11.1 NH: eq. (11.14) versus AH: eq. (11.15)

\[ F = \frac{5.30162/1}{6.11216/36} = 31.2260. \]

with a \( p \)-value on 1 and 36 df equal to \( 2.5 \times 10^{-6} \). Thus, there is strong evidence to reject the null hypothesis.

NH: eq. (11.14) versus AH: eq. (11.16)

\[ F = \frac{(11.4138 - 5.49940)/(37 - 35)}{5.49940/35} = 18.8206. \]

with a \( p \)-value on 2 and 35 df equal to \( 2.8 \times 10^{-6} \). Thus, there is strong evidence to reject the null hypothesis.

NH: eq. (11.14) versus AH: eq. (11.17)

\[ F = \frac{(11.4138 - 5.09933)/(37 - 34)}{5.09933/34} = 14.0340. \]

with a \( p \)-value on 3 and 34 df equal to \( 4.1 \times 10^{-6} \). Thus, there is strong evidence to reject the null hypothesis.

NH: eq. (11.14) versus AH: eq. (11.18)

\[ F = \frac{(11.4138 - 5.08693)/(37 - 33)}{5.08693/33} = 10.2609. \]

with a \( p \)-value on 4 and 33 df equal to \( 1.6 \times 10^{-5} \). Thus, there is strong evidence to reject the null hypothesis.

NH: eq. (11.15) versus AH: eq. (11.16)

\[ F = \frac{0.612764/1}{5.49940/35} = 3.8998. \]

with a \( p \)-value on 1 and 35 df equal to 0.0562. Thus, there is some evidence to reject the null hypothesis.

NH: eq. (11.15) versus AH: eq. (11.17)

\[ F = \frac{(6.11216 - 5.09933)/(36 - 34)}{5.09933/34} = 3.3765. \]

with a \( p \)-value on 2 and 34 df equal to 0.0460. Thus, there is some evidence to reject the null hypothesis.
NH: eq. (11.16) versus AH: eq. (11.17)

\[
F = \frac{0.400072/1}{5.09933/34} = 2.6675.
\]

with a \( p \)-value on 1 and 34 df equal to 0.1116. Thus, there is little evidence to reject the null hypothesis.

NH: eq. (11.17) versus AH: eq. (11.18)

\[
F = \frac{0.0123937/1}{5.08693/33} = 0.0804.
\]

with a \( p \)-value on 1 and 33 df equal to 0.7785. Thus, there is no evidence to reject the null hypothesis.

11.2

11.2.1. The predictors in the problem are \( D \) and \( Ht \), while the terms in (11.25) are the constant term 1, \( \log(D) \), and \( \log(Ht) \).

11.2.2. From Arc obtain:

Data set = Trees, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = log[Vol]
Terms = (log[D] log[Ht])

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>log[D]</td>
</tr>
<tr>
<td>log[Ht]</td>
</tr>
</tbody>
</table>

R Squared: 0.977678
Sigma hat: 0.0813861
Number of cases: 31
Degrees of freedom: 28

Summary Analysis of Variance Table
Source | df | SS | MS | F | p-value |
-------|----|----|----|---|---------|
Regression | 2 | 8.12323 | 4.06161 | 613.19 | 0.0000 |
Residual | 28 | 0.185463 | 0.00662369 | |

Then the \( t \)-test gives

\[
t = \frac{\hat{\eta}_1 - 2}{se(\hat{\eta}_1)} = \frac{1.98265 - 2}{0.0750106} = -0.231
\]

with a \( p \)-value on 28 df of 0.819.
Also obtain:

Data set = Trees, Name of Fit = L2
Normal Regression
Kernel mean function = Identity
Response = log[Vol]
Terms = (log[Ht])
Offset = off
Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.56750</td>
<td>0.737869</td>
<td>-8.901</td>
</tr>
<tr>
<td>log[Ht]</td>
<td>1.09205</td>
<td>0.170485</td>
<td>6.406</td>
</tr>
</tbody>
</table>

R Squared: 0.585899
Sigma hat: 0.0800469
Number of cases: 31
Degrees of freedom: 29

Summary Analysis of Variance Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.262908</td>
<td>0.262908</td>
<td>41.03</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>29</td>
<td>0.185818</td>
<td>0.00640751</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Then the $F$-test using offset, off, equal to $2\log(D)$ gives

$$F = \frac{(0.185818 - 0.185463)/(29 - 28)}{0.00662369} = 0.054$$

with a $p$-value on 1 and 28 df equal to 0.818. Thus, there is no evidence to reject the null hypothesis, and so it is reasonable to suppose that $\eta = 2$. Note that $t^2 = (-0.231)^2 = 0.054 = F$.

11.2.3. From Arc obtain:

Data set = Trees, Name of Fit = L3
Normal Regression
Kernel mean function = Identity
Response = log[Vol]
Terms = (Ones)
Offset = off2
With no intercept.
Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ones</td>
<td>-6.16917</td>
<td>0.0142061</td>
<td>-434.263</td>
</tr>
</tbody>
</table>

Sigma hat: 0.0790961
Number of cases: 31
Degrees of freedom: 30
Summary Analysis of Variance Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>340.155</td>
<td>340.155</td>
<td>54370.92</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>30</td>
<td>0.187686</td>
<td>0.00625619</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Then the $F$-test using offset, $\text{off2}$, equal to $2 \log(D) + \log(Ht)$ gives

$$F = \frac{(0.187686 - 0.185463)/(30 - 28)}{0.00662369} = 0.168$$

with a $p$-value on 2 and 28 df equal to 0.846. Thus, there is no evidence to reject the null hypothesis, and so it is reasonable to suppose that $\eta_1 = 2$ and $\eta_2 = 1$.

