Week_4_Problem_Set

Daniel Crownover

2024-02-12

Question 1:

The first question asked us to compare the relationship between Tree Height and Tree Diameter. We created a linear model and a system of tests, to test the assumptions of the correlation. I used figures 1.1.2 and 1.1.3 to show the paring matrix to see correlation across all variables. In Figure 1.1.4 we tested the linear correlation between mean tree height and tree diameter: with correlation coefficient (r =0.97) this indicates a strong positive linear relationship between Mean Diameter and Mean Height. The Null for the linear correlation is that there is no correlation between the two variables (in this case, Mean Diameter and Mean Height of trees), because (p = 0) there is statistic significance and we can assume correlation. Figure 1.1.5 is making linear model 1. The table and equation in 1.1.6 provide is the equation for a linear model for predicting the tree diameter based on the mean height of the tree. It suggests that the predicted tree diameter increases linearly with an increase in mean height, with a slope of 1.86315 and an intercept of -8.07896(see figure 1.1.5 and 1.1.6). The confidence intervals(97.5 and 2.5%) and predicted response variables of the linear model where plotted in 1.1.59 for a visualization. For time sake, a plot was not made again for the other re-run models until question 2.

The assumptions for Linear Model 1 plot. In plot linear model 1(figure 1.1.8) the residuals vs fitted plot shows a non linear horizontal line with distinct patterns and several outlying residuals noted. For the q-q plot in figure 1.1.8 the points show "significant departure" from the line. Suggesting we will need another model. The third model, standard resid vs fitted values on different scale(1.1.8) shows, the points distributed inside a triangular shape, with the scatter of the residuals increasing as the fitted values increase- suggesting we need to continue the data wrangling. The fourth plot in (1.1.8) is residuals-leverage and shows the plots that shows Cook's distance for each of the observed values. The point is to highlight those yi (response) values that have the biggest effect on parameter estimates. The idea is to verify that no single data point is so influential that leaving it out changes the structure of the model,I have done this and will be taken out later.

In figure 1.1.9 we are using the chi^2 to assess whether the variance of the errors in the linear regression model is constant across all levels of the predictor variable. Generally, we want Homoscedasticity (which is Lower error variance), which means that the spread of the residuals around the regression line remains constant across all levels of the predictor variable. Homoscedasticity is one of the assumptions of linear regression analysis and is the null hypothesis. When we have a high p value(p = 0.77) we fail to reject the null, or we don't have enough evidence to reject that the assumption of constant error variance is violated. This is good, however lets move onto the cook's model.

In figure 1.2.1, we identify all points of leverage using the cooks model and display them. Observations with large Cook's distances are potential outliers or influential points. These observations might have a significant impact on the estimated coefficients of the regression model. In 1.2.1 we see the top three with high residuals numbers, these have the potentials to be outliers, and when as i also saw this sames rows as problematic in the linear model plot, I am deciding to remove them in the re-run of the model. The p-value(p = 0.57) in the Durbin-Watson (figure 1.2.2) test assesses the presence of autocorrelation in the residuals of a regression model. Specifically, it tests whether there is a significant correlation between adjacent residuals, which could indicate a violation of the assumption of independence. We can assume is no significant autocorrelation detected in the residuals of the regression model. In figure 1.2.29, the shapiro wilks test has a p value(0.002) of high significance which means we reject that the null is that the residuals are normally distributed and need to wrangle the data to make it so we can fail to reject the null.

Linear model 2 I took out outliers identified by the various test in linear model 1 displayed in figure 1.2.4. Figure 1.2.49 makes linear model 2 without outliers. The predicted response variable equation shows -8.73904 intercept and 1.89 slope for mean height. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm2 show straighter line for residuals vs fitted(2.2.7) but slight variation indicating we need to log it. Plot 2 (fig. 2.2.7) shows a s shape, showing we need to log it. And plot 3 shows the same thing plot 1 did. Plot 4 shows we essentially need to log the data.

The P value increased in the Shapiro wilks (2.2.11)test ($p = 0.15$) which means that we fail to reject that the data is normally distributed(assuming it is normally distrubuted) but to make that p-value higher we should log the data. The non-constant variance test(2.28) p value is still high at($p = 0.6141$) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 2.2.9)helped us identify some more residuals, however, it showed that the ones removed may have helped increase that p value in the shapiro test but still we know that the largest outliers were removed AND the data still fails the assumptions of good regression from the linear plots, so LETS LOG the data to see if this helps even more!

Linear model 3 I logged the data frame without the outliers. Figure 2.3.2 makes linear model 3. The predicted response variable equation shows -0.766 intercept and 1.38 slope for mean height these values you can expect to see when the data is normilized. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm2 show MUCH straighter line for residuals vs fitted(2.3.5). Plot 2 (fig. 2.3.5) shows a MUCH straighter shape fitting to that line, showing that the logarithm proved sucsessfull And plot 3 shows the same thing plot 1 did. Plot 4 shows a much straighter line.

The P value increased in the Shapiro wilks (2.3.9)test ($p = 0.348$) which means that we fail to reject that the data is normally distributed(assuming it is normally distributed) and make it much higher than last time. The non-constant variance test(2.3.6) p value is still high at($p = 0.8$) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 2.3.7)helped us verify some more residuals, however, it showed that the ones removed and the logged values may have helped increase that p value in the shapiro test. We still don't know if logged with outliers made a difference. We are going to test this next.

Linear model 4 I logged comparison with the outliers. Figure 3.1.2 makes linear model 3. The predicted response variable equation shows -0.66 intercept and 1.35 slope for mean height these values you can expect to see when the data is normilized. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm2 show a straight line but it is still wonkier than the logged without outliers(3.3.5). Plot 2 (fig. 3.3.5) shows a slightly straighter shape fitting to that line, showing that the logarithm proved good but containing the outliers still disrupts the data.And plot 3 shows the same thing plot 1 did. Plot 4 shows a much straighter line.

The P value increased in the Shapiro wilks (2.3.9)test ($p = 0.0165$) which means that we reject that the data is normally distributed(assuming it is normally distributed) but sine this is much lower than the logged data without outliers, it suggest that there is more reason to assume that the outliers have a large effect on the linear correlation of the data frame and should be removed.

The non-constant variance test(3.3.6) p value is still high at($p = 0.84$) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 3.3.7)helped us verify some more residuals, however, it showed that the ones removed and the logged values may have helped increase that p value in the shapiro test. We now know that the logged values without outliers are the winner here and provide.

Appendix:

1. Question 1: Null and Alternative Hypothesis
    - 1.1.2 Pairing Matrix tdat
    - 1.1.3 Pairing Matrix tdat
    - 1.1.4 Simple Linear Correlation
    - 1.1.49 [Simple Linear Model 1] [raw data with outliers] [Below]
    - 1.1.5 Making Linear Model 1 [raw data with outliers]
    - 1.1.6 Predicted Response Variable in Liner Model 1 [raw data with outliers]
    - 1.1.7 Confidence Intervals of Liner Model 1 [raw data with outliers]
    - 1.79 Assumptions of Linear Model 1 [raw data with outliers] [Below]
    - 1.1.8 Plot Linear Model 1 [raw data with outliers]
    - 1.1.9 Non-constant Error Variance Test Linear Model 1 [raw data with outliers]
    - 1.2.1 Cook's Model: Identifying Points of Leverage Linear Model 1 [raw data with outliers]
    - 1.2.2 Durbin Watson Test Linear Model 1 [raw data with outliers]
    - 1.2.29 Shapiro Wilks Test Linear Model 1 [raw data with outliers]
    - 1.2.299 Report Results Simple Linear Model 1 [raw data with outliers]
    - 1.2.3 [Simple Linear Model 2] [filtered and excluding outliers] [Below]
    - 1.2.4 Filtering and Excluding Extemporaneous Data [filtered and excluding outliers]
    - 1.2.49 Making Simple Liner Model 2 [filtered and excluding outliers]
    - 1.2.5 Predicted Response Variable Linear Model 2 [filtered and excluding outliers]
    - 1.2.6 Confidence Intervals Simple Linear Model 2 [filtered and excluding outliers]
    - 1.2.69 Assumptions of Simple Linear Model 2 [filtered and excluding outliers] [Below]
    - 2.2.7 Plot Simple Linear Model 2 [filtered and excluding outliers]
    - 2.2.8 Non-constant Error Variance Test [filtered and excluding outliers]
    - 2.2.9 Cook's Model Linear Model 2 [filtered and excluding outliers]
    - 2.2.10 Durbin Watson Test Linear Model 2 [filtered and excluding outliers]
    - 2.2.11 Shapiro Wilks Test Linear Model 2 [filtered and excluding outliers]
    - 2.2.12 Report Results Simple Linear Model 2 [filtered and excluding outliers]
    - 2.3 [Simple Linear Model 3] [Log transformed with filtered and excluded outliers]
    - 2.3.1 Log transforming tdat to make simple linear model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.2 Making Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.3 Predicted Response Variable Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.4 Confidence Intervals Simple Linear Model 3 [Log transformed with filtered and excluded outliers]
    - Assumptions of Linear model 3: [Log transformed with filtered and excluded outliers]
    - 2.3.5 Plot linear model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.6 Non-constant Error Variance Test Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.7 Cook's Model: Identifying Points of Leverage Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.8 Durbin Watson Test Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 2.3.9 Shapiro Wilks Test Linear Model 3 [Log transformed with filtered and excluded outliers]
    - 3.0 [Simple Linear Model 4] [Logged comparison with outliers for comparison]
    - 3.1.1 log transforming original tdat data frame [Logged comparison with outliers for comparison]
    - 3.1.2 Making simple linear model 4 [Logged comparison with outliers for comparison]
    - 3.3.4 Predicted Response Variable Linear Model 4 [Logged comparison with outliers for comparison]
    - 3.3.4 Confidence Intervals Simple Linear Model 4 [Logged comparison with outliers for comparison]
    - Assumptions of Linear Model 4 [Logged comparison with outliers for comparison]
    - 3.3.5 Plot Linear Model 4 [Logged comparison with outliers for comparison]

- 3.3.6 Non-constant Error Variance Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.7 Cook's Model: Identifying Points of Leverage [Logged comparison with outliers for comparison]
- 3.3.9 Durbin Watson Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.10 Shapiro Wilks Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.11 Report Results Simple Linear Model 4 [Logged comparison with outliers for comparison]

2. Question 2: Null and Alternative Hypothesis

- 1.1.4 Simple Linear Correlation
- 1.1.49 [Simple Linear Model 1] [raw data with outliers] [Below]
- 1.1.5 Making Linear Model 1 [raw data with outliers]
- 1.1.6 Predicted Response Variable in Liner Model 1 [raw data with outliers]
- 1.1.7 Confidence Intervals of Liner Model 1 [raw data with outliers]
- 1.79 Assumptions of Linear Model 1 [raw data with outliers] [Below]
- 1.1.8 Plot Linear Model 1 [raw data with outliers]
- 1.1.9 Non-constant Error Variance Test Linear Model 1 [raw data with outliers]
- 1.2.2 Durbin Watson Test Linear Model 1 [raw data with outliers]
- 1.2.1 Cook's Model: Identifying Points of Leverage Linear Model 1 [raw data with outliers]
- 1.2.29 Shapiro Wilks Test Linear Model 1 [raw data with outliers]
- 1.2.299 Report Results Simple Linear Model 1 [raw data with outliers]
- 1.2.3 [Simple Linear Model 2] [filtered and excluding outliers] [Below]
- 1.2.4 Filtering and Excluding Extemporaneous Data [filtered and excluding outliers]
- 1.2.4 Making Simple Liner Model 2 [filtered and excluding outliers]
- 1.2.5 Predicted Response Variable Linear Model 2 [filtered and excluding outliers]
- 1.2.6 Confidence Intervals Simple Linear Model 2 [filtered and excluding outliers]
- 1.2.69 Assumptions of Simple Linear Model 2 [filtered and excluding outliers] [Below]
- 2.2.7 Plot Simple Linear Model 2 [filtered and excluding outliers]
- 2.2.8 Non-constant Error Variance Test [filtered and excluding outliers]
- 2.2.9 Cook's Model Linear Model 2 [filtered and excluding outliers]
- 2.2.10 Durbin Watson Test Linear Model 2 [filtered and excluding outliers]
- 2.2.11 Shapiro Wilks Test Linear Model 2 [filtered and excluding outliers]
- 2.11.12 Report Results Simple Linear Model 2 [filtered and excluding outliers]
- 2.3 [Simple Linear Model 3] [Log transformed with filtered and excluded outliers]
- 2.3.1 Log transforming tdat to make simple linear model 3 [Log transformed with filtered and excluded outliers]
- 2.3.2 Making Linear Model 3 [Log transformed with filtered and excluded outliers]
- 2.3.3 Predicted Response Variable Linear Model 3 [Log transformed with filtered and excluded outliers]
- 2.3.4 Confidence Intervals Simple Linear Model 3 [Log transformed with filtered and excluded outliers]
- Assumptions of Linear model 3: [Log transformed with filtered and excluded outliers]
- 2.3.5 Plot linear model 3 [Log transformed with filtered and excluded outliers]
- 2.3.6 Non-constant Error Variance Test Linear Model 3 [Log transformed with filtered and excluded outliers]
- 2.3.7 Cook's Model: Identifying Points of Leverage Linear Model 3 [Log transformed with filtered and excluded outliers]
- 2.3.8 Durbin Watson Test Linear Model 3 [Log transformed with filtered and excluded outliers]
- 2.3.9 Shapiro Wilks Test Linear Model 3 [Log transformed with filtered and excluded outliers]
- 3.0 [Simple Linear Model 4] [Logged comparison with outliers for comparison]
- 3.1.1 log transforming original tdat data frame [Logged comparison with outliers for comparison]
- 3.1.2 Making simple linear model 4 [Logged comparison with outliers for comparison]
- 3.3.4 Predicted Response Variable Linear Model 4 [Logged comparison with outliers for comparison]

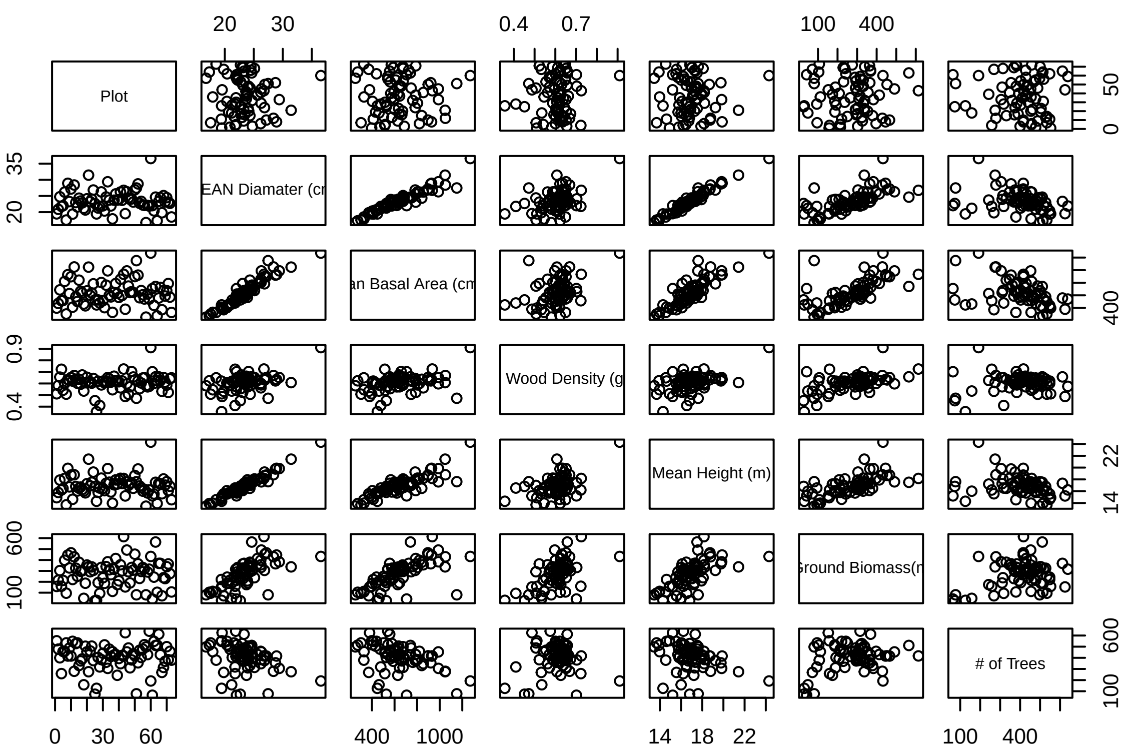
- 3.3.4 Confidence Intervals Simple Linear Model 4 [Logged comparison with outliers for comparison]
- Assumptions of Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.5 Plot Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.6 Non-constant Error Variance Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.7 Cook's Model: Identifying Points of Leverage [Logged comparison with outliers for comparison]
- 3.3.9 Durbin Watson Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.10 Shapiro Wilks Test Linear Model 4 [Logged comparison with outliers for comparison]
- 3.3.11 Report Results Simple Linear Model 4 [Logged comparison with outliers for comparison]

Question 1: Null and Alternative Hypothesis

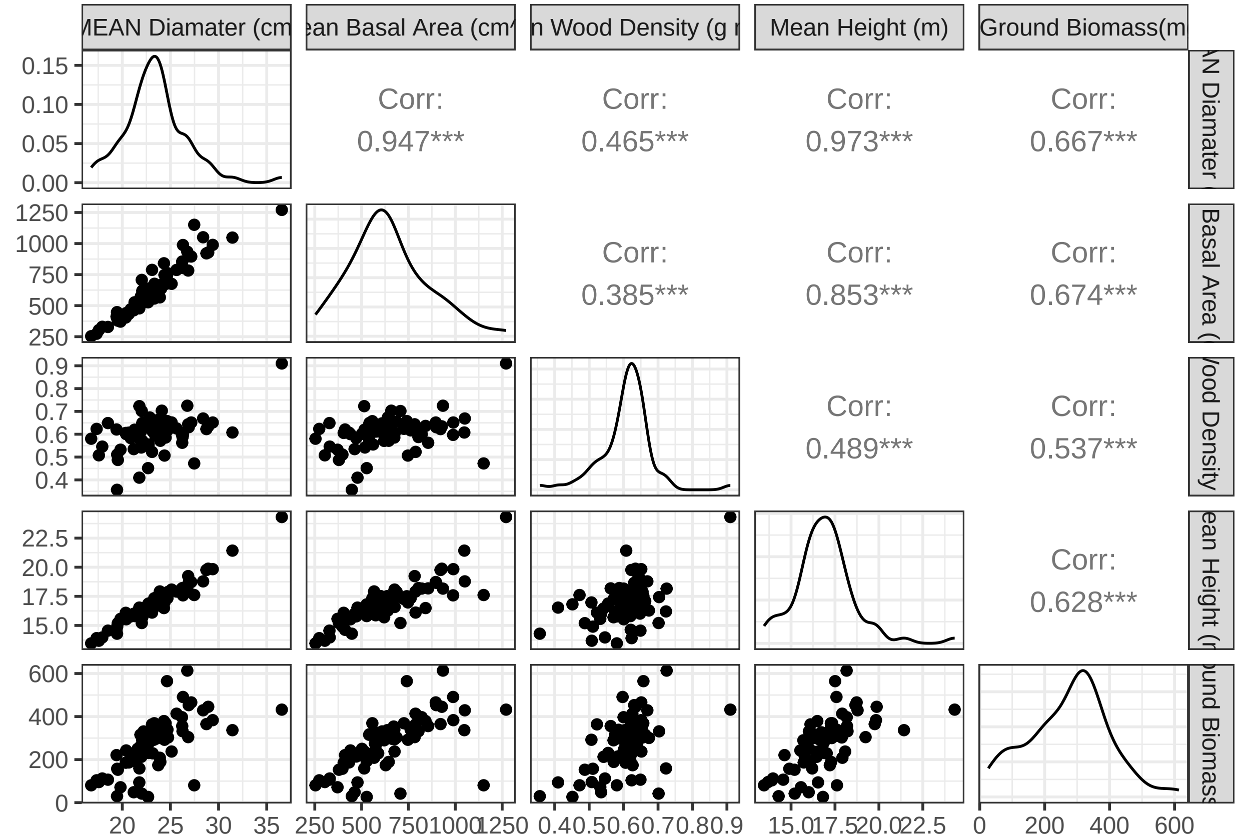
- H_0: Mean tree diameter does not change with mean tree height
- H_a: Mean tree diameter does change with mean tree height

plot scatter

1.1.2 Pairing Matrix tdat



1.1.3 Pairing Matrix tdat



1.1.4 Simple Linear Correlation

var1	var2	cor	statistic	p	conf.low	conf.high	method
MEAN Diamater (cm)	Mean Height (m)	0.97	35.86288	0	0.957981	0.9833245	Pearson

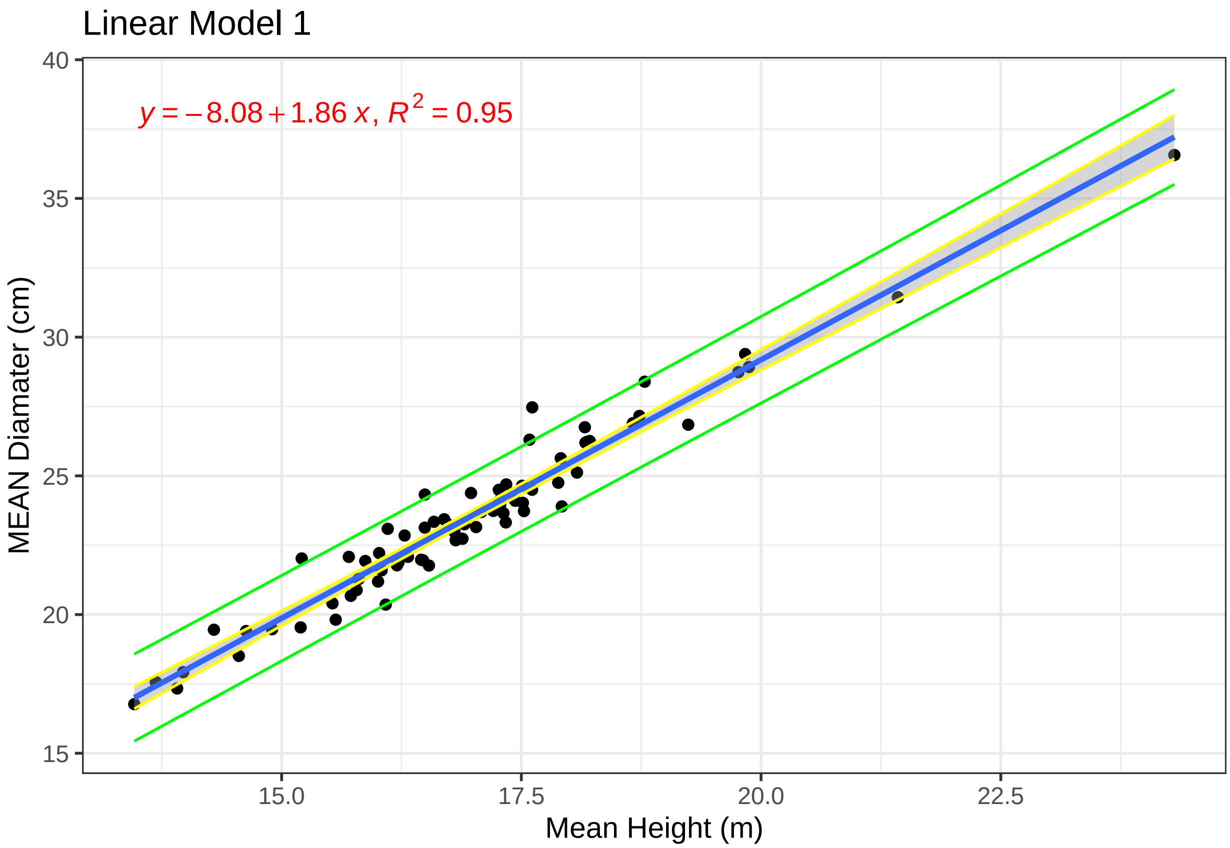
1.1.49 [Simple Linear Model 1] [raw data with outliers] [Below]

1.1.5 Making Linear Model 1 [raw data with outliers]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-8.078955	0.8840232	-9.138849	0
`Mean Height (m)`	1.863152	0.0519521	35.862875	0

1.1.58 Calculating prediction and confidence intervals

1.1.59 Plotting Prediction and Confidence Intervals linear model 1



1.1.6 Predicted Response Variable in Liner Model 1 [raw data with outliers]

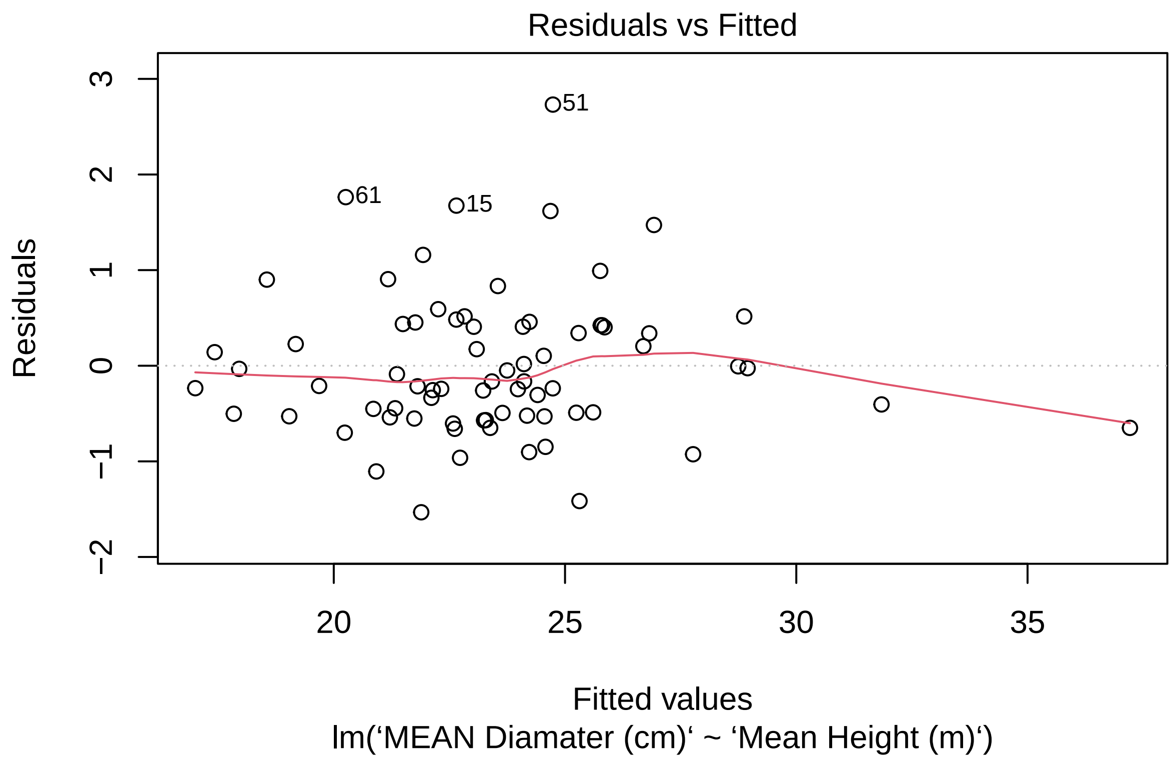
$$\hat{y}_i = -8.07896 + 1.86315\text{Ö}MeanHeight(m)$$

