

Brassicas_2021_morphology

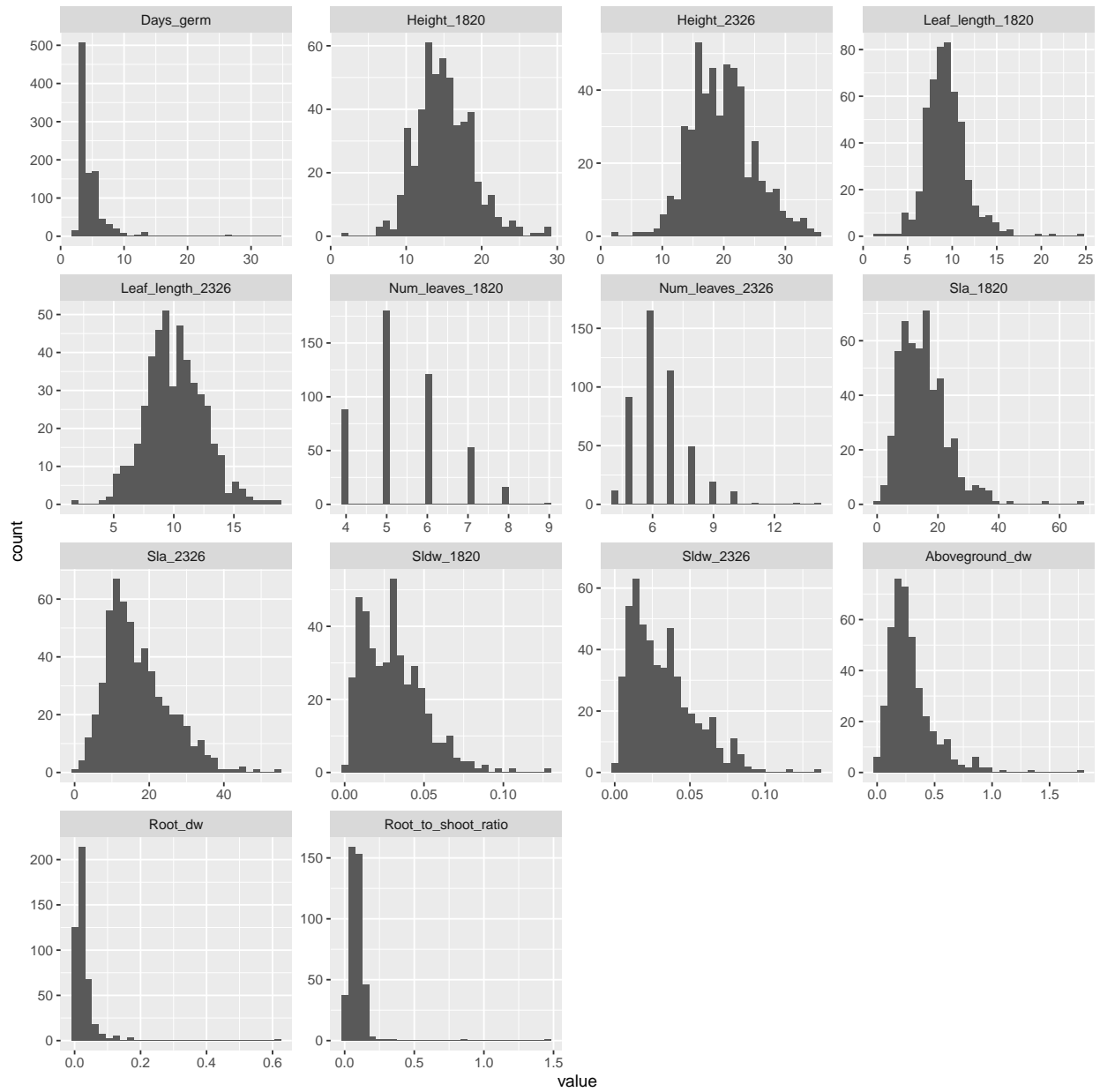
Benjamin A Taylor

20/09/2021

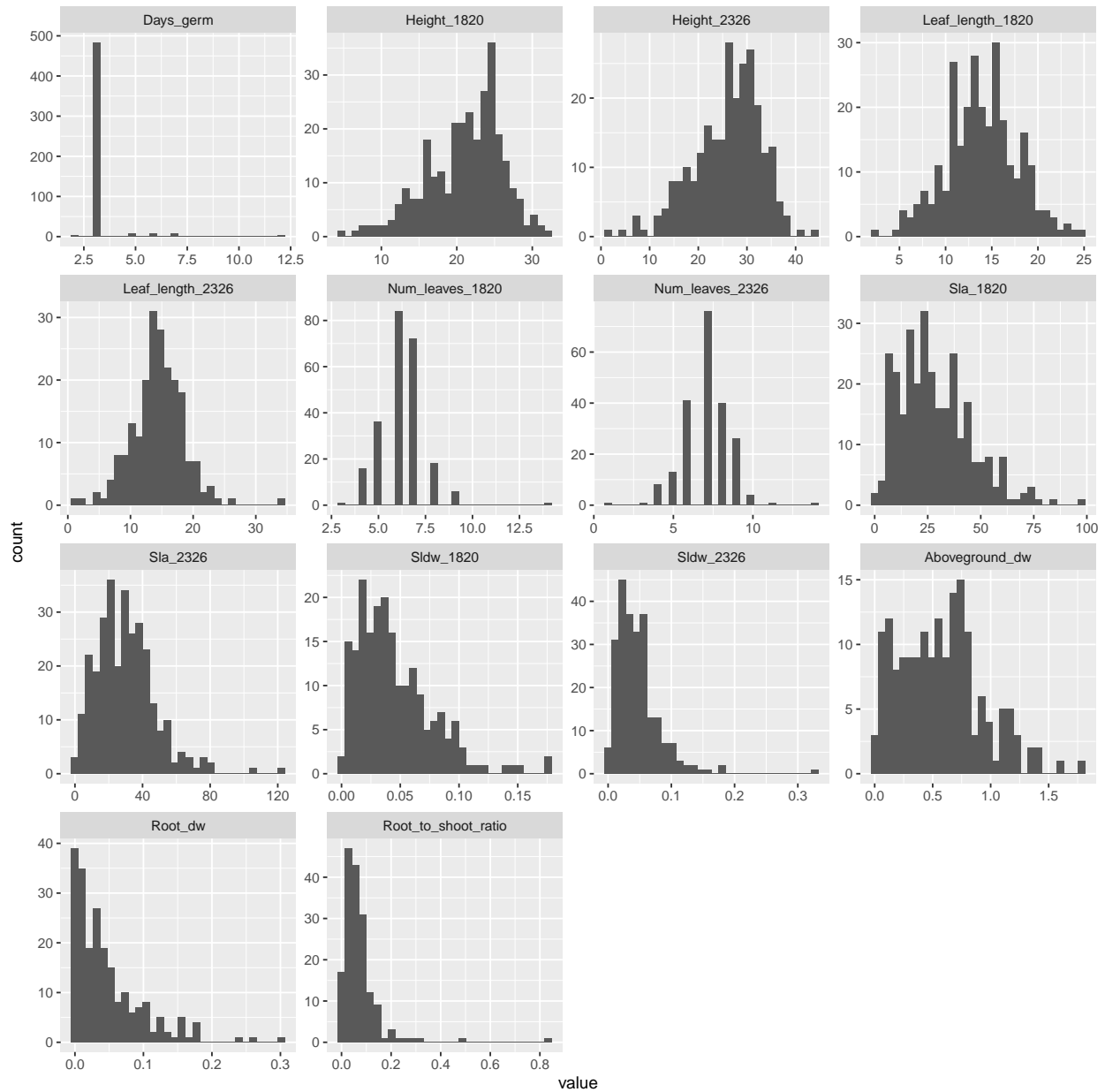
Data import, pre-filtering and QC

```
##
## 15 16 17 18 19 20 21 22 23 24 25 26 28
##  9 10  4 238 291 178 10  3 110 228 206 212  1
```

Check trait distributions for Brassica:



And for Raphanus:



What's going on with root dry weight in these plots? It looks unimodal, but the scale seems off, perhaps indicating outliers. Sure enough, if we look at the quantile distributions for dry weight there are definitely outliers here, which we can reasonably attribute to experimental error given the sensitivity of the weighing to e.g. residual moisture in the roots.

```
##      0%      25%      50%      75%      95%     100%
## 0.0001 0.0090 0.0180 0.0300 0.0670 0.6160
```

```
##      0%      25%      50%      75%      95%     100%
## 0.00010 0.00900 0.03200 0.06725 0.15325 0.30200
```

To be safe, we'll remove values for root dry weight that are more than 3 standard deviations above the mean.

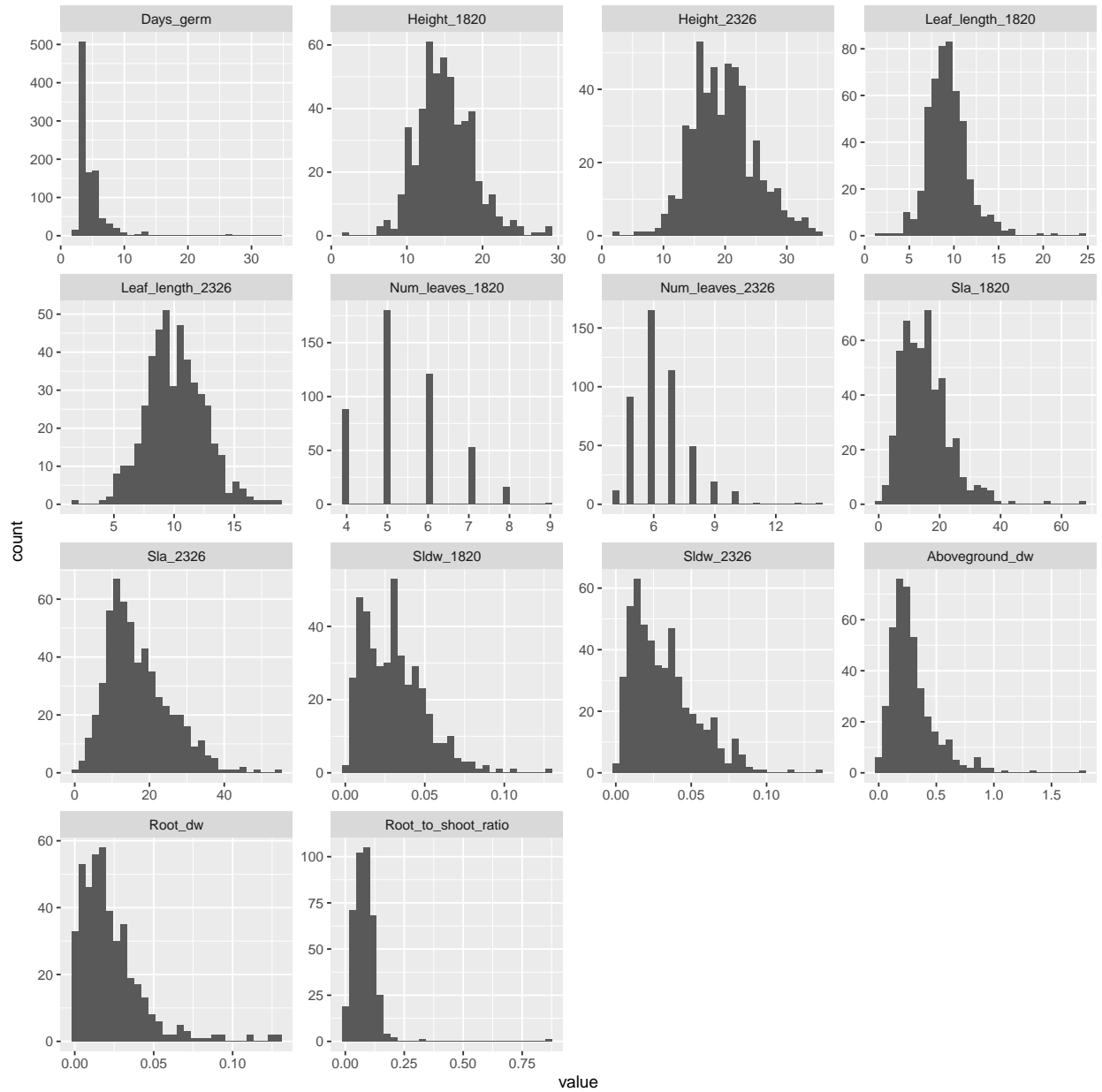
```
## [1] "Removing values: "
```

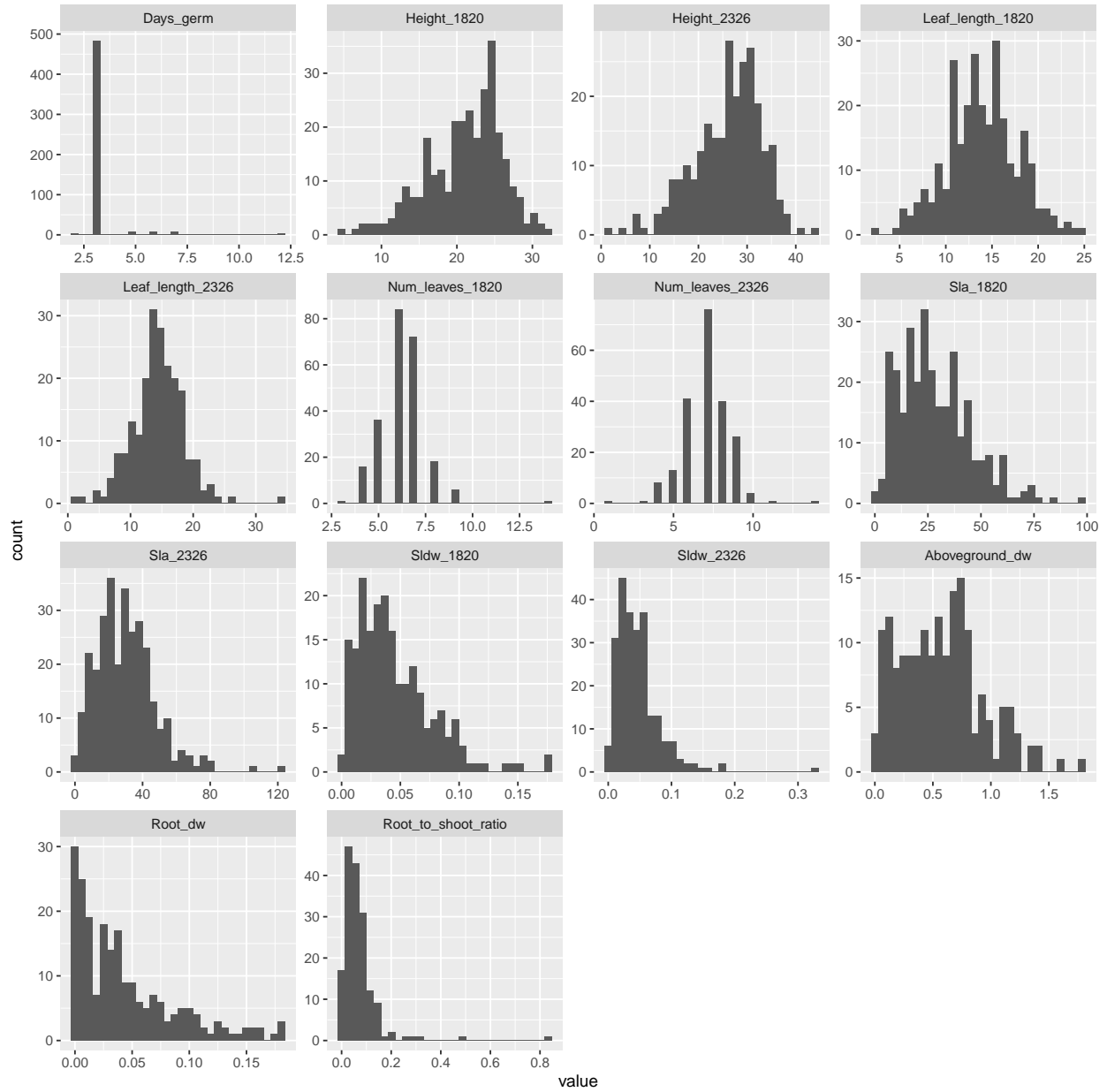
```
## [1] "0.616" "0.175" "0.162" "0.171" "0.137"
```

```
## [1] "Removing values: "
```

```
## [1] "0.302" "0.238" "0.257"
```