11.2.4. Starting with the mean function (11.25), conditions under which the mean function for the regression of $\log(Vol)$ on $\log(D)$ is a simple linear regression model are any of the following:

1. $\eta_2 = 0$
2. $E(\log(Ht)|\log(D)) = E(\log(Ht)) = \alpha_0$
3. $E(\log(Ht)|\log(D)) = \alpha_0 + \alpha_1 \log(D)$

Condition 1 is not satisfied because the $t$ statistic for $\eta_2 = 0$ is 5.464 with a $p$-value on 28 df of 0.0000. Condition 2 is not satisfied because the overall $F$ statistic for the regression of $\log(Ht)$ on $\log(D)$ is 11.34 with a $p$-value on 1 and 29 df of 0.0022. Condition 3 may be satisfied—the regression of $\log(Ht)$ on $\log(D)$ appears to fit reasonably well, and a plot of $\log(Ht)$ versus $\log(D)$ shows that although a lowess mean smooth indicates a possibly nonlinear mean function, a linear mean function does not appear that unreasonable, particularly given the relatively small sample size.

The conditions under which the variance is constant in the submodel are either of the following:

1. $\eta_2 = 0$
2. $\text{Var}(\log(Ht)|\log(D)) = \gamma^2$

Condition 1 is not satisfied as explained above. Condition 2 is possibly satisfied—a plot of $\log(Ht)$ versus $\log(D)$ shows that although a lowess variance smooth indicates a possibly nonconstant variance function, a constant variance function does not appear that unreasonable, particularly given the relatively small sample size.

Little information appears to be lost by predicting volume from diameter alone—$R^2$ remains very high (0.954 for a mean function with only $\log(D)$, compared with 0.978 for mean function (11.25)), a plot of $\log(Vol)$ versus
the fitted values from the model with only \( \log(D) \) shows points fairly tightly clustered about the 45° line, and a plot of the fitted values from model (11.25) versus the fitted values from the model with only \( \log(D) \) shows points fairly tightly clustered about the 45° line.

11.3 From Arc we obtain:

Data set = Constructed, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = Y
Terms = (X1 X2 X3)
Coefficient Estimates
Label   Estimate   Std. Error t-value
Constant -984.780 44.9680 -21.900
X1     0.984681    0.0449623 21.900
X2     0.984726    0.0449424 21.911
X3     0.0181494   0.0355275  0.511

R Squared: 0.99982
Sigma hat: 0.0643099
Number of cases: 5
Degrees of freedom: 1

Summary Analysis of Variance Table
Source df SS MS F p-value
Regression 3 22.9736 7.65786 1851.62 0.0171
Residual 1 0.004137 0.004136

The output indicates that variable \( x_3 \) is not statistically significant for the model.

From the model menu, select the “Examine Submodels...” and then select forward selection. The output obtained is shown below:

Data set = Constructed, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = Y
Terms = (X1 X2 X3)
Forward Selection: Sequentially add terms that minimize the value of \( C_I \).
All fits include an intercept.

Base terms: Intercept
<table>
<thead>
<tr>
<th>df</th>
<th>RSS</th>
<th>k</th>
<th>C_I</th>
</tr>
</thead>
</table>
Add: X3 | 3     | 2.0661 | 2  498.568
Repeating the process, but now selecting backward selection, we obtain:

Data set = Constructed, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = Y
Terms = (X1 X2 X3)
Backward Elimination: Sequentially remove terms that give the smallest change in C_I. All fits include an intercept.

Current terms: (X1 X2 X3)

<table>
<thead>
<tr>
<th>df</th>
<th>RSS</th>
<th>k</th>
<th>C_I</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.00521509</td>
<td>3</td>
<td>2.261</td>
</tr>
<tr>
<td>2</td>
<td>1.98772</td>
<td>3</td>
<td>481.617</td>
</tr>
<tr>
<td>2</td>
<td>1.98965</td>
<td>3</td>
<td>482.085</td>
</tr>
</tbody>
</table>

Current terms: (X1 X2)

<table>
<thead>
<tr>
<th>df</th>
<th>RSS</th>
<th>k</th>
<th>C_I</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>14.3637</td>
<td>2</td>
<td>3472.054</td>
</tr>
<tr>
<td>3</td>
<td>14.4692</td>
<td>2</td>
<td>3497.551</td>
</tr>
</tbody>
</table>

Using forward selection, seems that there are no good candidates for sub-model mean functions: all $C_I$ values are far larger than $k$. The only subset selected that has an acceptable $C_I$ is the full model using all three predictors. Backward selection indicates that a good candidate submodel includes $x_1$ and $x_2$ but not $x_3$. A closer look at the data shows that to four decimals $x_1$ and $x_2$ have correlation -1:

Data set = Constructed, Sample Correlations

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1.0000</td>
<td>-1.0000</td>
<td>0.6858</td>
</tr>
<tr>
<td>X2</td>
<td>-1.0000</td>
<td>1.0000</td>
<td>-0.6826</td>
</tr>
<tr>
<td>X3</td>
<td>0.6858</td>
<td>-0.6826</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

X1  X2  X3
This feature is known as ‘multicollinearity’. There are some who would argue that when predictors are this collinear you should only use one of them and not both. This though is contradicted by the fact that if you did try using only one of them you would lose the very strong relationship between $y$ and the pair $x_1, x_2$. These predictors show the phenomenon of ‘suppression’; neither predictor on its own has much predictive power, but there is a synergy when the pair are used together.