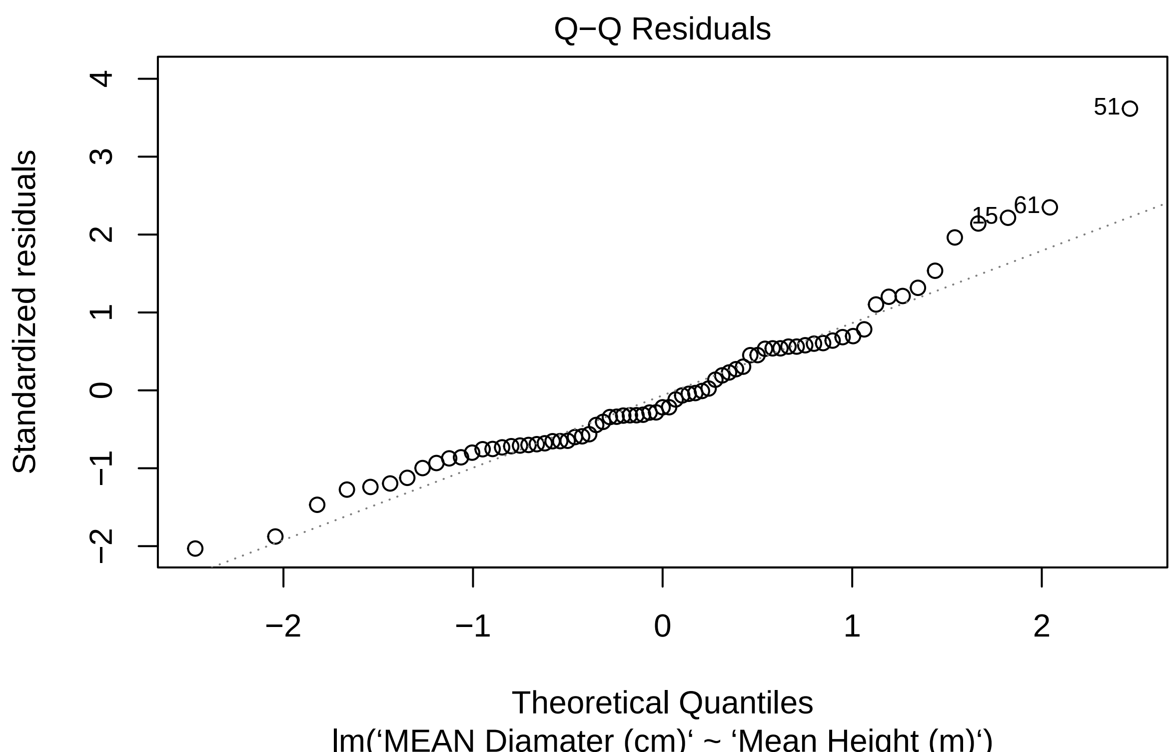
1.1.7 Confidence Intervals of Liner Model 1 [raw data with outliers]

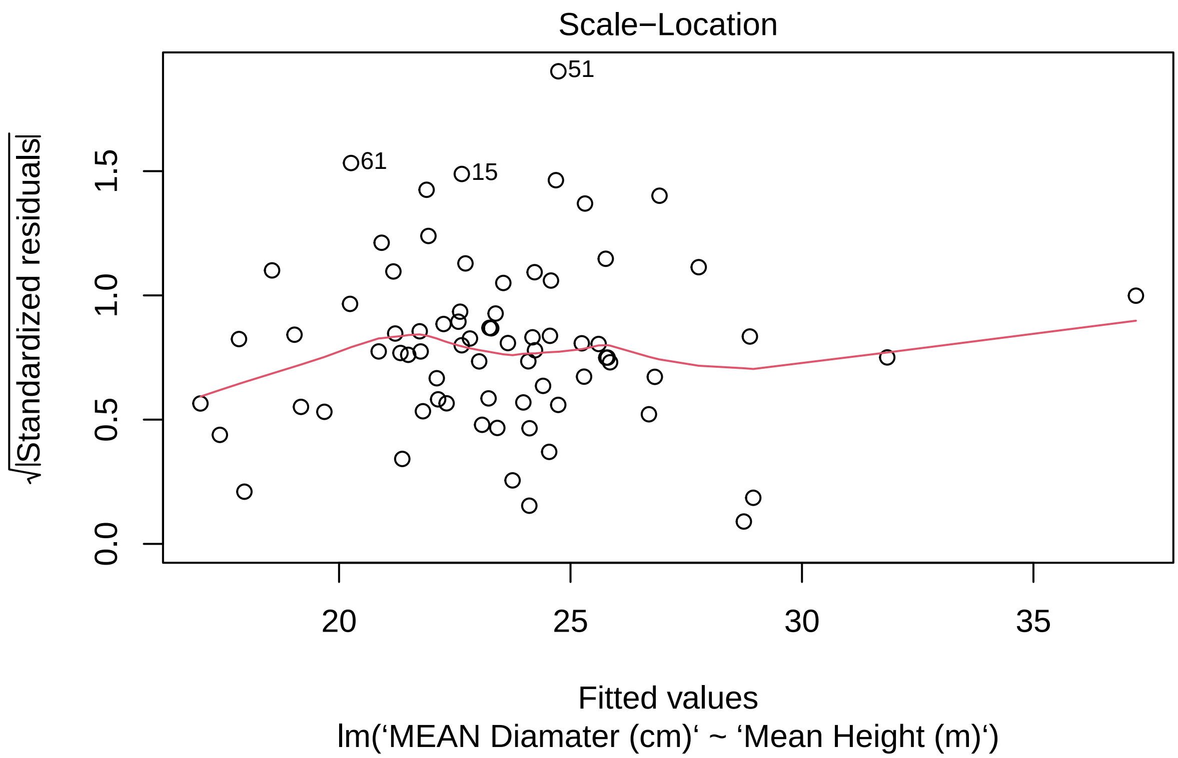
	2.5 %	97.5 %
MEAN Diamater (cm)	NA	NA

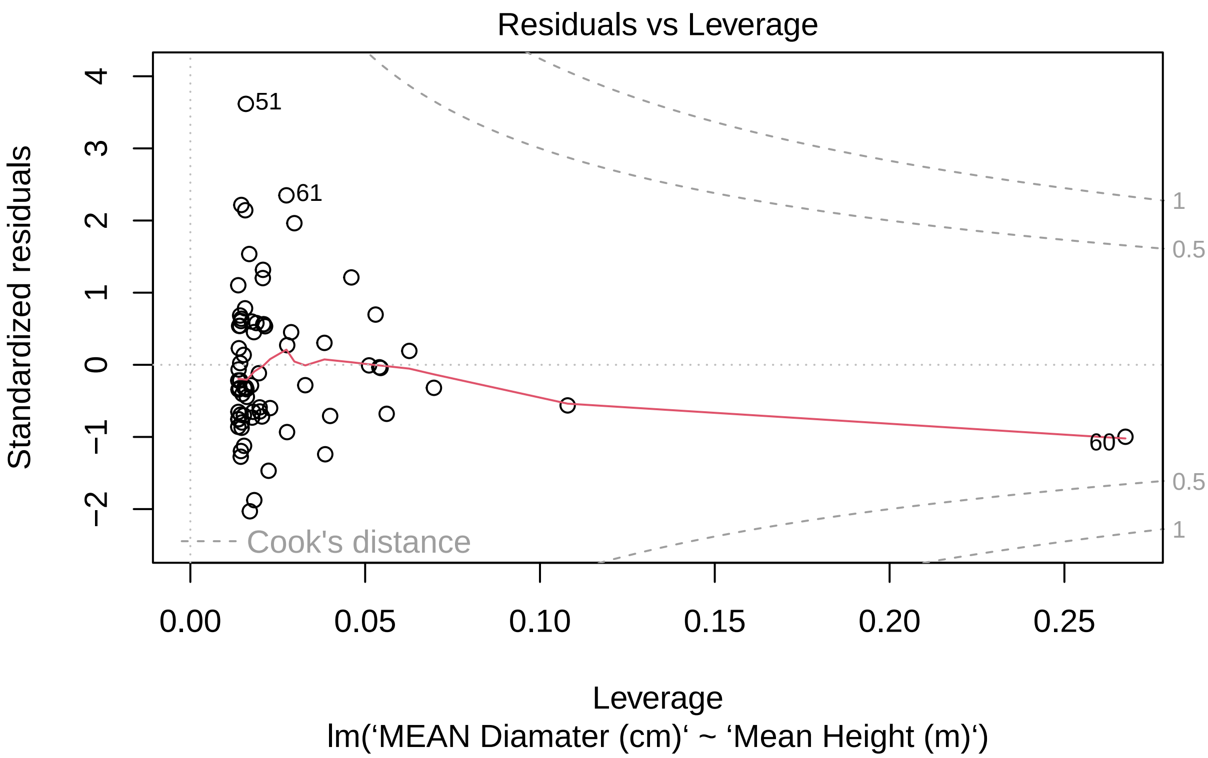
1.79 Assumptions of Linear Model 1 [raw data with outliers] [Below]

1.1.8 Plot Linear Model 1 [raw data with outliers]






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
51
61
15
20
25
30
35
Fitted values
lm(‘MEAN Diamater (cm)‘ ~ ‘Mean Height (m)‘)



1.1.9 Non-constant Error Variance Test Linear Model 1 [raw data with outliers] Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.07979587, Df = 1, p = 0.77757

`Breusch-Pagan test`

data: lm2 BP = 0.079796, df = 1, p-value = 0.7776

`studentized Breusch-Pagan test`

data: lm2 BP = 0.044867, df = 1, p-value = 0.8322

1.2.1 Cook's Model: Identifying Points of Leverage Linear Model 1 [raw data with outliers]

MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
27.46883	17.61353	24.73772	2.7311099	0.0158783	0.6923821	0.1055316	3.6168342
36.56458	24.31013	37.21450	- 0.6499206	0.2674576	0.7612039	0.1816803	- 0.9976031
22.02188	15.20942	20.25850	1.7633811	0.0274811	0.7362010	0.0779702	2.3491513

MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
19.47160	14.90057	19.68306	-0.2114646	0.0328750	0.7661646	0.0013564	-0.2824946
20.88178	15.78266	21.32653	-0.4447519	0.0198254	0.7647129	0.0035225	-0.5901738
24.75175	17.88486	25.24325	-0.4914990	0.0179504	0.7643001	0.0038802	-0.6515831
21.77481	16.20334	22.11032	-0.3355125	0.0161550	0.7655287	0.0016213	-0.4443848
21.97505	16.45494	22.57909	-0.6040414	0.0147477	0.7631372	0.0047837	-0.7994786
26.19477	18.16854	25.77179	0.4229821	0.0208503	0.7648911	0.0033578	0.5615796
17.56665	13.68853	17.42485	0.1417996	0.0626294	0.7663957	0.0012368	0.1924124
28.92236	19.87320	28.94783	-0.0254677	0.0540653	0.7665892	0.0000338	-0.0344012
23.26352	16.72912	23.08993	0.1735897	0.0138857	0.7663108	0.0003713	0.2296541
27.16010	18.73177	26.82117	0.3389293	0.0288299	0.7654927	0.0030302	0.4518303
19.40383	14.62893	19.17696	0.2268719	0.0383535	0.7660967	0.0018422	0.3039392
28.39427	18.78610	26.92240	1.4718743	0.0297560	0.7455004	0.0590951	1.9631074
22.97026	16.80397	23.22939	-0.2591272	0.0137720	0.7659609	0.0008205	-0.3427979
21.59683	16.04309	21.81175	-0.2149225	0.0173588	0.7661575	0.0007166	-0.2848382
24.32598	16.49388	22.65164	1.6743375	0.0145826	0.7396131	0.0363311	2.2158825
23.43494	16.69550	23.02729	0.4076488	0.0139537	0.7650237	0.0020581	0.5393261
22.84949	16.28270	22.25818	0.5913078	0.0156474	0.7632788	0.0048727	0.7829826
21.18988	16.00568	21.74205	-0.5521720	0.0176742	0.7636981	0.0048192	-0.7319146
20.67302	15.72165	21.21286	-0.5398410	0.0204947	0.7638183	0.0053723	-0.7165992
24.12819	17.27680	24.11034	0.0178489	0.0142605	0.7665926	0.0000040	0.0236180
31.43781	21.42687	31.84255	-0.4047404	0.1079206	0.7648827	0.0191711	-0.5629743
23.25433	16.90569	23.41891	-0.1645770	0.0137013	0.7663397	0.0003292	-0.2177101
24.02615	17.51584	24.55571	-0.5295589	0.0153002	0.7639374	0.0038187	-0.7010939
26.84346	19.24051	27.76903	-0.9255704	0.0385780	0.7582476	0.0308552	-1.2401273
22.67808	16.81473	23.24943	-0.5713547	0.0137600	0.7635052	0.0039853	-0.7558373
19.45317	14.29385	18.55265	0.9005167	0.0460585	0.7586335	0.0354199	1.2112806
21.27790	15.80383	21.36597	-0.0880748	0.0196013	0.7665219	0.0001365	-0.1168596
21.76793	16.53633	22.73073	-0.9628033	0.0144187	0.7577812	0.0118745	-1.2741049
26.25678	18.21303	25.85468	0.4021005	0.0213731	0.7650546	0.0031139	0.5339983
21.95601	16.47457	22.61567	-0.6596551	0.0146627	0.7624696	0.0056712	-0.8730483
20.35872	16.08570	21.89114	-1.5324214	0.0170153	0.7440032	0.0356863	-2.0305733
24.69139	17.34250	24.23275	0.4586398	0.0144932	0.7646042	0.0027089	0.6069540

MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
29.39016	19.83373	28.87429	0.5158709	0.0529900	0.7639727	0.0135695	0.6964301
23.72949	17.52765	24.57771	-0.8482227	0.0153654	0.7597566	0.0098403	-1.1230166
21.93059	15.87263	21.49416	0.4364303	0.0189018	0.7647845	0.0032278	0.5788587
17.92286	13.97336	17.95553	-0.0326719	0.0544068	0.7665851	0.0000561	-0.0441403
23.69882	17.08241	23.74816	-0.0493431	0.0138076	0.7665726	0.0000298	-0.0652769
25.63419	17.91138	25.29266	0.3415302	0.0181898	0.7654878	0.0018994	0.4528237
23.73620	17.20744	23.98111	-0.2449129	0.0140585	0.7660285	0.0007486	-0.3240410
23.89797	17.92227	25.31295	-1.4149795	0.0182899	0.7473510	0.0327902	-1.8761708
26.22478	18.18454	25.80160	0.4231817	0.0210362	0.7648892	0.0033922	0.5618979
23.65654	17.31352	24.17876	-0.5222161	0.0143856	0.7640131	0.0034851	-0.6910517
26.75227	18.16255	25.76063	0.9916424	0.0207813	0.7571808	0.0183917	1.3165249
19.53745	15.19823	20.23765	-0.7002003	0.0276610	0.7618829	0.0123787	-0.9328832
26.30278	17.58526	24.68505	1.6177312	0.0157019	0.7414081	0.0366023	2.1421844
24.38176	16.97513	23.54828	0.8334758	0.0137083	0.7600045	0.0084481	1.1025649
23.95129	17.27926	24.11492	-0.1636345	0.0142685	0.7663424	0.0003393	-0.2165257
24.10158	17.43608	24.40710	-0.3055239	0.0148941	0.7657122	0.0012363	-0.4044060
23.13280	16.49305	22.65010	0.4827039	0.0145860	0.7643892	0.0030203	0.6388299
26.89809	18.66346	26.69390	0.2041912	0.0277047	0.7661959	0.0010545	0.2720519
27.46883	17.61353	24.73772	2.7311099	0.0158783	0.6923821	0.1055316	3.6168342
28.73804	19.76382	28.74404	-0.0059962	0.0511212	0.7665952	0.0000018	-0.0080870
22.73226	16.88591	23.38205	-0.6497938	0.0137075	0.7625963	0.0051345	-0.8595805
23.34391	16.58835	22.82765	0.5162555	0.0142408	0.7640722	0.0033707	0.6831139
22.21547	16.01698	21.76311	0.4523644	0.0175776	0.7646523	0.0032162	0.5995883
22.08166	16.31768	22.32336	-0.2416952	0.0154423	0.7660425	0.0008031	-0.3200084
16.77089	13.46335	17.00531	-0.2344160	0.0696650	0.7660450	0.0038169	-0.3192871
22.07949	15.70088	21.17416	0.9053266	0.0207304	0.7587570	0.0152901	1.2018990
21.88512	16.21957	22.14056	-0.2554414	0.0160464	0.7659774	0.0009333	-0.3383124
36.56458	24.31013	37.21450	-0.6499206	0.2674576	0.7612039	0.1816803	-0.9976031
22.02188	15.20942	20.25850	1.7633811	0.0274811	0.7362010	0.0779702	2.3491513
20.40529	15.53043	20.85659	-0.4512992	0.0228168	0.7646511	0.0041998	-0.5997779
24.64344	17.50733	24.53985	0.1035865	0.0152540	0.7664941	0.0001457	0.1371371
17.33599	13.91049	17.83840	-0.5024055	0.0561568	0.7640997	0.0137312	-0.6793887
23.32092	17.33791	24.22420	-0.9032783	0.0144757	0.7588423	0.0104941	-1.1953683

MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
24.49696	17.26535	24.08901	0.4079520	0.0142241	0.7650209	0.0021022	0.5398011
19.81413	15.56425	20.91960	-1.1054710	0.0223813	0.7548583	0.0246967	-1.4688466
25.11956	18.08022	25.60723	-0.4876743	0.0198671	0.7643313	0.0042445	-0.6471445
23.15366	17.02831	23.64737	-0.4937066	0.0137441	0.7642893	0.0029722	-0.6531125
24.49861	17.61200	24.73487	-0.2362595	0.0158686	0.7660669	0.0007892	-0.3128791
23.08956	16.10686	21.93057	1.1589944	0.0168511	0.7537575	0.0202092	1.5356262
22.71812	16.83497	23.28714	-0.5690249	0.0137403	0.7635305	0.0039470	-0.7527477
18.50897	14.55408	19.03750	-0.5285312	0.0399839	0.7638795	0.0104584	-0.7086719

1.2.2 Durbin Watson Test Linear Model 1 [raw data with outliers] lag Autocorrelation D-W Statistic p-value 1 -0.06953094 2.131184 0.57 Alternative hypothesis: rho != 0

1.2.29 Shapiro Wilks Test Linear Model 1 [raw data with outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm2 W = 0.94273, p-value = 0.002396

1.2.299 Report Results Simple Linear Model 1 [raw data with outliers] The correlation coefficient (r = 0.97) from Figure 1.1.4 suggests a strong positive linear relationship between Mean Diameter and Mean Height, supported by a statistically significant p-value (p = 0). Linear Model 1, presented in Figure 1.1.5, illustrates a predictive equation for tree diameter based on mean height, with a slope of 1.86315 and an intercept of -8.07896. However, assessments of assumptions in Linear Model 1 reveal non-linear patterns and influential outliers, prompting data wrangling and outlier removal. While the chi-squared test indicates homoscedasticity (p = 0.77), the Cook's model identifies influential outliers. Further tests, including the Durbin-Watson test and Shapiro-Wilks test, reveal no significant autocorrelation but highlight the need for data normalization to achieve normal distribution.

1.2.3 [Simple Linear Model 2] [filtered and excluding outliers] [Below]

1.2.4 Filtering and Excluding Extemporaneous Data [filtered and excluding outliers]

Plot	MEAN Diamater (cm)	Mean Basal Area (cm^2)	Mean Wood Density (g m^3)	Mean Height (m)	Above Gr
1	19.47160	397.4730	0.51084	14.90057	
2	20.88178	471.2643	0.58283	15.78266	
3	24.75175	686.6133	0.65501	17.88486	
4	21.77481	513.7697	0.72304	16.20334	
5	21.97505	517.5035	0.54183	16.45494	
6	26.19477	820.6032	0.59995	18.16854	

1.2.49 Making Simple Liner Model 2 [filtered and excluding outliers]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-8.739044	0.8944387	-9.770422	0
`Mean Height (m)`	1.899090	0.0529098	35.892996	0

1.2.5 Predicted Response Variable Linear Model 2 [filtered and excluding outliers]

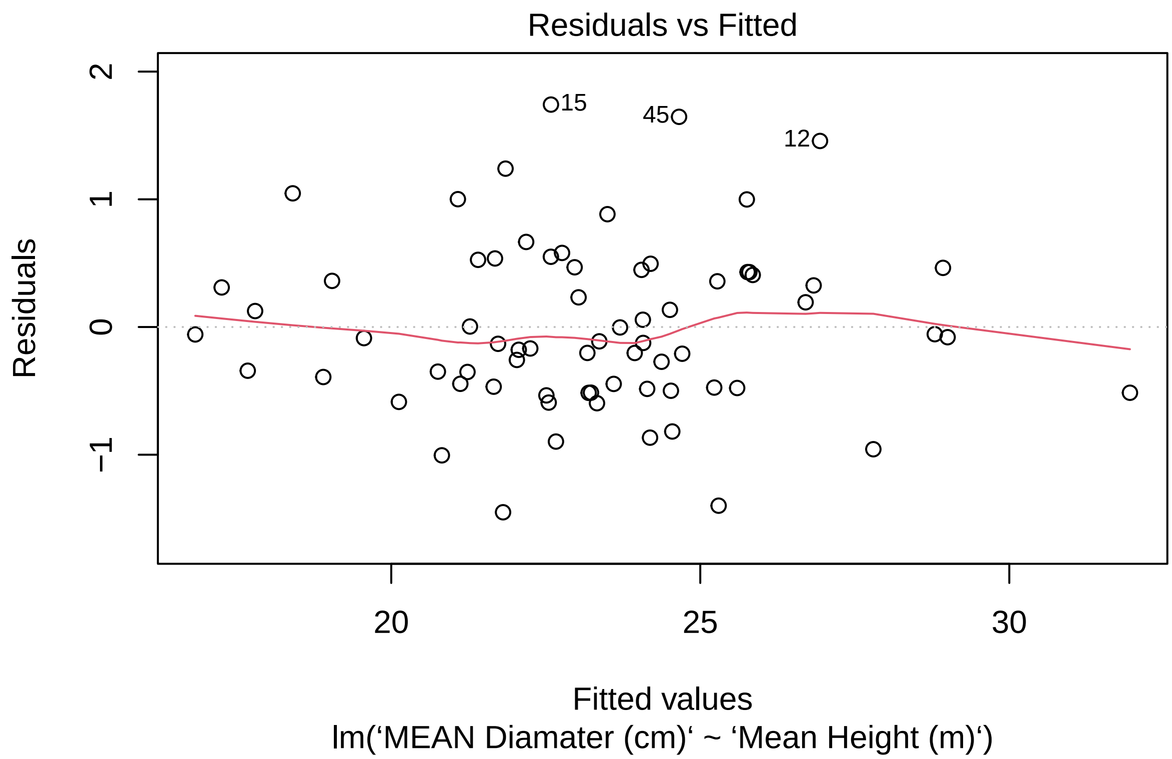
$$\hat{y_i} = -8.73904 + 1.89909\text{Ö}MeanHeight(m)$$

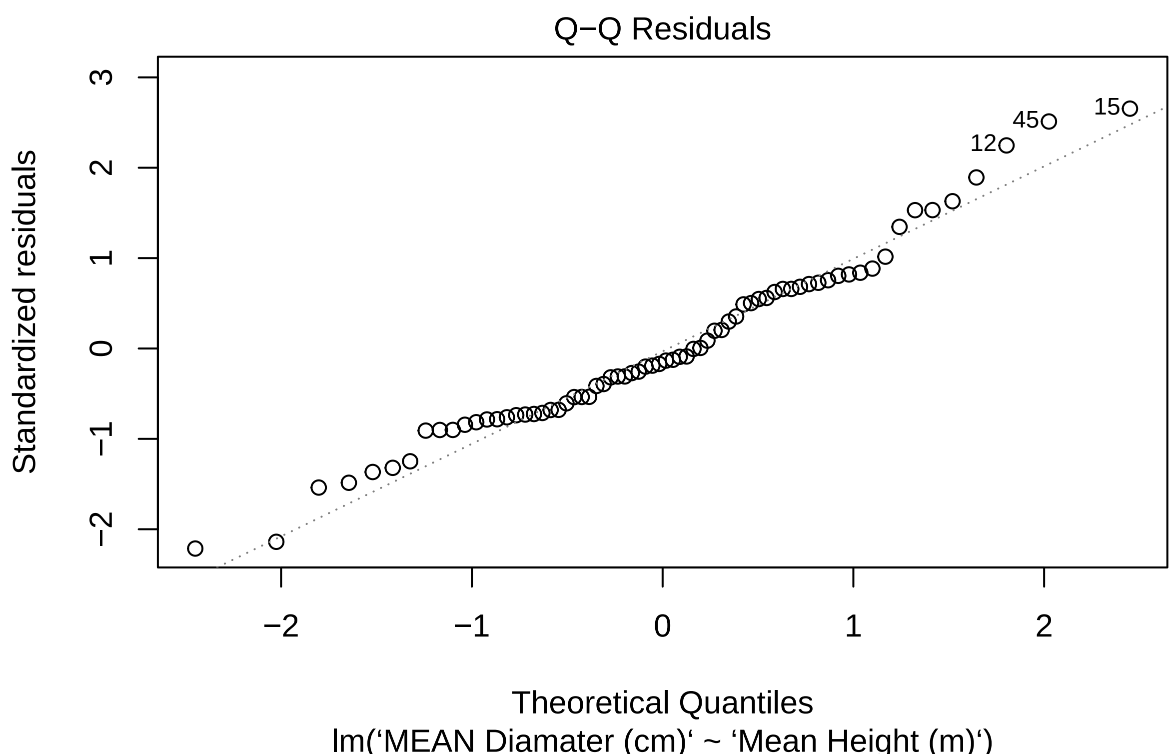
1.2.6 Confidence Intervals Simple Linear Model 2 [filtered and excluding outliers]