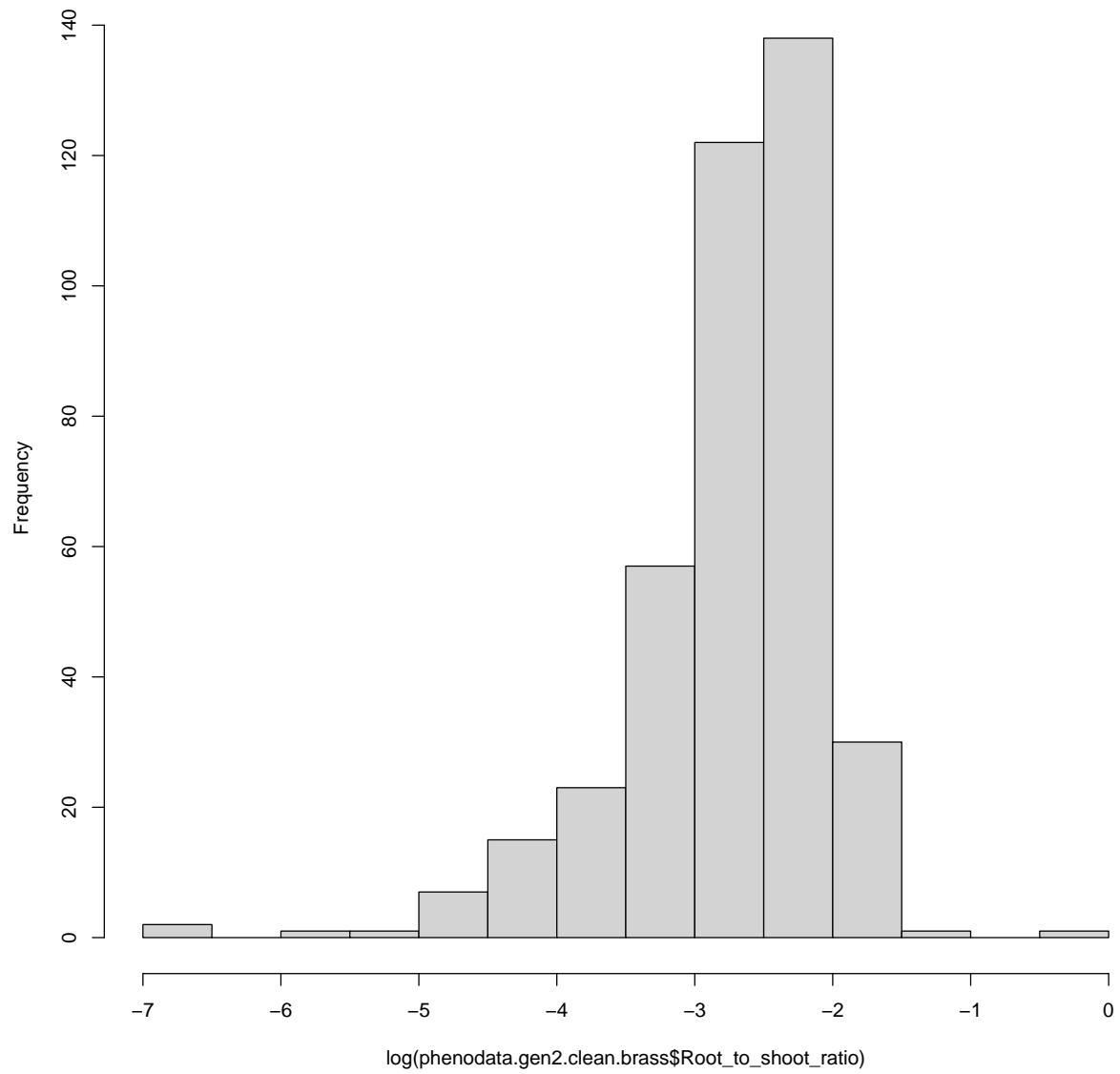
How do the distributions look now for Brassica and for Raphanus?

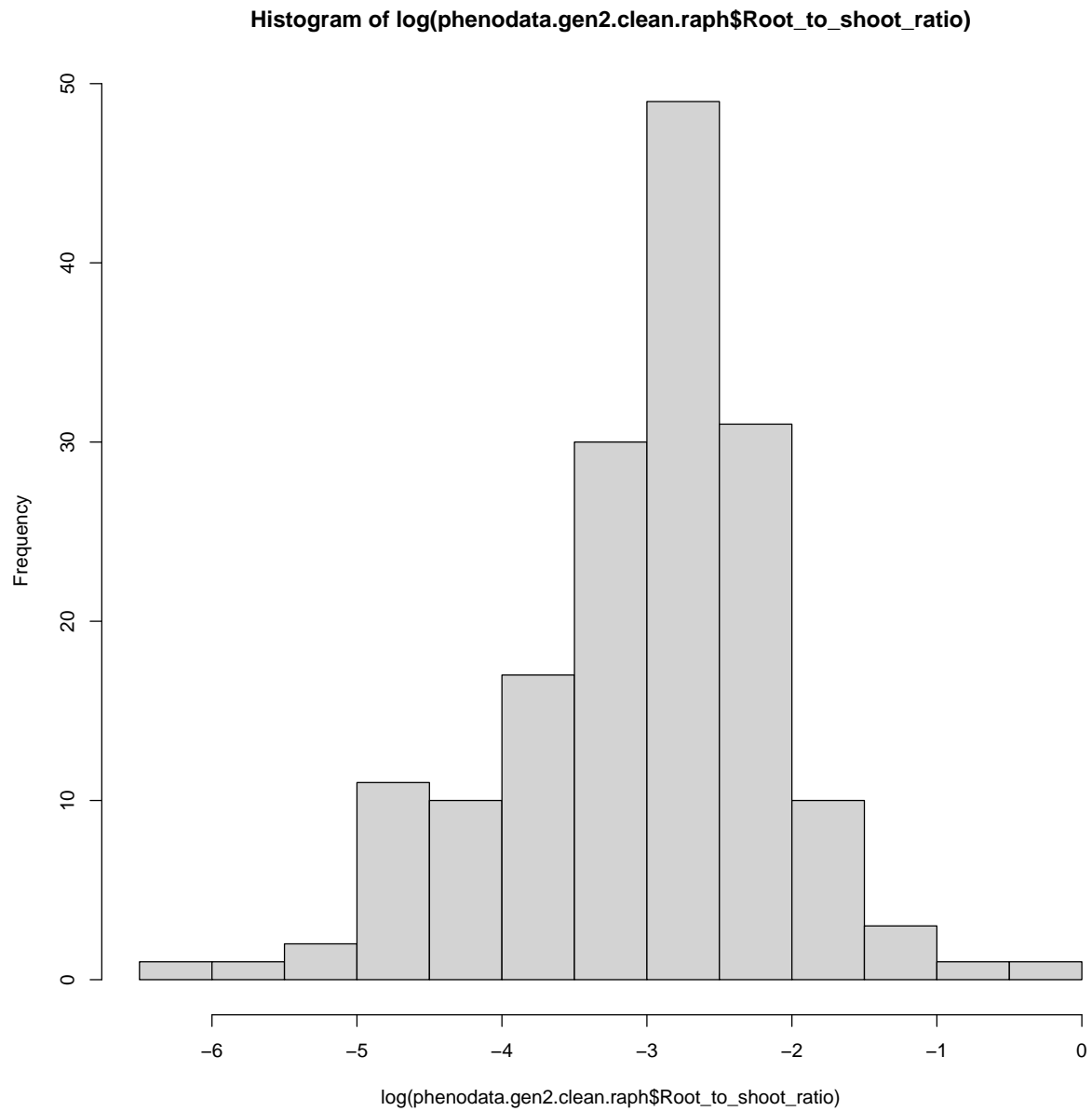




Root dry weight looks better now, but root-shoot ratio is still very skewed- not surprising for a ratio like this. If we log the ratio, it looks much better:

Histogram of $\log(\text{phenodata.gen2.clean.brass}\$Root_to_shoot_ratio)$





We'll use logged root-shoot ratio instead of the unmodified ratio henceforth. We'll also drop days to germination, since this is clearly a highly unimodal trait.

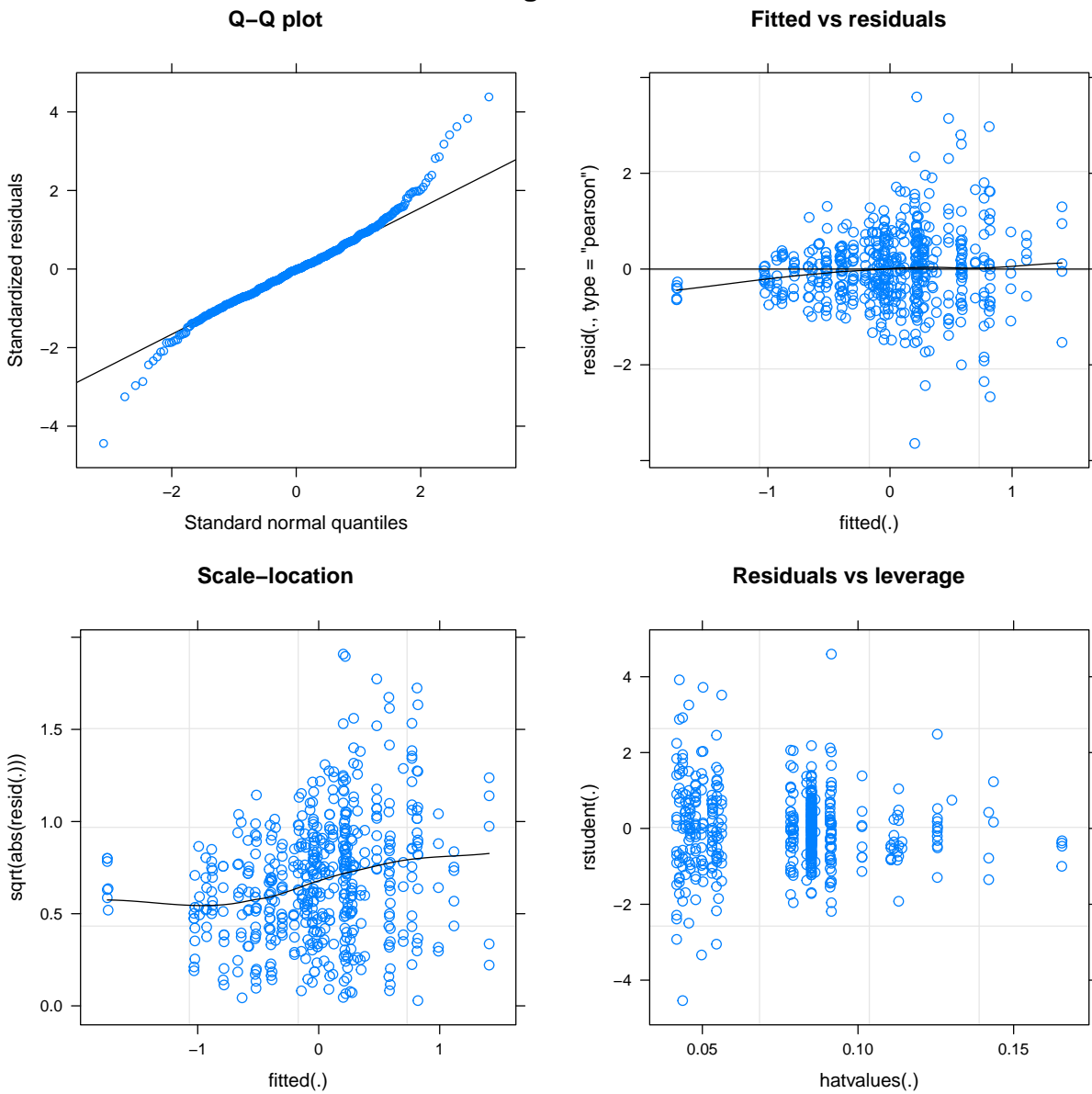
Trait model testing

We'd now like to fit the data for each trait so that we can plug our data into mixed models down the line. In an ideal world each trait would be fit reasonably well by a normal distribution- let's see if that's the case. What we're looking for:

1. Points fitted nicely along the QQ plot abline
2. Even distribution about abline in fitted vs residuals plot
3. Roughly horizontal fit for scale-location plot
4. No outliers in residuals vs leverage plot

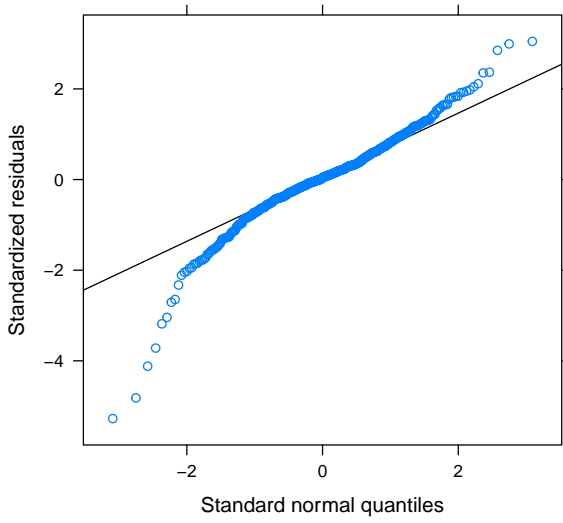
First for Brassica traits:

Height_1820

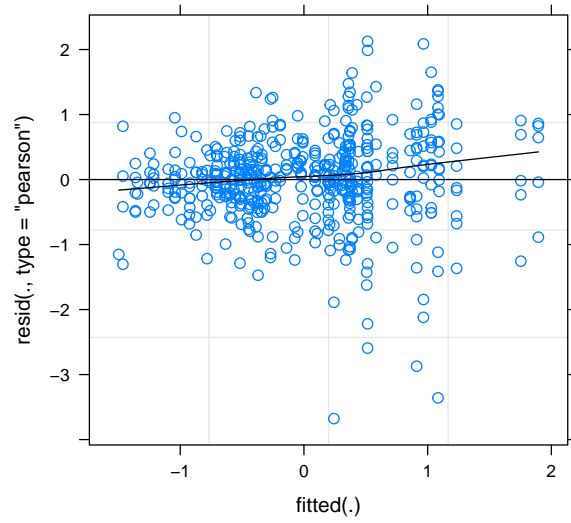


Height_2326

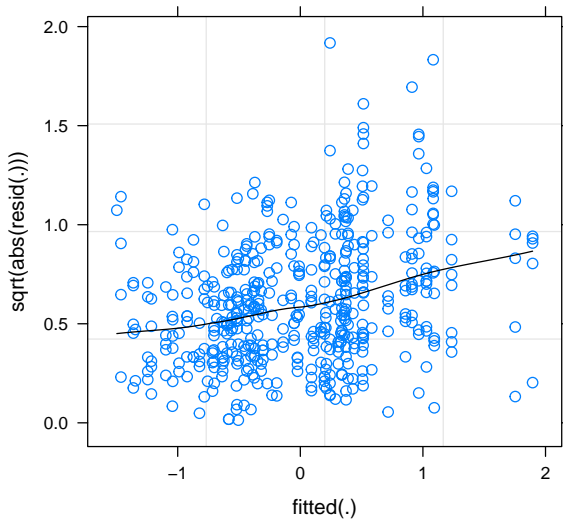
Q-Q plot



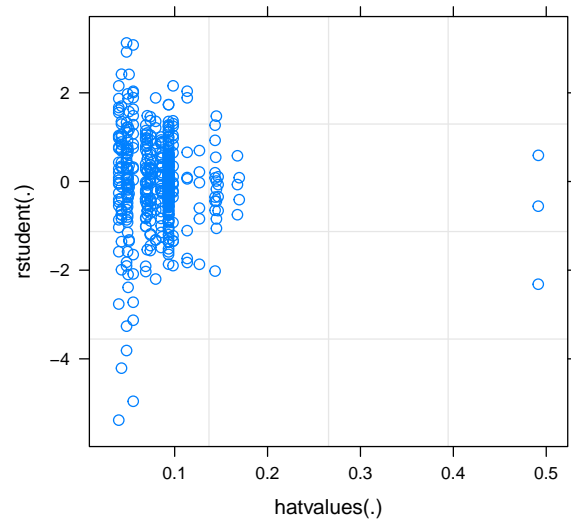
Fitted vs residuals



Scale-location

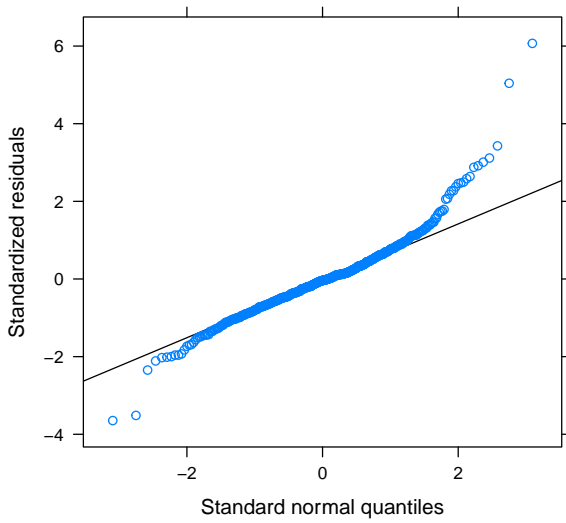


Residuals vs leverage

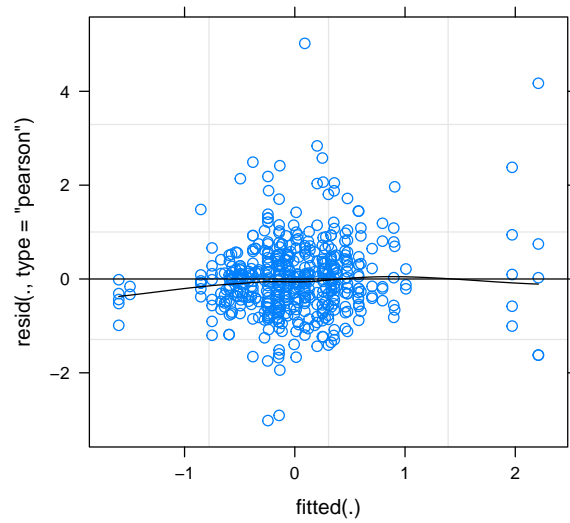


Leaf_length_1820

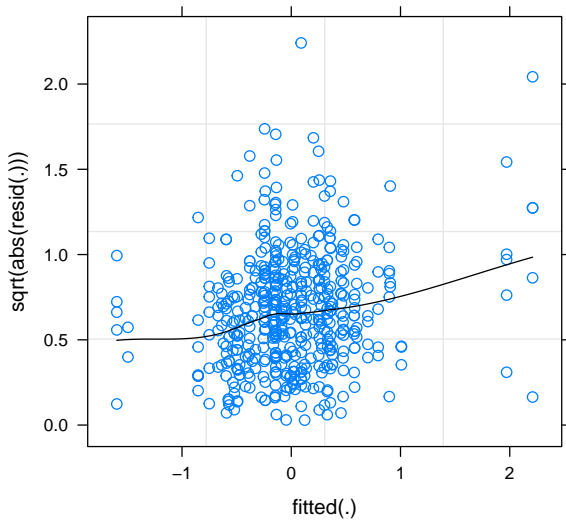
Q-Q plot



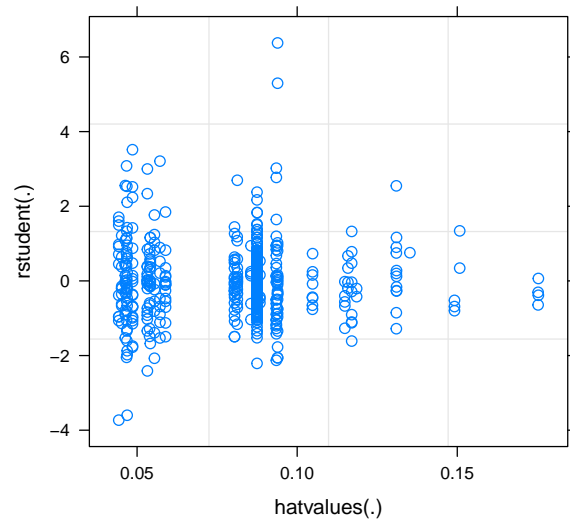
Fitted vs residuals



Scale-location

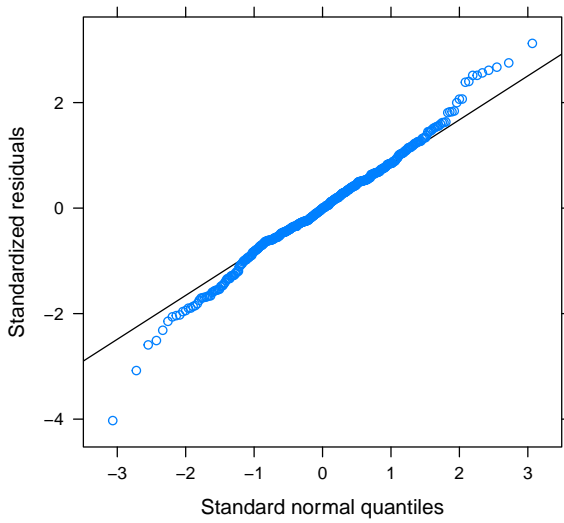


Residuals vs leverage

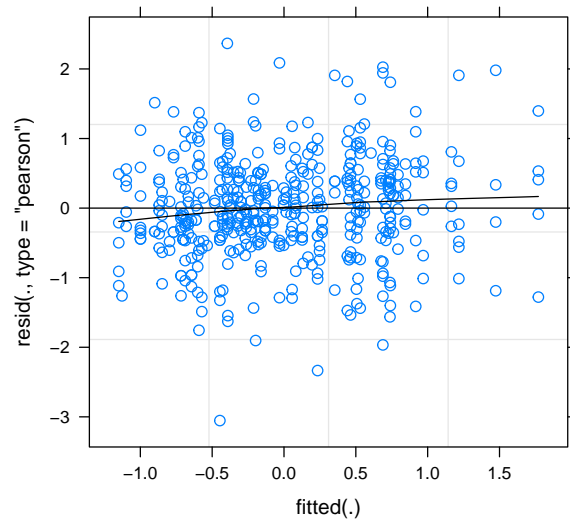


Leaf_length_2326

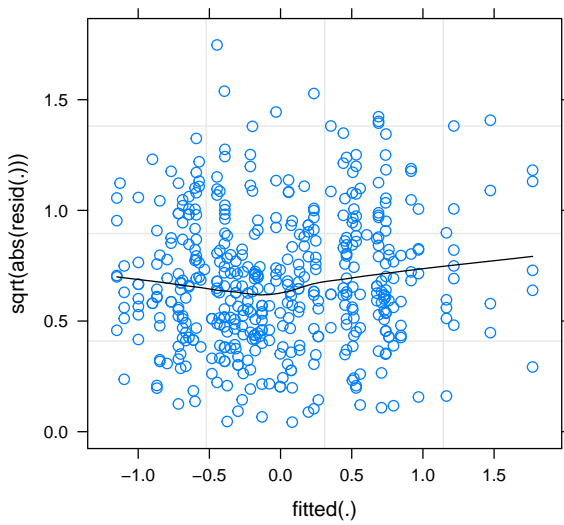
Q-Q plot



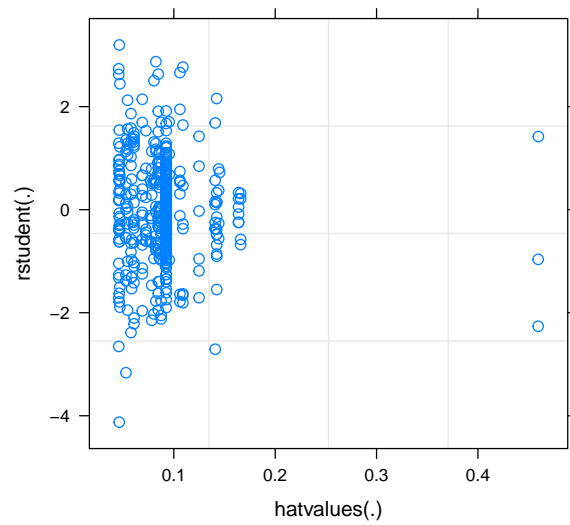
Fitted vs residuals



Scale-location

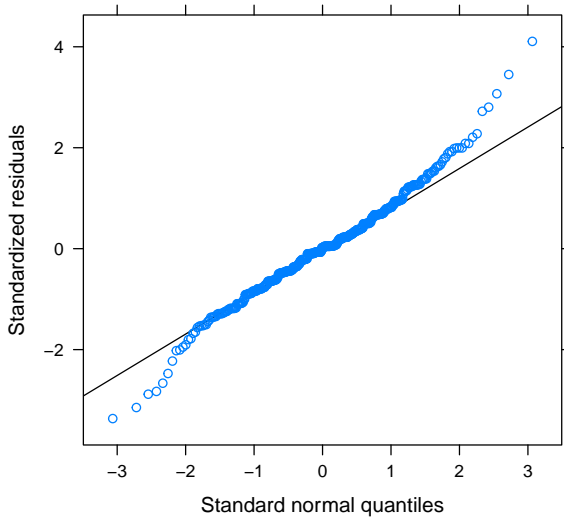


Residuals vs leverage

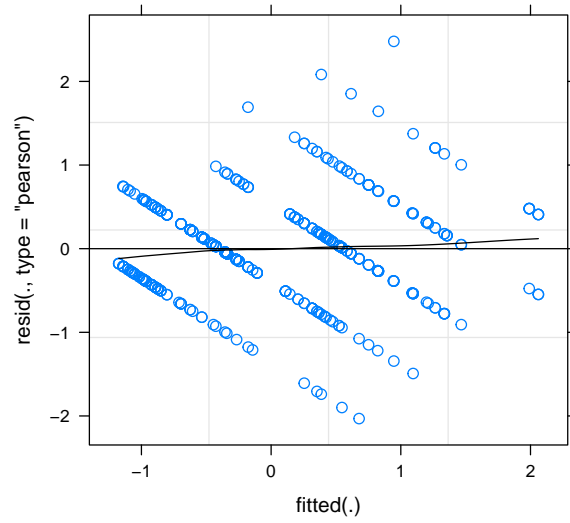


Num_leaves_1820

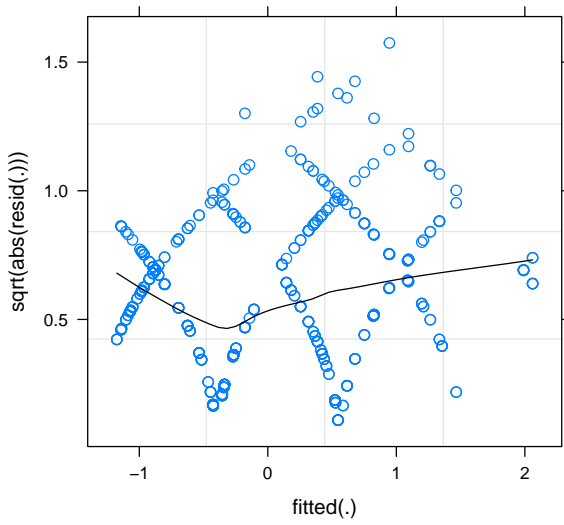
Q-Q plot



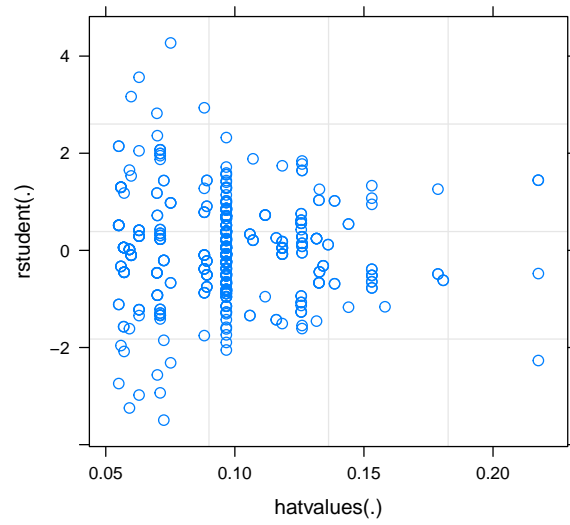
Fitted vs residuals



