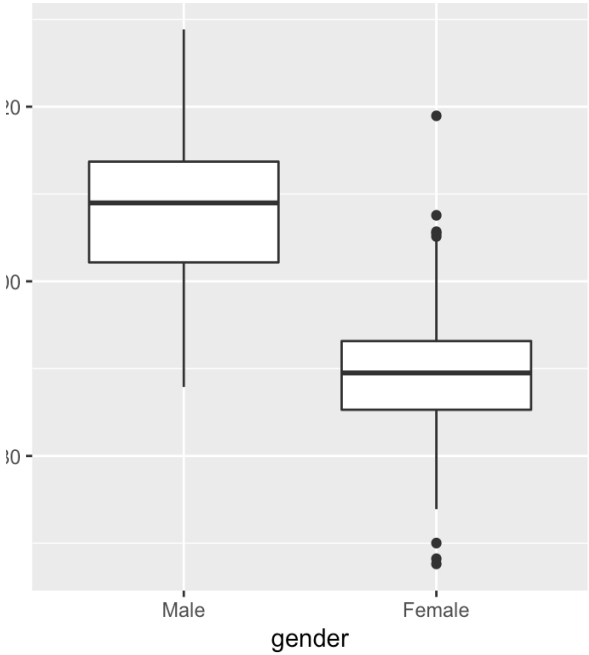
DATA DESCRIPTION &QUESTION

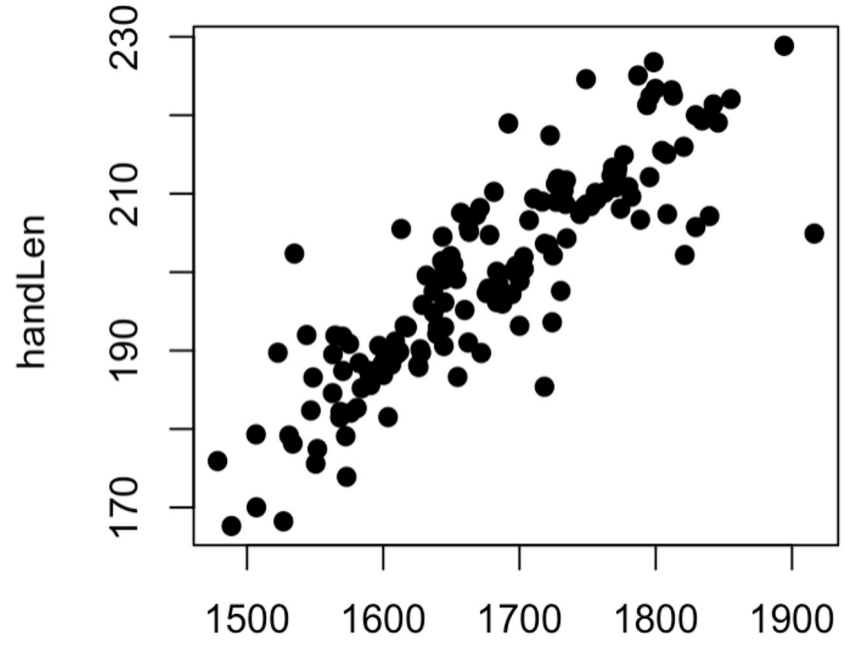
- The primary question is to identify the association between the hand length (y) and the variables that are closely related

- IdGen
- gender(X1): male or female
- height(X2): height in cm
- footLen(X3): foot length in cm
- handLen(Y): hand length in cm

EXPLORATORY ANALYSIS

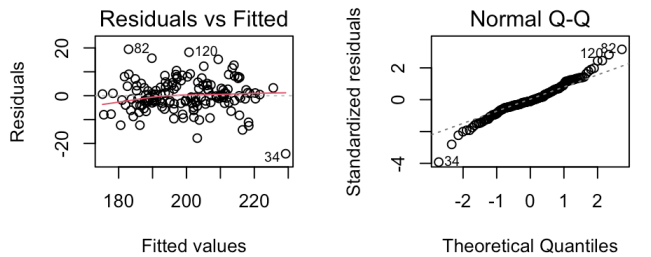






FULL MODEL AND ANALYSIS

- Here the statistics of regression coefficients is shown for the full model
- We also see the residuals that show us the spread of the data as well