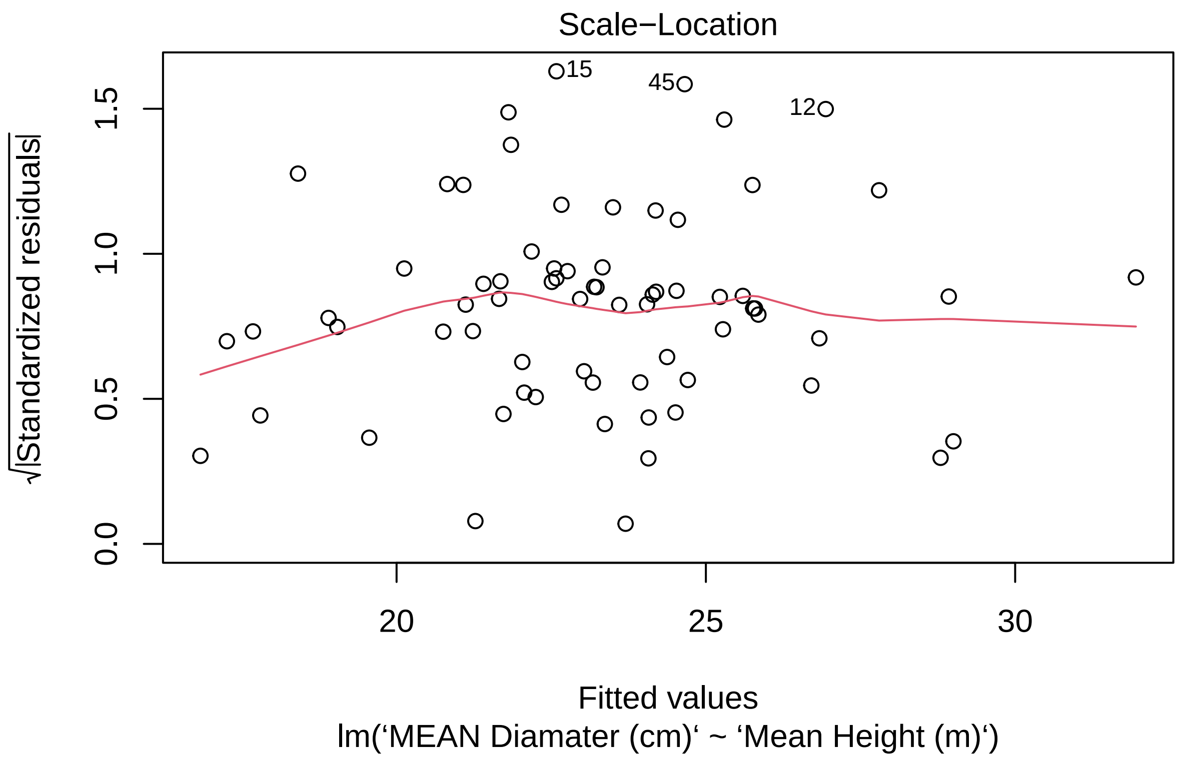
	2.5 %	97.5 %
MEAN Diamater (cm)	NA	NA

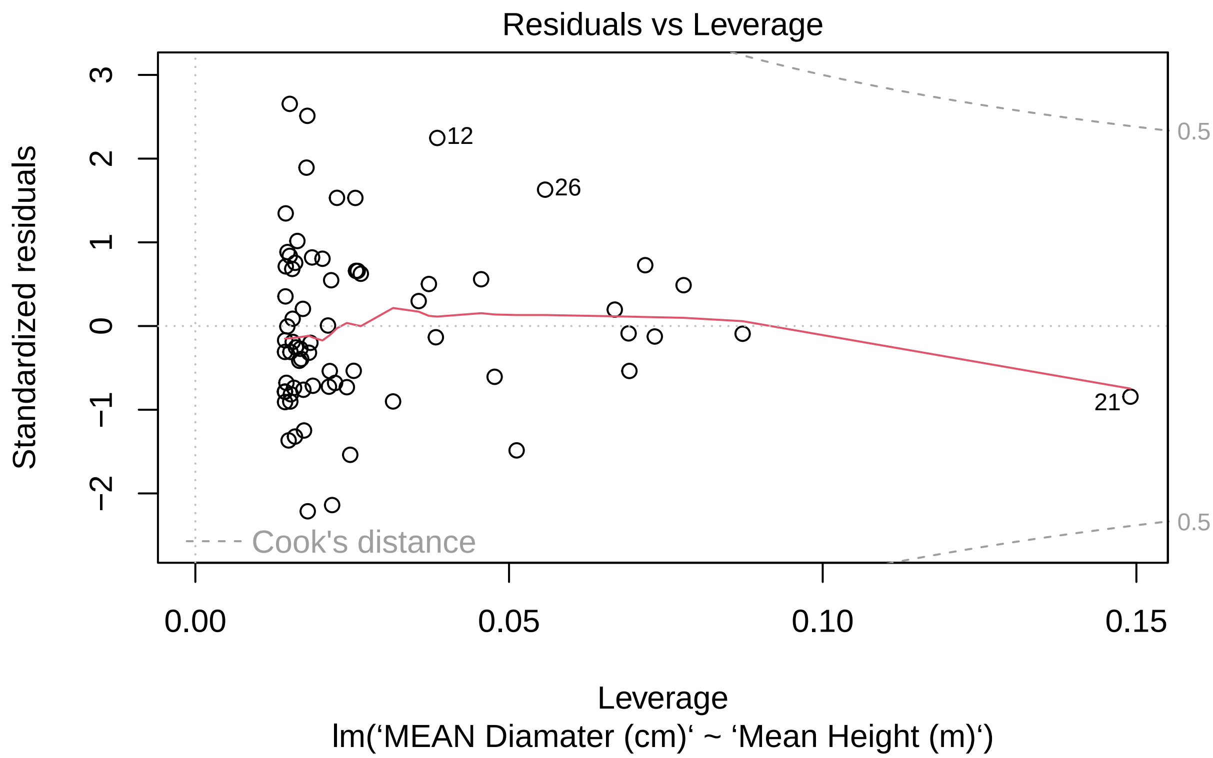
1.2.69 Assumptions of Simple Linear Model 2 [filtered and excluding outliers] [Below]

2.2.7 Plot Simple Linear Model 2 [filtered and excluding outliers]






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
45
12
20
25
30
Fitted values
lm(‘MEAN Diamater (cm)‘ ~ ‘Mean Height (m)‘)



2.2.8 Non-constant Error Variance Test [filtered and excluding outliers]

Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.2542306, Df = 1, p = 0.61411

```
Breusch-Pagan test
```

data: lm3 BP = 0.25423, df = 1, p-value = 0.6141

```
studentized Breusch-Pagan test
```

data: lm3 BP = 0.22639, df = 1, p-value = 0.6342

2.2.9 Cook's Model Linear Model 2 [filtered and excluding outliers]

.rownames	MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
12	28.39427	18.78610	26.93745	1.456815	0.0385620	0.6409604	0.1012430	2.2468731
21	31.43781	21.42687	31.95251	-0.514705	0.1490549	0.6626691	0.0623590	-0.8438054
26	19.45317	14.29385	18.40627	1.046903	0.0557543	0.6530335	0.0783719	1.6292913

.rownames	MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
1	19.47160	14.90057	19.55848	-0.0868827	0.0383392	0.6660779	0.0003579	-0.1339853
2	20.88178	15.78266	21.23365	-0.3518712	0.0214279	0.6647470	0.0031681	-0.5379253
3	24.75175	17.88486	25.22592	-0.4741686	0.0212907	0.6635874	0.0057146	-0.7248374
4	21.77481	16.20334	22.03256	-0.2577504	0.0168715	0.6654084	0.0013261	-0.3931235
5	21.97505	16.45494	22.51037	-0.5353215	0.0152294	0.6628979	0.0051461	-0.8157965
6	26.19477	18.16854	25.76465	0.4301175	0.0256055	0.6640355	0.0057053	0.6589529
7	17.56665	13.68853	17.25671	0.3099406	0.0778265	0.6649979	0.0100531	0.4880977
8	28.92236	19.87320	29.00196	-0.0795956	0.0732339	0.6660893	0.0006177	-0.1250370
9	23.26352	16.72912	23.03106	0.2324559	0.0143628	0.6655514	0.0009135	0.3540925
10	27.16010	18.73177	26.83428	0.3258229	0.0372262	0.6649295	0.0048753	0.5021740
11	19.40383	14.62893	19.04261	0.3612162	0.0455535	0.6646327	0.0074609	0.5591471
12	28.39427	18.78610	26.93745	1.4568154	0.0385620	0.6409604	0.1012430	2.2468731
13	22.97026	16.80397	23.17321	-0.2029510	0.0142935	0.6656976	0.0006929	-0.3091377
14	21.59683	16.04309	21.72823	-0.1314012	0.0183400	0.6659688	0.0003758	-0.2005643
15	24.32598	16.49388	22.58432	1.7416579	0.0150477	0.6307227	0.0538026	2.6539328
16	23.43494	16.69550	22.96722	0.4677233	0.0144173	0.6636747	0.0037129	0.7124874
17	22.84949	16.28270	22.18327	0.6662178	0.0162660	0.6610922	0.0085310	1.0158090
18	21.18988	16.00568	21.65719	-0.4673063	0.0187301	0.6636681	0.0048574	-0.7134147
19	20.67302	15.72165	21.11779	-0.4447677	0.0222768	0.6638955	0.0052714	-0.6802365
20	24.12819	17.27680	24.07116	0.0570322	0.0155137	0.6661288	0.0000595	0.0869261
21	31.43781	21.42687	31.95251	-0.5147050	0.1490549	0.6626691	0.0623590	-0.8438054
22	23.25433	16.90569	23.36639	-0.1120564	0.0143143	0.6660231	0.0002115	-0.1706877
23	24.02615	17.51584	24.52512	-0.4989663	0.0172200	0.6633219	0.0050758	-0.7611631
24	26.84346	19.24051	27.80042	-0.9569602	0.0512142	0.6552640	0.0595774	-1.4857457
25	22.67808	16.81473	23.19365	-0.5155652	0.0142894	0.6631381	0.0044702	-0.7853144
26	19.45317	14.29385	18.40627	1.0469033	0.0557543	0.6530335	0.0783719	1.6292913
27	21.27790	15.80383	21.27385	0.0040451	0.0211445	0.6661657	0.0000004	0.0061831
28	21.76793	16.53633	22.66494	-0.8970085	0.0148717	0.6569523	0.0140996	-1.3667366
29	26.25678	18.21303	25.84914	0.4076370	0.0263757	0.6642512	0.0052870	0.6247590
30	21.95601	16.47457	22.54765	-0.5916407	0.0151353	0.6621723	0.0062459	-0.9015803
31	20.35872	16.08570	21.80915	-1.4504315	0.0179174	0.6417202	0.0446903	-2.2133897
32	24.69139	17.34250	24.19593	0.4954620	0.0159097	0.6633655	0.0046116	0.7553140
33	29.39016	19.83373	28.92700	0.4631615	0.0717103	0.6635720	0.0204136	0.7269849

.rownames	MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
34	23.72949	17.52765	24.54754	-0.8180546	0.0173232	0.6584927	0.0137281	-1.2479915
35	21.93059	15.87263	21.40451	0.5260777	0.0202629	0.6629938	0.0066807	0.8037663
36	17.92286	13.97336	17.79763	0.1252327	0.0668561	0.6659776	0.0013770	0.1960551
37	23.69882	17.08241	23.70199	-0.0031737	0.0146655	0.6661657	0.0000002	-0.0048351
38	25.63419	17.91138	25.27628	0.3579075	0.0216504	0.6646975	0.0033133	0.5472156
39	23.73620	17.20744	23.93944	-0.2032369	0.0151555	0.6656959	0.0007380	-0.3097087
40	23.89797	17.92227	25.29696	-1.3989935	0.0218007	0.6433617	0.0509902	-2.1391278
41	26.22478	18.18454	25.79504	0.4297421	0.0258795	0.6640387	0.0057595	0.6584703
42	23.65654	17.31352	24.14089	-0.4843524	0.0157282	0.6634904	0.0043552	-0.7383098
43	26.75227	18.16255	25.75328	0.9989931	0.0255037	0.6545928	0.0306484	1.5304074
44	19.53745	15.19823	20.12377	-0.5863159	0.0315187	0.6621776	0.0132096	-0.9009916
45	26.30278	17.58526	24.65695	1.6458288	0.0178526	0.6345179	0.0573267	2.5114875
46	24.38176	16.97513	23.49826	0.8835007	0.0144046	0.6572338	0.0132360	1.3458364
47	23.95129	17.27926	24.07583	-0.1245396	0.0155275	0.6659893	0.0002842	-0.1898192
48	24.10158	17.43608	24.37364	-0.2720649	0.0165693	0.6653222	0.0014501	-0.4148922
49	23.13280	16.49305	22.58275	0.5500541	0.0150513	0.6627157	0.0053678	0.8381723
50	26.89809	18.66346	26.70455	0.1935398	0.0356004	0.6657306	0.0016395	0.2980413
52	28.73804	19.76382	28.79423	-0.0561931	0.0690607	0.6661279	0.0002877	-0.0880759
53	22.73226	16.88591	23.32882	-0.5965624	0.0142999	0.6621089	0.0059896	-0.9086951
54	23.34391	16.58835	22.76373	0.5801808	0.0146875	0.6623277	0.0058233	0.8839161
55	22.21547	16.01698	21.67865	0.5368240	0.0186104	0.6628682	0.0063676	0.8194942
56	22.08166	16.31768	22.24970	-0.1680424	0.0160248	0.6658443	0.0005344	-0.2561896
57	16.77089	13.46335	16.82907	-0.0581823	0.0872347	0.6661243	0.0004053	-0.0920971
58	22.07949	15.70088	21.07834	1.0011464	0.0225767	0.6545776	0.0270851	1.5314081
59	21.88512	16.21957	22.06338	-0.1782627	0.0167411	0.6658037	0.0006292	-0.2718699
62	20.40529	15.53043	20.75464	-0.3493537	0.0252465	0.6647617	0.0037084	-0.5351217
63	24.64344	17.50733	24.50896	0.1344849	0.0171467	0.6659597	0.0003671	0.2051464
64	17.33599	13.91049	17.67823	-0.3422415	0.0691882	0.6647547	0.0106958	-0.5364588
65	23.32092	17.33791	24.18721	-0.8662912	0.0158803	0.6575677	0.0140711	-1.3206101
66	24.49696	17.26535	24.04941	0.4475468	0.0154503	0.6638829	0.0036507	0.6821098
67	19.81413	15.56425	20.81887	-1.0047409	0.0246872	0.6544680	0.0299592	-1.5385683
68	25.11956	18.08022	25.59692	-0.4773648	0.0241517	0.6635447	0.0066088	-0.7307923
69	23.15366	17.02831	23.59925	-0.4455929	0.0145155	0.6639050	0.0033935	-0.6788097

.rownames	MEAN Diamater (cm)	Mean Height (m)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
70	24.49861	17.61200	24.70773	-0.2091228	0.0181127	0.6656667	0.0009395	-0.3191577
71	23.08956	16.10686	21.84934	1.2402238	0.0177162	0.6483867	0.0322951	1.8924144
72	22.71812	16.83497	23.23208	-0.5139628	0.0142858	0.6631570	0.0044413	-0.7828722
73	18.50897	14.55408	18.90047	-0.3914969	0.0477074	0.6643604	0.0092202	-0.6067054

2.2.10 Durbin Watson Test Linear Model 2 [filtered and excluding outliers] lag Autocorrelation D-W Statistic p-value 1 -0.0349286 2.064448 0.79 Alternative hypothesis: rho != 0

2.2.11 Shapiro Wilks Test Linear Model 2 [filtered and excluding outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm3 W = 0.97432, p-value = 0.1597

2.2.12 Report Results Simple Linear Model 2 [filtered and excluding outliers] Linear Model 2, constructed after removing outliers detected in Linear Model 1 (Figure 1.2.4), exhibits an intercept of -8.73904 and a slope of 1.89 for mean height. While confidence intervals remain normal, plots indicate slight deviations necessitating data logging to address non-linearity. The Shapiro-Wilks test yields a p-value of 0.15, indicating non-rejection of normality assumptions, yet data logging could further enhance this. Despite a high p-value (p = 0.6141) in the non-constant variance test and residual identification by the Cook's model, data normalization remains imperative for improved regression model assumptions.

2.3 [Simple Linear Model 3] [Log transformed with filtered and excluded outliers]

2.3.1 Log transforming tdat to make simple linear model 3 [Log transformed with filtered and excluded outliers]

2.3.2 Making Linear Model 3 [Log transformed with filtered and excluded outliers]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.7658777	0.1080780	-7.086343	0
log_height	1.3844570	0.0383099	36.138340	0

2.3.3 Predicted Response Variable Linear Model 3 [Log transformed with filtered and excluded outliers]

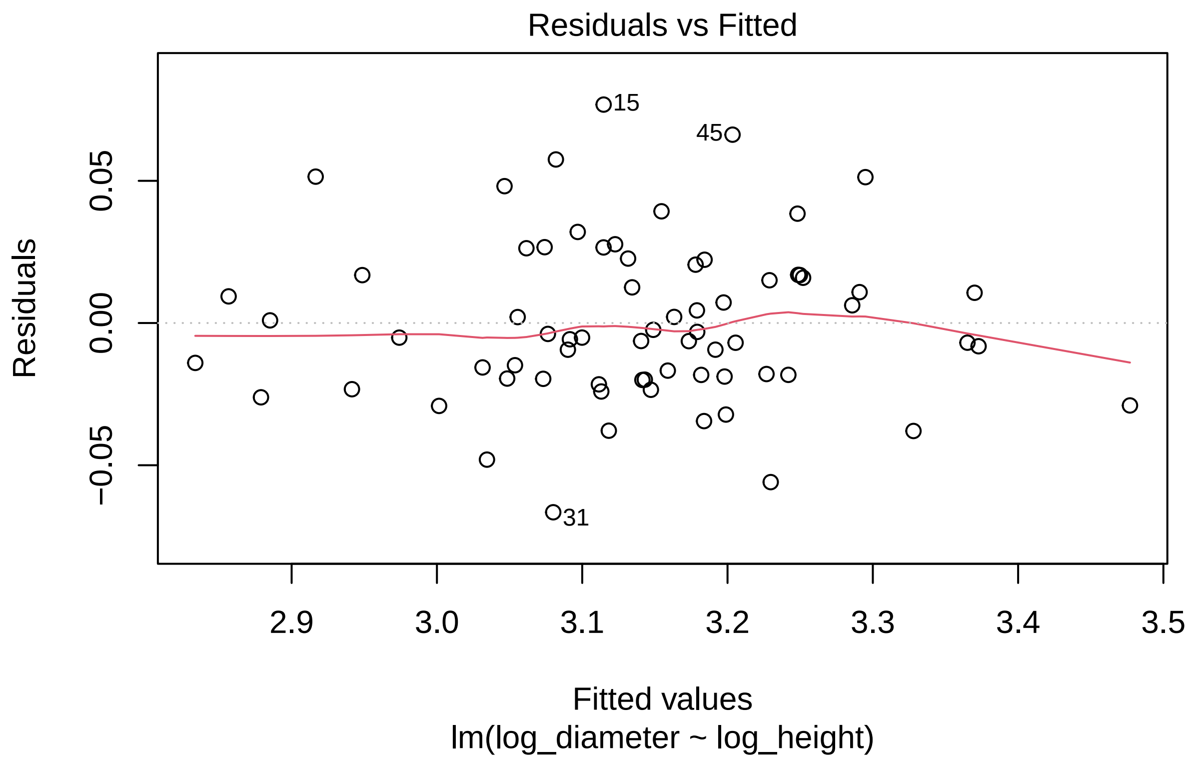
$$\hat{y_i} = -0.7658777 + 1.3844570 \text{Ö} MeanHeight(m)$$

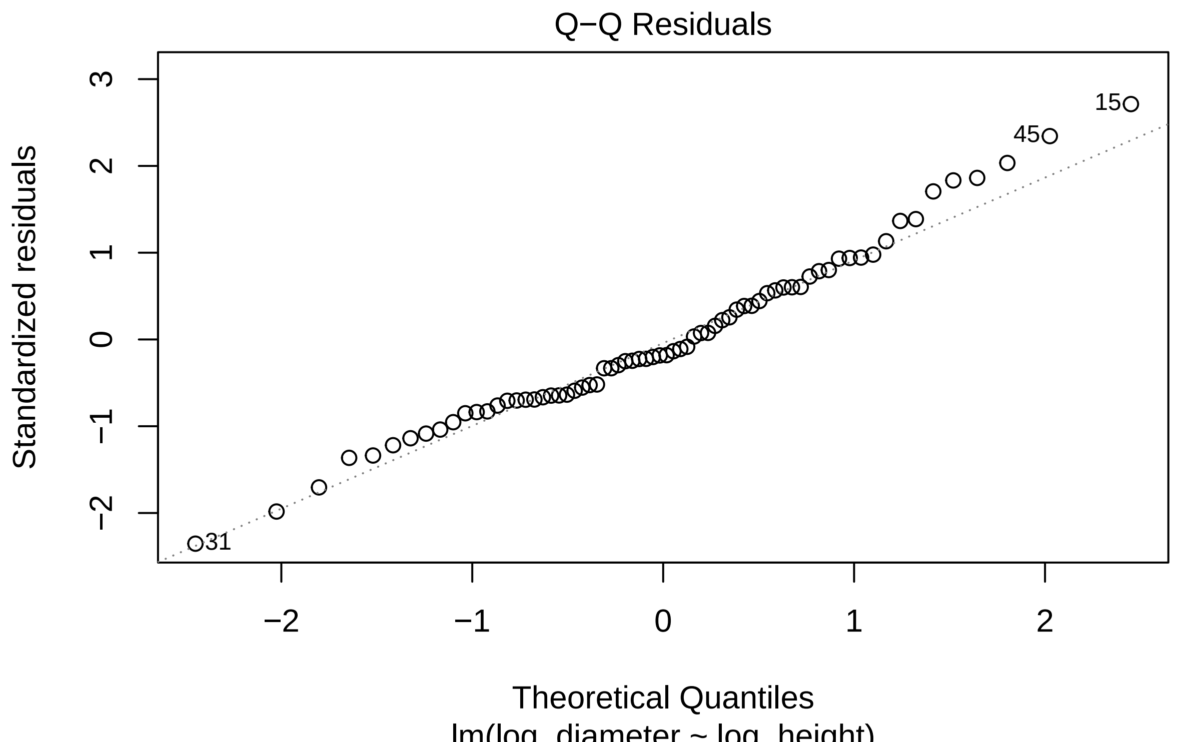
2.3.4 Confidence Intervals Simple Linear Model 3 [Log transformed with filtered and excluded outliers]

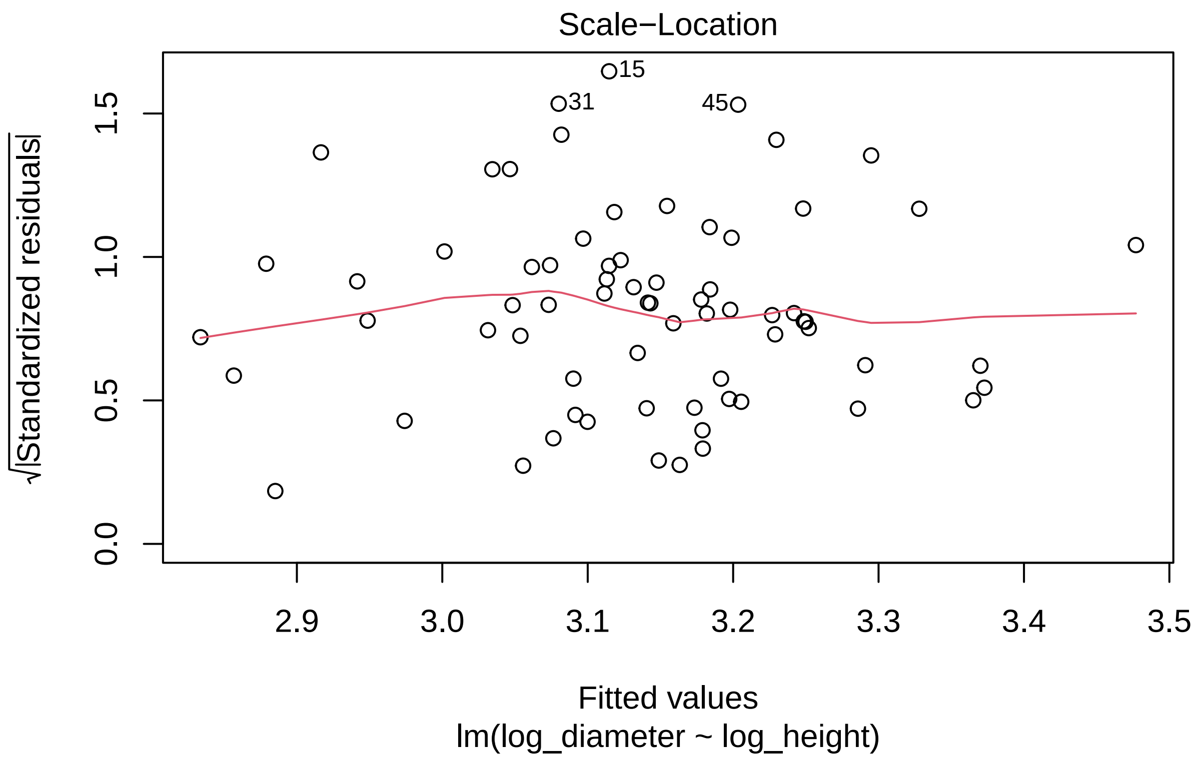
	2.5 %	97.5 %
MEAN Diamater (cm)	NA	NA

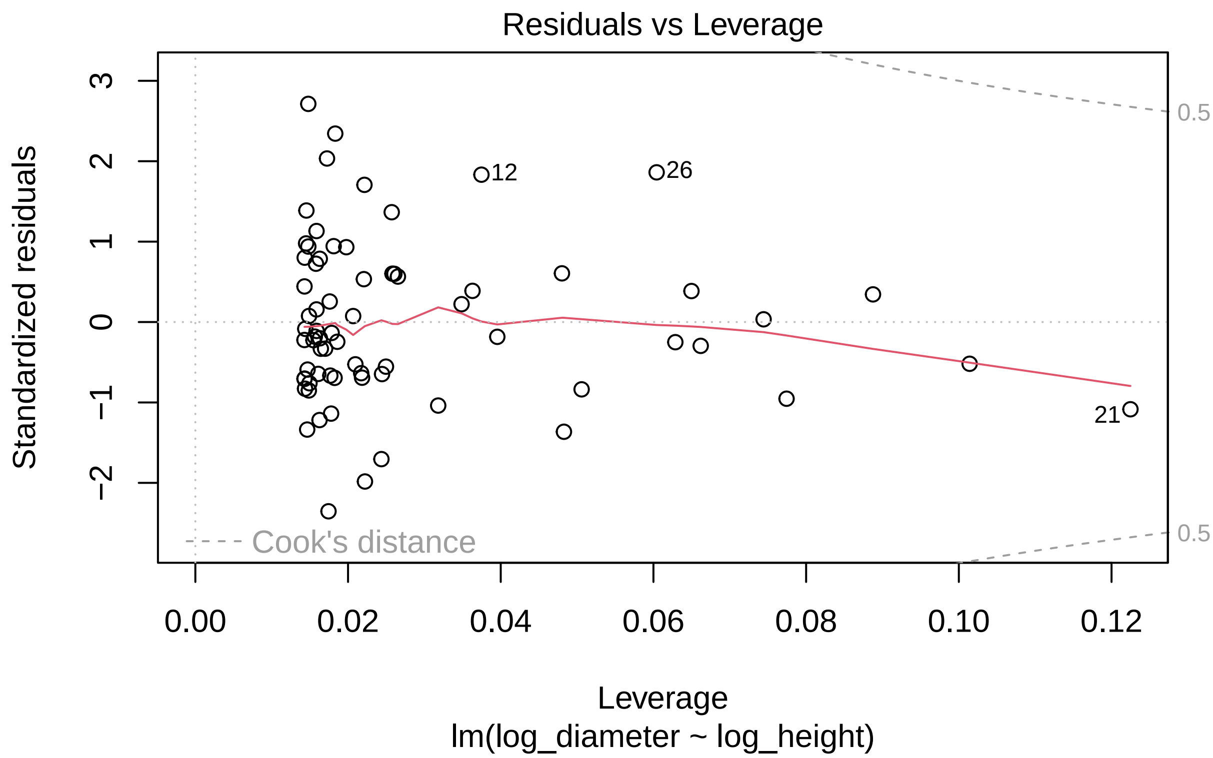
Assumptions of Linear model 3: [Log transformed with filtered and excluded outliers]

2.3.5 Plot linear model 3 [Log transformed with filtered and excluded outliers]






Scale−Location
15
31
45
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.9
3.0
3.1
3.2
3.3
3.4
3.5
Fitted values
lm(log_diameter ~ log_height)



2.3.6 Non-constant Error Variance Test Linear Model 3 [Log transformed with filtered and excluded outliers]

Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.06193583, Df = 1, p = 0.80346

```
Breusch-Pagan test
```

data: lm4 BP = 0.061936, df = 1, p-value = 0.8035

```
studentized Breusch-Pagan test
```

data: lm4 BP = 0.056989, df = 1, p-value = 0.8113

2.3.7 Cook's Model: Identifying Points of Leverage Linear Model 3 [Log transformed with filtered and excluded outliers]

.rownames	log_diameter	log_height	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
12	3.346187	2.933117	3.294897	0.0512904	0.0374720	0.0280168	0.0653907	1.832847
21	3.448011	3.064646	3.476993	-0.0289812	0.1224782	0.0284859	0.0820994	-1.084637