Scale-location

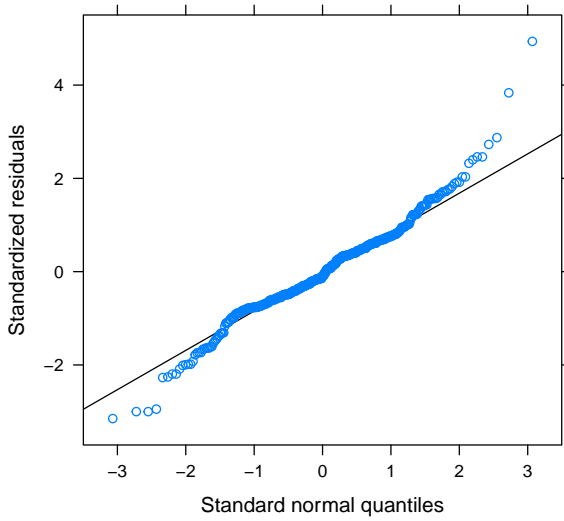


Residuals vs leverage

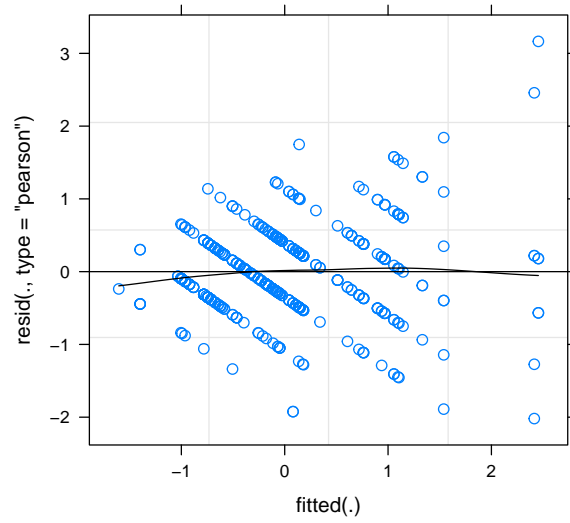


Num_leaves_2326

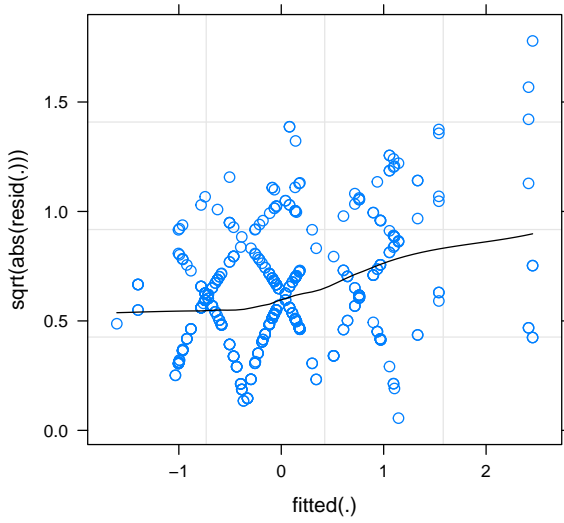
Q-Q plot



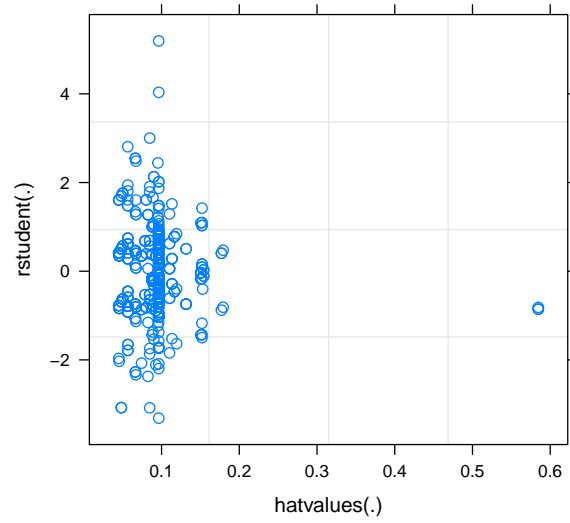
Fitted vs residuals



Scale-location

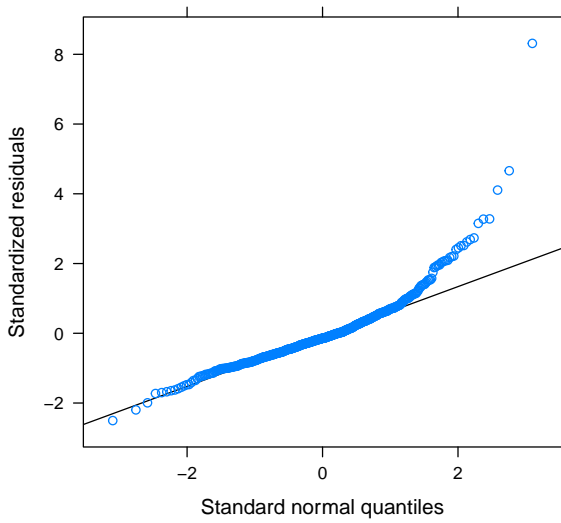


Residuals vs leverage

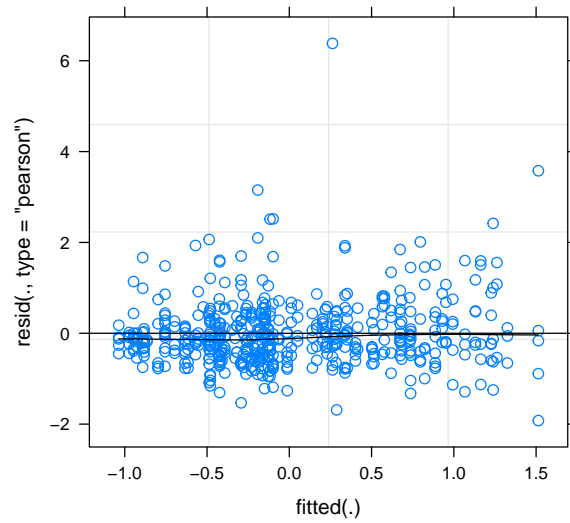


Sla_1820

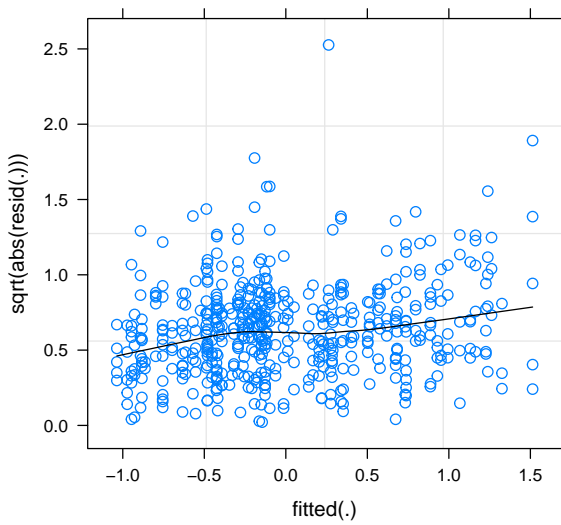
Q-Q plot



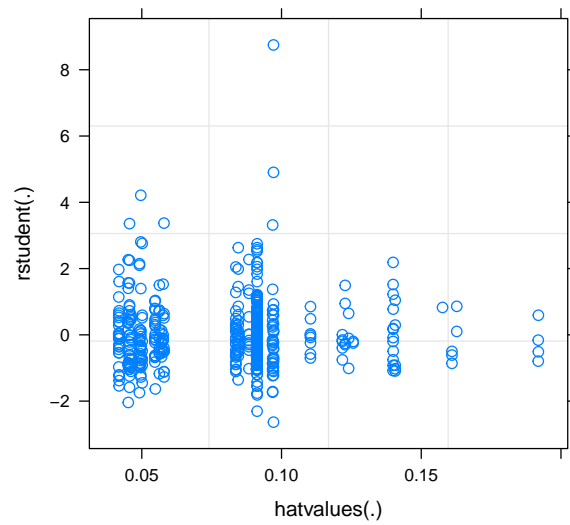
Fitted vs residuals



Scale-location

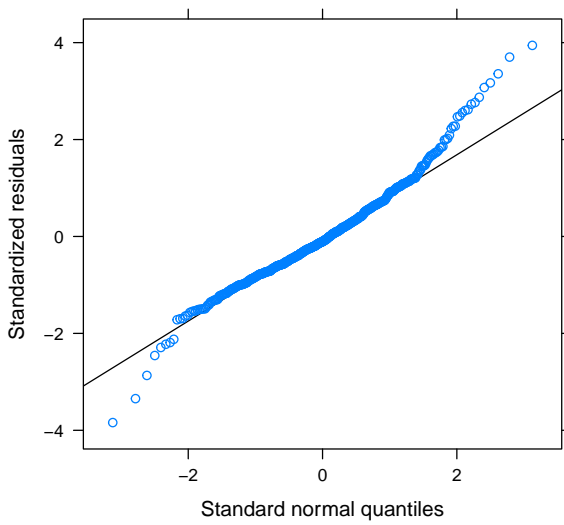


Residuals vs leverage

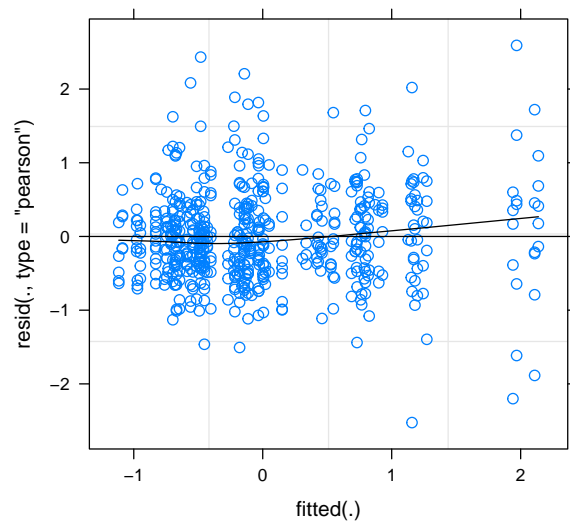


Sla_2326

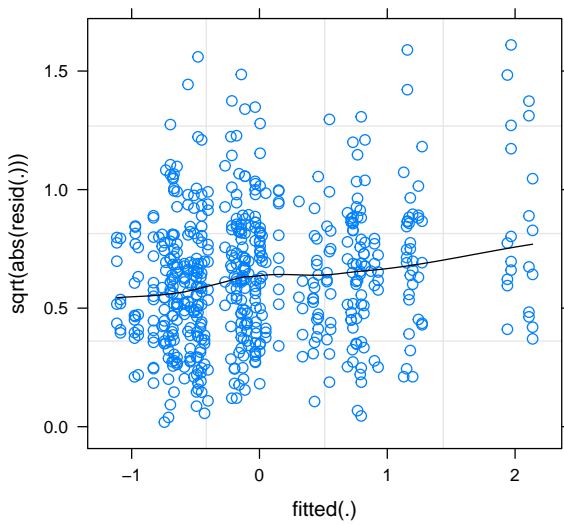
Q-Q plot



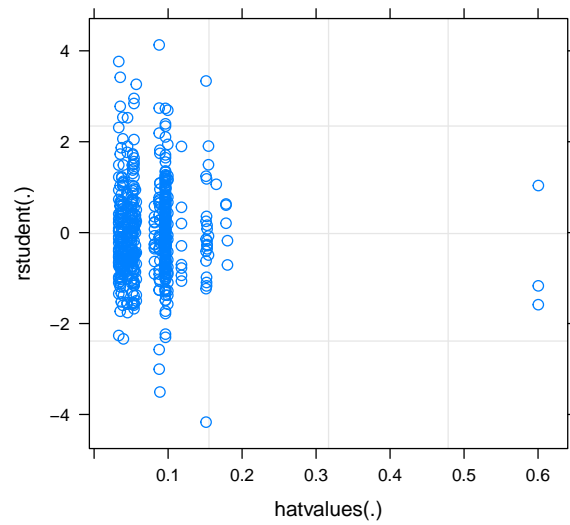
Fitted vs residuals



Scale-location

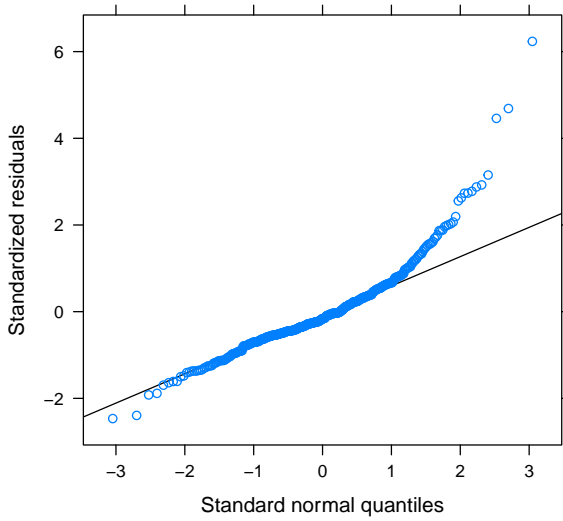


Residuals vs leverage

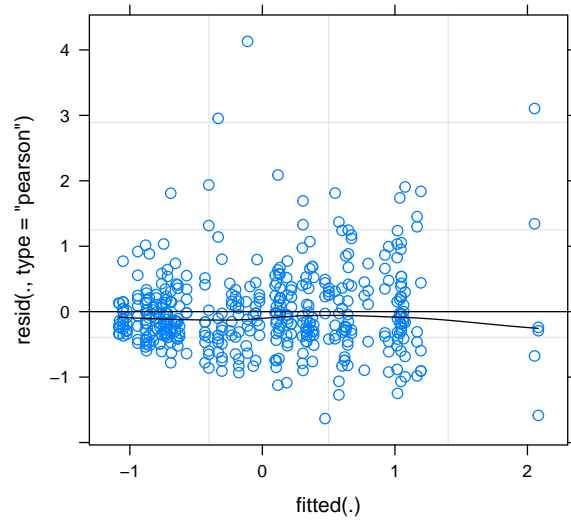


Slw_1820

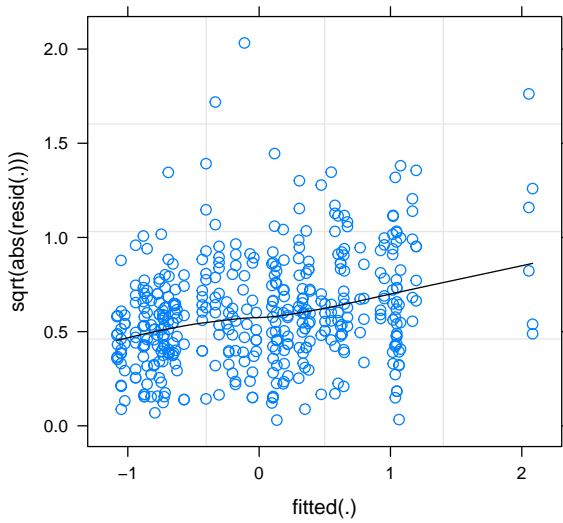
Q-Q plot