Residual plots of the full model

Statistics of Regression Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.7838792	16.6341692	0.1072419	0.9147404
idGen	0.0201191	0.0234713	0.8571777	0.3927141
genderFemale	-1.0803455	1.7963585	-0.6014086	0.5484757
height	0.1093853	0.0133900	8.1691775	0.0000000
footLen	0.0559712	0.0608566	0.9197221	0.3591947

TRANSFORMATIONS

- Here we transform the models by using a square root transformation and a log transformation

4.1 Square root transformation

log-transformed model

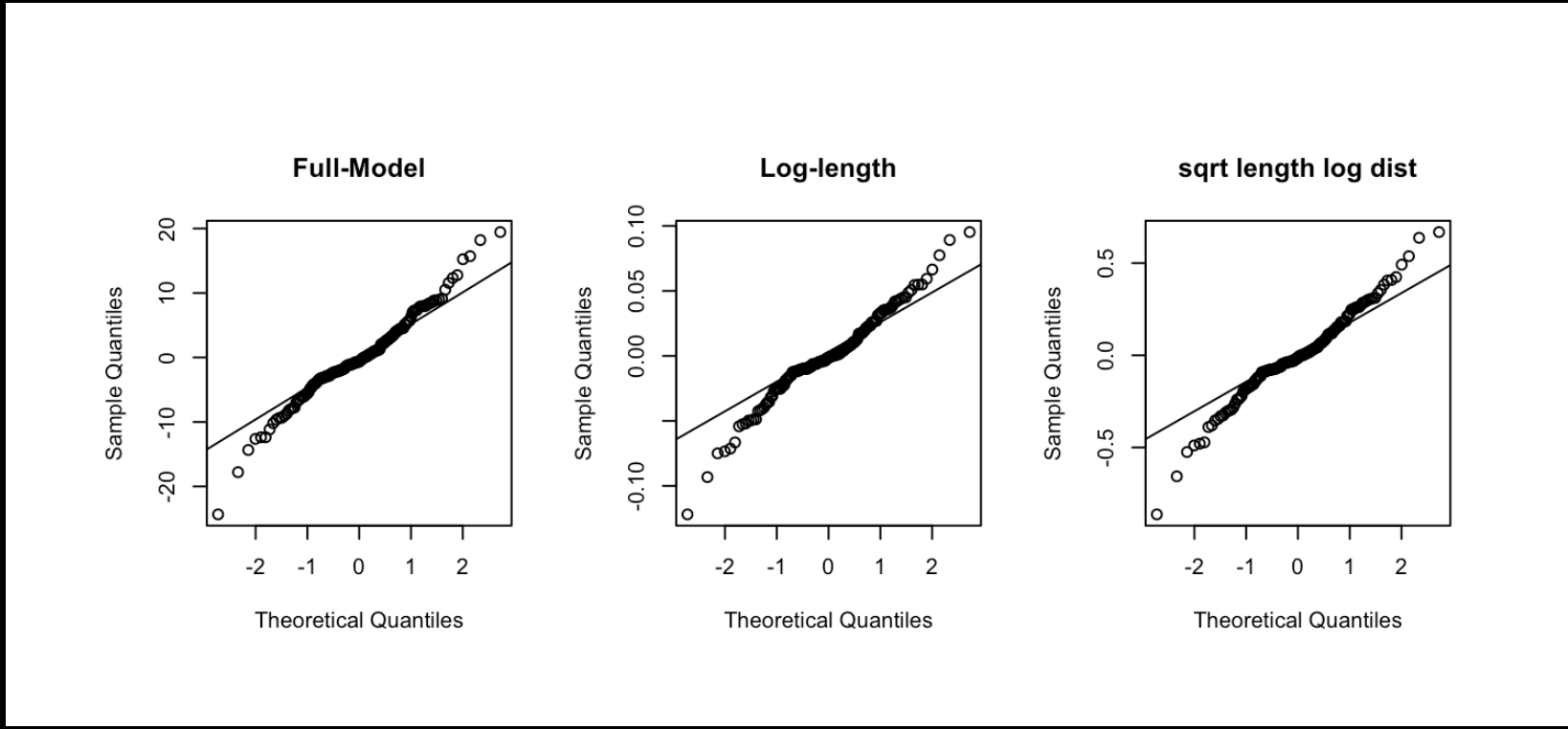
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.1362183	0.5890258	12.1152899	0.0000000
height	0.0038642	0.0004741	8.1506511	0.0000000
footLen	0.0020755	0.0021552	0.9630288	0.3370727
genderFemale	-0.0399784	0.0636355	-0.6282414	0.5307949