.rownames	log__diameter	log__height	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
26	2.968010	2.659829	2.916542	0.0514684	0.0604171	0.0279938	0.1114125	1.861528

.rownames	log__diameter	log__height	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
1	2.968957	2.701399	2.974094	-0.0051367	0.0395506	0.0287285	0.0006952	-0.1837565
2	3.038877	2.758912	3.053717	-0.0148401	0.0209614	0.0286771	0.0029598	-0.5258166
3	3.208896	2.883955	3.226833	-0.0179371	0.0217231	0.0286501	0.0044882	-0.6357979
4	3.080754	2.785217	3.090136	-0.0093822	0.0164362	0.0287123	0.0009191	-0.3316654
5	3.089908	2.800626	3.111468	-0.0215605	0.0149451	0.0286128	0.0044001	-0.7615960
6	3.265560	2.899691	3.248620	0.0169393	0.0258154	0.0286590	0.0047968	0.6016871
7	2.866002	2.616558	2.856635	0.0093675	0.0887599	0.0287106	0.0057646	0.3440378
8	3.364615	2.989372	3.372779	-0.0081644	0.0661905	0.0287171	0.0031095	-0.2962044
9	3.146886	2.817151	3.134347	0.0125399	0.0142979	0.0286941	0.0014221	0.4428102
10	3.301749	2.930221	3.290887	0.0108617	0.0363025	0.0287038	0.0028341	0.3879046
11	2.965470	2.683001	2.948622	0.0168486	0.0480168	0.0286581	0.0092432	0.6054039
12	3.346187	2.933117	3.294897	0.0512904	0.0374720	0.0280168	0.0653907	1.8328470
13	3.134200	2.821615	3.140527	-0.0063268	0.0142920	0.0287251	0.0003619	-0.2234126
14	3.072547	2.775278	3.076376	-0.0038291	0.0178525	0.0287317	0.0001668	-0.1354582
15	3.191545	2.802989	3.114741	0.0768043	0.0147921	0.0271361	0.0552472	2.7128052
16	3.154228	2.815139	3.131562	0.0226666	0.0143240	0.0285999	0.0046551	0.8004163
17	3.128929	2.790103	3.096900	0.0320286	0.0158707	0.0284636	0.0103307	1.1319023
18	3.053524	2.772944	3.073144	-0.0196198	0.0182369	0.0286336	0.0044760	-0.6942065
19	3.028830	2.755039	3.048355	-0.0195255	0.0218385	0.0286342	0.0053477	-0.6921390
20	3.183381	2.849365	3.178945	0.0044359	0.0158683	0.0287304	0.0001981	0.1567652
21	3.448011	3.064646	3.476993	-0.0289812	0.1224782	0.0284859	0.0820994	-1.0846370
22	3.146491	2.827650	3.148883	-0.0023911	0.0143985	0.0287341	0.0000521	-0.0844395
23	3.179143	2.863106	3.197969	-0.0188261	0.0176774	0.0286417	0.0039902	-0.6659311
24	3.290022	2.957018	3.327987	-0.0379643	0.0482786	0.0283396	0.0472115	-1.3643219
25	3.121399	2.822255	3.141413	-0.0200145	0.0142971	0.0286299	0.0036225	-0.7067561
26	2.968010	2.659829	2.916542	0.0514684	0.0604171	0.0279938	0.1114125	1.8615281
27	3.057669	2.760252	3.055573	0.0020961	0.0206704	0.0287344	0.0000582	0.0742567
28	3.080438	2.805560	3.118299	-0.0378614	0.0146487	0.0283553	0.0132915	-1.3372036
29	3.267924	2.902137	3.252006	0.0159177	0.0265316	0.0286679	0.0043596	0.5656090
30	3.089041	2.801818	3.113119	-0.0240779	0.0148654	0.0285824	0.0054574	-0.8504864

.rownames	log__diameter	log__height	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
31	3.013509	2.777931	3.080048	-0.0665385	0.0174397	0.0275405	0.0491508	-2.3533710
32	3.206455	2.853160	3.184200	0.0222548	0.0162999	0.0286045	0.0051271	0.7866638
33	3.380660	2.987384	3.370027	0.0106329	0.0649810	0.0287042	0.0051643	0.3855133
34	3.166719	2.863780	3.198902	-0.0321835	0.0177836	0.0284604	0.0117337	-1.1384814
35	3.087883	2.764596	3.061587	0.0262956	0.0197721	0.0285518	0.0087443	0.9311418
36	2.886077	2.637153	2.885147	0.0009303	0.0744280	0.0287354	0.0000462	0.0338994
37	3.165425	2.838049	3.163279	0.0021458	0.0148901	0.0287344	0.0000434	0.0757952
38	3.243927	2.885436	3.228885	0.0150423	0.0220703	0.0286754	0.0032091	0.5332822
39	3.167001	2.845342	3.173376	-0.0063744	0.0154677	0.0287249	0.0003985	-0.2252276
40	3.173793	2.886044	3.229726	-0.0559327	0.0222150	0.0278923	0.0446740	-1.9830823
41	3.266705	2.900572	3.249839	0.0168656	0.0260707	0.0286596	0.0048047	0.5991484
42	3.163640	2.851488	3.181884	-0.0182448	0.0161034	0.0286476	0.0034030	-0.6448527
43	3.286619	2.899362	3.248164	0.0384553	0.0257205	0.0283387	0.0246257	1.3658772
44	2.972333	2.721179	3.001478	-0.0291444	0.0318111	0.0285069	0.0177147	-1.0384180
45	3.269675	2.867061	3.203445	0.0662296	0.0183244	0.0275508	0.0512580	2.3435005
46	3.193835	2.831749	3.154558	0.0392778	0.0145457	0.0283261	0.0142010	1.3871590
47	3.176022	2.849507	3.179142	-0.0031199	0.0158836	0.0287330	0.0000981	-0.1102597
48	3.182277	2.858542	3.191650	-0.0093728	0.0170010	0.0287124	0.0009499	-0.3314296
49	3.141252	2.802939	3.114671	0.0265806	0.0147952	0.0285487	0.0066185	0.9388551
50	3.292055	2.926568	3.285829	0.0062260	0.0348704	0.0287252	0.0008918	0.2221843
52	3.358222	2.983853	3.365138	-0.0069168	0.0628679	0.0287223	0.0021047	-0.2504949
53	3.123785	2.826480	3.147262	-0.0234766	0.0143675	0.0285900	0.0050094	-0.8290385
54	3.150336	2.808701	3.122647	0.0276886	0.0145057	0.0285329	0.0070372	0.9778464
55	3.100789	2.773649	3.074121	0.0266683	0.0181186	0.0285469	0.0082141	0.9435441
56	3.094747	2.792249	3.099871	-0.0051238	0.0156495	0.0287287	0.0002606	-0.1810567
57	2.819645	2.599971	2.833671	-0.0140259	0.1014156	0.0286787	0.0151850	-0.5187393
58	3.094649	2.753717	3.046525	0.0481244	0.0221502	0.0281138	0.0329707	1.7061847
59	3.085807	2.786218	3.091522	-0.0057151	0.0163133	0.0287270	0.0003384	-0.2020184
62	3.015794	2.742801	3.031413	-0.0156186	0.0249654	0.0286706	0.0039368	-0.5545346
63	3.204511	2.862620	3.197296	0.0072146	0.0176018	0.0287218	0.0005834	0.2551927
64	2.852785	2.632643	2.878904	-0.0261189	0.0774353	0.0285429	0.0381435	-0.9533536
65	3.149351	2.852895	3.183833	-0.0344825	0.0162682	0.0284200	0.0122842	-1.2188701
66	3.198549	2.848702	3.178027	0.0205219	0.0157983	0.0286243	0.0042212	0.7252234
67	2.986395	2.744977	3.034424	-0.0480291	0.0243701	0.0281149	0.0362960	-1.7047404

.rownames	log_diameter	log_height	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
68	3.223647	2.894818	3.241874	-0.0182272	0.0244527	0.0286470	0.0052461	-0.6469833
69	3.142153	2.834877	3.158888	-0.0167351	0.0146988	0.0286617	0.0026059	-0.5910711
70	3.198616	2.868581	3.205549	-0.0069322	0.0185879	0.0287229	0.0005700	-0.2453263
71	3.139381	2.779245	3.081868	0.0575127	0.0172445	0.0278478	0.0362954	2.0339393
72	3.123163	2.823458	3.143079	-0.0199160	0.0143106	0.0286309	0.0035904	-0.7032810
73	2.918255	2.677871	2.941520	-0.0232646	0.0505950	0.0285872	0.0186706	-0.8370779

2.3.8 Durbin Watson Test Linear Model 3 [Log transformed with filtered and excluded outliers]

lag Autocorrelation D-W Statistic p-value 1 -0.03907805 2.067896 0.722 Alternative hypothesis: rho != 0

2.3.9 Shapiro Wilks Test Linear Model 3 [Log transformed with filtered and excluded outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm4 W = 0.9806, p-value = 0.3498

Report Results Linear Model 3 [Log transformed with filtered and excluded outliers] 2.3.99

Linear Model 3 involved logging the data frame after outlier removal, as depicted in Figure 2.3.2. The predicted response variable equation presents an intercept of -0.766 and a slope of 1.38 for mean height, reflecting normalized data. Confidence intervals remain consistent. Plots indicate significantly straighter lines for residuals vs fitted values (2.3.5), affirming the success of logarithmic transformation. The Shapiro-Wilks test yielded a p-value of 0.348, supporting the assumption of normal distribution, while the non-constant variance test (p = 0.8) suggests constant error assumptions remain unviolated. Despite residual verification by the Cook's model, the impact of logged outliers on model assumptions warrants further investigation.

3.0 [Simple Linear Model 4] [Logged comparison with outliers for comparisson]

3.1.1 log transforming original tdat data frame [Logged comparison with outliers for comparisson]

3.1.2 Making simple linear model 4 [Logged comparison with outliers for comparisson]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.6686391	0.1107058	-6.039782	1e-07
log_height_2	1.3507256	0.0391759	34.478442	0e+00

3.3.4 Predicted Response Variable Linear Model 4 [Logged comparison with outliers for comparisson]

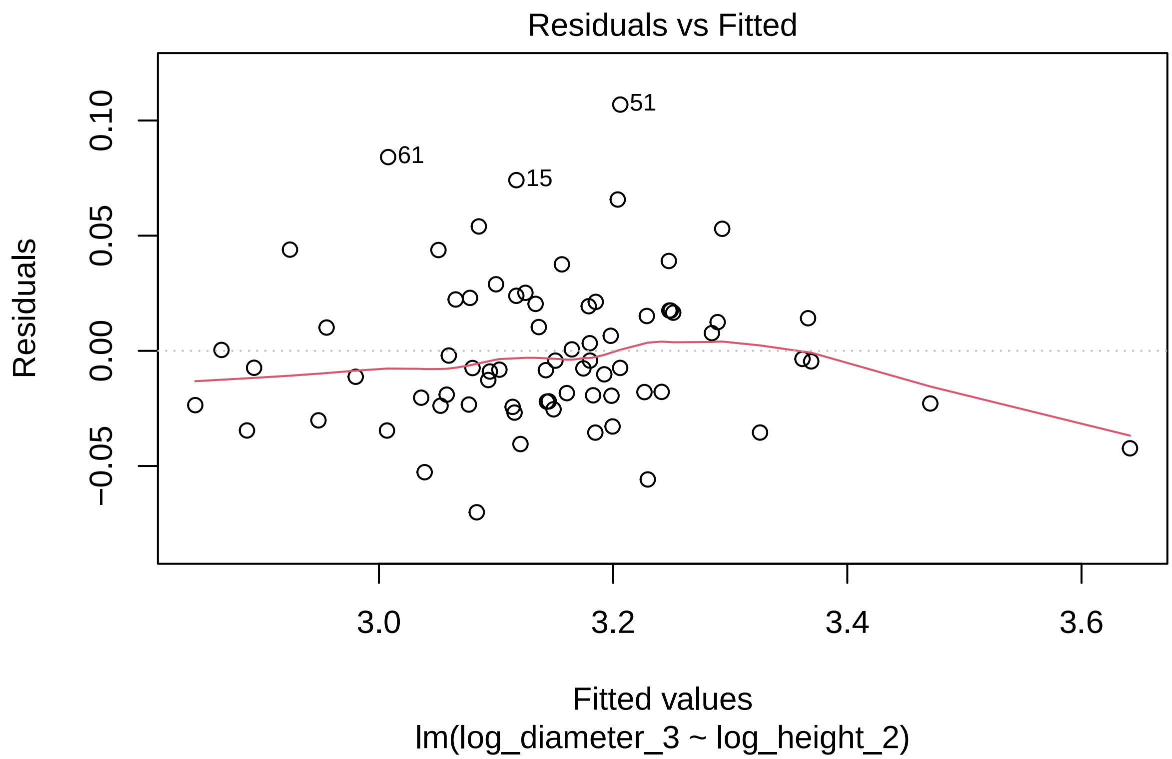
$$\hat{y_i} = -0.6686391 + 1.3507256\text{Ö}MeanHeight(m)$$

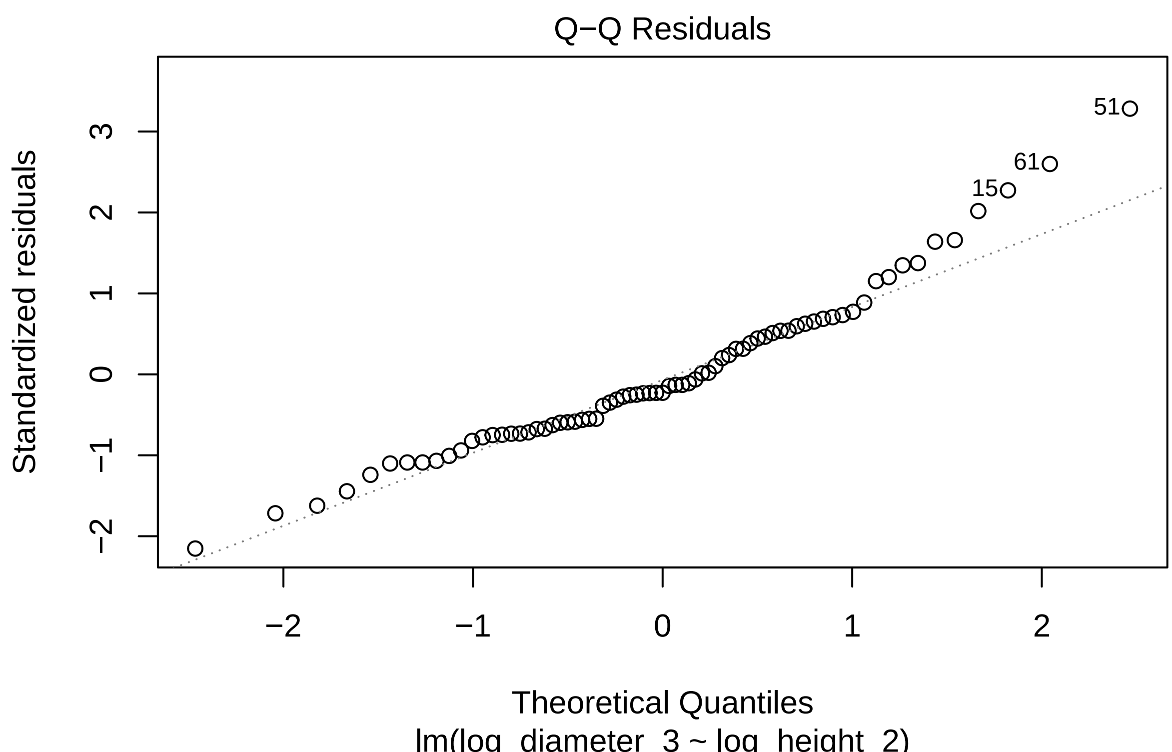
3.3.4 Confidence Intervals Simple Linear Model 4 [Logged comparison with outliers for comparisson]

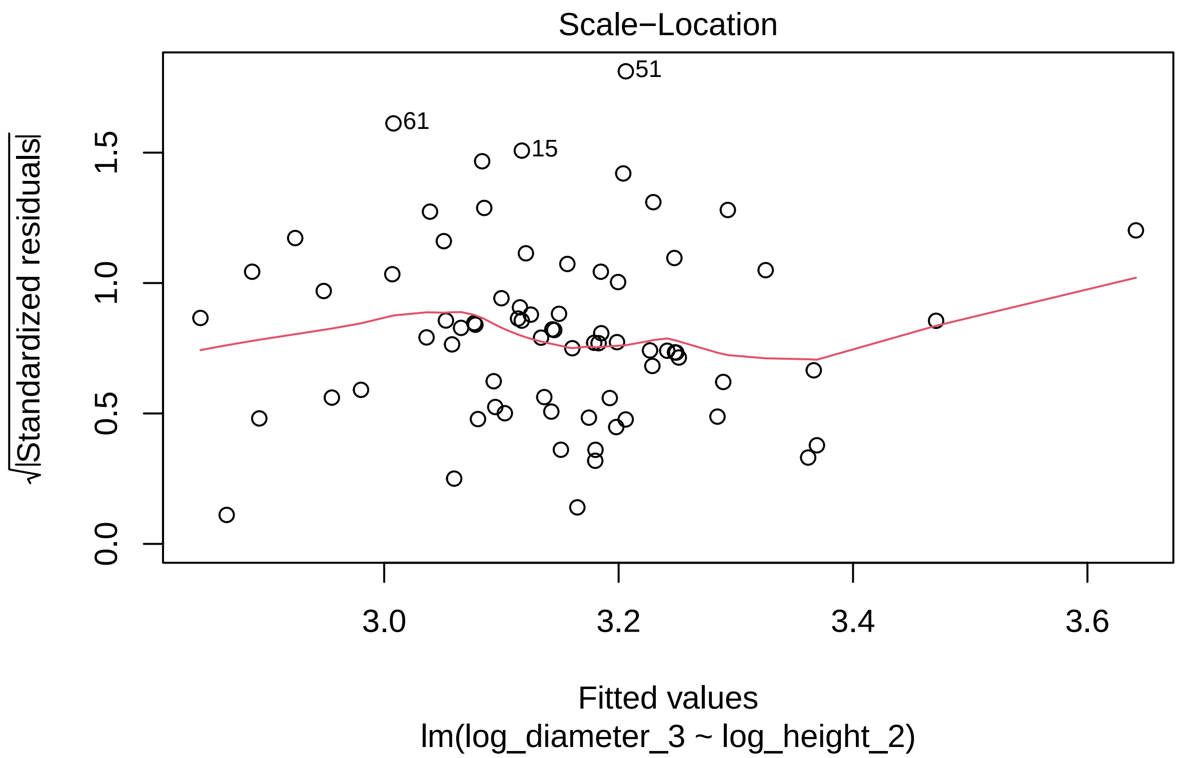
	2.5 %	97.5 %
MEAN Diamater (cm)	NA	NA

Assumptions of Linear Model 4 [Logged comparison with outliers for comparisson]

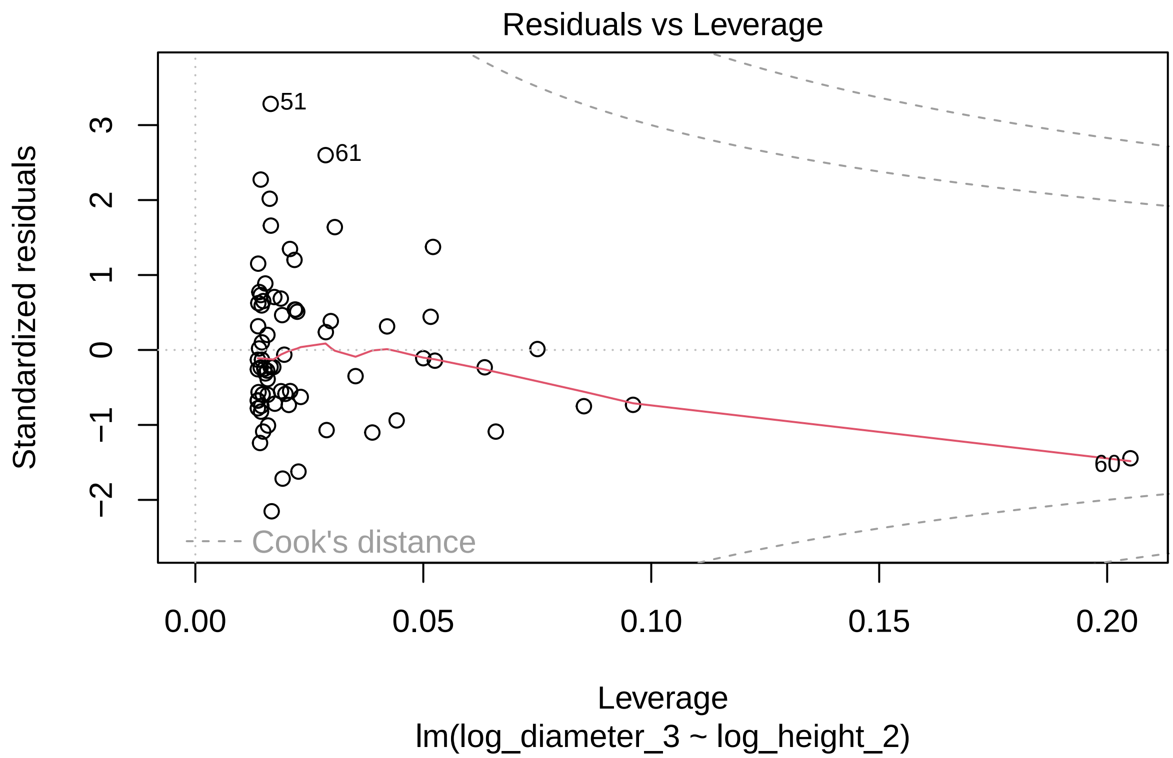
3.3.5 Plot Linear Model 4 [Logged comparison with outliers for comparisson]







Scale−Location
$\sqrt{|\text{Standardized residuals}|}$
0.0
0.5
1.0
1.5
51
61
15
3.0
3.2
3.4
3.6
Fitted values
lm(log_diameter_3 ~ log_height_2)



3.3.6 Non-constant Error Variance Test Linear Model 4 [Logged comparison with outliers for comparisson]

Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.03714213, Df = 1, p = 0.84718

```
Breusch-Pagan test
```

data: lm5_with_outliers BP = 0.037142, df = 1, p-value = 0.8472

```
studentized Breusch-Pagan test
```

data: lm5_with_outliers BP = 0.024904, df = 1, p-value = 0.8746

3.3.7 Cook's Model: Identifying Points of Leverage [Logged comparison with outliers for comparisson]

log_diameter_3	log_height_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
3.313052	2.868667	3.206143	0.1069087	0.0165180	0.0304593	0.0904995	3.282789
3.599080	3.190893	3.641382	-0.0423018	0.2051102	0.0325827	0.2693324	-1.444838
3.092036	2.721915	3.007921	0.0841154	0.0285765	0.0314601	0.0993437	2.598871

log_diameter_3	log_height_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
2.968957	2.701399	2.980210	-0.0112532	0.0351461	0.0330441	0.0022167	-0.3488674
3.038877	2.758912	3.057894	-0.0190167	0.0197575	0.0329927	0.0034477	-0.5849002
3.208896	2.883955	3.226792	-0.0178959	0.0187873	0.0330019	0.0028976	-0.5501539
3.080754	2.785217	3.093425	-0.0126715	0.0158568	0.0330372	0.0012188	-0.3889652
3.089908	2.800626	3.114238	-0.0243300	0.0144868	0.0329425	0.0040938	-0.7463178
3.265560	2.899691	3.248048	0.0175114	0.0218182	0.0330047	0.0032420	0.5391677
2.866002	2.616558	2.865613	0.0003892	0.0750358	0.0330725	0.0000062	0.0123223
3.364615	2.989372	3.369182	-0.0045672	0.0525454	0.0330677	0.0005661	-0.1428850
3.146886	2.817151	3.136559	0.0103278	0.0137685	0.0330491	0.0007001	0.3166878
3.301749	2.930221	3.289285	0.0124636	0.0297084	0.0330379	0.0022728	0.3853069
2.965470	2.683001	2.955359	0.0101114	0.0420567	0.0330494	0.0021726	0.3145973
3.346187	2.933117	3.293197	0.0529900	0.0305947	0.0324409	0.0423858	1.6389082
3.134200	2.821615	3.142589	-0.0083883	0.0137078	0.0330571	0.0004597	-0.2572087
3.072547	2.775278	3.080000	-0.0074536	0.0170991	0.0330603	0.0004559	-0.2289423
3.191545	2.802989	3.117430	0.0741146	0.0143364	0.0318461	0.0375827	2.2732774
3.154228	2.815139	3.133841	0.0203866	0.0138144	0.0329813	0.0027372	0.6251433
3.128929	2.790103	3.100025	0.0289042	0.0153493	0.0328887	0.0061325	0.8870184
3.053524	2.772944	3.076847	-0.0233231	0.0174317	0.0329527	0.0045539	-0.7165033
3.028830	2.755039	3.052662	-0.0238327	0.0204981	0.0329470	0.0056266	-0.7333042
3.183381	2.849365	3.180070	0.0033104	0.0146028	0.0330701	0.0000764	0.1015517
3.448011	3.064646	3.470856	-0.0228450	0.0960077	0.0329476	0.0284285	-0.7316803
3.146491	2.827650	3.150740	-0.0042491	0.0137160	0.0330685	0.0001180	-0.1302879
3.179143	2.863106	3.198631	-0.0194880	0.0158574	0.0329890	0.0028830	-0.5982085
3.290022	2.957018	3.325481	-0.0354584	0.0388202	0.0327888	0.0244954	-1.1013635
3.121399	2.822255	3.143453	-0.0220545	0.0137038	0.0329658	0.0031770	-0.6762483
2.968010	2.659829	2.924060	0.0439496	0.0521310	0.0326294	0.0519644	1.3746557
3.057669	2.760252	3.059704	-0.0020353	0.0195111	0.0330716	0.0000390	-0.0625926
3.080438	2.805560	3.120902	-0.0404644	0.0141910	0.0327118	0.0110859	-1.2410536
3.267924	2.902137	3.251352	0.0165723	0.0223525	0.0330118	0.0029780	0.5103938
3.089041	2.801818	3.115848	-0.0268072	0.0144090	0.0329146	0.0049424	-0.8222733
3.013509	2.777931	3.083583	-0.0700736	0.0167401	0.0319757	0.0394209	-2.1519564
3.206455	2.853160	3.185197	0.0212573	0.0148956	0.0329733	0.0032159	0.6522006
3.380660	2.987384	3.366497	0.0141630	0.0516161	0.0330268	0.0053374	0.4428712
3.166719	2.863780	3.199541	-0.0328227	0.0159328	0.0328352	0.0082184	-1.0075701
3.087883	2.764596	3.065572	0.0223107	0.0187477	0.0329627	0.0044938	0.6858619
2.886077	2.637153	2.893430	-0.0073534	0.0634699	0.0330600	0.0018143	-0.2313892
3.165425	2.838049	3.164787	0.0006386	0.0139732	0.0330724	0.0000027	0.0195846
3.243927	2.885436	3.228793	0.0151336	0.0190426	0.0330220	0.0021014	0.4652960
3.167001	2.845342	3.174637	-0.0076356	0.0143372	0.0330597	0.0003989	-0.2342025
3.173793	2.886044	3.229614	-0.0558209	0.0191491	0.0323791	0.0287564	-1.7163614
3.266705	2.900572	3.249237	0.0174674	0.0220086	0.0330050	0.0032552	0.5378656
3.163640	2.851488	3.182938	-0.0192986	0.0147616	0.0329907	0.0026260	-0.5920650
3.286619	2.899362	3.247603	0.0390163	0.0217475	0.0327347	0.0160397	1.2012523
2.972333	2.721179	3.006927	-0.0345938	0.0287915	0.0328053	0.0169368	-1.0689446
3.269675	2.867061	3.203974	0.0657011	0.0163181	0.0321108	0.0337523	2.0172436
3.193835	2.831749	3.156277	0.0375582	0.0137806	0.0327621	0.0092668	1.1516779
3.176022	2.849507	3.180263	-0.0042406	0.0146131	0.0330685	0.0001255	-0.1300881
3.182277	2.858542	3.192466	-0.0101888	0.0153811	0.0330497	0.0007636	-0.3126806
3.141252	2.802939	3.117362	0.0238891	0.0143395	0.0329472	0.0039055	0.7327402
3.292055	2.926568	3.284351	0.0077047	0.0286245	0.0330593	0.0008350	0.2380535
3.313052	2.868667	3.206143	0.1069087	0.0165180	0.0304593	0.0904995	3.2827888
3.358222	2.983853	3.361727	-0.0035058	0.0499933	0.0330697	0.0003157	-0.1095297

log_diameter_3	log_height_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
3.123785	2.826480	3.149159	-0.0253740	0.0137063	0.0329312	0.0042061	-0.7780362
3.150336	2.808701	3.125145	0.0251915	0.0140387	0.0329332	0.0042492	0.7725679
3.100789	2.773649	3.077800	0.0229888	0.0173295	0.0329561	0.0043974	0.7061969
3.094747	2.792249	3.102923	-0.0081759	0.0151478	0.0330578	0.0004840	-0.2508776
2.819645	2.599971	2.843208	-0.0235638	0.0852289	0.0329411	0.0262210	-0.7502442
3.094649	2.753717	3.050877	0.0437726	0.0207607	0.0326472	0.0192338	1.3470105
3.085807	2.786218	3.094778	-0.0089706	0.0157473	0.0330548	0.0006065	-0.2753469
3.599080	3.190893	3.641382	-0.0423018	0.2051102	0.0325827	0.2693324	-1.4448378
3.092036	2.721915	3.007921	0.0841154	0.0285765	0.0314601	0.0993437	2.5988712
3.015794	2.742801	3.036133	-0.0203386	0.0231189	0.0329809	0.0046465	-0.6266336
3.204511	2.862620	3.197974	0.0065363	0.0158039	0.0330631	0.0003232	0.2006332
2.852785	2.632643	2.887339	-0.0345547	0.0658991	0.0327952	0.0418123	-1.0887399
3.149351	2.852895	3.184840	-0.0354889	0.0148739	0.0327952	0.0089500	-1.0888297
3.198549	2.848702	3.179175	0.0193740	0.0145559	0.0329901	0.0026086	0.5943157
2.986395	2.744977	3.039071	-0.0526757	0.0226219	0.0324536	0.0304664	-1.6225282
3.223647	2.894818	3.241466	-0.0178195	0.0208043	0.0330024	0.0031945	-0.5483698
3.142153	2.834877	3.160502	-0.0183492	0.0138621	0.0329987	0.0022253	-0.5626811
3.198616	2.868581	3.206026	-0.0074095	0.0165070	0.0330604	0.0004344	-0.2275196
3.139381	2.779245	3.085359	0.0540220	0.0165696	0.0324252	0.0231826	1.6588696
3.123163	2.823458	3.145078	-0.0219153	0.0136993	0.0329671	0.0031360	-0.6719808
2.918255	2.677871	2.948430	-0.0301748	0.0441552	0.0328661	0.0204029	-0.9398604