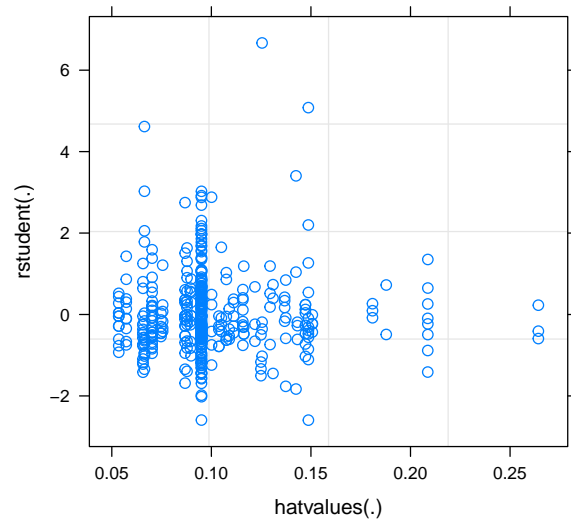
Fitted vs residuals



Scale-location

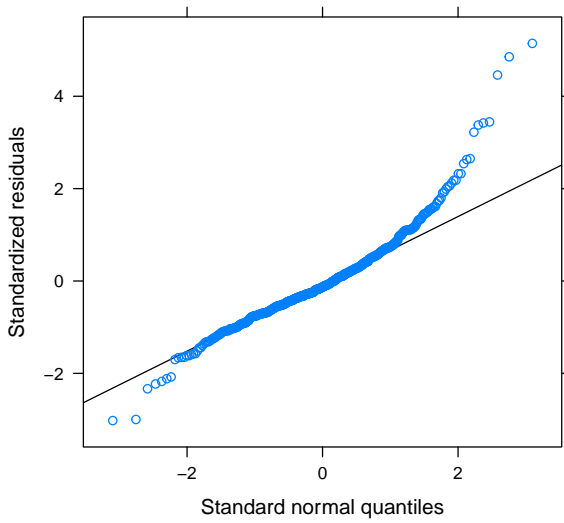


Residuals vs leverage

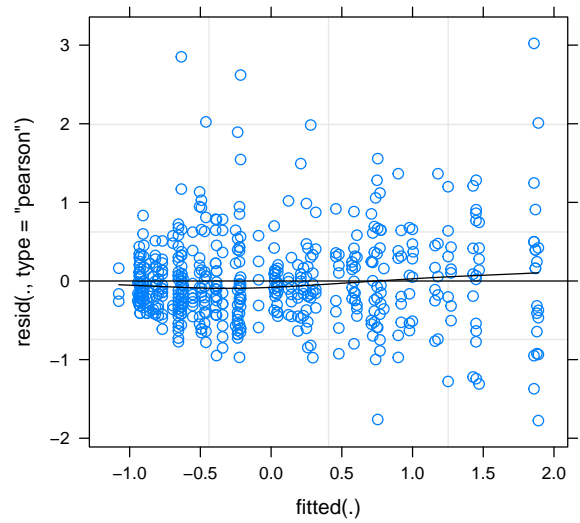


Slw_2326

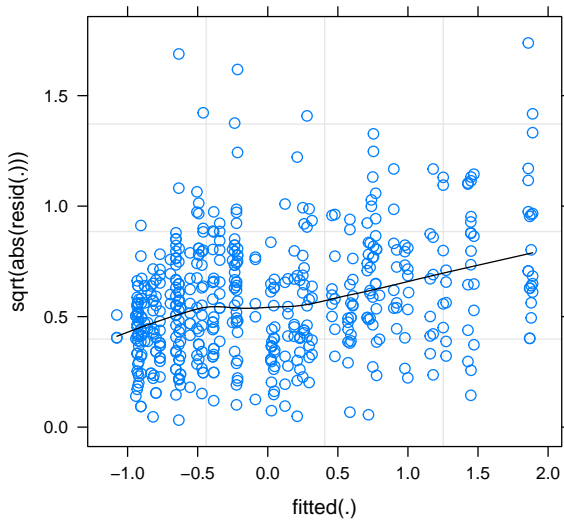
Q-Q plot



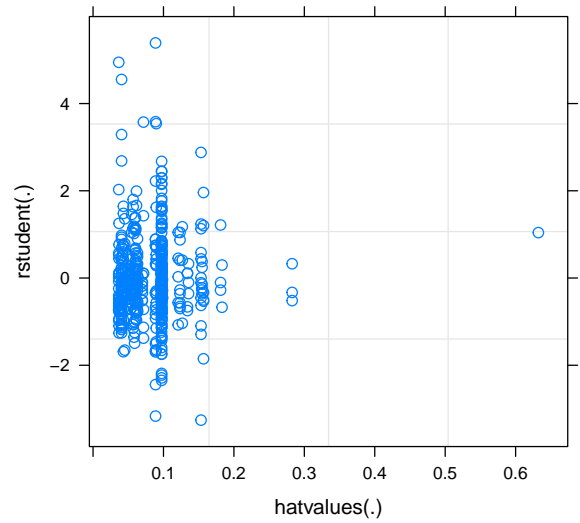
Fitted vs residuals



Scale-location

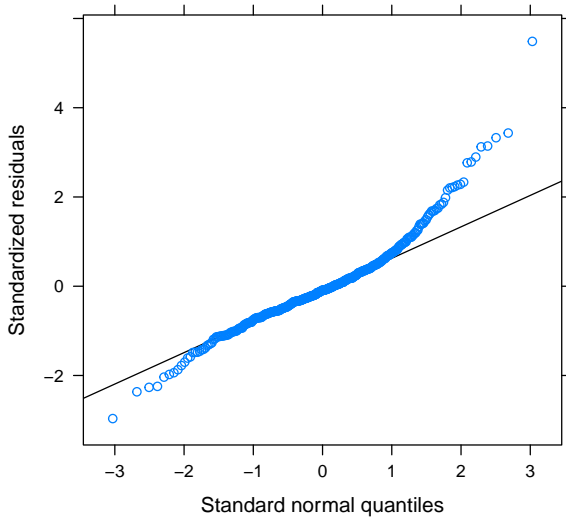


Residuals vs leverage

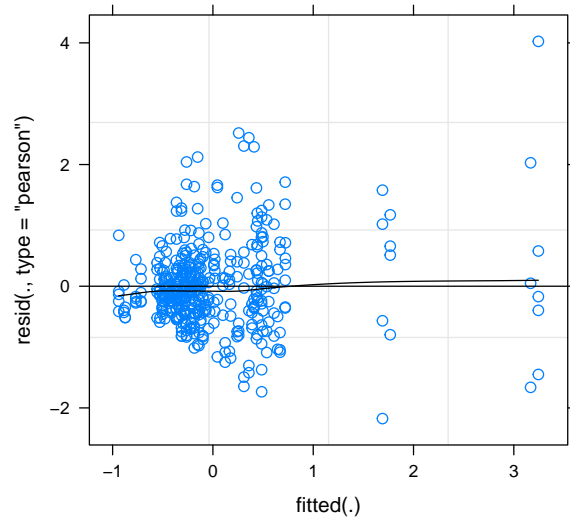


Aboveground_dw

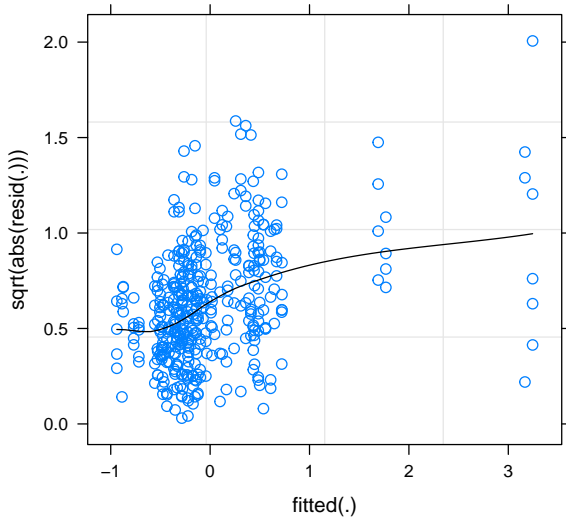
Q-Q plot



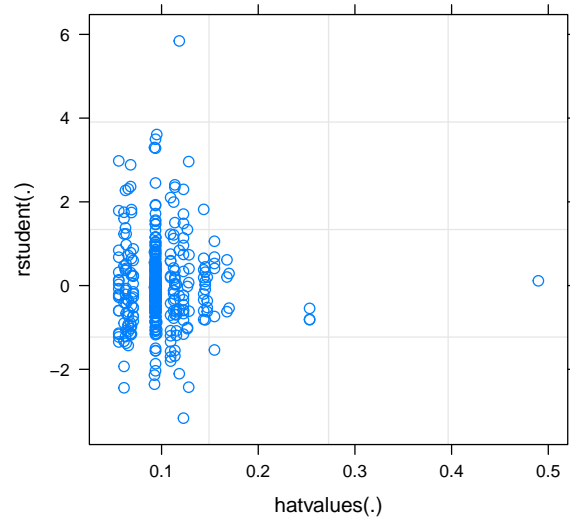
Fitted vs residuals



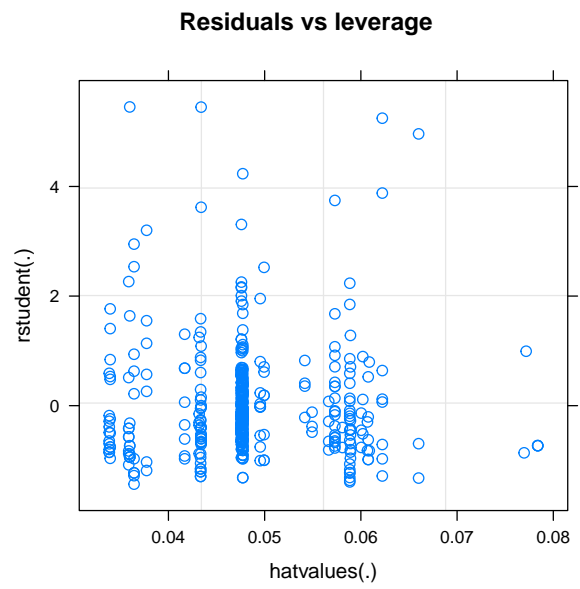
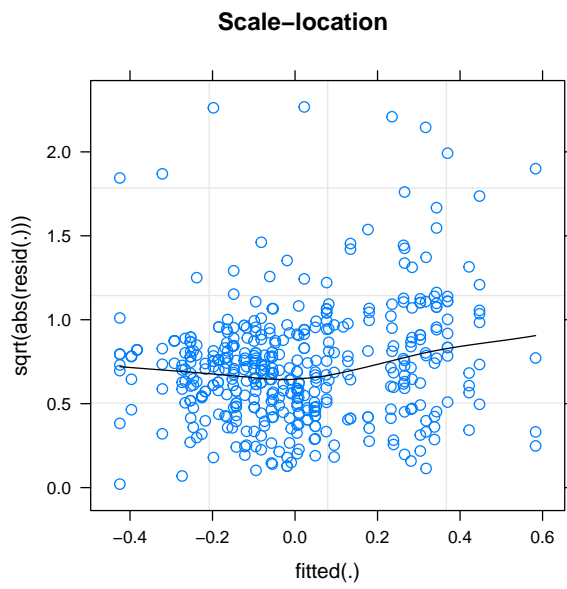
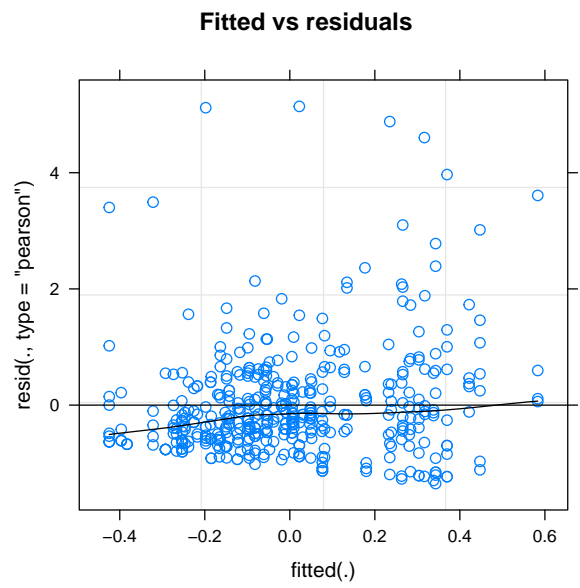
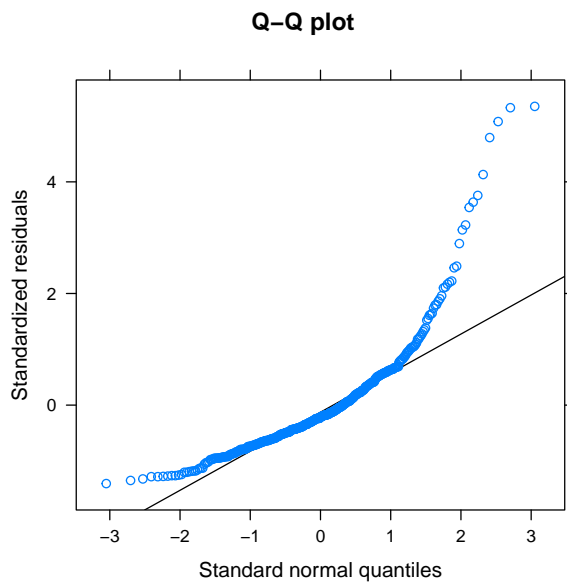
Scale-location



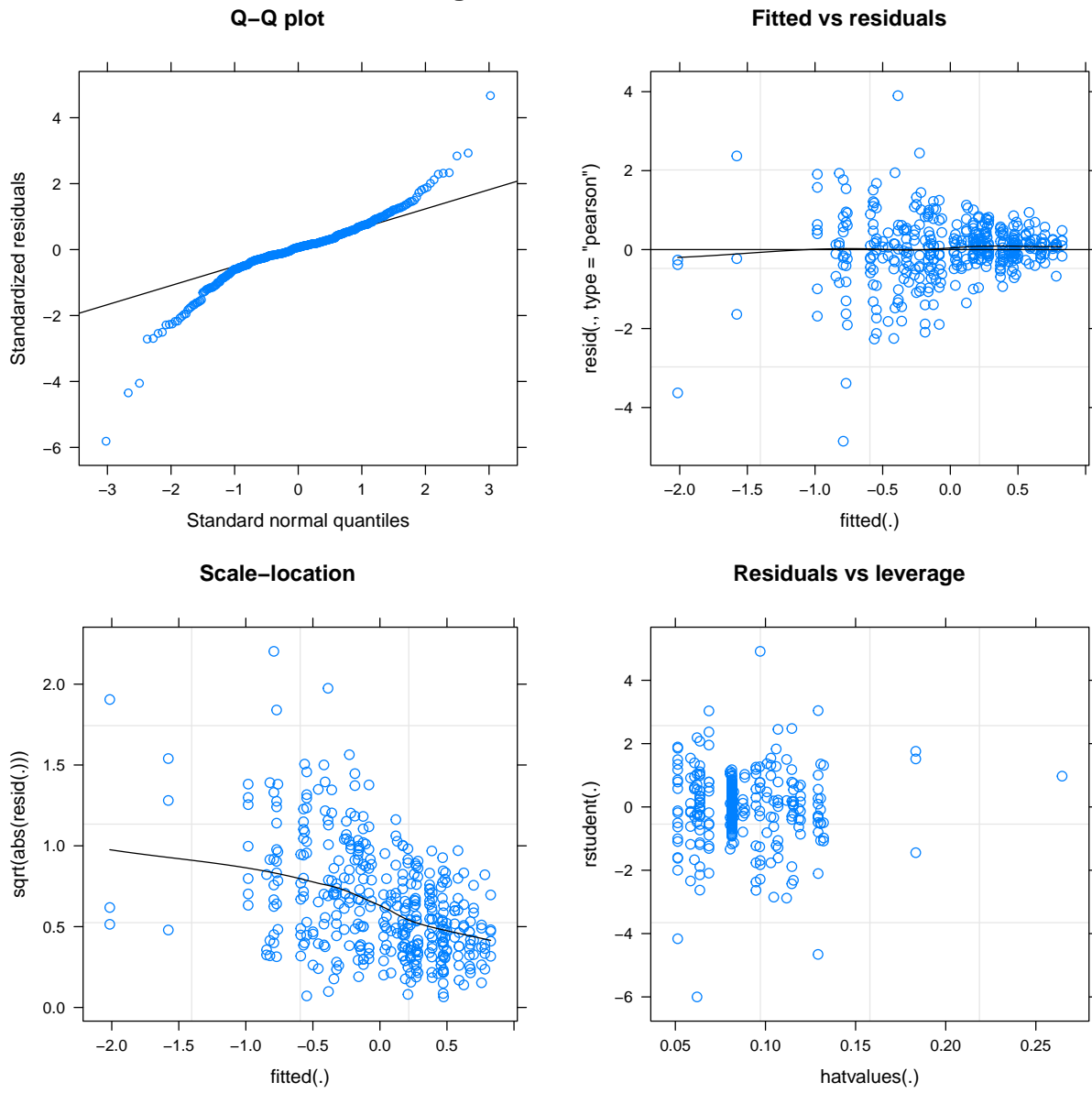
Residuals vs leverage



Root_dw



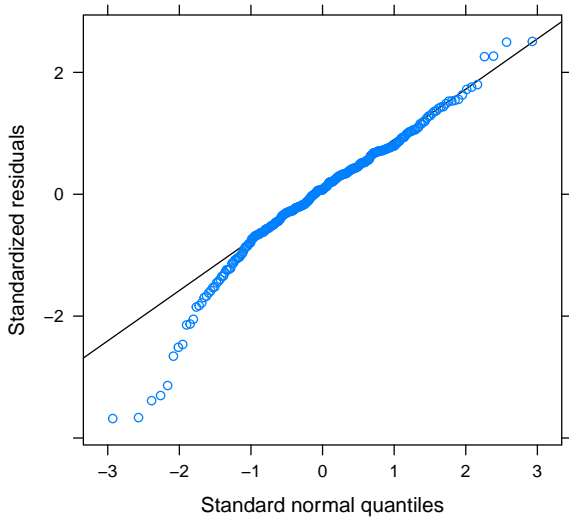
Log_root_shoot_ratio



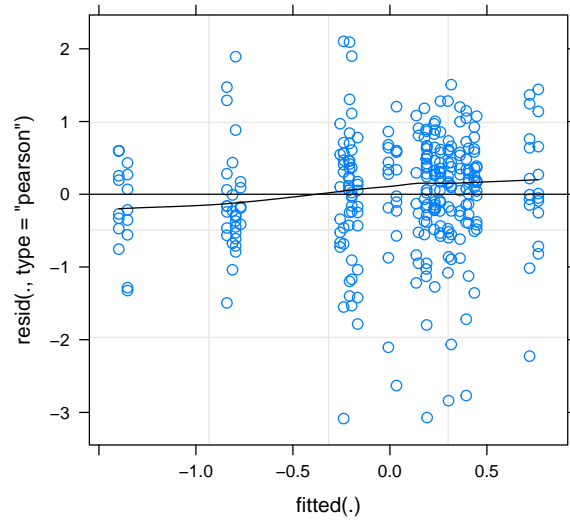
Then for Raphanus traits:

Height_1820

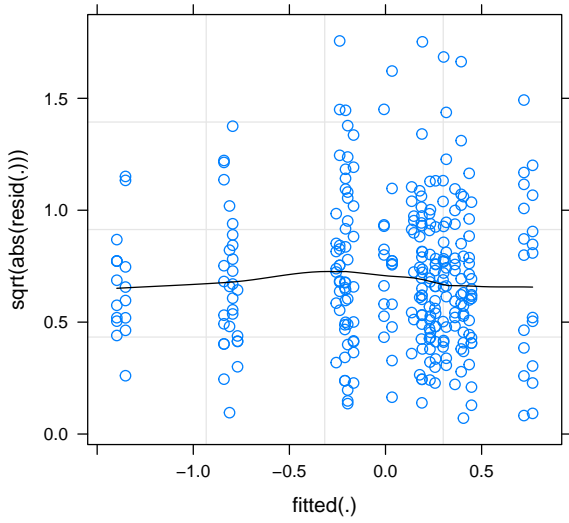
Q-Q plot



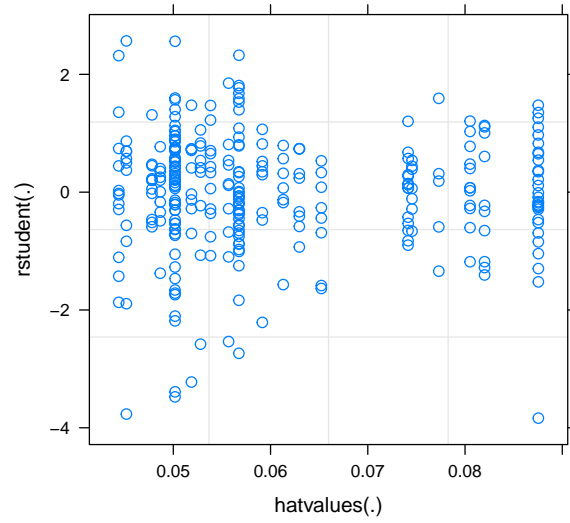
Fitted vs residuals



Scale-location

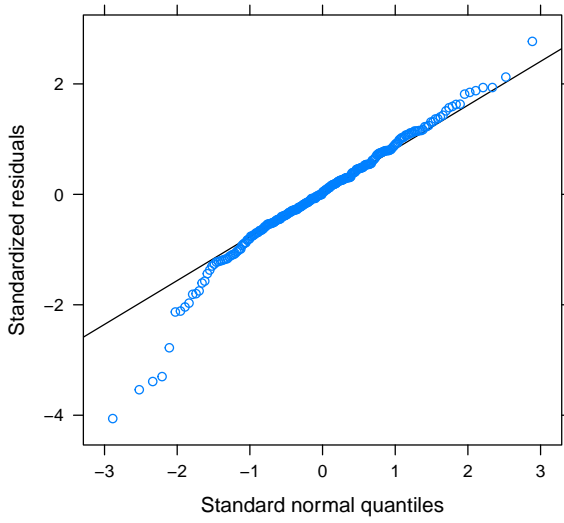


Residuals vs leverage

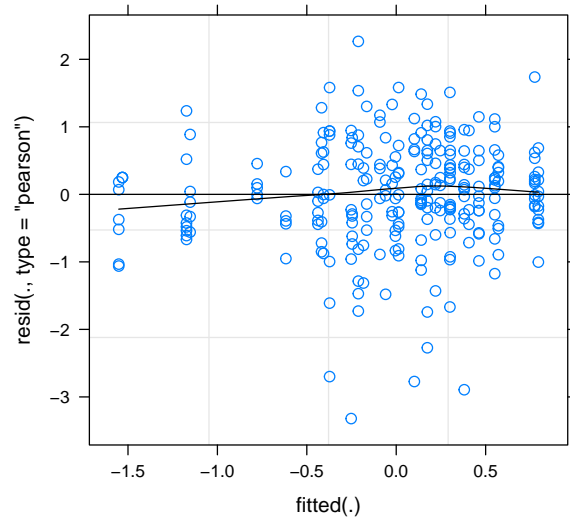


Height_2326

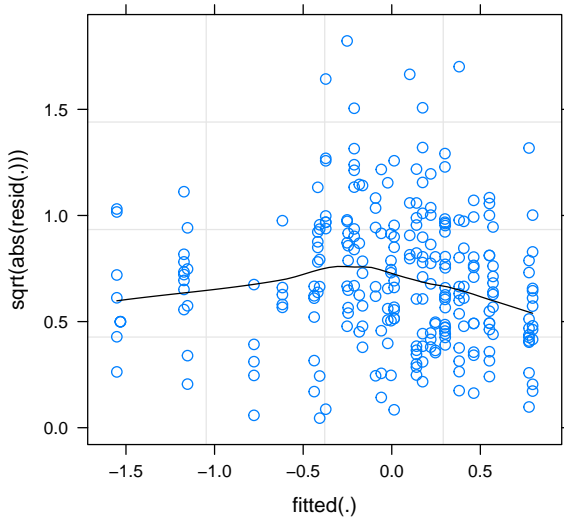
Q-Q plot



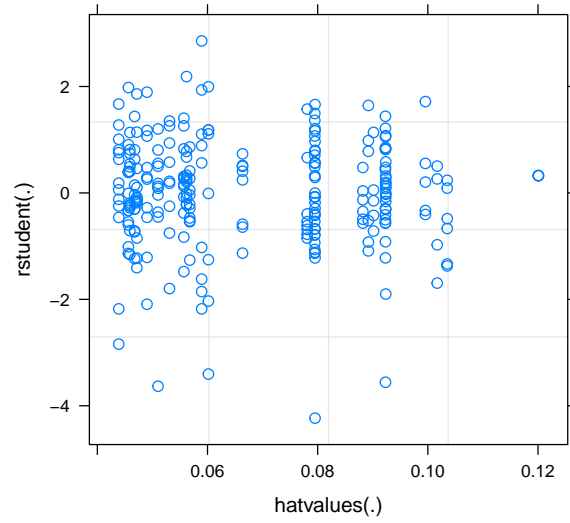
Fitted vs residuals



Scale-location

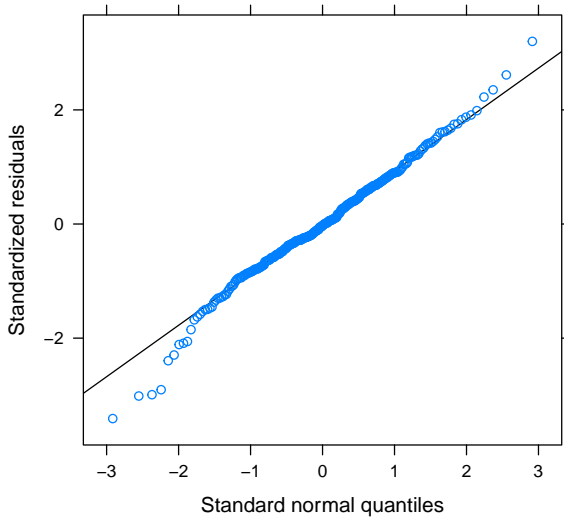


Residuals vs leverage

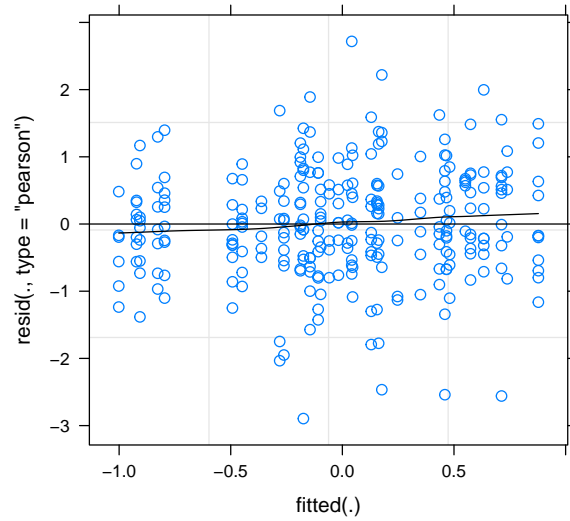


Leaf_length_1820

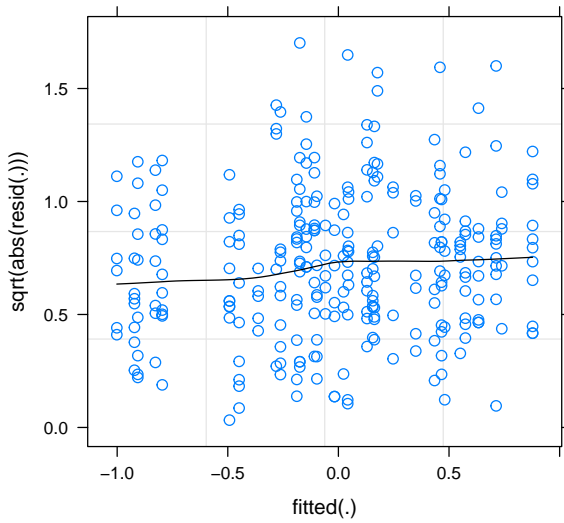
Q-Q plot



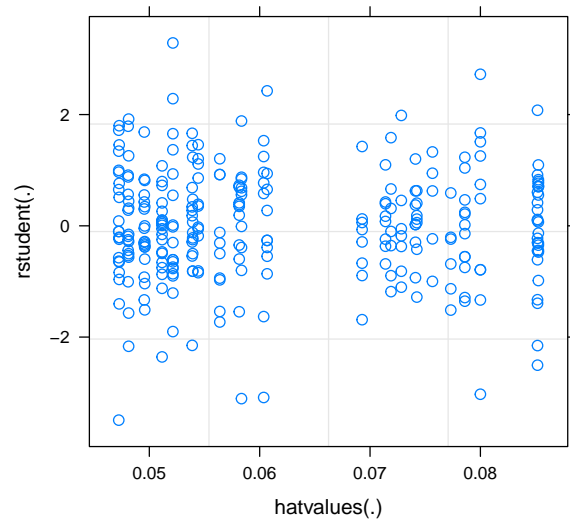
Fitted vs residuals



Scale-location

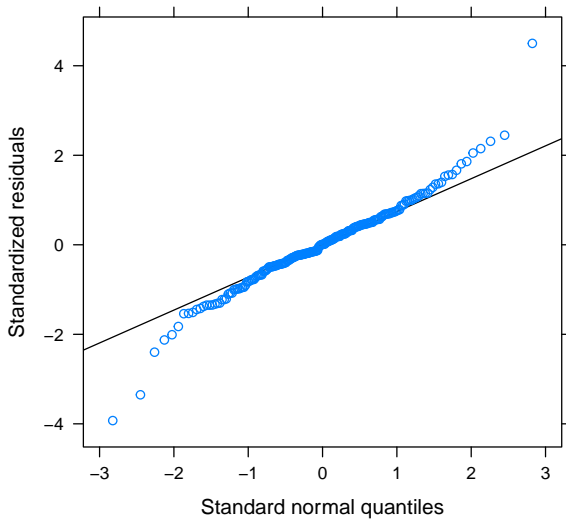


Residuals vs leverage

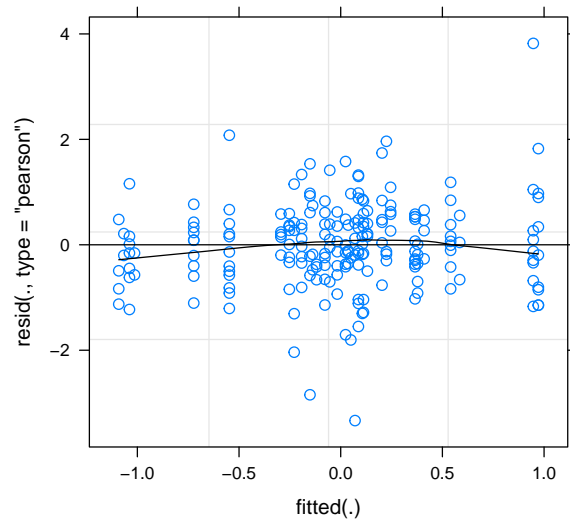


Leaf_length_2326

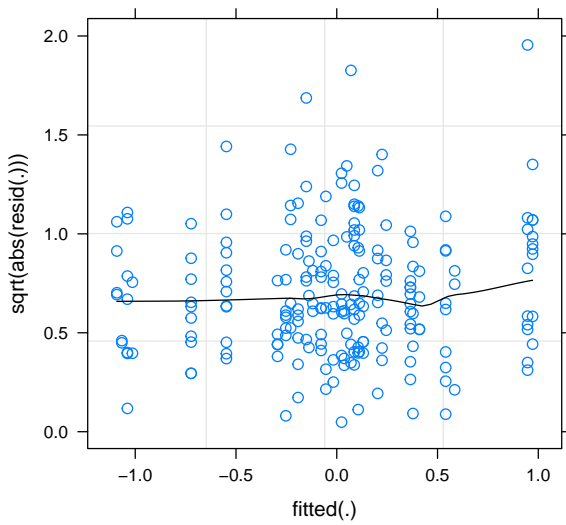
Q-Q plot



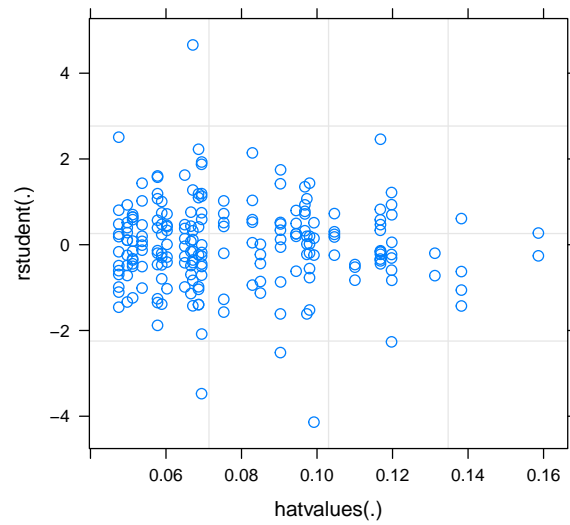
Fitted vs residuals



Scale-location

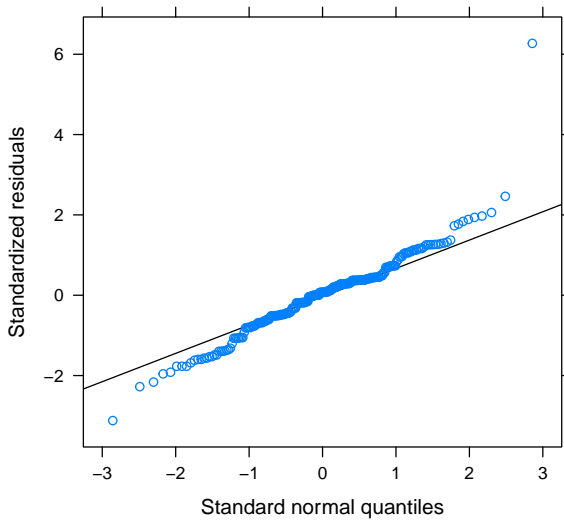


Residuals vs leverage

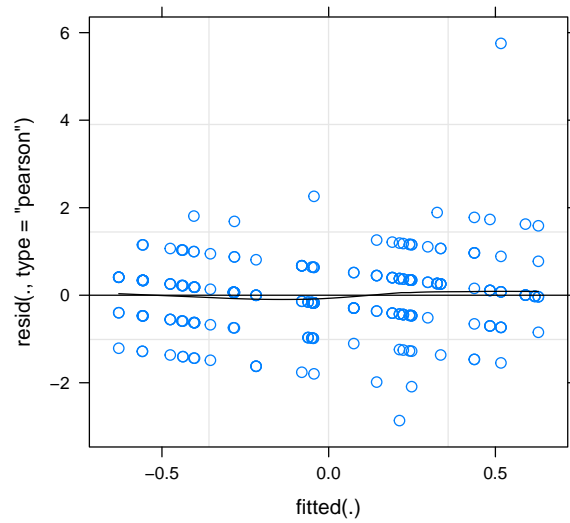


Num_leaves_1820

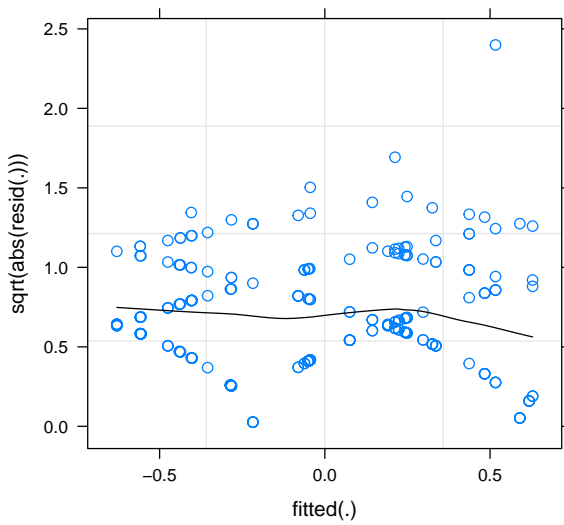
Q-Q plot



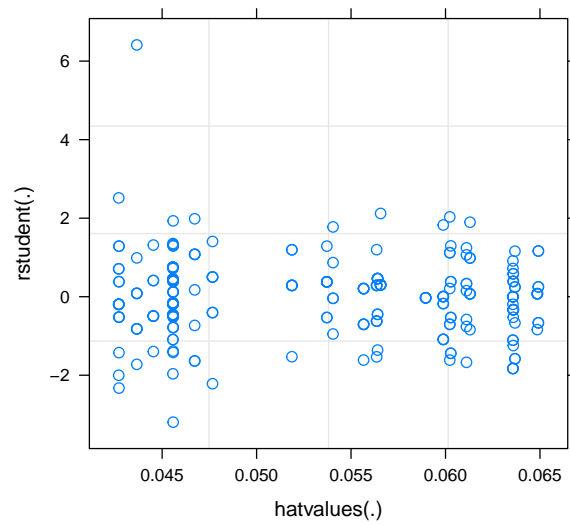