4.2 Log transformation

log-transformed model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.3042058	0.0837839	51.372728	0.0000000
height	0.0005444	0.0000674	8.073456	0.0000000
footLen	0.0003162	0.0003066	1.031457	0.3039757
genderFemale	-0.0057154	0.0090516	-0.631420	0.5287206

REGRESSION ANALYSIS



GOODNESS OF FIT

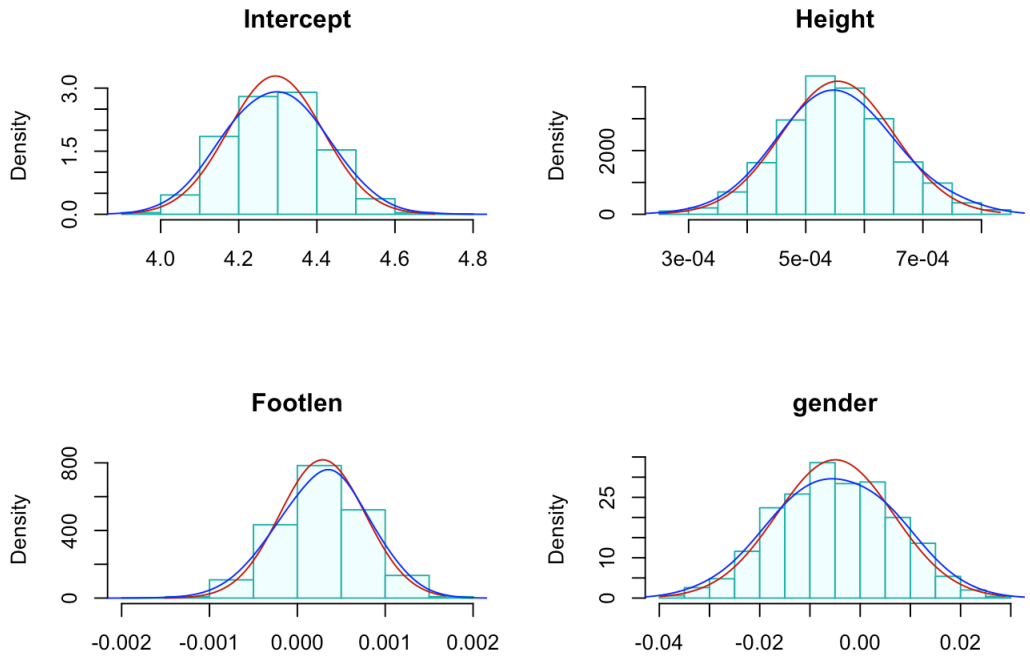
Goodness-of-fit Measures of Candidate Models

	SSE	R.sq	R.adj	Cp	AIC	SBC	PRESS
full.model	6420.4589478	0.7657848	0.7595391	5	587.1921	602.4092	7160.4399877
sqrt.length.log.dist	8.1141691	0.7645480	0.7598701	4	-449.2211	-437.0474	8.8925687
log.length	0.1641709	0.7637833	0.7590903	4	-1053.7922	-1041.6185	0.1798398

BOOTSTRAPPING REGRESSION

Regression Coefficient Matrix

	Estimate	Std. Error	t value	Pr(>|t|)	btc.ci.95
(Intercept)	4.3042	0.0838	51.3727	0.0000	[4.06 , 4.5359]
height	0.0005	0.0001	8.0735	0.0000	[4e-04 , 7e-04]
footLen	0.0003	0.0003	1.0315	0.3040	[-7e-04 , 0.0012]
genderFemale	-0.0057	0.0091	-0.6314	0.5287	[-0.0285 , 0.0176]



FINAL MODEL

$$\log(handLen) = 4.3042058 + 0.0005444 \times height + 0.0003162 \times footLen - 0.0057154 \times genderFemale$$

FINAL THOUGHTS

- The log transformation Is the best model for determining correlation between hand length and closely related variables
- Regression analysis wasn’t helpful
- Bootstrapping isn’t necessary for the final model because the data is normal
- Kept the hand length and gender even though they weren’t “highly correlated” because they are important and you can see in exploratory analysis that it has correlation.