3.3.9 Durbin Watson Test Linear Model 4 [Logged comparison with outliers for comparisson]

lag Autocorrelation D-W Statistic p-value 1 -0.1055016 2.197457 0.406 Alternative hypothesis: rho != 0

3.3.10 Shapiro Wilks Test Linear Model 4 [Logged comparison with outliers for comparisson]

`Shapiro-Wilk normality test`

data: residuals_lm5_with_outliers W = 0.95824, p-value = 0.0165

3.3.11 Report Results Simple Linear Model 4 [Logged comparison with outliers for comparisson] Figure 3.1.2 illustrates the creation of this model. The predicted response variable equation yields an intercept of -0.66 and a slope of 1.35 for mean height, indicating normalized data. Confidence intervals remain consistent. Plots reveal a straighter line but with noticeable irregularities compared to the logged data without outliers (3.3.5). While the Shapiro-Wilks test resulted in a p-value of 0.0165, indicating rejection of normal distribution, the non-constant variance test ($p = 0.84$) suggests constant error assumptions are unviolated. Although the Cook's model verified additional residuals, the persistence of high p-values suggests the superiority of logged values without outliers for modeling purposes.

Question 2:

The second question asked us to compare the relationship between Tree Height and Tree Density. We created a linear model and a system of tests, to test the assumptions of the correlation. Figure 1.1.4 we tested the linear correlation between mean tree height and tree density: with correlation coefficient (r =0.49) this indicates a positive linear relationship between Mean Diameter and Mean Height. The Null for the linear correlation is that there is no correlation between the two variables (in this case, Mean Diameter and Mean Height of trees), because (p = 1.15e-05) there is statistic significance and we can assume correlation. Figure 1.1.5 is making linear model 1. The table and equation in 1.1.6 provide is the equation for a linear model for predicting the tree diameter based on the mean height of the tree. It suggests that the predicted tree height increases linearly with an increase in density, with a slope of 11.24 and an intercept of 10.11(see figure 1.1.5 and 1.1.6). The confidence intervals(97.5 and 2.5%) and predicted response variables of the linear model where plotted in 1.1.59 for a visualization.

The assumptions for Linear Model 1 plot. In plot linear model 1(figure 1.1.8) the residuals vs fitted plot shows a non linear "horizonta"l line with distinct patterns and several outlying residuals noted. For the q-q plot in figure 1.1.8 the points show "significant departure" from the line. Suggesting we will need another model. The third model, standard resid vs fitted values on different scale(1.1.8) shows, the points distributed inside a triangular shape, with the scatter of the residuals increasing as the fitted values increase- suggesting we need to continue the data wrangling. The fourth plot in (1.1.8) is residuals-leverage and shows the plots that shows Cook's distance for each of the observed values. The point is to highlight those yi (response) values that have the biggest effect on parameter estimates. The idea is to verify that no single data point is so influential that leaving it out changes the structure of the model,I have done this and will be taken out later.

In figure 1.1.9 we are using the chiˆ2 to assess whether the variance of the errors in the linear regression model is constant across all levels of the predictor variable. Generally, we want Homoscedasticity (which is Lower error variance), which means that the spread of the residuals around the regression line remains constant across all levels of the predictor variable. Homoscedasticity is one of the assumptions of linear regression analysis and is the null hypothesis. When we have a p value(p = 0.020) and we fail to reject the null, or we don't have enough evidence to reject that the assumption of constant error variance is violated. This is good, however lets move onto the cook's model.

In figure 1.2.1, we identify all points of leverage using the cooks model and display them. Observations with large Cook's distances are potential outliers or influential points. These observations might have a significant impact on the estimated coefficients of the regression model. In 1.2.1 we see the top three with high residuals numbers, these have the potentials to be outliers, and also observed this same rows as problematic in the linear model plot, so it was decided to remove them in the re-run of the model. The p-value(p = 0.03) in the Durbin-Watson (figure 1.2.2) test assesses the presence of autocorrelation in the residuals of a regression model. Specifically, it tests whether there is a significant correlation between adjacent residuals, which could indicate a violation of the assumption of independence. We can assume is no significant autocorrelation detected in the residuals of the regression model. In figure 1.2.29, the shapiro wilks test has a p value(0.2871) of low significance which means we fail reject that the null is that the residuals are normally distributed, however we can make this P value higher so lets wrangle the data a bit.

Linear model 2 I took out outliers identified by the various test in linear model 1 displayed in figure 1.2.4. Figure 1.2.49 makes linear model 2 without outliers. The predicted response variable equation shows 11.29 intercept and 9.09 slope for mean height. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm2 show straighter line for residuals vs fitted(2.2.7) but slight variation indicating we need to log it. Plot 2 (fig. 2.2.7) shows a s shape, showing we need to log it. And plot 3 shows the same thing plot 1 did. Plot 4 shows we essentially need to log the data.

The P value increased in the Shapiro wilks (2.2.11)test (p = 0.4368) which means that we fail to reject that the data is normally distributed(assuming it is normally distrubuted) but to make that p-value higher we should log the data. The non-constant variance test(2.2.8) p value is still high at(p = 0.46) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 2.2.9)helped us identify some more residuals, however, it showed that the ones removed may have helped increase that p value in the shapiro test but still we know that the largest outliers were removed AND the data still fails the assumptions of good regression from the linear plots, so LETS LOG the data to see if this helps even more!

Linear model 3 I logged the data frame without the outliers. Figure 2.3.2 makes linear model 3. The predicted response variable equation shows 2.97 intercept and .31 slope for mean height these values you can expect to see when the data is normilized. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm2 show MUCH straighter line for residuals vs fitted(2.3.5). Plot 2 (fig. 2.3.5) shows a MUCH straighter shape fitting to that line, showing that the logarithm proved successful and plot 3 shows the same thing plot 1 did. Plot 4 shows a much straighter line! These are all great signs.

The P value increased in the Shapiro wilks (2.3.9)test (p = 0.174) which means that we fail to reject that the data is normally distributed(assuming it is normally distributed) and make it much higher than last time. But this could be a higher p value, next lets try it with outliers and logged.

The non-constant variance test(2.3.6) p value is still high at(p = 0.14) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 2.3.7)helped us verify some more residuals, however, it showed that the ones removed and the logged values may have helped increase that p value in the shapiro test. We still don't know if logged with outliers made a difference. We are going to test this next.

Linear model 4 I logged comparison with the outliers. Figure 3.1.2 makes linear model 4. The predicted response variable equation shows 2.9 intercept and .335 slope for mean height these values you can expect to see when the data is normalized. Confidence intervals remain normal at 97.5 and 2.5%. Plots for lm4 show a straight line but it is still wonkier than the logged without outliers(3.3.5). Plot 2 (fig. 3.3.5) shows a slightly straighter shape fitting to that line, showing that the logarithm proved good but containing the outliers still disrupts the data.And plot 3 shows the same thing plot 1 did. Plot 4 shows a much straighter line.

The P value increased in the Shapiro wilks (2.3.9)test (p = 0.34) which means that we fail to reject that the data is normally distributed(assuming it is normally distributed) but sine this is much lower than the logged data without outliers, it suggest that there is more reason to assume that the outliers have a large effect on the linear correlation of the data frame and should be removed.

The non-constant variance test(3.3.6) p value is still high at(p = 0.84) suggesting we still don't have enough evidence to reject that the assumption of constant error is violated. The cooks model (figure 3.3.7)helped us verify some more residuals, however, it showed that the Although the logged values with outliers had a higher shapiro p value than logged values without, their plots for the linear model 4 looked very wonky and non linear compared to that of linear model 3, making the logged data WITHOUT outliers, the final winner for what we will use for the correlation model.

Question 2: Null and Alternative Hypothesis

- H_0: Tree density does not change with mean tree height
- H_a: Tree density does change with mean tree height

1.1.4 Simple Linear Correlation

var1	var2	cor	statistic	p	conf.low	conf.high	method
Mean Height (m)	Mean Wood Density (g m^3)	0.49	4.719028	1.15e-05	0.2913157	0.6460706	Pearson

1.1.49 [Simple Linear Model 1] [raw data with outliers] [Below]

1.1.5 Making Linear Model 1 [raw data with outliers]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.11059	1.45585	6.944800	0.00e+00
`Mean Wood Density (g m^3)`	11.24190	2.38225	4.719028	1.15e-05

1.1.6 Predicted Response Variable in Liner Model 1 [raw data with outliers]

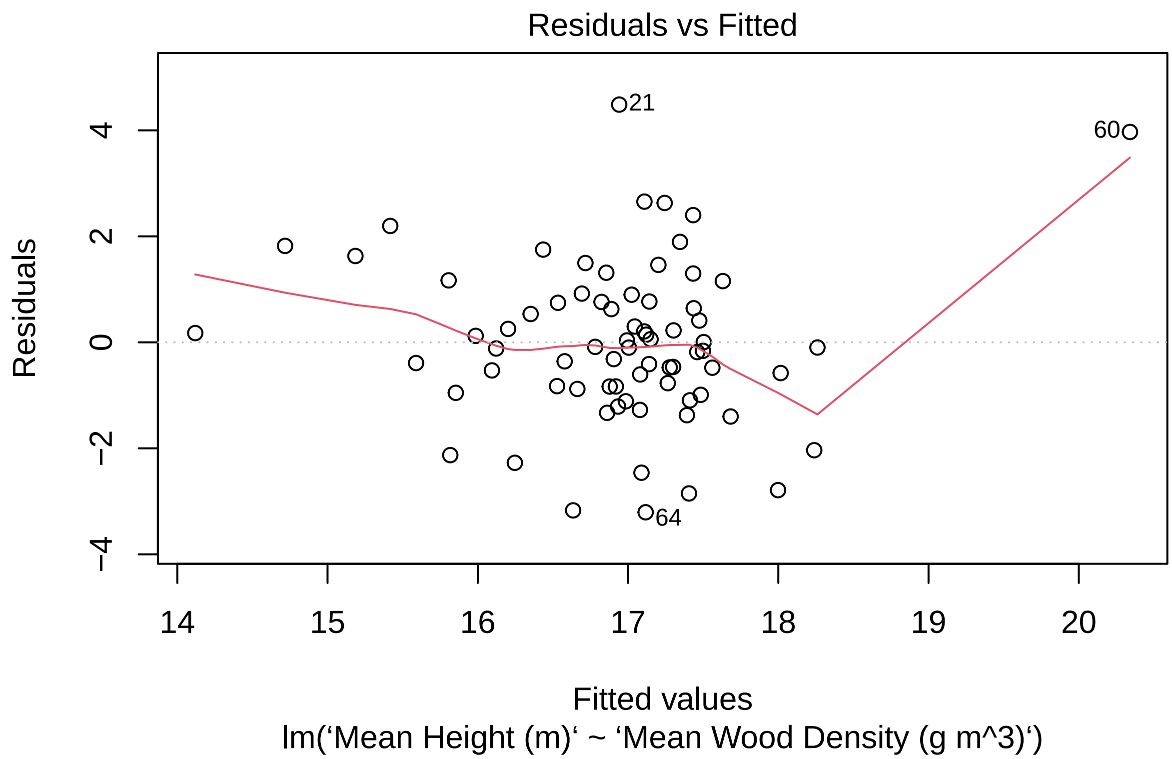
$$\hat{y}_i = 10.11059 + 11.24190\text{Ö}MeanHeight(m)$$

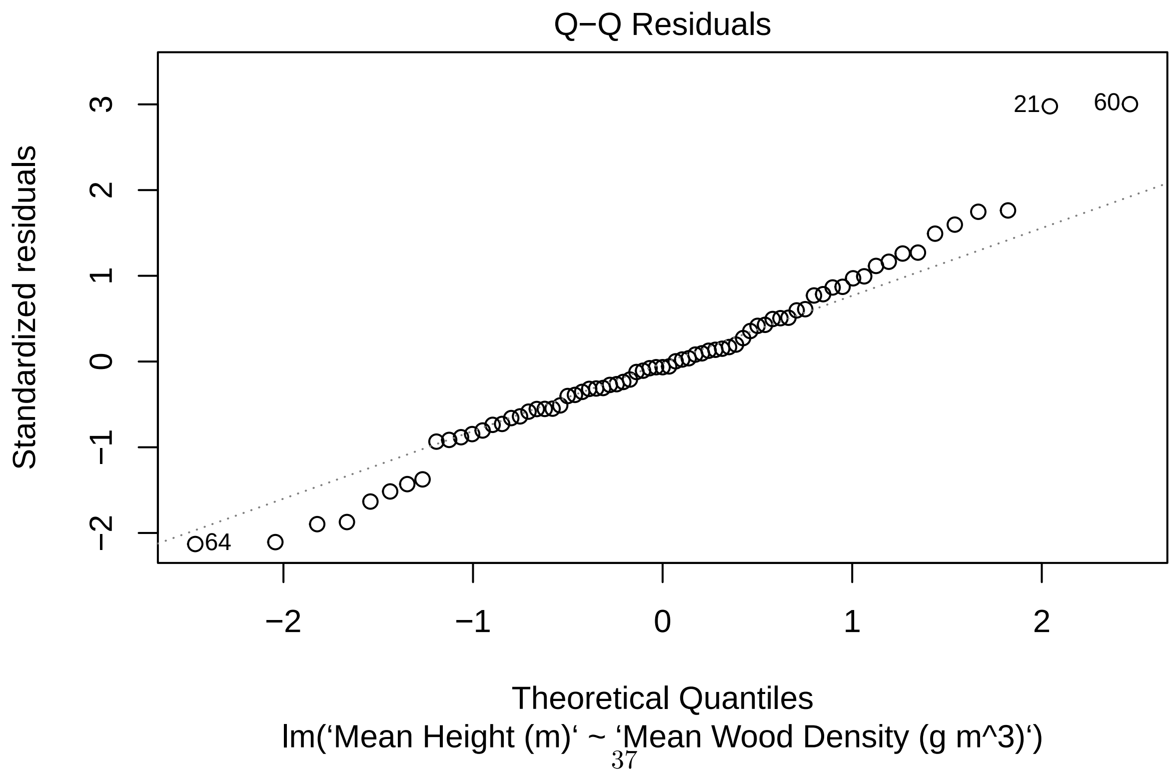
1.1.7 Confidence Intervals of Liner Model 1 [raw data with outliers]

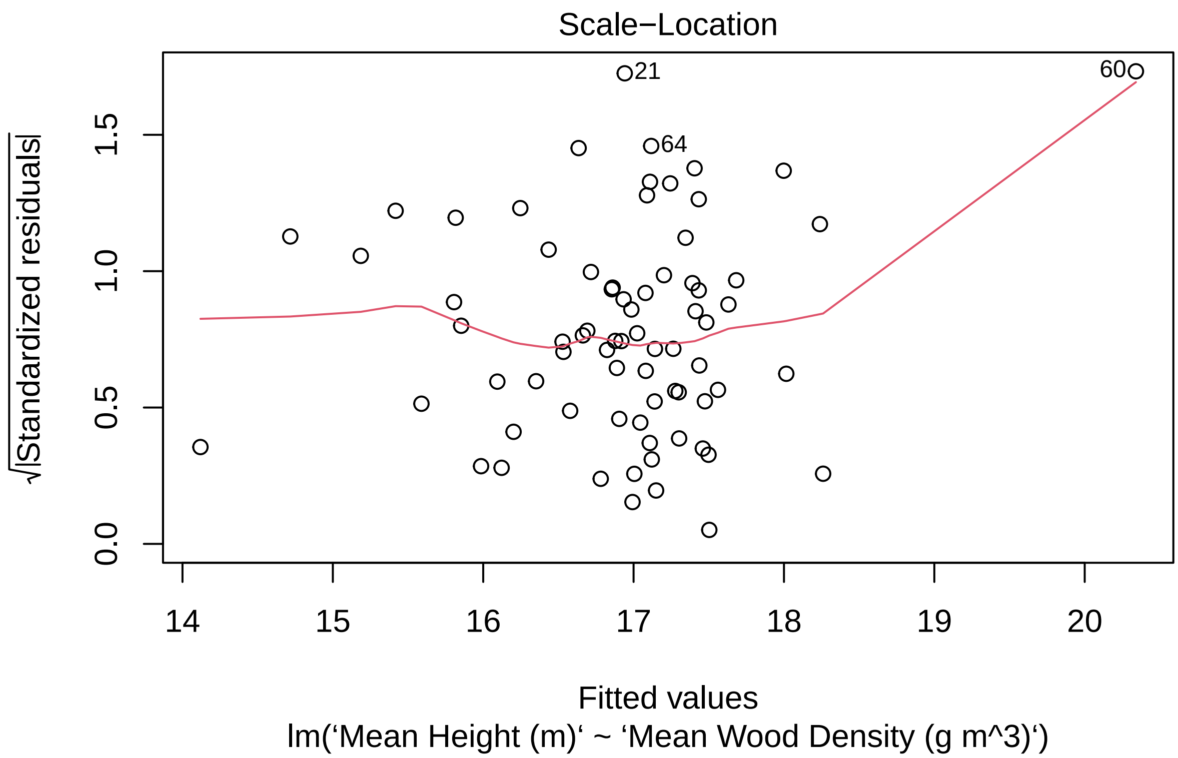
	2.5 %	97.5 %
MEAN Height (m)	NA	NA

1.79 Assumptions of Linear Model 1 [raw data with outliers] [Below]

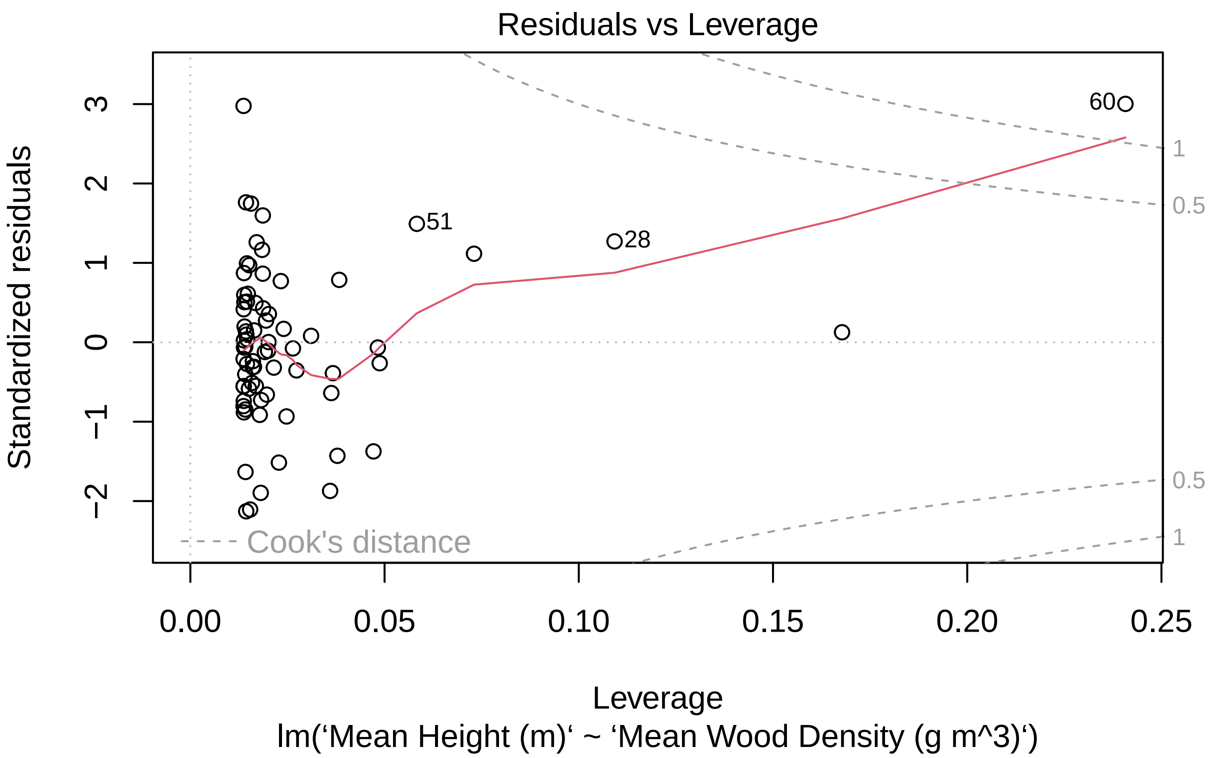
1.1.8 Plot Linear Model 1 [raw data with outliers]







Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
14
15
16
17
18
19
20
21
64
60
Fitted values
lm(‘Mean Height (m)‘ ~ ‘Mean Wood Density (g m^3)‘)



1.1.9 Non-constant Error Variance Test Linear Model 1 [raw data with outliers] Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 5.397114, Df = 1, p = 0.02017

```
Breusch-Pagan test
```

data: lm22 BP = 5.3971, df = 1, p-value = 0.02017

```
studentized Breusch-Pagan test
```

data: lm22 BP = 4.109, df = 1, p-value = 0.04266

1.2.1 Cook's Model: Identifying Points of Leverage Linear Model 1 [raw data with outliers]

Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
16.53633	0.40975	14.71696	1.819372	0.1092084	1.510425	0.0989667	1.270631
17.61353	0.47205	15.41733	2.196202	0.0583121	1.503761	0.0689019	1.491780
24.31013	0.91000	20.34072	3.969410	0.2407312	1.427590	1.4293395	3.002716

Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
14.90057	0.51084	15.85340	-0.9528318	0.0362912	1.523486	0.0077070	-0.6397782
15.78266	0.58283	16.66271	-0.8800465	0.0150873	1.524216	0.0026168	-0.5845112
17.88486	0.65501	17.47415	0.4107129	0.0194863	1.527092	0.0007427	0.2733994
16.20334	0.72304	18.23893	-2.0355939	0.0471519	1.507430	0.0467494	-1.3745653
16.45494	0.54183	16.20179	0.2531516	0.0240305	1.527590	0.0003512	0.1689074
18.16854	0.59995	16.85517	1.3133721	0.0138064	1.519698	0.0053195	0.8717517
13.68853	0.50758	15.81675	-2.1282232	0.0378563	1.505729	0.0402379	-1.4301556
19.87320	0.63450	17.24358	2.6296243	0.0156233	1.494696	0.0242203	1.7470250
16.72912	0.62528	17.13993	-0.4108053	0.0145626	1.527096	0.0005498	-0.2727769
18.73177	0.65138	17.43334	1.2984310	0.0186516	1.519844	0.0070933	0.8639594
14.62893	0.62078	17.08934	-2.4604068	0.0141971	1.498914	0.0192122	-1.6334203
18.78610	0.66894	17.63075	1.1553531	0.0232930	1.521494	0.0070806	0.7705817
16.80397	0.63749	17.27719	-0.4732190	0.0160573	1.526833	0.0008069	-0.3144585
16.04309	0.60195	16.87765	-0.8345617	0.0137511	1.524592	0.0021391	-0.5539254
16.49388	0.63632	17.26404	-0.7701560	0.0158822	1.525077	0.0021131	-0.5117304
16.69550	0.59338	16.78131	-0.0858086	0.0141271	1.527862	0.0000232	-0.0569648
16.28270	0.67354	17.68246	-1.3997596	0.0247602	1.518476	0.0110810	-0.9342946
16.00568	0.53475	16.12220	-0.1165158	0.0264142	1.527832	0.0000822	-0.0778366
15.72165	0.60692	16.93352	-1.2118740	0.0136989	1.520920	0.0044929	-0.8043384
17.27680	0.65379	17.46043	-0.1836320	0.0191985	1.527736	0.0001462	-0.1222205
21.42687	0.60759	16.94106	4.4858139	0.0137012	1.429339	0.0615698	2.9773035
16.90569	0.61328	17.00502	-0.0993325	0.0138099	1.527850	0.0000304	-0.0659321
17.51584	0.60294	16.88878	0.6270588	0.0137310	1.526032	0.0012058	0.4161948
19.24051	0.64358	17.34565	1.8948578	0.0170776	1.510723	0.0137874	1.2598043
16.81473	0.45144	15.18563	1.6290973	0.0730309	1.514453	0.0490018	1.1153225
14.29385	0.35659	14.11934	0.1745119	0.1677721	1.527726	0.0016026	0.1260927
15.80383	0.61987	17.07911	-1.2752766	0.0141354	1.520165	0.0051383	-0.8466070
16.53633	0.40975	14.71696	1.8193722	0.1092084	1.510425	0.0989667	1.2706314
18.21303	0.58759	16.71622	1.4968120	0.0145861	1.517231	0.0073110	0.9939030
16.47457	0.62003	17.08091	-0.6063353	0.0141459	1.526153	0.0011624	-0.4025248
16.08570	0.60560	16.91868	-0.8329847	0.0137009	1.524605	0.0021230	-0.5528646
17.34250	0.61680	17.04459	0.2979060	0.0139571	1.527476	0.0002768	0.1977504
19.83373	0.65138	17.43334	2.4003910	0.0186516	1.500197	0.0242423	1.5971895

Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
17.52765	0.63976	17.30271	0.2249419	0.0164162	1.527656	0.0001865	0.1495033
15.87263	0.61152	16.98524	-1.1126067	0.0137593	1.522018	0.0038041	-0.7384759
13.97336	0.54581	16.24653	-2.2731712	0.0227990	1.502973	0.0268013	-1.5157455
17.08241	0.66275	17.56116	-0.4787495	0.0214833	1.526801	0.0011172	-0.3190144
17.91138	0.62545	17.14184	0.7695435	0.0145783	1.525085	0.0019314	0.5109851
17.20744	0.62615	17.14971	0.0577342	0.0146448	1.527881	0.0000109	0.0383374
17.92227	0.61495	17.02380	0.8984735	0.0138721	1.524065	0.0025017	0.5963823
18.18454	0.56258	16.43506	1.7494820	0.0184683	1.513249	0.0127462	1.1639744
17.31352	0.62239	17.10744	0.2060838	0.0143164	1.527696	0.0001360	0.1368236
18.16255	0.72496	18.26052	-0.0979683	0.0482638	1.527850	0.0001111	-0.0661932
15.19823	0.48733	15.58910	-0.3908747	0.0487519	1.527146	0.0017882	-0.2641659
17.58526	0.59712	16.82335	0.7619067	0.0139184	1.525142	0.0018051	0.5057448
16.97513	0.50661	15.80585	1.1692814	0.0383321	1.521236	0.0123110	0.7859458
17.27926	0.57133	16.53342	0.7458354	0.0167592	1.525250	0.0020949	0.4957916
17.43608	0.70316	18.01544	-0.5793648	0.0367071	1.526267	0.0028846	-0.3890980
16.49305	0.65584	17.48348	-0.9904279	0.0196864	1.523212	0.0043654	-0.6593658
18.66346	0.63079	17.20187	1.4615918	0.0151461	1.517723	0.0072469	0.9707922
17.61353	0.47205	15.41733	2.1962016	0.0583121	1.503761	0.0689019	1.4917801
19.76382	0.62253	17.10901	2.6548099	0.0143274	1.494095	0.0225793	1.7625976
16.88591	0.55516	16.35164	0.5342670	0.0202134	1.526534	0.0013057	0.3557777
16.58835	0.60437	16.90486	-0.3165071	0.0137105	1.527422	0.0003067	-0.2100716
16.01698	0.64766	17.39152	-1.3745392	0.0178635	1.518877	0.0076011	-0.9142338
16.31768	0.64947	17.41187	-1.0941870	0.0182384	1.522185	0.0049215	-0.7279049
13.46335	0.58032	16.63449	-3.1711393	0.0153966	1.479380	0.0346956	-2.1065450
15.70088	0.57081	16.52758	-0.8266988	0.0168503	1.524644	0.0025882	-0.5495707
16.21957	0.57529	16.57794	-0.3583726	0.0161099	1.527287	0.0004643	-0.2381483
24.31013	0.91000	20.34072	3.9694097	0.2407312	1.427590	1.4293395	3.0027159
15.20942	0.70163	17.99824	-2.7888247	0.0359840	1.489703	0.0654223	-1.8722558
15.53043	0.60043	16.86056	-1.3301340	0.0137913	1.519487	0.0054501	-0.8828707
17.50733	0.65761	17.50338	0.0039539	0.0201242	1.527897	0.0000001	0.0026329
13.91049	0.62324	17.11699	-3.2065019	0.0143845	1.478325	0.0330739	-2.1289420
17.33791	0.65716	17.49832	-0.1604072	0.0200114	1.527774	0.0001165	-0.1068070
17.26535	0.62359	17.12093	0.1444235	0.0144136	1.527798	0.0000672	0.0958907
15.56425	0.53223	16.09387	-0.5296162	0.0273223	1.526548	0.0017597	-0.3539670

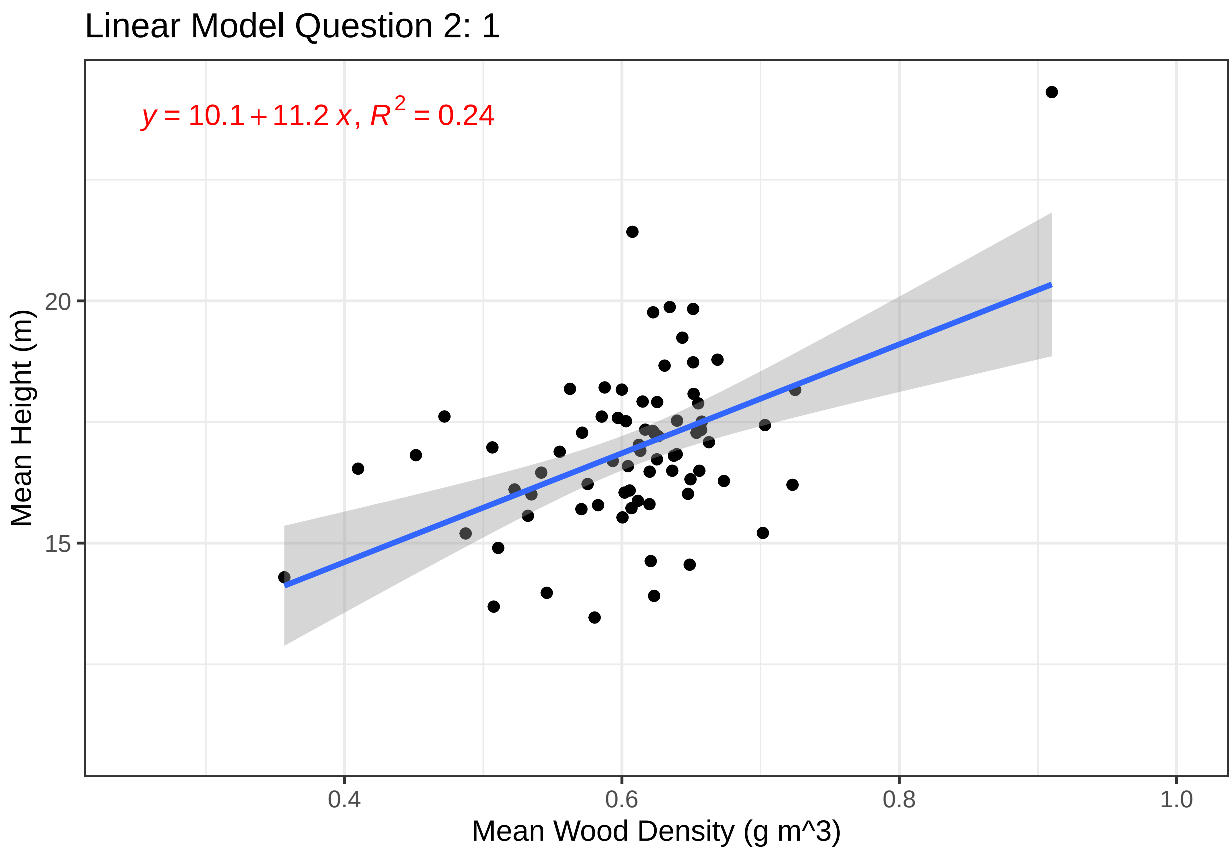
Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
18.08022	0.65172	17.43716	0.6430587	0.0187270	1.525926	0.0017472	0.4278995
17.02831	0.61220	16.99288	0.0354288	0.0137770	1.527891	0.0000039	0.0235155
17.61200	0.58546	16.69227	0.9197273	0.0147966	1.523878	0.0028014	0.6107763
16.10686	0.52260	15.98561	0.1212534	0.0310809	1.527826	0.0001057	0.0811963
16.83497	0.63951	17.29990	-0.4649276	0.0163754	1.526869	0.0007948	-0.3089988
14.55408	0.64891	17.40557	-2.8514915	0.0181207	1.488680	0.0332005	-1.8968331

1.2.2 Durbin Watson Test Linear Model 1 [raw data with outliers] lag Autocorrelation D-W Statistic p-value 1 0.03159474 1.881497 0.604 Alternative hypothesis: rho != 0

1.1.58 Calculating prediction and confidence intervals

1.1.59 Plotting Prediction and Confidence Intervals linear model 1

[1] "Plot" "MEAN Diamater (cm)"
[3] "Mean Basal Area (cm^2)" "Mean Wood Density (g m^3)"
[5] "Mean Height (m)" " Above Ground Biomass(mg ha-1)" [7] "# of Trees"
[1] "Plot" "MEAN Diamater (cm)"
[3] "Mean Basal Area (cm^2)" "Mean Wood Density (g m^3)"
[5] "Mean Height (m)" " Above Ground Biomass(mg ha-1)" [7] "# of Trees"



1.2.29 Shapiro Wilks Test Linear Model 1 [raw data with outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm22 W = 0.97968, p-value = 0.2871

1.2.299 Report Results Simple Linear Model 1 [raw data with outliers] The analysis for Linear Model 1 between Tree Height and Tree Density reveals a positive linear relationship with a correlation coefficient ($r = 0.49$), indicating statistical significance ($p = 1.15e-05$). The linear model equation suggests that tree height increases linearly with density, with a slope of 11.24 and an intercept of 10.11. Confidence intervals and predicted response variables were visualized in Figure 1.1.59. Assumptions for Linear Model 1 were assessed. The residuals vs fitted plot (Figure 1.1.8) displays a non-linear pattern with distinct outliers, suggesting the need for another model. The q-q plot also indicates significant deviations from linearity. The standard residuals vs fitted values plot shows a triangular distribution, implying data wrangling is required. The residuals-leverage plot (Figure 1.1.8) using the Cook's model highlights potential outliers, aiding in their identification and potential removal for model refinement. In Figure 1.1.9, the chiˆ2 test assesses the constancy of error variance across predictor variable levels. A p-value of 0.020 suggests that we fail to reject the null hypothesis of constant error variance, indicating homoscedasticity. Moving to the Durbin-Watson test (Figure 1.2.2), the p-value ($p = 0.03$) indicates no significant autocorrelation in the residuals, supporting the assumption of independence. However, the Shapiro-Wilks test (Figure 1.2.29) suggests non-normality in residuals ($p = 0.2871$), necessitating further data manipulation for normal distribution attainment.

1.2.3 [Simple Linear Model 2] [filtered and excluding outliers] [Below]

1.2.4 Filtering and Excluding Extemporaneous Data [filtered and excluding outliers]

Plot	MEAN Diamater (cm)	Mean Basal Area (cm^2)	Mean Wood Density (g m^3)	Mean Height (m)	Above Gr
1	19.47160	397.4730	0.51084	14.90057	
2	20.88178	471.2643	0.58283	15.78266	
3	24.75175	686.6133	0.65501	17.88486	
4	21.77481	513.7697	0.72304	16.20334	
5	21.97505	517.5035	0.54183	16.45494	
6	26.19477	820.6032	0.59995	18.16854	

1.2.4 Making Simple Liner Model 2 [filtered and excluding outliers] [1] "Plot" "MEAN Diamater (cm)"
[3] "Mean Basal Area (cmˆ2)" "Mean Wood Density (g mˆ3)"
[5] "Mean Height (m)" " Above Ground Biomass(mg ha-1)" [7] "# of Trees"

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.232336	1.584147	7.090463	0.0000000
`Mean Wood Density (g m^3)`	9.095957	2.596932	3.502578	0.0008255

1.2.5 Predicted Response Variable Linear Model 2 [filtered and excluding outliers]

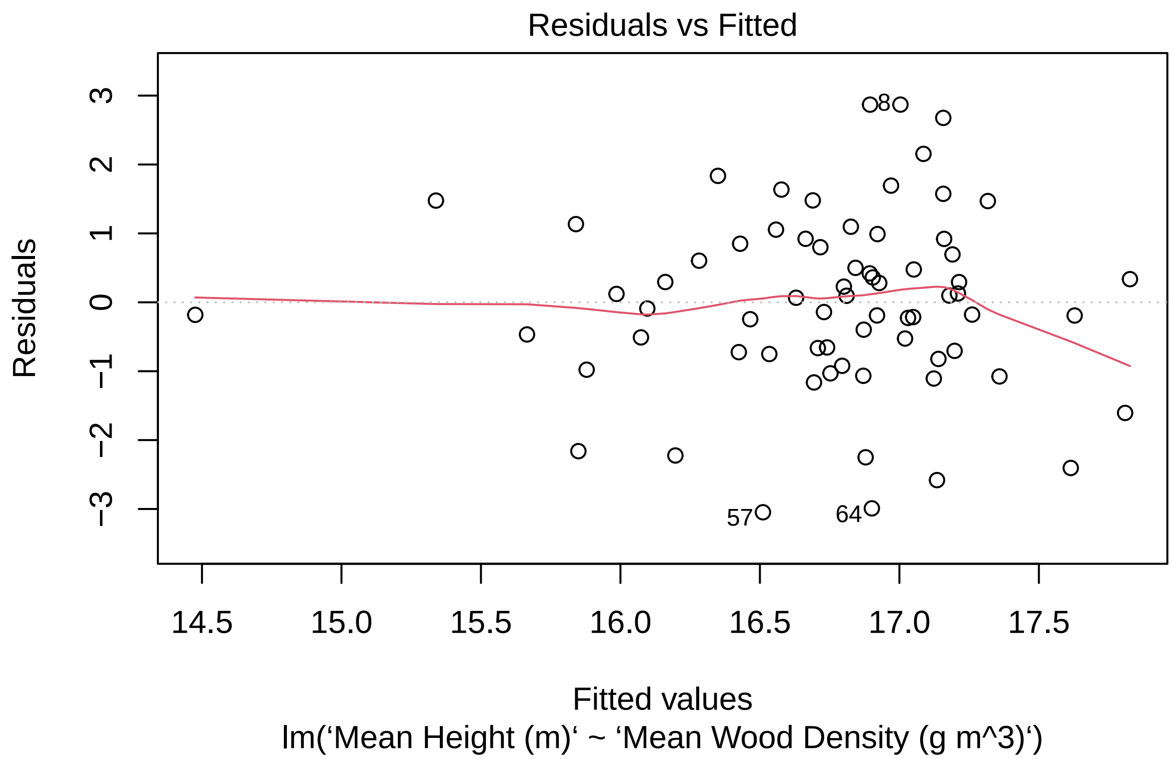
$$\hat{y}_i = 11.292064 + 9.107418\text{Ö}MeanHeight(m)$$

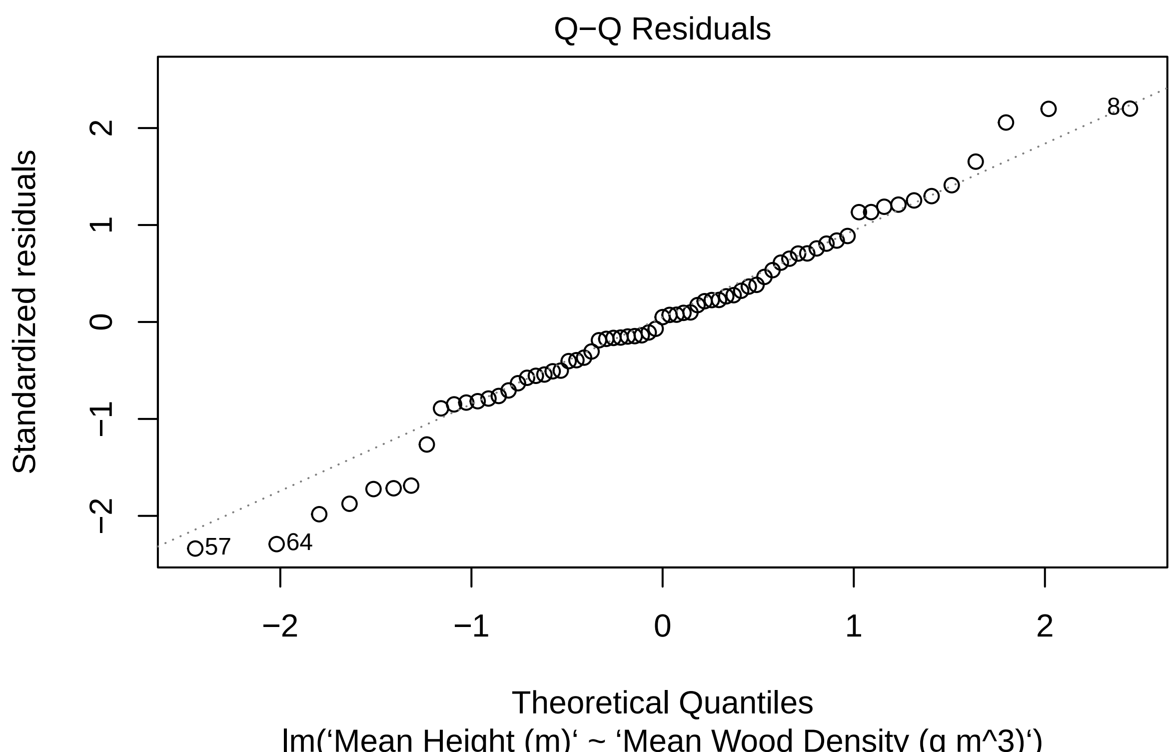
1.2.6 Confidence Intervals Simple Linear Model 2 [filtered and excluding outliers]

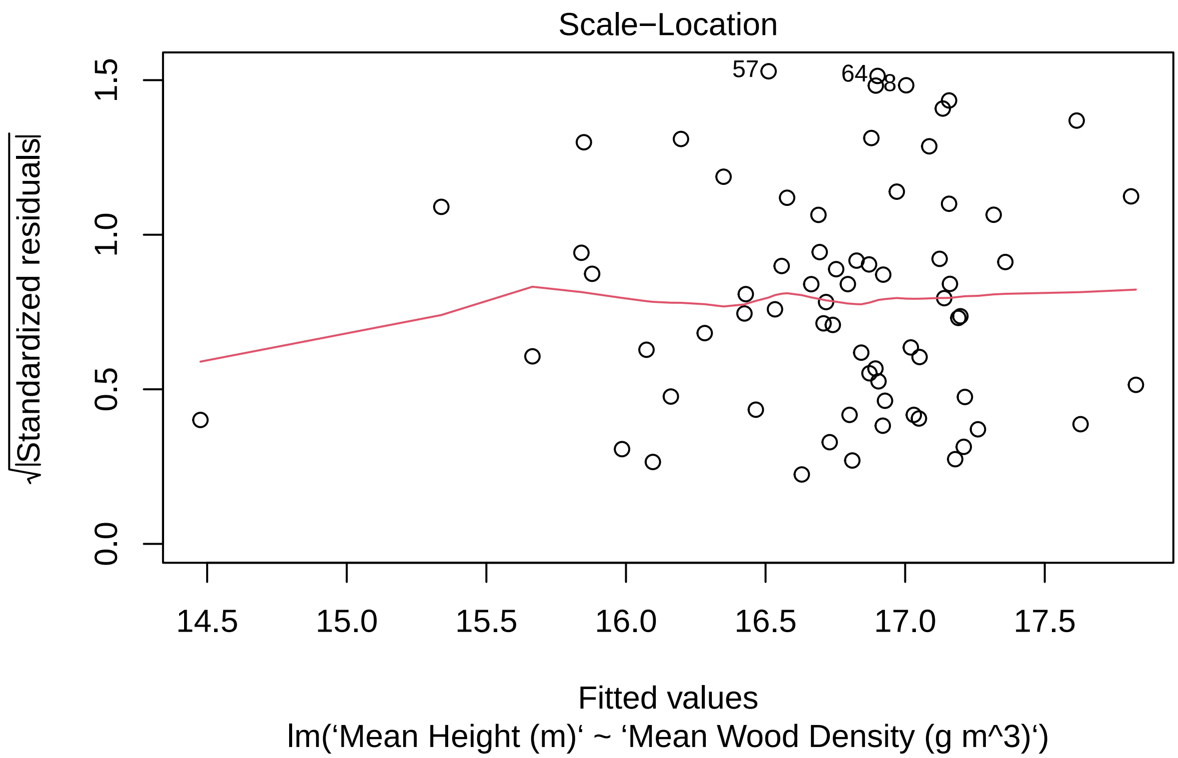
	2.5 %	97.5 %
Mean Height (m)	NA	NA

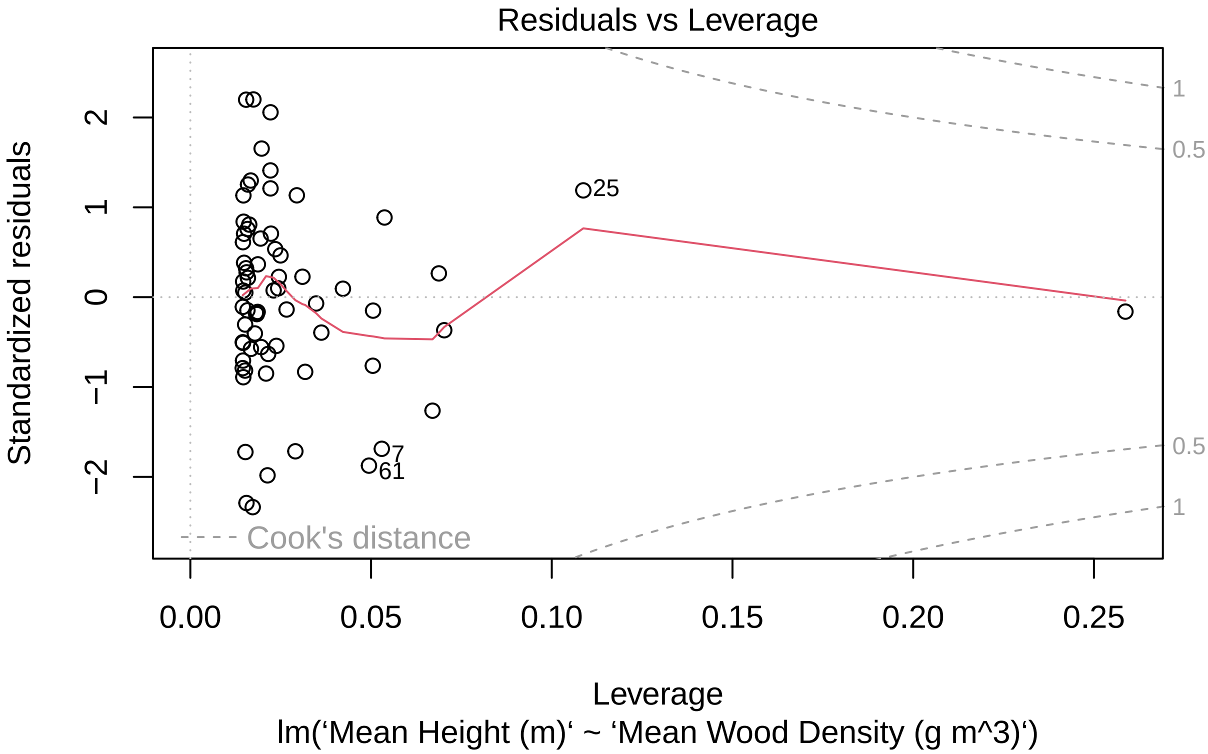
2.2.69 Assumptions of Simple Linear Model 2 [filtered and excluding outliers] [Below]

2.2.7 Plot Simple Linear Model 2 [filtered and excluding outliers]






Scale−Location
57
64
8
1.5
1.0
0.5
0.0
√|Standardized residuals|
14.5
15.0
15.5
16.0
16.5
17.0
17.5
Fitted values
lm(‘Mean Height (m)‘ ~ ‘Mean Wood Density (g m^3)‘)



2.2.8 Non-constant Error Variance Test [filtered and excluding outliers]

Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.5136245, Df = 1, p = 0.47357

```
Breusch-Pagan test
```

data: lm33 BP = 0.51362, df = 1, p-value = 0.4736

```
studentized Breusch-Pagan test
```

data: lm33 BP = 0.49798, df = 1, p-value = 0.4804

2.2.9 Cook's Model Linear Model 2 [filtered and excluding outliers]

.rownames	Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
7	13.68853	0.50758	15.84926	-2.160731	0.0529684	1.297054	0.0796541	-1.687690
25	16.81473	0.45144	15.33861	1.476116	0.1087230	1.311489	0.0861499	1.188470
61	15.20942	0.70163	17.61433	-2.404912	0.0494212	1.290294	0.0913805	-1.874905

.rownames	Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
1	14.90057	0.51084	15.87891	-0.9783442	0.0504854	1.319761	0.0154833	-0.7631589
2	15.78266	0.58283	16.53373	-0.7510721	0.0167597	1.322252	0.0028251	-0.5757394
3	17.88486	0.65501	17.19028	0.6945817	0.0234924	1.322709	0.0034335	0.5342686
4	16.20334	0.72304	17.80908	-1.6057362	0.0670042	1.309646	0.0573336	-1.2635967
5	16.45494	0.54183	16.16080	0.2941421	0.0310165	1.325025	0.0008256	0.2271293
6	18.16854	0.59995	16.68945	1.4790851	0.0146837	1.312784	0.0095585	1.1326076
7	13.68853	0.50758	15.84926	-2.1607314	0.0529684	1.297054	0.0796541	-1.6876901
8	19.87320	0.63450	17.00372	2.8694798	0.0174500	1.276743	0.0429943	2.2003917
9	16.72912	0.62528	16.91986	-0.1907355	0.0158018	1.325324	0.0001714	-0.1461384
10	18.73177	0.65138	17.15726	1.5745101	0.0221842	1.310965	0.0166165	1.2102945
11	14.62893	0.62078	16.87892	-2.2499937	0.0152379	1.295821	0.0229797	-1.7234147
12	18.78610	0.66894	17.31698	1.4691151	0.0294655	1.312764	0.0195039	1.1335077
13	16.80397	0.63749	17.03092	-0.2269471	0.0181268	1.325235	0.0002798	-0.1740889
14	16.04309	0.60195	16.70765	-0.6645568	0.0145902	1.322971	0.0019169	-0.5088594
15	16.49388	0.63632	17.02027	-0.5263948	0.0178537	1.323922	0.0014816	-0.4037362
16	16.69550	0.59338	16.62969	0.0658055	0.0152103	1.325510	0.0000196	0.0504040
17	16.28270	0.67354	17.35883	-1.0761263	0.0317701	1.318682	0.0113372	-0.8312816
18	16.00568	0.53475	16.09640	-0.0907185	0.0348048	1.325487	0.0000888	-0.0701880
19	15.72165	0.60692	16.75285	-1.0312037	0.0144928	1.319354	0.0045839	-0.7895665
20	17.27680	0.65379	17.17918	0.0976188	0.0230413	1.325479	0.0000665	0.0750705
22	16.90569	0.61328	16.81070	0.0949860	0.0146488	1.325483	0.0000393	0.0727341
23	17.51584	0.60294	16.71665	0.7991882	0.0145554	1.321826	0.0027655	0.6119375
24	19.24051	0.64358	17.08631	2.1541985	0.0197206	1.298198	0.0275113	1.6538074
25	16.81473	0.45144	15.33861	1.4761156	0.1087230	1.311489	0.0861499	1.1884698
26	14.29385	0.35659	14.47586	-0.1820129	0.2587244	1.325280	0.0045061	-0.1606891
27	15.80383	0.61987	16.87065	-1.0668164	0.0151431	1.318914	0.0051329	-0.8171038
29	18.21303	0.58759	16.57703	1.6360011	0.0159533	1.309898	0.0127381	1.2535737
30	16.47457	0.62003	16.87210	-0.3975317	0.0151593	1.324618	0.0007135	-0.3044829
31	16.08570	0.60560	16.74085	-0.6551471	0.0144999	1.323044	0.0018512	-0.5016313
32	17.34250	0.61680	16.84272	0.4997782	0.0148707	1.324085	0.0011057	0.3827409
33	19.83373	0.65138	17.15726	2.6764701	0.0221842	1.282982	0.0480145	2.0573492
34	17.52765	0.63976	17.05156	0.4760851	0.0186871	1.324215	0.0012706	0.3653044
35	15.87263	0.61152	16.79470	-0.9220651	0.0145741	1.320595	0.0036862	-0.7060310
36	13.97336	0.54581	16.19700	-2.2236398	0.0290584	1.296103	0.0440284	-1.7153077