Fitted vs residuals



Scale-location

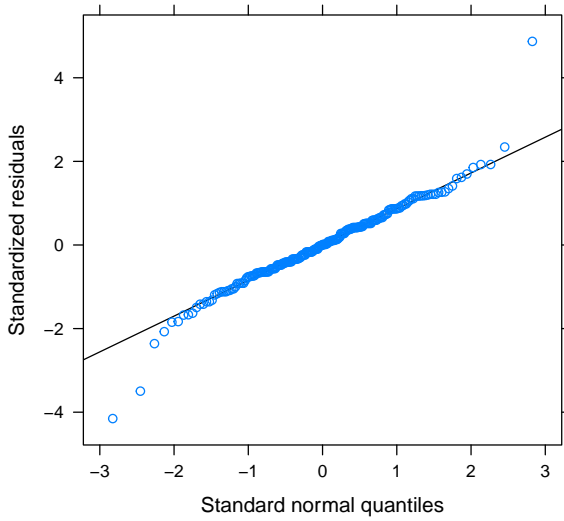


Residuals vs leverage

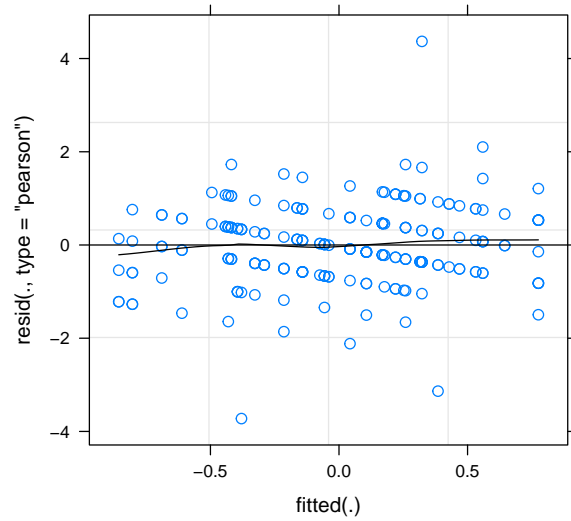


Num_leaves_2326

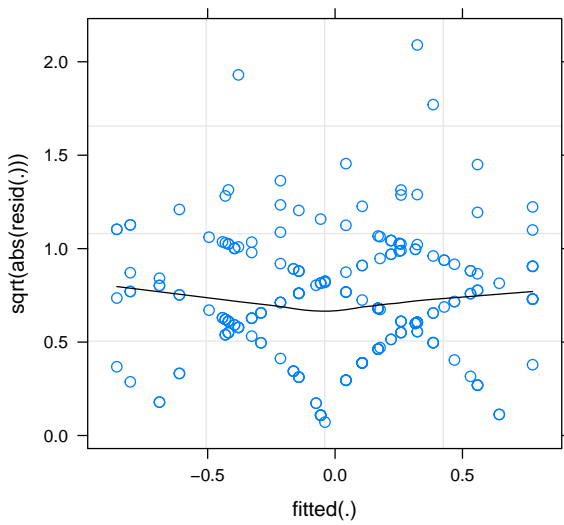
Q-Q plot



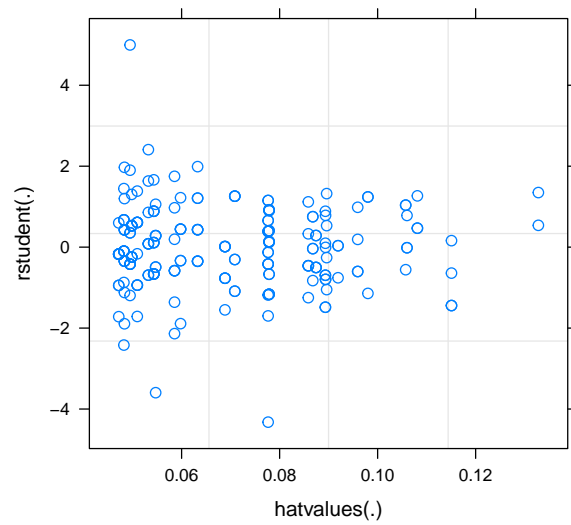
Fitted vs residuals



Scale-location

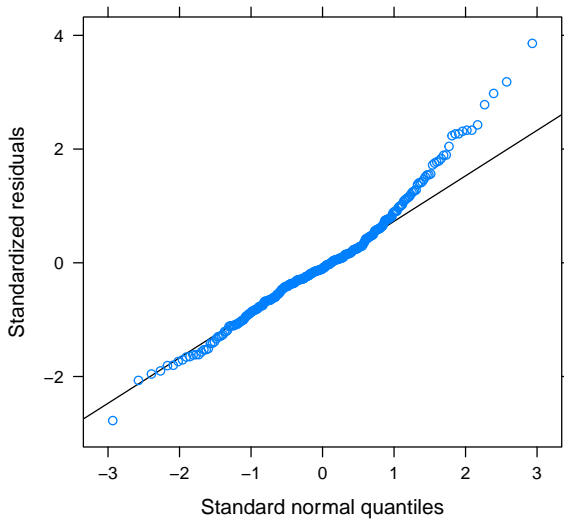


Residuals vs leverage

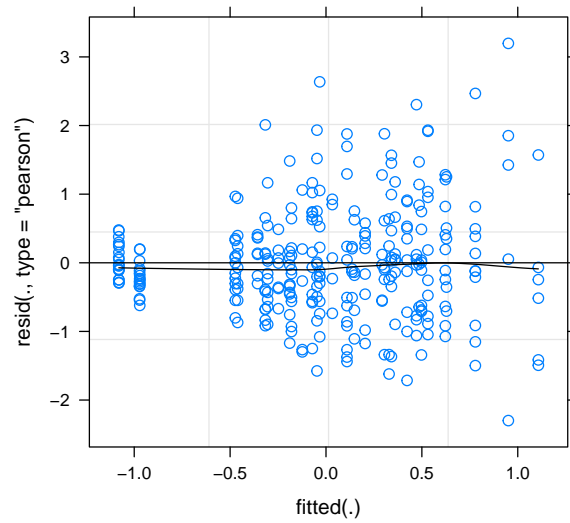


Sla_1820

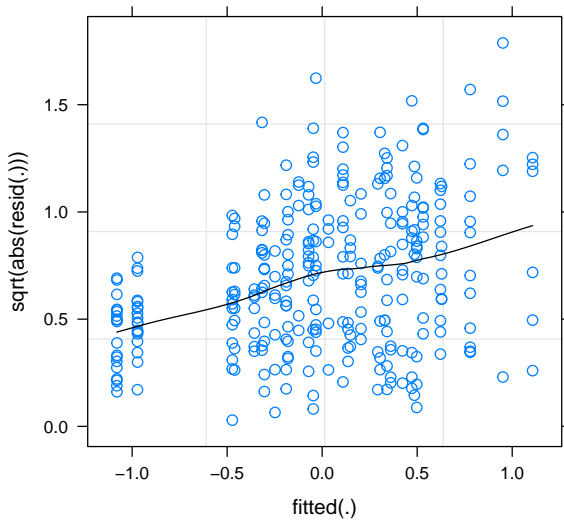
Q-Q plot



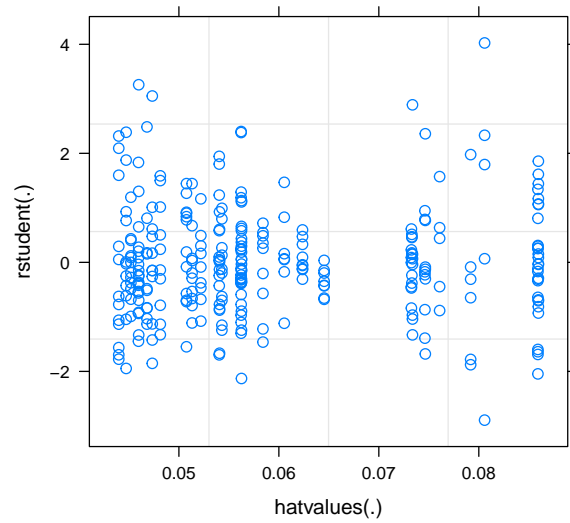
Fitted vs residuals



Scale-location

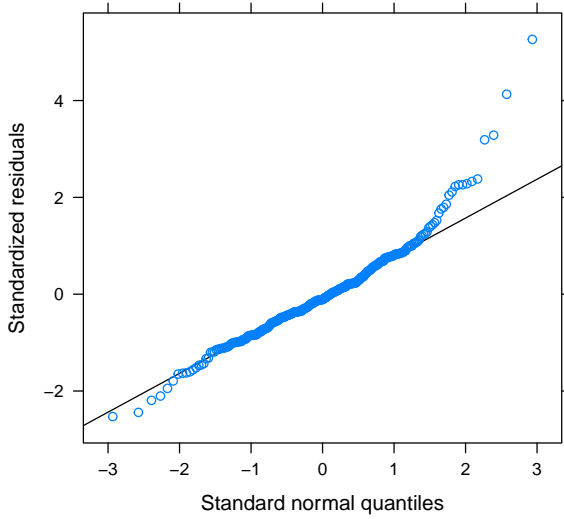


Residuals vs leverage

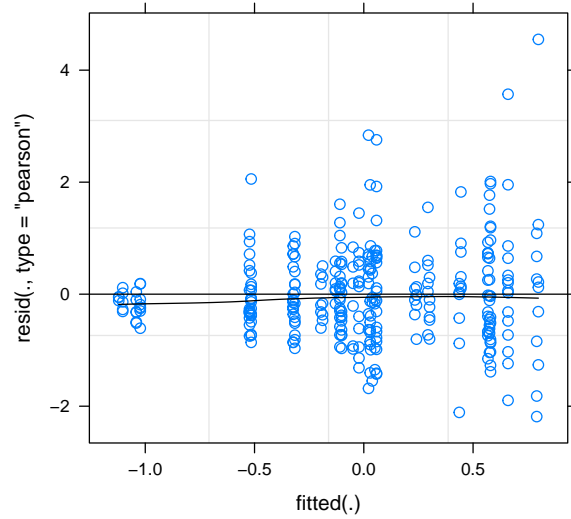


Sla_2326

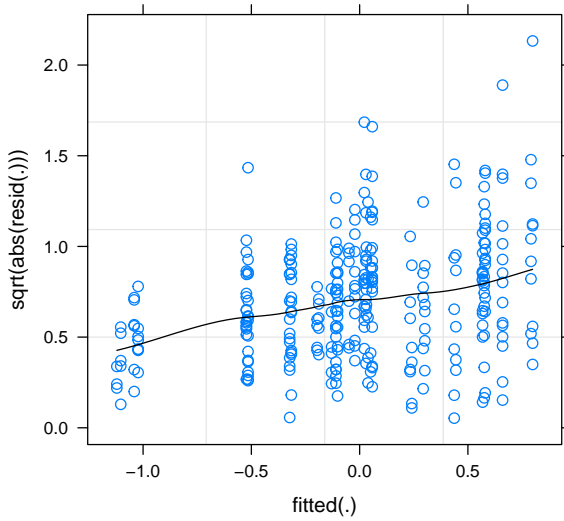
Q-Q plot



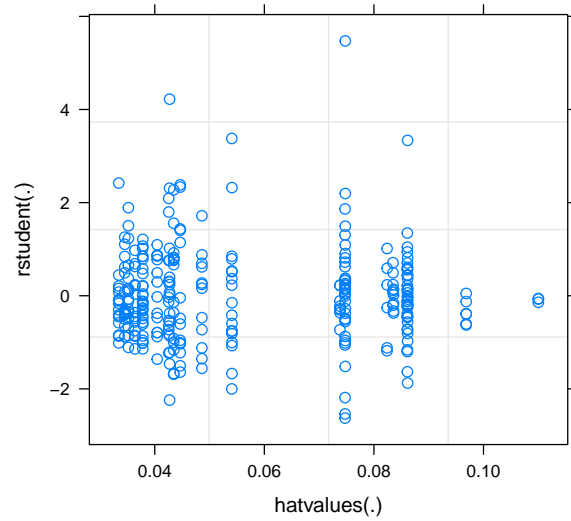
Fitted vs residuals



Scale-location

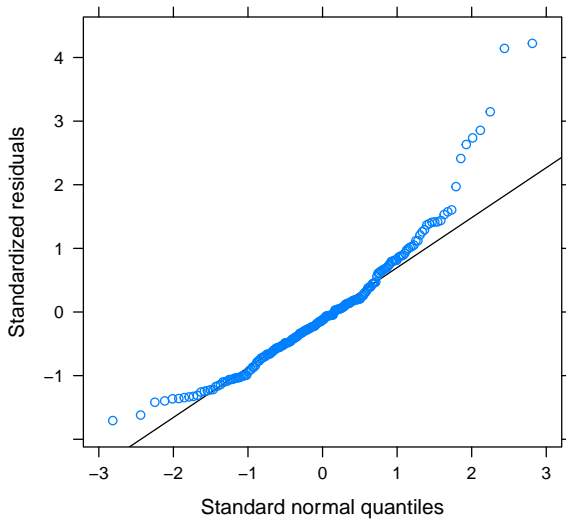


Residuals vs leverage

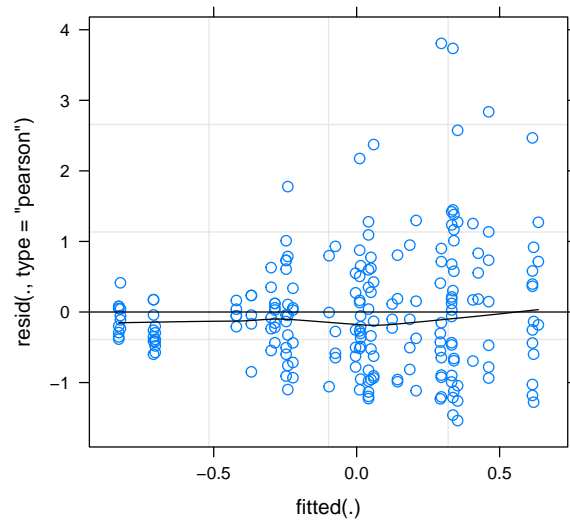


Sldw_1820

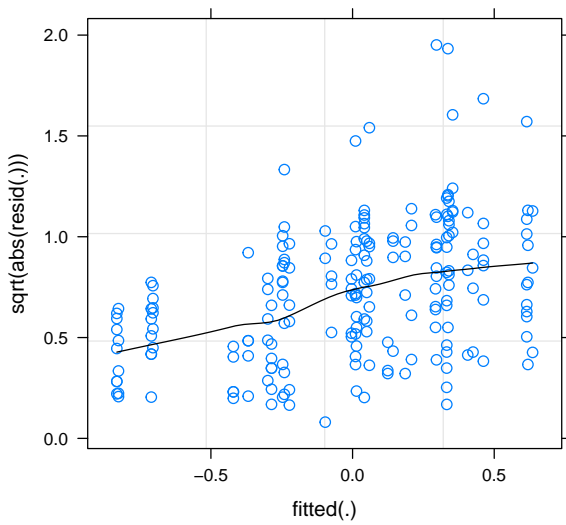
Q-Q plot



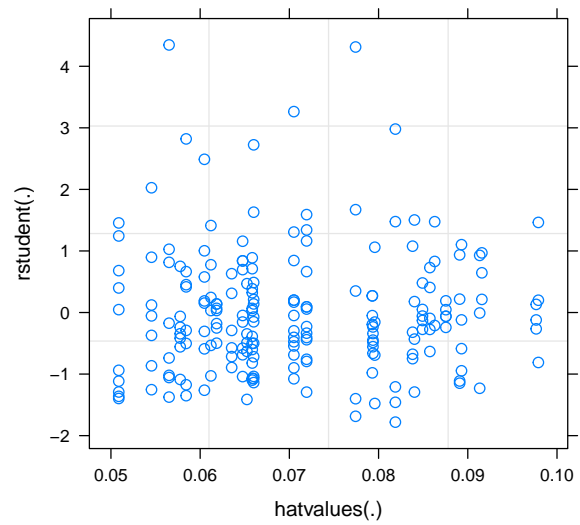
Fitted vs residuals



Scale-location

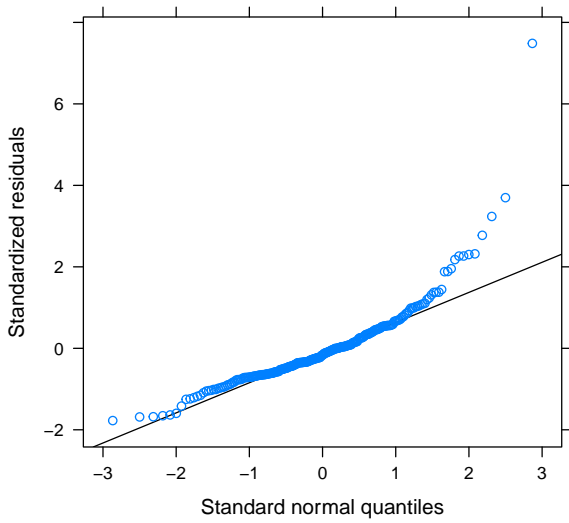


Residuals vs leverage

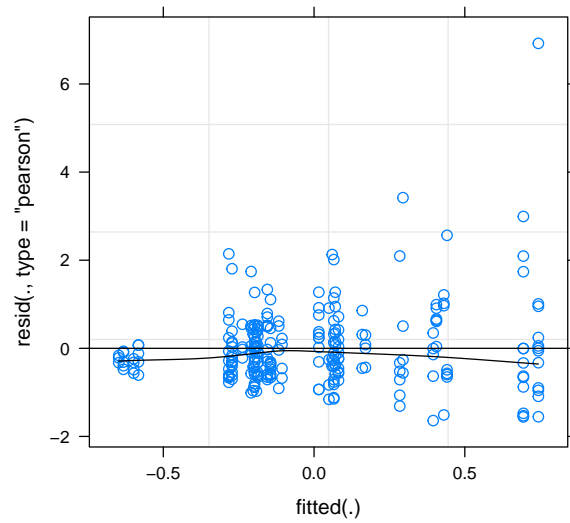


