

Laboratory 5: Biomass equations in R

Felipe Bravo Oviedo
Universidad de Valladolid, Spain

In this Lab we will learn how to:

- Fit biomass equations (independently)
- Select the best equations between different alternatives

Preparatory tasks

As usual, first we must define our working directory.

```
# setting the working directory  
  
setwd('C:/your_desired_working_directoryR')
```

Now we can import the dataset (in this lab we will use the biomass.csv dataset) by using this code:

```
# import the data set  
biomass <- read.csv2("biomassData.csv")
```

Descriptive analysis (including correlation graph)

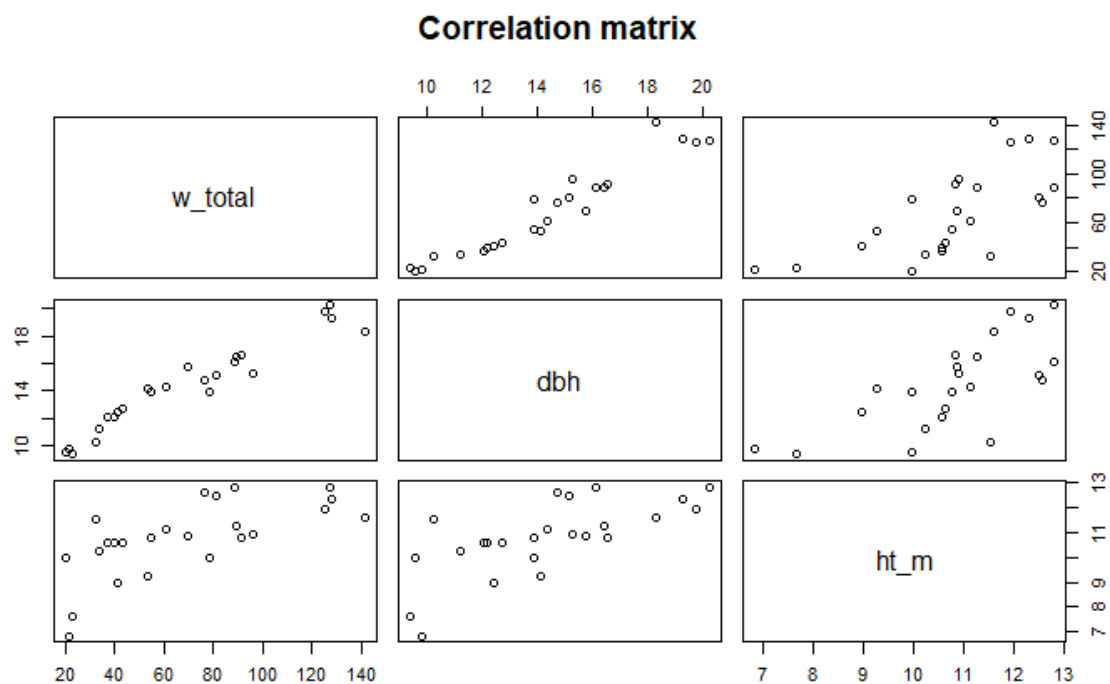
In this part, we'll obtain the main measures for the selected (or the whole) dataset variables. These measures can be classified as of central tendency (mean, median and mode), variability (standard deviation, variance, range, kurtosis and skewness).

```
# Descriptive analysis (including basic graphs)  
  
summary(biomass)  
  
# Analyzing the correlation between variables by  
# a correlation graph (lattice package needed)  
  
library(lattice) #if not already uploaded into your R version you must  
# install.packages(lattice)  
  
corr.biomass <- pairs(~w_total+dbh+ht_m, data=biomass,  
#correlation matrix for, ~variable1+variable2+...,  
main="Correlation matrix") #main= Graph heading
```

You will obtain the following summary for the whole dataset and this plot for the selected variables:

ht_m		dbh		w_branch		w_roots	
Min.	: 6.83	Min.	: 9.36	Min.	: 1.570	Min.	: 2.230
1st Qu.	: 10.17	1st Qu.	: 12.11	1st Qu.	: 5.740	1st Qu.	: 6.277
Median	: 10.83	Median	: 14.22	Median	: 9.255	Median	: 11.395
Mean	: 10.76	Mean	: 14.29	Mean	: 10.791	Mean	: 13.001
3rd Qu.	: 11.67	3rd Qu.	: 16.19	3rd Qu.	: 14.443	3rd Qu.	: 17.152
Max.	: 12.80	Max.	: 20.24	Max.	: 28.000	Max.	: 31.260

w_leaves		w_stem		w_total	
Min.	: 0.830	Min.	: 12.40	Min.	: 19.93
1st Qu.	: 2.542	1st Qu.	: 23.78	1st Qu.	: 39.06
Median	: 4.125	Median	: 40.31	Median	: 65.50
Mean	: 4.789	Mean	: 40.36	Mean	: 68.94
3rd Qu.	: 6.527	3rd Qu.	: 52.30	3rd Qu.	: 89.81
Max.	: 11.170	Max.	: 81.29	Max.	: 141.51



Now you try repeat the plot for the other variables in the dataset.