.rownames	Mean Height (m)	Mean Wood Density (g m^3)	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
37	17.08241	0.66275	17.26068	-0.1782710	0.0266246	1.325349	0.0002580	-0.1373456
38	17.91138	0.62545	16.92140	0.9899782	0.0158262	1.319832	0.0046260	0.7585145
39	17.20744	0.62615	16.92777	0.2796710	0.0159290	1.325081	0.0003717	0.2142932
40	17.92227	0.61495	16.82589	1.0963758	0.0147421	1.318544	0.0052735	0.8395733
41	18.18454	0.56258	16.34954	1.8350010	0.0221639	1.305706	0.0225479	1.4105141
42	17.31352	0.62239	16.89357	0.4199518	0.0154216	1.324511	0.0008105	0.3216981
43	18.16255	0.72496	17.82654	0.3360096	0.0687555	1.324842	0.0025858	0.2646634
44	15.19823	0.48733	15.66507	-0.4668383	0.0702476	1.324195	0.0051162	-0.3680077
45	17.58526	0.59712	16.66371	0.9215467	0.0148693	1.320599	0.0037589	0.7057397
46	16.97513	0.50661	15.84044	1.1346917	0.0537233	1.317736	0.0223152	0.8866311
47	17.27926	0.57133	16.42913	0.8501314	0.0194367	1.321316	0.0042205	0.6525629
48	17.43608	0.70316	17.62825	-0.1921686	0.0505592	1.325313	0.0005983	-0.1499072
49	16.49305	0.65584	17.19783	-0.7047779	0.0238059	1.322624	0.0035846	-0.5421985
50	18.66346	0.63079	16.96997	1.6934858	0.0167072	1.308760	0.0143159	1.2981183
52	19.76382	0.62253	16.89484	2.8689784	0.0154385	1.276862	0.0378697	2.1977586
53	16.88591	0.55516	16.28205	0.6038630	0.0249441	1.323396	0.0027638	0.4648339
54	16.58835	0.60437	16.72966	-0.1413090	0.0145187	1.325419	0.0000862	-0.1081982
55	16.01698	0.64766	17.12342	-1.1064430	0.0209501	1.318369	0.0077295	-0.8499645
56	16.31768	0.64947	17.13989	-0.8222067	0.0215371	1.321581	0.0043932	-0.6318049
57	13.46335	0.58032	16.51090	-3.0475513	0.0172561	1.270375	0.0479383	-2.3367109
58	15.70088	0.57081	16.42440	-0.7235187	0.0195821	1.322480	0.0030807	-0.5554159
59	16.21957	0.57529	16.46515	-0.2455786	0.0183986	1.325184	0.0003327	-0.1884070
61	15.20942	0.70163	17.61433	-2.4049118	0.0494212	1.290294	0.0913805	-1.8749049
62	15.53043	0.60043	16.69382	-1.1633910	0.0146584	1.317661	0.0059031	-0.8908538
63	17.50733	0.65761	17.21393	0.2934022	0.0244925	1.325031	0.0006401	0.2257991
64	13.91049	0.62324	16.90130	-2.9908097	0.0155266	1.272547	0.0413967	-2.2911892
65	17.33791	0.65716	17.20983	0.1280754	0.0243156	1.325439	0.0001210	0.0985565
66	17.26535	0.62359	16.90448	0.3608667	0.0155715	1.324779	0.0006045	0.2764578
67	15.56425	0.53223	16.07348	-0.5092267	0.0362474	1.323997	0.0029234	-0.3942780
68	18.08022	0.65172	17.16035	0.9198674	0.0223024	1.320580	0.0057031	0.7071265
69	17.02831	0.61220	16.80088	0.2274296	0.0146001	1.325235	0.0002247	0.1741466
70	17.61200	0.58546	16.55765	1.0543455	0.0162923	1.319061	0.0054067	0.8080236
71	16.10686	0.52260	15.98588	0.1209773	0.0422162	1.325448	0.0001946	0.0939603
72	16.83497	0.63951	17.04929	-0.2143209	0.0186234	1.325268	0.0002566	-0.1644451
73	14.55408	0.64891	17.13479	-2.5807129	0.0213528	1.286053	0.0428943	-1.9828998

2.2.10 Durbin Watson Test Linear Model 2 [filtered and excluding outliers]

lag Autocorrelation D-W Statistic p-value 1 0.06438939 1.805536 0.432 Alternative hypothesis: rho != 0

2.2.11 Shapiro Wilks Test Linear Model 2 [filtered and excluding outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm33 W = 0.98231, p-value = 0.4368

2.11.12 Report Results Simple Linear Model 2 [filtered and excluding outliers]

For Linear Model 2, outliers identified in Linear Model 1 were removed, as seen in Figure 1.2.4. Linear Model 2, constructed without outliers, is depicted in Figure 1.2.49, revealing a predicted response variable equation with an intercept of 11.29 and a slope of 9.09 for mean height. Confidence intervals at 97.5% and 2.5% remain consistent.

Plots for Linear Model 2 exhibit a relatively straighter line for residuals vs fitted (Figure 2.2.7), indicating a need for logarithmic transformation due to slight variation. Figure 2 (2.2.7) displays an "S" shape, suggesting logarithmic transformation is required. Similarly, Plot 3 depicts characteristics akin to Plot 1, signifying the need for data transformation. Plot 4 underscores the necessity of data logging.

The Shapiro-Wilks test (2.2.11) demonstrates an increased p-value (p = 0.4368), suggesting a failure to reject the assumption of normal distribution in the data (assuming normality), though further improvement can be achieved through data logging.

The non-constant variance test (2.2.8) yields a high p-value (p = 0.46), indicating insufficient evidence to reject the assumption of constant error variance violation. The Cook's model (Figure 2.2.9) identifies additional residuals, suggesting that the removal of outliers may have contributed to the increased p-value in the Shapiro test. However, despite these adjustments, the data still fails regression assumptions based on linear plots, warranting further data transformation through logging for potential improvement.

2.3 [Simple Linear Model 3] [Log transformed with filtered and excluded outliers]

2.3.1 Log transforming tdat to make simple linear model 3 [Log transformed with filtered and excluded outliers]

2.3.2 Making Linear Model 3 [Log transformed with filtered and excluded outliers]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.9716255	0.0447965	66.336106	0.000000
log_density_question_2_pt1	0.3101906	0.0866692	3.579019	0.000647

2.3.3 Predicted Response Variable Linear Model 3 [Log transformed with filtered and excluded outliers]

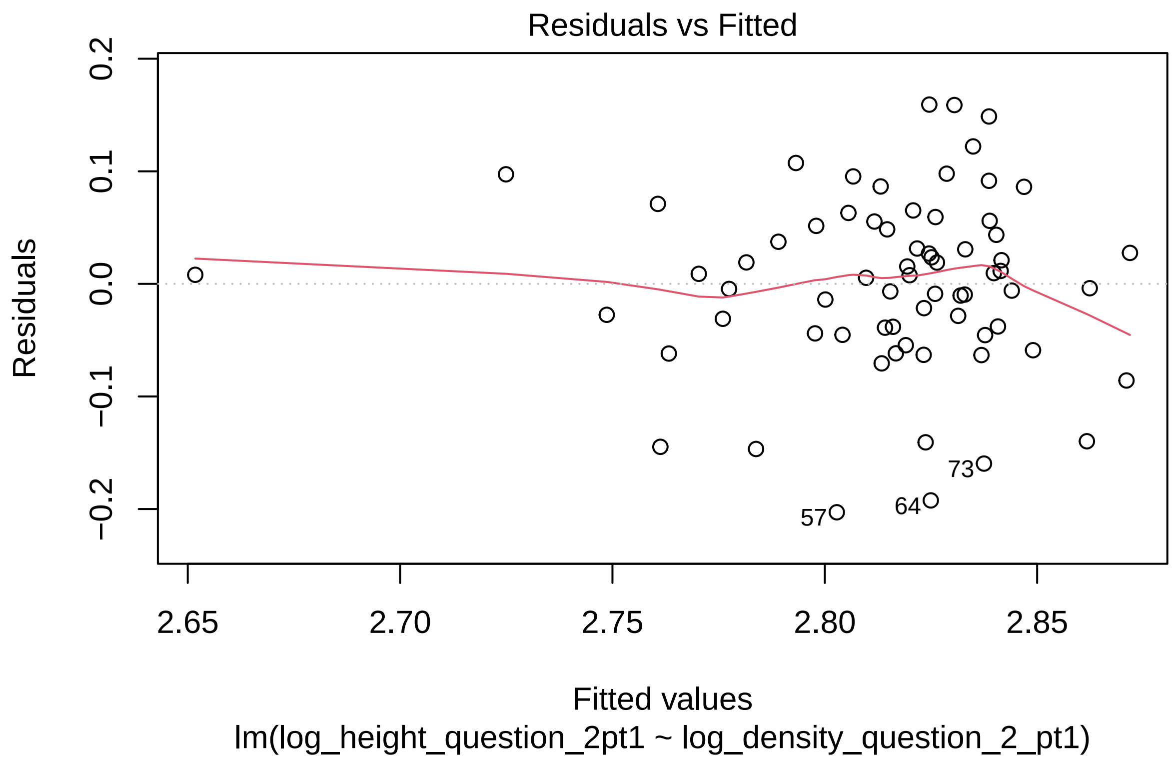
$$\hat{y}_i = 2.9761525 + 0.3121518\text{Ö}MeanHeight(m)$$

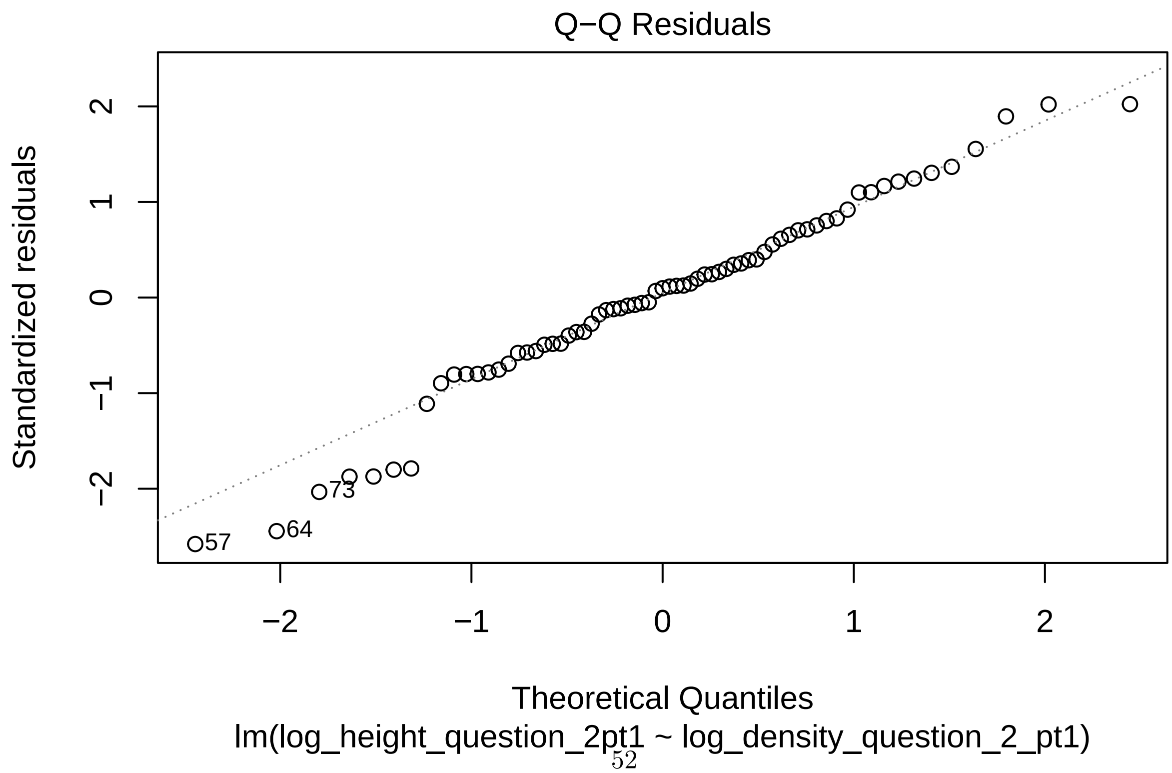
2.3.4 Confidence Intervals Simple Linear Model 3 [Log transformed with filtered and excluded outliers]

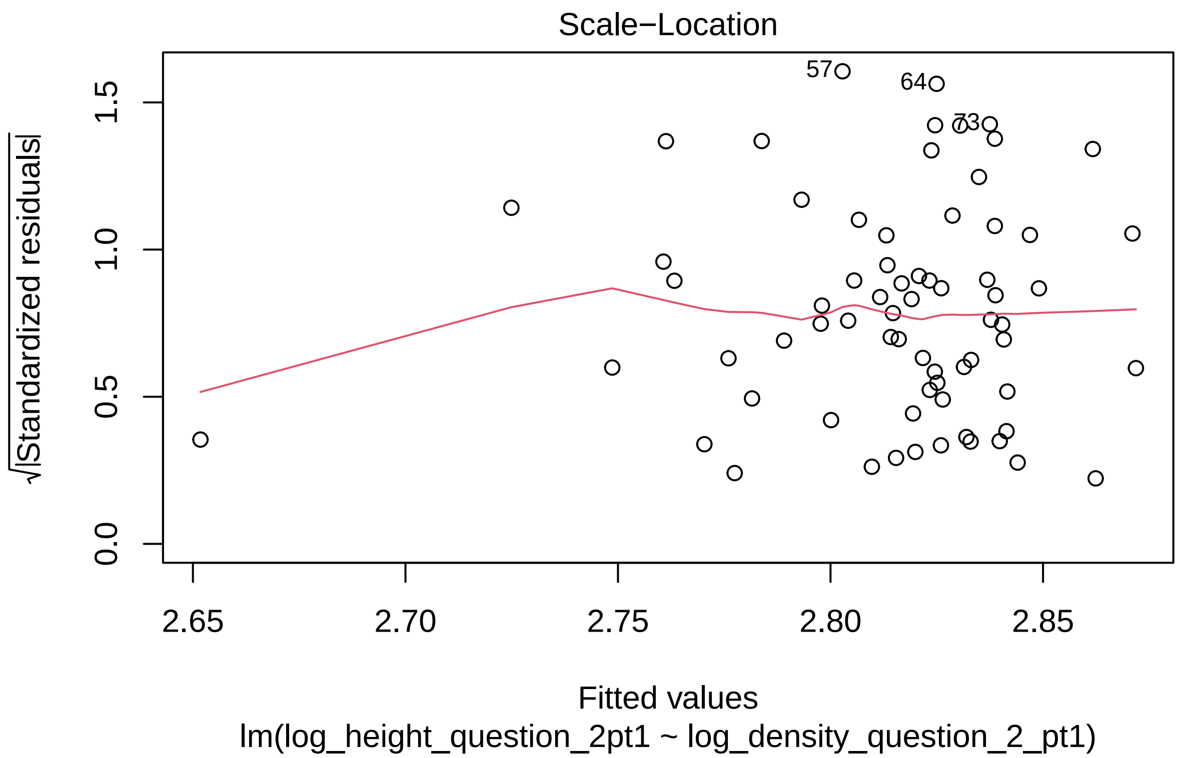
	2.5 %	97.5 %
log_height_question_2pt1	NA	NA

Assumptions of Linear model 3: [Log transformed with filtered and excluded outliers]

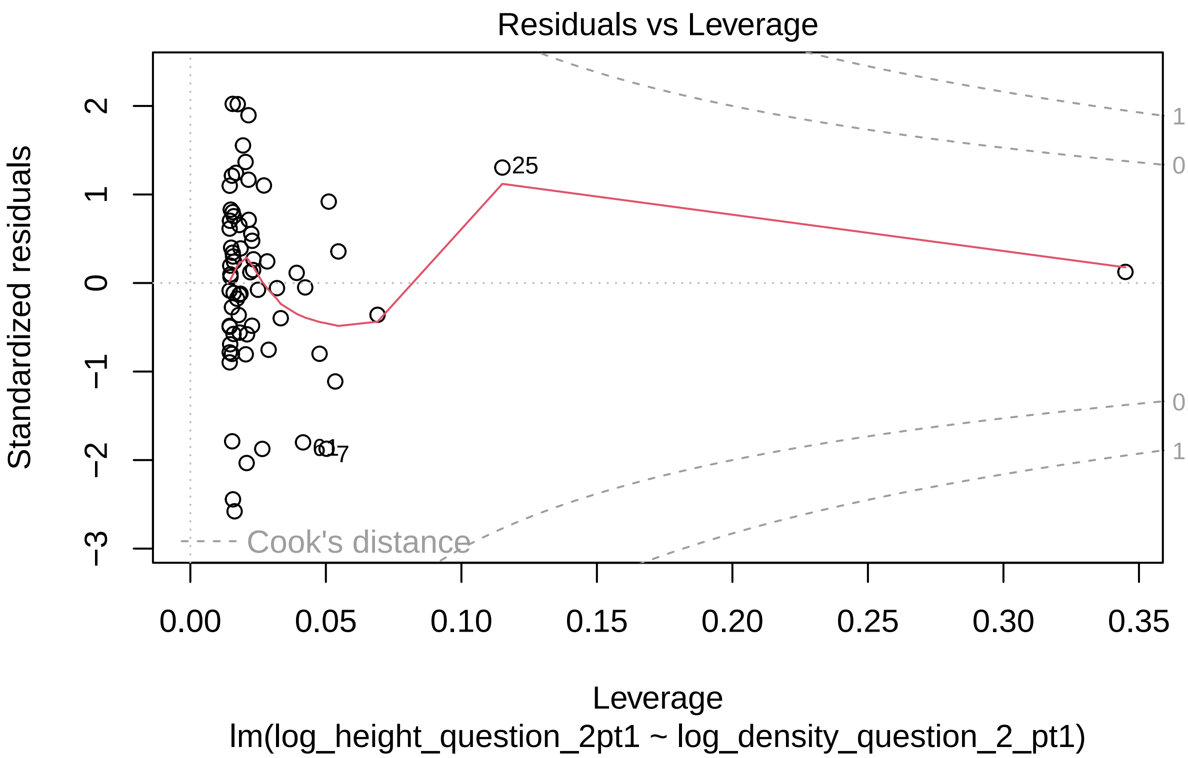
2.3.5 Plot linear model 3 [Log transformed with filtered and excluded outliers]







Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.65
2.70
2.75
2.80
2.85
57
64
73
Fitted values
lm(log_height_question_2pt1 ~ log_density_question_2_pt1)



2.3.6 Non-constant Error Variance Test Linear Model 3 [Log transformed with filtered and excluded outliers]

Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 0.1193919, Df = 1, p = 0.72969

```
Breusch-Pagan test
```

data: lm44 BP = 0.11939, df = 1, p-value = 0.7297

```
studentized Breusch-Pagan test
```

data: lm44 BP = 0.10837, df = 1, p-value = 0.742

2.3.7 Cook's Model: Identifying Points of Leverage Linear Model 3 [Log transformed with filtered and excluded outliers]

.rownames	log_height_question_2pt1	log_density_question_2_pt1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
7	2.616558	-0.6781009	2.761285	-0.1447267	0.0502685	0.0778036	0.0927592	-1.872174
25	2.822255	-0.7953128	2.724927	0.0973283	0.1151154	0.0789010	0.1106636	1.304349

.rownames	log_height_question_2_pt1	log_density_question_2_pt1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
61	2.721915	-0.3543491	2.861710	-0.1397948	0.0415819	0.0779655	0.0702978	-1.800161

.rownames	log_height_question_2_pt1	log_density_question_2_pt1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
1	2.701399	-0.6716988	2.763271	-0.0618714	0.0476713	0.0795403	0.0159893	-0.7992715
2	2.758912	-0.5398597	2.804166	-0.0452542	0.0159445	0.0797247	0.0026795	-0.5751053
3	2.883955	-0.4231048	2.840382	0.0435722	0.0224967	0.0797379	0.0035520	0.5555817
4	2.785217	-0.3242907	2.871034	-0.0858162	0.0534712	0.0791813	0.0349266	-1.1119883
5	2.800626	-0.6128030	2.781540	0.0190859	0.0283696	0.0798867	0.0008698	0.2440962
6	2.899691	-0.5109090	2.813146	0.0865452	0.0145346	0.0791985	0.0089078	1.0990567
7	2.616558	-0.6781009	2.761285	-0.1447267	0.0502685	0.0778036	0.0927592	-1.8721742
8	2.989372	-0.4549180	2.830514	0.1588579	0.0174855	0.0774493	0.0363231	2.0203992
9	2.817151	-0.4695557	2.825974	-0.0088228	0.0159914	0.0799147	0.0001022	-0.1121260
10	2.930221	-0.4286621	2.838659	0.0915625	0.0214472	0.0791059	0.0149211	1.1668720
11	2.683001	-0.4767785	2.823733	-0.1407322	0.0154427	0.0779921	0.0250723	-1.7880143
12	2.933117	-0.4020609	2.846910	0.0862072	0.0271395	0.0791948	0.0169338	1.1018346
13	2.821615	-0.4502167	2.831973	-0.0103574	0.0180739	0.0799119	0.0001598	-0.1317673
14	2.775278	-0.5075809	2.814179	-0.0389005	0.0145008	0.0797765	0.0017954	-0.4939973
15	2.802989	-0.4520537	2.831403	-0.0284133	0.0178377	0.0798443	0.0011863	-0.3614332
16	2.815139	-0.5219203	2.809731	0.0054085	0.0148350	0.0799194	0.0000355	0.0686939
17	2.790103	-0.3952079	2.849036	-0.0589325	0.0288796	0.0795825	0.0084513	-0.7539050
18	2.772944	-0.6259559	2.777460	-0.0045162	0.0319619	0.0799202	0.0000553	-0.0578663
19	2.755039	-0.4993583	2.816729	-0.0616905	0.0145306	0.0795553	0.0045248	-0.7834203
20	2.849365	-0.4249691	2.839804	0.0095605	0.0221364	0.0799133	0.0001681	0.1218815
22	2.827650	-0.4889337	2.819963	0.0076874	0.0148004	0.0799165	0.0000716	0.0976368
23	2.863106	-0.5059376	2.814688	0.0484172	0.0144938	0.0796964	0.0027799	0.6148481
24	2.957018	-0.4407089	2.834922	0.1220962	0.0194251	0.0784679	0.0239316	1.5543892
25	2.822255	-0.7953128	2.724927	0.0973283	0.1151154	0.0789010	0.1106636	1.3043493
26	2.659829	-1.0311686	2.651767	0.0080627	0.3450116	0.0799128	0.0041542	0.1255913
27	2.760252	-0.4782455	2.823278	-0.0630259	0.0153465	0.0795389	0.0049963	-0.8007105
29	2.902137	-0.5317259	2.806689	0.0954481	0.0153463	0.0790403	0.0114587	1.2126168
30	2.801818	-0.4779874	2.823358	-0.0215403	0.0153630	0.0798775	0.0005842	-0.2736606
31	2.777931	-0.5015356	2.816054	-0.0381232	0.0145070	0.0797823	0.0017251	-0.4841285
32	2.853160	-0.4832105	2.821738	0.0314220	0.0150589	0.0798271	0.0012179	0.3991408
33	2.987384	-0.4286621	2.838659	0.1487255	0.0214472	0.0777501	0.0393675	1.8953576

.rownames	log_height_question_2pt1	log_density_question_2pt1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
34	2.863780	-0.4466622	2.833075	0.0307045	0.0185538	0.0798311	0.0014430	0.3907214
35	2.764596	-0.4918076	2.819071	-0.0544752	0.0147001	0.0796362	0.0035706	-0.6918506
36	2.637153	-0.6054843	2.783810	-0.1466573	0.0265496	0.0777997	0.0478854	-1.8738927
37	2.838049	-0.4113574	2.844026	-0.0059770	0.0249581	0.0799187	0.0000745	-0.0763082
38	2.885436	-0.4692839	2.826058	0.0593782	0.0160145	0.0795818	0.0046340	0.7546241
39	2.845342	-0.4681653	2.826405	0.0189368	0.0161113	0.0798877	0.0004743	0.2406756
40	2.886044	-0.4862143	2.820806	0.0652377	0.0149135	0.0795116	0.0051975	0.8286273
41	2.900572	-0.5752219	2.793197	0.1073747	0.0203816	0.0787987	0.0194578	1.3676390
42	2.851488	-0.4741884	2.824537	0.0269510	0.0156251	0.0798522	0.0009307	0.3424459
43	2.899362	-0.3216388	2.871856	0.0275056	0.0546237	0.0798463	0.0036743	0.3566291
44	2.721179	-0.7188138	2.748656	-0.0274773	0.0690746	0.0798453	0.0047819	-0.3590162
45	2.867061	-0.5156372	2.811680	0.0553813	0.0146282	0.0796266	0.0036718	0.7033336
46	2.831749	-0.6800138	2.760692	0.0710577	0.0510634	0.0794162	0.0227522	0.9195822
47	2.849507	-0.5597883	2.797984	0.0515225	0.0180779	0.0796655	0.0039551	0.6554753
48	2.858542	-0.3521708	2.862385	-0.0038438	0.0423710	0.0799207	0.0000542	-0.0495178
49	2.802939	-0.4218384	2.840775	-0.0378361	0.0227462	0.0797832	0.0027094	-0.4825039
50	2.926568	-0.4607823	2.828695	0.0978724	0.0168255	0.0789933	0.0132493	1.2443512
52	2.983853	-0.4739635	2.824606	0.1592465	0.0156417	0.0774417	0.0325300	2.0234442
53	2.826480	-0.5884989	2.789079	0.0374009	0.0228184	0.0797864	0.0026562	0.4769710
54	2.808701	-0.5035687	2.815423	-0.0067226	0.0144952	0.0799179	0.0000536	-0.0853701
55	2.773649	-0.4343894	2.836882	-0.0632326	0.0204426	0.0795344	0.0067690	-0.8054229
56	2.792249	-0.4315986	2.837748	-0.0454985	0.0209223	0.0797215	0.0035903	-0.5796774
57	2.599971	-0.5441756	2.802827	-0.2028562	0.0163261	0.0758533	0.0551724	-2.5784613
58	2.753717	-0.5606989	2.797702	-0.0439852	0.0181980	0.0797352	0.0029024	-0.5596196
59	2.786218	-0.5528810	2.800127	-0.0139085	0.0172311	0.0799035	0.0002742	-0.1768691
61	2.721915	-0.3543491	2.861710	-0.1397948	0.0415819	0.0779655	0.0702978	-1.8001613
62	2.742801	-0.5101092	2.813394	-0.0705931	0.0145241	0.0794414	0.0059223	-0.8964730
63	2.862620	-0.4191432	2.841611	0.0210084	0.0232900	0.0798794	0.0008562	0.2679839
64	2.632643	-0.4728236	2.824960	-0.1923168	0.0157277	0.0762772	0.0477129	-2.4437549
65	2.852895	-0.4198278	2.841399	0.0114965	0.0231502	0.0799094	0.0002548	0.1466397
66	2.848702	-0.4722622	2.825134	0.0235674	0.0157712	0.0798687	0.0007186	0.2994754
67	2.744977	-0.6306796	2.775995	-0.0310180	0.0333528	0.0798278	0.0027289	-0.3977207
68	2.894818	-0.4281403	2.838820	0.0559981	0.0215426	0.0796178	0.0056069	0.7136746
69	2.834877	-0.4906963	2.819416	0.0154611	0.0147366	0.0798992	0.0002884	0.1963642
70	2.868581	-0.5353574	2.805563	0.0630178	0.0155938	0.0795389	0.0050780	0.8007079
71	2.779245	-0.6489389	2.770331	0.0089145	0.0392304	0.0799144	0.0002684	0.1146531

.rownames	log_height_question_2_1	log_density_question_2_1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
72	2.823458	-0.4470530	2.832954	-0.0094956	0.0184995	0.0799135	0.0001376	-0.1208300
73	2.677871	-0.4324612	2.837480	-0.1596087	0.0207721	0.0774170	0.0438522	-2.0333526

2.3.8 Durbin Watson Test Linear Model 3 [Log transformed with filtered and excluded outliers]

lag Autocorrelation D-W Statistic p-value 1 0.05749272 1.815507 0.434 Alternative hypothesis: rho != 0

2.3.9 Shapiro Wilks Test Linear Model 3 [Log transformed with filtered and excluded outliers]

```
Shapiro-Wilk normality test
```

data: residuals_lm44 W = 0.97468, p-value = 0.174

2.9.99 Report Results Linear Model 3: In Linear Model 3, the dataset was logged to remove outliers. Figure 2.3.2 illustrates the construction of Linear Model 3, revealing a predicted response variable equation with an intercept of 2.97 and a slope of 0.31 for mean height, indicating normalized data. Confidence intervals remain consistent at 97.5% and 2.5%. Plots for Linear Model 3 depict significantly straighter lines for residuals vs fitted (2.3.5), indicating successful logarithmic transformation. Plot 2 (Figure 2.3.5) demonstrates a pronounced straight shape fitting to the line, while Plot 3 resembles the pattern observed in Plot 1. Plot 4 showcases a considerably straighter line, indicating positive progress. The Shapiro-Wilks test (2.3.9) yields an increased p-value ($p = 0.174$), suggesting a failure to reject the assumption of normal distribution, though further improvement is sought. The non-constant variance test (2.3.6) maintains a high p-value ($p = 0.14$), indicating insufficient evidence to reject the assumption of constant error variance violation. The Cook's model (Figure 2.3.7) identifies additional residuals, indicating potential improvements through further exploration of logged data with outliers. Further testing is warranted to ascertain the impact of logged data with outliers on model performance.

3.0 [Simple Linear Model 4] [Logged comparison with outliers for comparisson]

3.1.1 log transforming original tdat data frame [Logged comparison with outliers for comparisson]

3.1.2 Making simple linear model 4 [Logged comparison with outliers for comparisson]

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.9947200	0.0424153	70.604754	0.00e+00
log_density_originaltdat_question_2_1	0.3357901	0.0809328	4.149002	9.14e-05

3.3.4 Predicted Response Variable Linear Model 4 [Logged comparison with outliers for comparisson]

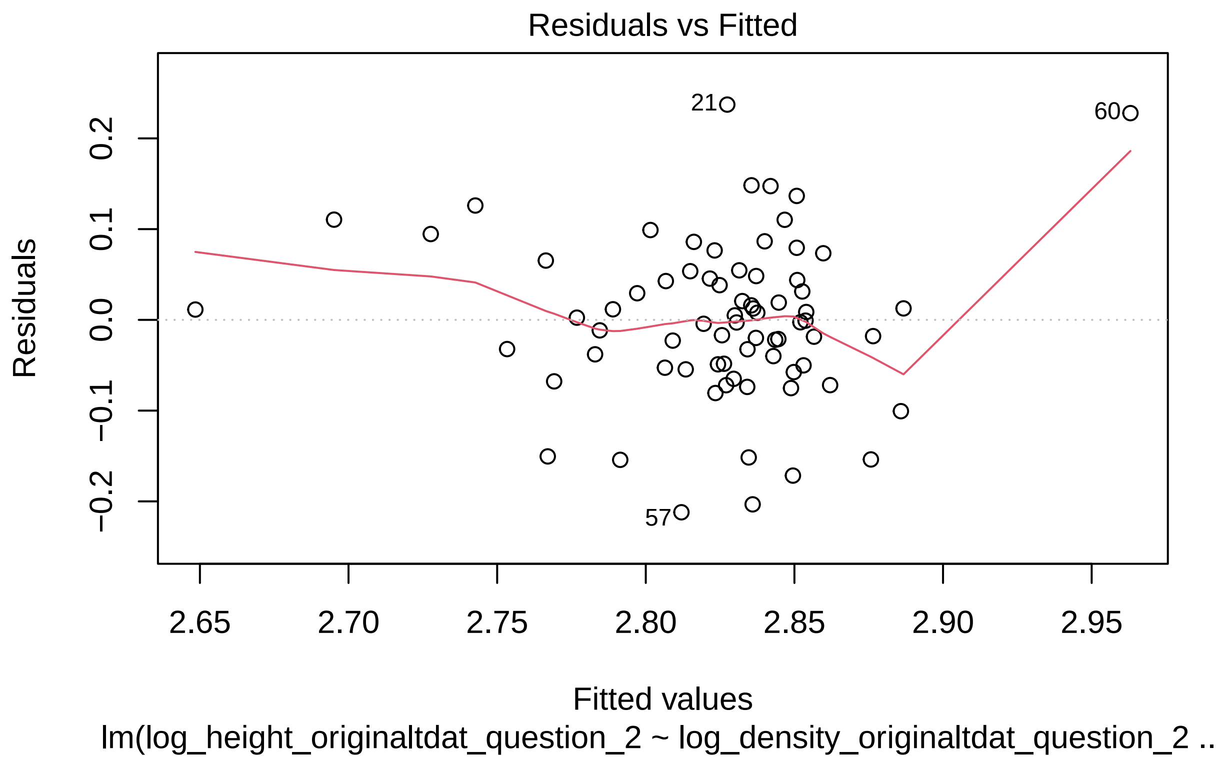
$$\hat{y_i} = 2.9947200 - 0.3357901Ö MeanHeight(m)$$

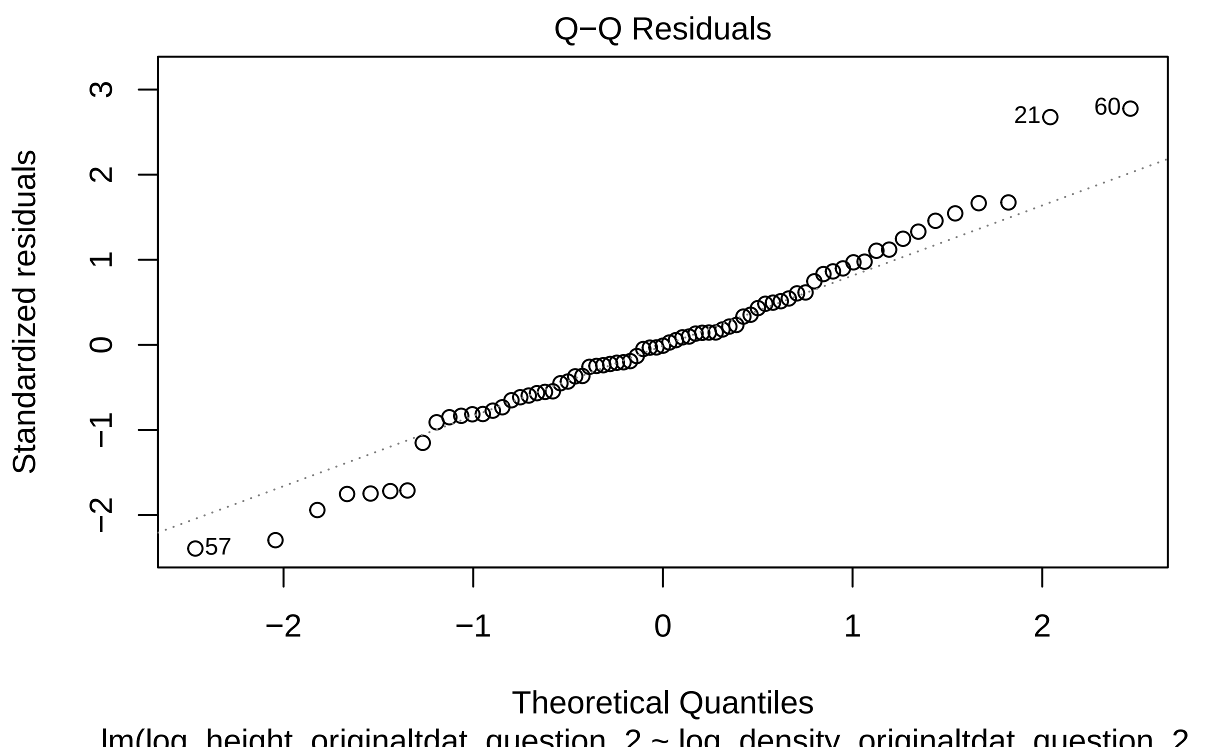
3.3.4 Confidence Intervals Simple Linear Model 4 [Logged comparison with outliers for comparisson]

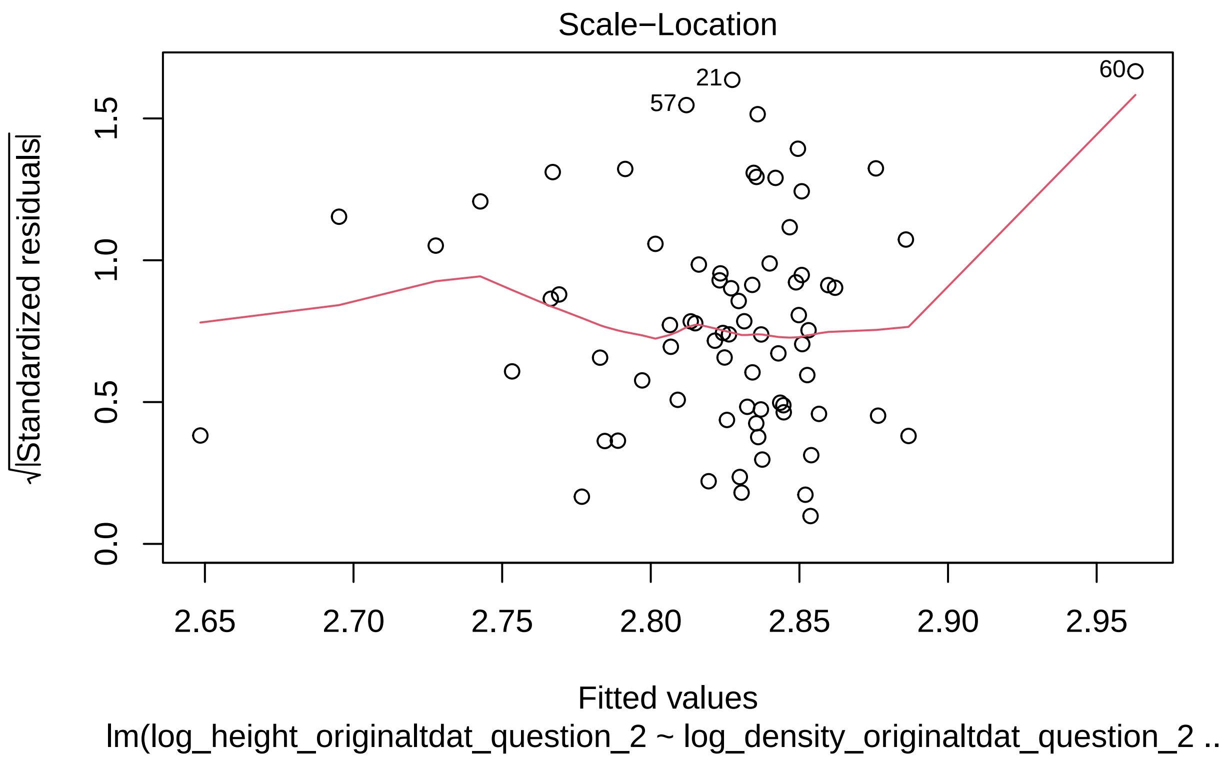
	2.5 %	97.5 %
Mean height (m)	NA	NA

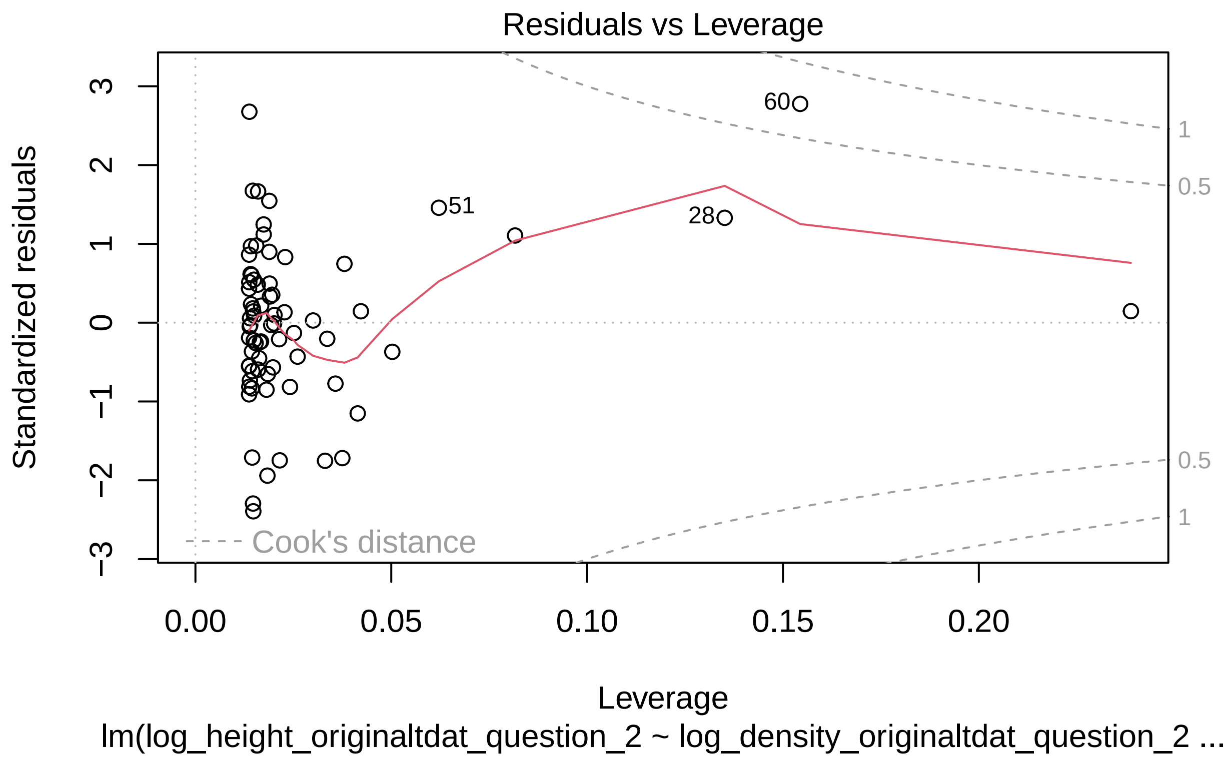
Assumptions of Linear Model 4 [Logged comparison with outliers for comparisson]

3.3.5 Plot Linear Model 4 [Logged comparison with outliers for comparisson]






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.65
2.70
2.75
2.80
2.85
2.90
2.95
21
57
60
Fitted values
lm(log_height_originaltdat_question_2 ~ log_density_originaltdat_question_2 ...



3.3.6 Non-constant Error Variance Test Linear Model 4 [Logged comparison with outliers for comparisson] Non-constant Variance Score Test Variance formula: ~ fitted.values Chisquare = 2.158654, Df = 1, p = 0.14177

```
Breusch-Pagan test
```

data: lm55_with_outliers BP = 2.1587, df = 1, p-value = 0.1418

```
studentized Breusch-Pagan test
```

data: lm55_with_outliers BP = 1.7358, df = 1, p-value = 0.1877

3.3.7 Cook's Model: Identifying Points of Leverage [Logged comparison with outliers for comparisson]

log_height_originaltdat_question_2	log_density_originaltdat_question_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
2.805560	-0.8922081	2.695125	0.1104345	0.1351278	0.0887554	0.1383007	1.330548
2.868667	-0.7506704	2.742652	0.1260151	0.0621498	0.0885273	0.0704355	1.458001
3.190893	-0.0943107	2.963051	0.2278417	0.1543921	0.0848644	0.7036027	2.776200

log_height_originaltdat_questions_2	log_density_originaltdat_question_2_1	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
2.701399	-0.6716988	2.769170	-0.0677707	0.0357516	0.0895038	0.0110860	-0.7733025
2.758912	-0.5398597	2.813440	-0.0545286	0.0145365	0.0896430	0.0027938	-0.6154684
2.883955	-0.4231048	2.852646	0.0313090	0.0196169	0.0898036	0.0012559	0.3543013
2.785217	-0.3242907	2.885826	-0.1006090	0.0414336	0.0890400	0.0286521	-1.1514039
2.800626	-0.6128030	2.788947	0.0116789	0.0227414	0.0898720	0.0002039	0.1323734
2.899691	-0.5109090	2.823162	0.0765297	0.0137059	0.0894100	0.0051800	0.8634337
2.616558	-0.6781009	2.767020	-0.1504621	0.0375096	0.0879941	0.0575411	-1.7184269
2.989372	-0.4549180	2.841963	0.1474091	0.0160104	0.0881107	0.0225551	1.6650644
2.817151	-0.4695557	2.837048	-0.0198969	0.0149102	0.0898512	0.0003818	-0.2246206
2.930221	-0.4286621	2.850779	0.0794415	0.0188669	0.0893705	0.0077645	0.8986398
2.683001	-0.4767785	2.834623	-0.1516214	0.0144971	0.0880098	0.0215408	-1.7113294
2.933117	-0.4020609	2.859712	0.0734053	0.0229173	0.0894438	0.0081195	0.8320779
2.821615	-0.4502167	2.843542	-0.0219265	0.0164386	0.0898443	0.0005128	-0.2477255
2.775278	-0.5075809	2.824279	-0.0490011	0.0136987	0.0896894	0.0021225	-0.5528448
2.802989	-0.4520537	2.842925	-0.0399354	0.0162670	0.0897542	0.0016828	-0.4511508
2.815139	-0.5219203	2.819464	-0.0043251	0.0138594	0.0898816	0.0000167	-0.0488009
2.790103	-0.3952079	2.862013	-0.0719099	0.0241492	0.0894610	0.0082317	-0.8156409
2.772944	-0.6259559	2.784530	-0.0115865	0.0251522	0.0898722	0.0002230	-0.1314880
2.755039	-0.4993583	2.827040	-0.0720017	0.0137592	0.0894644	0.0046035	-0.8123686
2.849365	-0.4249691	2.852020	-0.0026550	0.0193597	0.0898825	0.0000089	-0.0300409
3.064646	-0.4982550	2.827411	0.2372348	0.0137758	0.0852276	0.0500374	2.6766570
2.827650	-0.4889337	2.830541	-0.0028906	0.0139957	0.0898824	0.0000076	-0.0326179
2.863106	-0.5059376	2.824831	0.0382745	0.0137019	0.0897650	0.0012953	0.4318244
2.957018	-0.4407089	2.846734	0.1102837	0.0174155	0.0888940	0.0137719	1.2466035
2.822255	-0.7953128	2.727662	0.0945935	0.0816106	0.0891055	0.0543488	1.1059867
2.659829	-1.0311686	2.648464	0.0113656	0.2388306	0.0898696	0.0033426	0.1459671
2.760252	-0.4782455	2.834130	-0.0738776	0.0144237	0.0894419	0.0050874	-0.8338146
2.805560	-0.8922081	2.695125	0.1104345	0.1351278	0.0887554	0.1383007	1.3305485
2.902137	-0.5317259	2.816172	0.0859656	0.0141639	0.0892854	0.0067608	0.9701174
2.801818	-0.4779874	2.834217	-0.0323986	0.0144364	0.0897984	0.0009793	-0.3656667
2.777931	-0.5015356	2.826309	-0.0483786	0.0137324	0.0896943	0.0020741	-0.5458309
2.853160	-0.4832105	2.832463	0.0206974	0.0142015	0.0898486	0.0003930	0.2335741
2.987384	-0.4286621	2.850779	0.1366045	0.0188669	0.0883587	0.0229588	1.5452659
2.863780	-0.4466622	2.844735	0.0190444	0.0167864	0.0898538	0.0003953	0.2152014

log_height_originaltdat_question_2	log_density_originaltdat_question_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
2.764596	-0.4918076	2.829576	-0.0649796	0.0139126	0.0895422	0.0037923	-0.7331983
2.637153	-0.6054843	2.791404	-0.1542517	0.0215233	0.0879294	0.0335769	-1.7472565
2.838049	-0.4113574	2.856590	-0.0185409	0.0213695	0.0898552	0.0004815	-0.2100026
2.885436	-0.4692839	2.837139	0.0482972	0.0149274	0.0896947	0.0022525	0.5452424
2.845342	-0.4681653	2.837515	0.0078271	0.0149995	0.0898782	0.0000595	0.0883665
2.886044	-0.4862143	2.831454	0.0545900	0.0140868	0.0896426	0.0027110	0.6160220
2.900572	-0.5752219	2.801566	0.0990056	0.0174213	0.0890868	0.0111030	1.1191243
2.851488	-0.4741884	2.835492	0.0159955	0.0146354	0.0898625	0.0002421	0.1805514
2.899362	-0.3216388	2.886717	0.0126449	0.0422404	0.0898698	0.0004622	0.1447738
2.721179	-0.7188138	2.753349	-0.0321704	0.0502666	0.0897965	0.0036205	-0.3698778
2.867061	-0.5156372	2.821574	0.0454869	0.0137474	0.0897162	0.0018357	0.5132095
2.831749	-0.6800138	2.766378	0.0653713	0.0380479	0.0895294	0.0110299	0.7468136
2.849507	-0.5597883	2.806749	0.0427583	0.0159094	0.0897353	0.0018854	0.4829535
2.858542	-0.3521708	2.876464	-0.0179229	0.0336518	0.0898567	0.0007267	-0.2042882
2.802939	-0.4218384	2.853071	-0.0501317	0.0197949	0.0896791	0.0032503	-0.5673569
2.926568	-0.4607823	2.839994	0.0865737	0.0155273	0.0892760	0.0075377	0.9776569
2.868667	-0.7506704	2.742652	0.1260151	0.0621498	0.0885273	0.0704355	1.4580009
2.983853	-0.4739635	2.835568	0.1482853	0.0146479	0.0880919	0.0208239	1.6738029
2.826480	-0.5884989	2.797108	0.0293717	0.0190355	0.0898132	0.0010712	0.3322800
2.808701	-0.5035687	2.825627	-0.0169260	0.0137143	0.0898600	0.0002535	-0.1909651
2.773649	-0.4343894	2.848856	-0.0752069	0.0181472	0.0894242	0.0066835	-0.8504261
2.792249	-0.4315986	2.849793	-0.0575443	0.0184911	0.0896146	0.0039898	-0.6508140
2.599971	-0.5441756	2.811991	-0.2120200	0.0147784	0.0861810	0.0429624	-2.3933818
2.753717	-0.5606989	2.806443	-0.0527261	0.0159877	0.0896583	0.0028814	-0.5955625
2.786218	-0.5528810	2.809068	-0.0228495	0.0153596	0.0898410	0.0005192	-0.2580116
3.190893	-0.0943107	2.963051	0.2278417	0.1543921	0.0848644	0.7036027	2.7762004
2.721915	-0.3543491	2.875733	-0.1538181	0.0330976	0.0879170	0.0525800	-1.7527428
2.742801	-0.5101092	2.823430	-0.0806290	0.0137025	0.0893578	0.0057483	-0.9096819
2.862620	-0.4191432	2.853976	0.0086438	0.0201826	0.0898770	0.0000986	0.0978442
2.632643	-0.4728236	2.835951	-0.2033073	0.0147127	0.0864851	0.0393229	-2.2949515
2.852895	-0.4198278	2.853746	-0.0008505	0.0200830	0.0898830	0.0000009	-0.0096273
2.848702	-0.4722622	2.836139	0.0125626	0.0147453	0.0898704	0.0001505	0.1418100
2.744977	-0.6306796	2.782944	-0.0379674	0.0260874	0.0897654	0.0024888	-0.4310751
2.894818	-0.4281403	2.850955	0.0438638	0.0189352	0.0897271	0.0023761	0.4962031
2.834877	-0.4906963	2.829949	0.0049282	0.0139431	0.0898812	0.0000219	0.0556085

log_height_originaltdat_question_2	log_density_originaltdat_question_2	.fitted	.resid	.hat	.sigma	.cooksd	.std.resid
2.868581	-0.5353574	2.814952	0.0536282	0.0143168	0.0896509	0.0026603	0.6052390
2.779245	-0.6489389	2.776813	0.0024326	0.0300476	0.0898826	0.0000119	0.0276752
2.823458	-0.4470530	2.844604	-0.0211457	0.0167472	0.0898470	0.0004862	-0.2389418
2.677871	-0.4324612	2.849504	-0.1716324	0.0183835	0.0874658	0.0352790	-1.9410219

3.3.8 Durbin Watson Test Linear Model 4 [Logged comparison with outliers for comparisson] lag Autocorrelation D-W Statistic p-value 1 0.02285574 1.894078 0.692 Alternative hypothesis: rho != 0 #### 3.3.9 Shapiro Wilks Test Linear Model 4 [Logged comparison with outliers for comparisson]

```
Shapiro-Wilk normality test
```

data: residuals_lm55_with_outliers W = 0.98157, p-value = 0.3636

3.3.11 Report Results Simple Linear Model 4 [Logged comparison with outliers for comparisson] In Linear Model 4, the dataset was logged for comparison with outliers. Figure 3.1.2 illustrates the construction of Linear Model 4, revealing a predicted response variable equation with an intercept of 2.9 and a slope of 0.335 for mean height, indicative of normalized data. Confidence intervals remain consistent at 97.5% and 2.5%. Plots for Linear Model 4 depict a straighter line compared to the unlogged data, albeit still exhibiting some distortion due to outliers (3.3.5). Plot 2 (Figure 3.3.5) shows a slightly improved shape, suggesting that logarithmic transformation is effective, yet outliers continue to disrupt data integrity. Plot 3 mirrors the findings of Plot 1, while Plot 4 showcases a notably straighter line. The Shapiro-Wilks test (2.3.9) yields an increased p-value (p = 0.34), indicating a failure to reject the assumption of normal distribution. However, the lower p-value compared to the logged data without outliers suggests that outliers significantly impact linear correlation. The non-constant variance test (3.3.6) maintains a high p-value (p = 0.84), indicating insufficient evidence to reject the assumption of constant error variance violation. While the Cook's model (Figure 3.3.7) identifies additional residuals, the plots for Linear Model 4 appear non-linear compared to those of Linear Model 3, affirming that the logged data without outliers is the preferred choice for correlation modeling.