Fitting candidate biomass equations

We will fit two alternative biomass equations for the stem compartment (w_stem). Later you can repeat for the other compartments and total biomass. The 3 candidate modes are the following:

model1 $\rightarrow \log(w_stem) \sim b_0 + b_1 \cdot \log(dbh) + b_2 \cdot \log(ht_m)$

model2 $\rightarrow w_stem \sim b_0 + b_1 \cdot dbh + b_2 \cdot ht_m$

We will need to define the log transformation of the variables and insert in the model the proper variable transformations (multiplication of two variables and the power of a variable). To do that we'll use the `lm()` function.

```
## Fitting candidate biomass models

# log transformations

biomass$logw_stem <- log(biomass$w_stem)
biomass$log_dbh <- log(biomass$dbh)
biomass$loght_m <- log(biomass$ht_m)

# fitting the model1 by linear regression

model1.fit <- lm(logw_stem ~ log_dbh + loght_m, data=biomass)

# we can check the outcome

summary(model1.fit)
```

the outcome for the model1 fit is the following:

```
Call:
lm(formula = logw_stem ~ log_dbh + loght_m, data = biomass)

Residuals:
    Min       1Q   Median       3Q      Max
-0.20646 -0.05754 -0.02199  0.03302  0.30911

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.1372     0.3866  -8.114 6.56e-08 ***
log_dbh       2.1185     0.1551  13.663 6.42e-12 ***
loght_m       0.4731     0.2296   2.060  0.052 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1195 on 21 degrees of freedom
Multiple R-squared:  0.9562, Adjusted R-squared:  0.9521
F-statistic: 229.5 on 2 and 21 DF, p-value: 5.384e-15
```

Now you must repeat for the model2 and the model3 with this script:

```
# fitting the model2 and model3 by linear regression

model2.fit <- lm(w_stem ~ dbh + ht_m, data=biomass)
# we can check the outcomes

summary(model2.fit)
```

In the two model we observed that the total height (ht_m) is not significant so we must fit again the models but deleting this non significant variable. We can use the following script:

```
# fitting the reduced models (modell1 and model2) by linear regression
```

```
modell1r.fit <- lm (logw_stem ~ log_dbh , data=biomass)
model2r.fit <- lm (w_stem ~ dbh , data=biomass)
```

```
# we can check the outcomes
```

```
summary(modell1r.fit)
plot (residuals(modell1r.fit), xlab="dbh")
```

```
summary(model2r.fit)
plot (residuals(model2r.fit), xlab="dbh")
```

For the model2r.fit we will obtain the following table and plot (observe the results for modell1r.fit)

Call:

```
lm(formula = w_stem ~ dbh, data = biomass)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.1129	-3.3936	-0.1968	3.7192	9.5705

Coefficients:

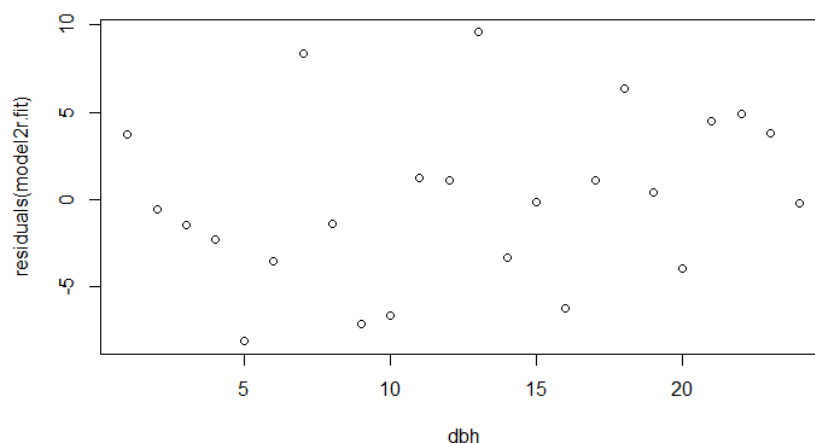
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-47.2706	4.6971	-10.06	1.08e-09 ***
dbh	6.1300	0.3211	19.09	3.51e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.888 on 22 degrees of freedom

Multiple R-squared: 0.9431, Adjusted R-squared: 0.9405

F-statistic: 364.5 on 1 and 22 DF, p-value: 3.514e-15



You can add a horizontal line in the plot by adding these `abline()` instructions after the plot instructions:

```
plot (residuals(model1r.fit), xlab="dbh")
abline(h=0, col'blue', lty=2)
```

```
plot (residuals(model2r.fit), xlab="dbh")
abline(h=0, col'blue', lty=2)
```

Comparing the candidate biomass equations

Now we are to compare the two reduced models by using different goodness of fit estimators that asses the relative quality of the tested equation for a given dataset. We'll use the following estimators:

- AIC
- BIC
- R^2
- R^2 adjusted

To compare the fitted equations, we'll need the *performance* and *see* packages and the following script.

```
## Fitting candidate biomass models

#comparing model performance by using a table and graphs

library(performance)
library(see)

#if not already upload to your R version import
# you must before install.packages("performance")
# and install.packages("see")

compare_performance(model1r.fit, model2r.fit)
```

you will obtain the following outcome:

Comparison of Model Performance Indices

Model	Type	AIC	BIC	R2	R2_adjusted	RMSE	BF
model 1r. fit	lm	-26.65	-23.12	0.95	0.94	0.12	
model 2r. fit	lm	148.19	151.72	0.94	0.94	4.68	0

With this outcome we should select the `model1r.fit` because shows the lower AIC and BIC and the higher R^2 .

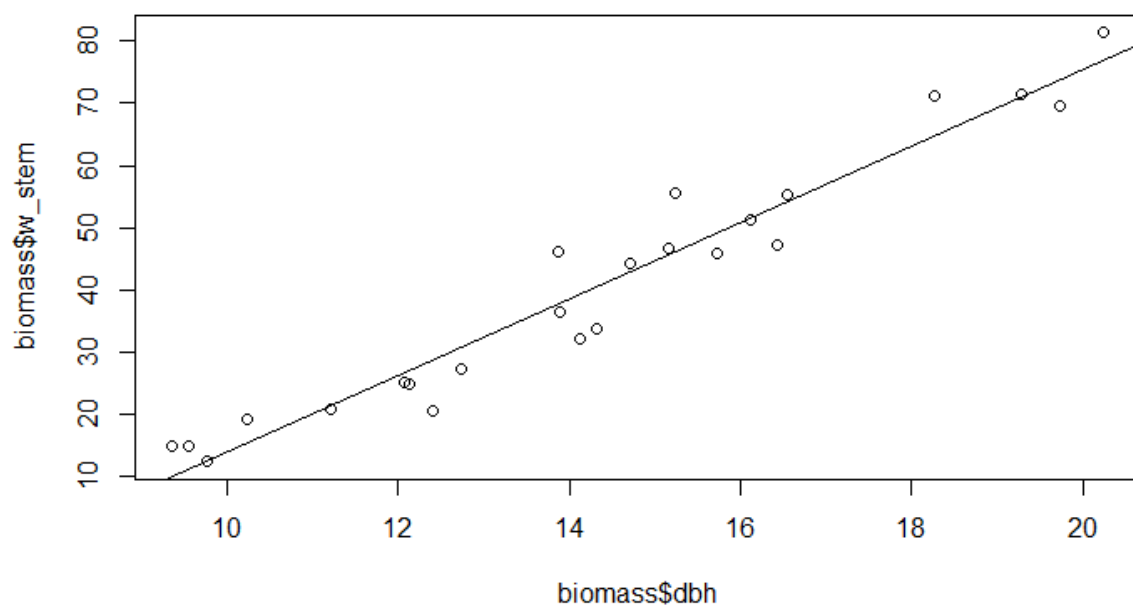
Plotting the fitted model over the original data

Now we will plot the original variables used to fit our two equations and over this plot we will overlap the 2 lines. We will use this script

```
## Plotting the two equations
```

```
plot (biomass$dbh, biomass$w_stem)  
abline (model2r.fit)
```

and we will obtain the following plot



We did not include the line `modellr.fit` because we have to detransformed the log transformation and the variable to obtain the correct estimation of the stem biomass (`w_stem`) and beside that we must apply the log bias correction by using the factor proposed by Baskerville (1972) that is calculated as the standard error (SE) of the regression to the power of 2 divided by 2, that is, $cf = RSE^2/2$

Where the RSE (Residual Standard Error) value can be obtained from the summary (`modellr.fit`) If we use `anova(modellr.fit)` we can obtain the MSE (Mean Square Error). From our basic statistics courses we know that $MSE = RSE^2$ Let's compute it with R:

```
## computing the bias correction factor for modellr.fit
```

```
cf <- exp(0.128^2/2)  
cf
```

obtaining a value for cf equal to 1.008226 as bias correction factor so now we correct the predicted values from model1r.fit multiplying the outcome for this factor.

Full script

You can copy and paste into your R environment (changing your working directory)

```
# Script to fit biomass equations [linear models]
# developed by Felipe Bravo 2020/05/02
# with R version 4.0.0 (2020-04-24) -- "Arbor Day"
# Platform: x86_64-w64-mingw32/x64 (64-bit)
#####

## setting the working directory
setwd('C:/your_desired_working_directoryR')

## import the data set
biomass <- read.csv2("biomassData.csv")

## Descriptive analysis (including basic graphs)

summary(biomass)

# Analyzing the correlation between variables
# correlation graph (lattice package needed)

library(lattice)
#if not already upload to your R version import
# you must before install.packages(lattice)

corr.biomass <- pairs(~w_total+dbh+ht_m,data=biomass, #correlation matrix for,
~variable1+variable2+..,
main="Correlation matrix") #main= Graph heading

## Fitting candidate biomass models

# log transformations

biomass$logw_stem <- log(biomass$w_stem)
biomass$log_dbh <- log(biomass$dbh)
biomass$loght_m <- log(biomass$ht_m)

#### fitting the model 1 by linear regression

model1r.fit <- lm (logw_stem ~ log_dbh + loght_m, data=biomass)

# we can check the outcome
```

```

summary(model1.fit)

# fitting the model2 by linear regression

model2.fit <- lm (w_stem ~ dbh + ht_m, data=biomass)

# we can check the outcome

summary(model2.fit)

# fitting the reduced models (model1 and model2) by linear regression

model1r.fit <- lm (logw_stem ~ log_dbh , data=biomass)
model2r.fit <- lm (w_stem ~ dbh , data=biomass)

# we can check the outcomes

summary(model1r.fit)
plot (residuals(model1r.fit), xlab="dbh")
abline(h=0, col'blue', lty=2)

summary(model2r.fit)
plot (residuals(model2r.fit), xlab="dbh")
abline(h=0, col'blue', lty=2)

#comparing model performance by using a table and graphs

library(performance)
library(see)

#if not already upload to your R version import
# you must before install.packages("performance")
# and install.packages("see")

compare_performance(model1r.fit, model2r.fit)

## Plotting the two equations

plot (biomass$dbh, biomass$w_stem)
abline (model2r.fit)
lines(biomass$dbh, biomass$w_stempred.corr.model1r)

## computing the bias correction factor for model1r.fit

cf <- exp(0.128^2/2)
cf

```


Further work

Now you can do the following tasks:

1. Fit for each compartment and the total tree the following biomass equations:

```
model1 → w_comp~ b0+ b1*dbh + b2*ht_m
model2 → w_comp~ b0+ b1*dbh + b2*dbh^2
model3 → w_comp~ b0+ b1*dbh + b2* dbh^2 + (b3*dbh^2*ht_m)
model4 → w_comp~ b0+ b1*dbh^2 + b2*ht_m
model5 → w_comp~ b0+ b1*dbh^2 + b2*dbh*ht_m
model6 → w_comp~ b0+ b1*dbh^2*ht_m + b2*dbh*ht_m
model7 → w_comp~ b0+ b1*dbh^2 + b2*dbh^2*ht_m
model8 → w_comp~ b0+ b1*dbh^2 + b2*ht_m+b3*dbh^2*ht_m
model9 → w_comp~ b0*(dbh^b1)*(ht_m^b2)
model10 → w_comp~ b0*(dbh^b1)
model11 → w_comp~ b0*(dbh*ht_m)^b1
```

Replace w_comp by the right compartment or the total biomass. Models 9, 10 and 11 need log transformation

2. Compare the performance of the fitted models
3. Check if the sum of the biomass estimated by compartment is equal to the total biomass estimation with the fitted equation for the whole tree.

References

- Baskerville, G.L. 1972 Use of logarithmic regression in the estimation of plant biomass. Canadian Journal of Forest Research 2: 49-53 [available at <https://www.esf.edu/quest/documents/Baskerville-1972-CJFR.pdf>]
- James, G., Witten, D., Hastie, T., Tibshirani, R. 2013 An introduction to statistical learning with applications in R. Springer 426 pages [available at <http://faculty.marshall.usc.edu/gareth-james/ISL/>]
- Newman, M.C. 1993 Regression analysis of log-transformed data: statistical bias and its correction. *Environmental Toxicology and Chemistry* 12:1129-1133 [available at <http://materias.df.uba.ar/11a2018c2/files/2012/07/Newman1993.pdf>]