Slw_2326

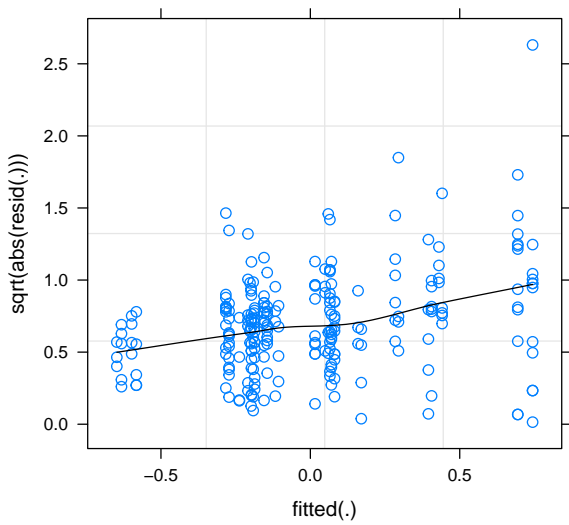
Q-Q plot



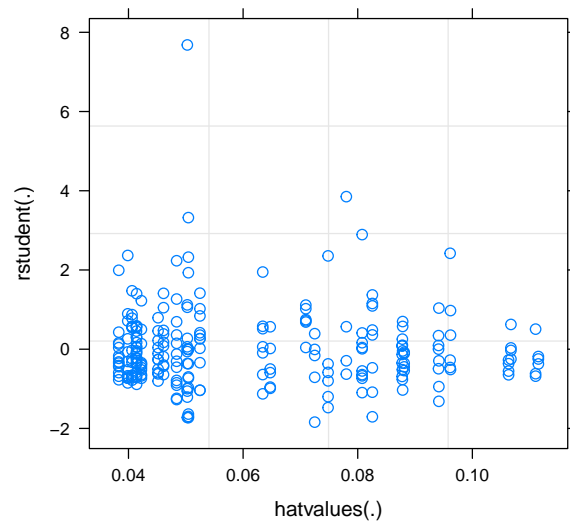
Fitted vs residuals



Scale-location

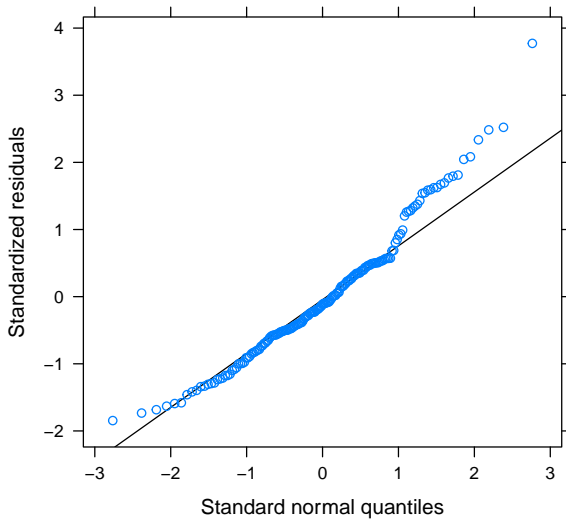


Residuals vs leverage

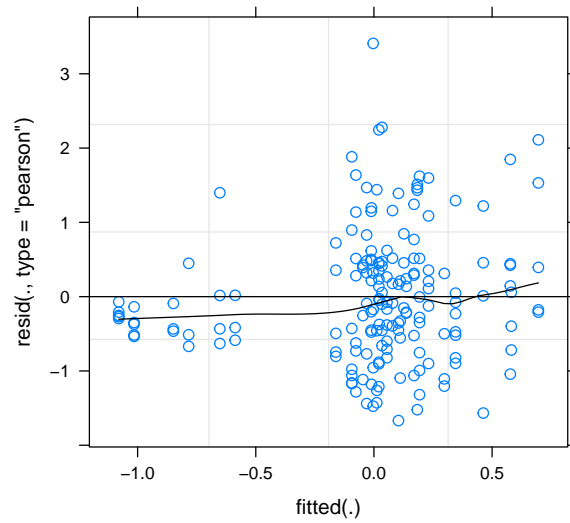


Aboveground_dw

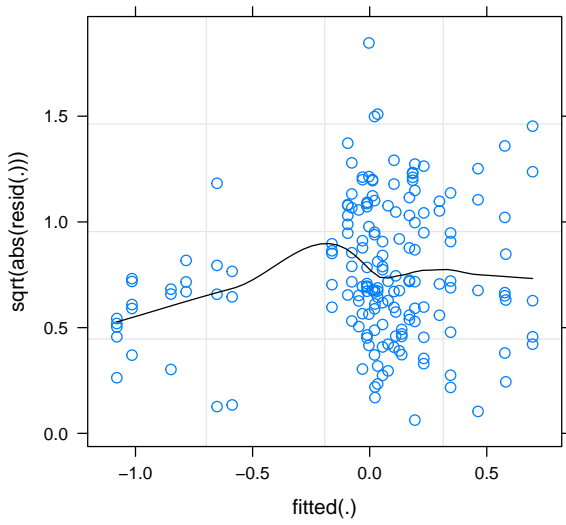
Q-Q plot



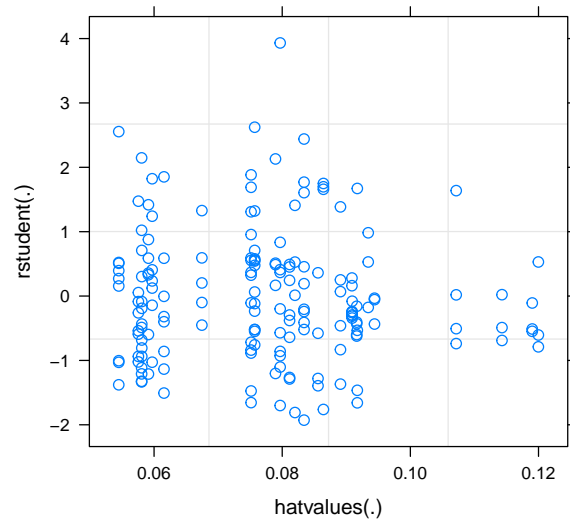
Fitted vs residuals



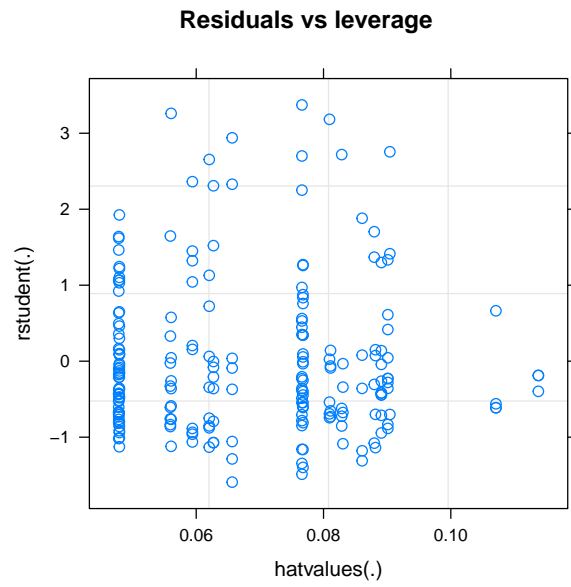
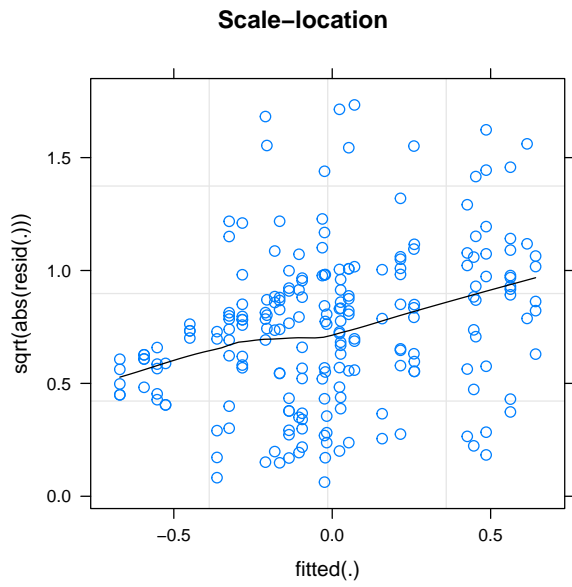
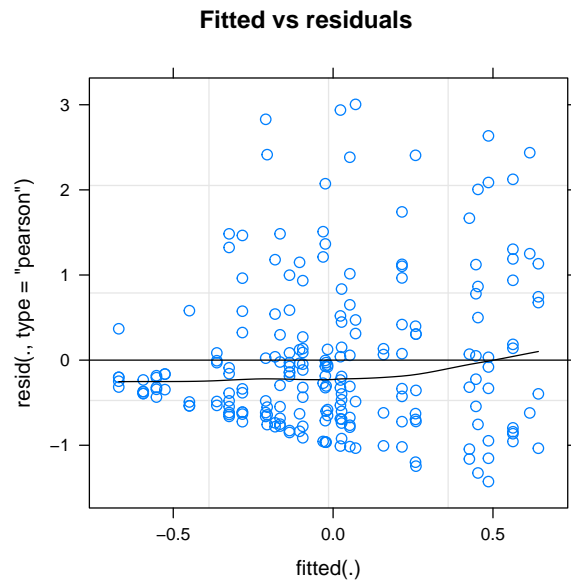
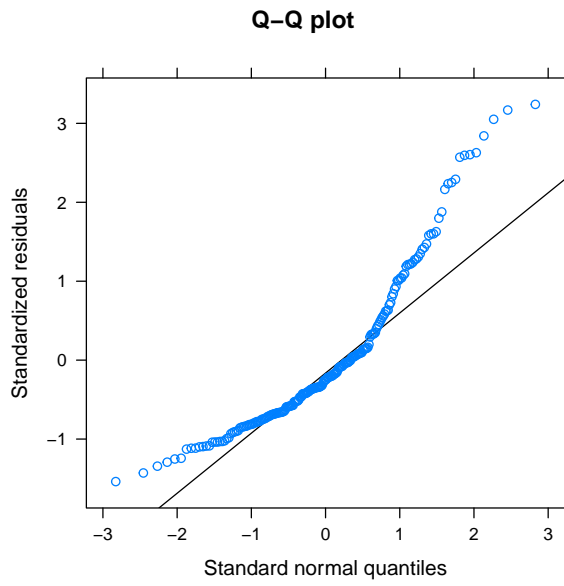
Scale-location

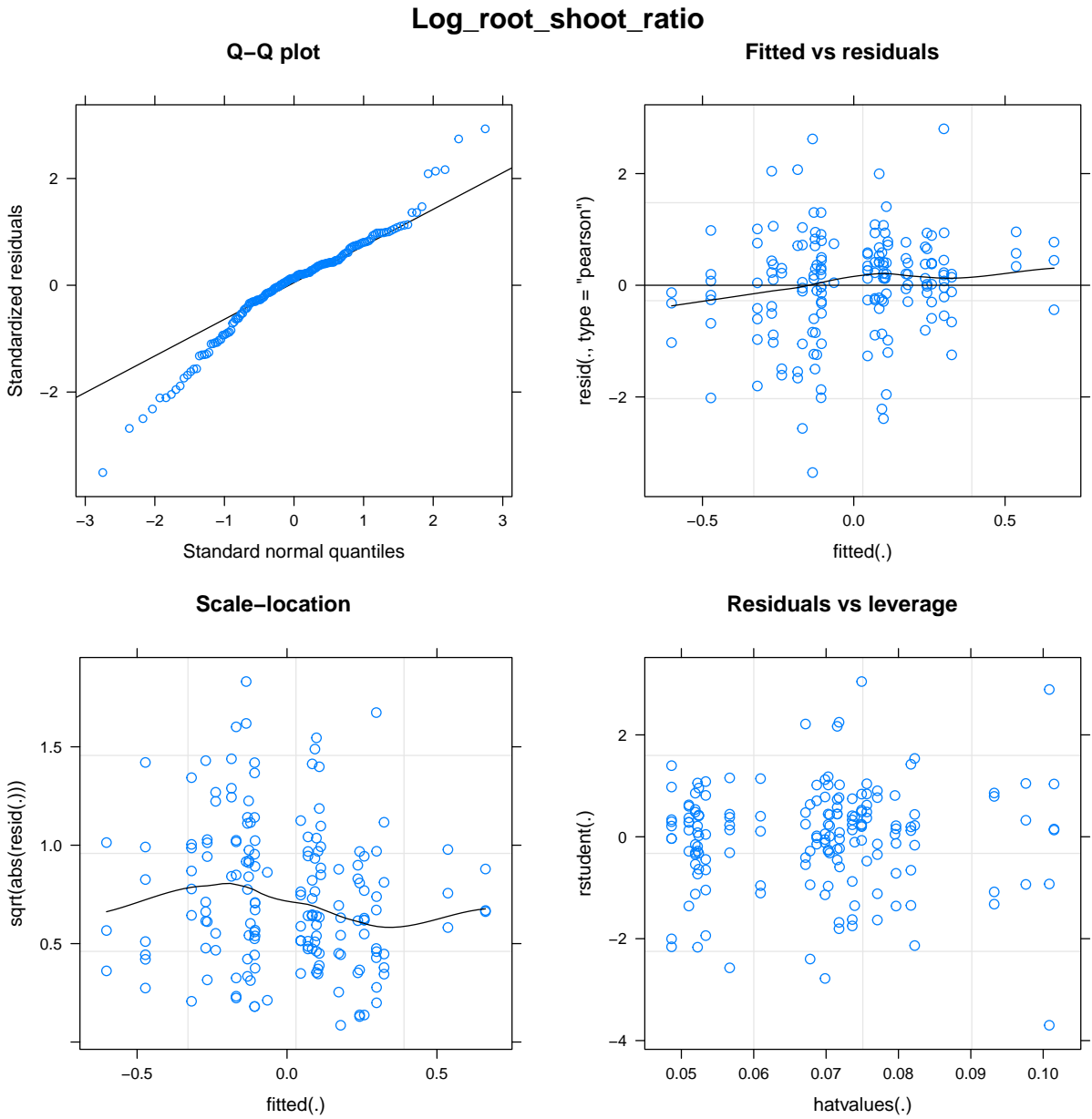


Residuals vs leverage



Root_dw

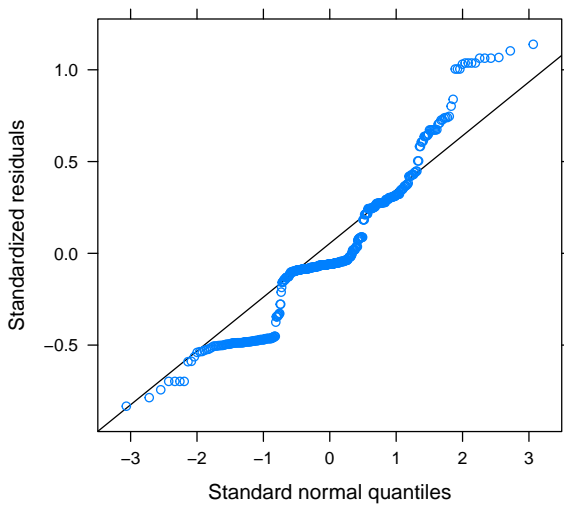




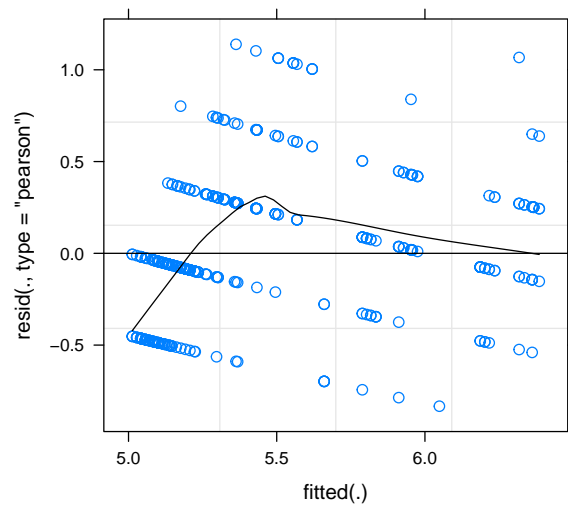
While none of these fits are perfect, the only massive issues are with number of leaves, since this is a discrete integer trait and thus fits weirdly to a gaussian distribution. For this trait, we might be better off fitting a Poisson distribution:

Num_leaves_1820

Q-Q plot

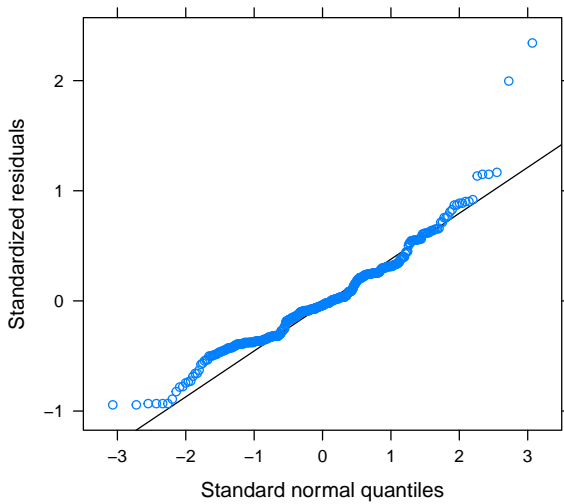


Fitted vs residuals

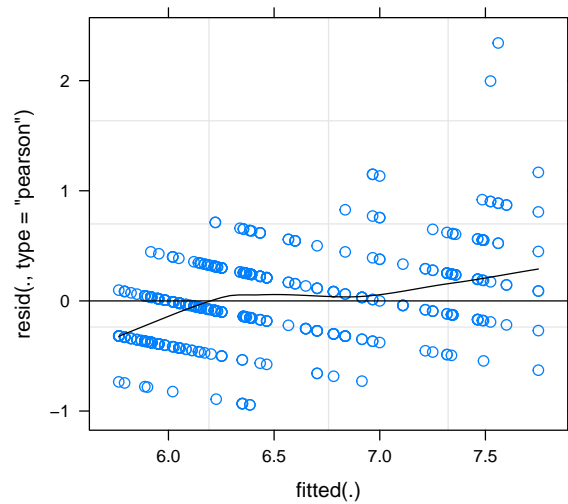


Num_leaves_2326

Q-Q plot

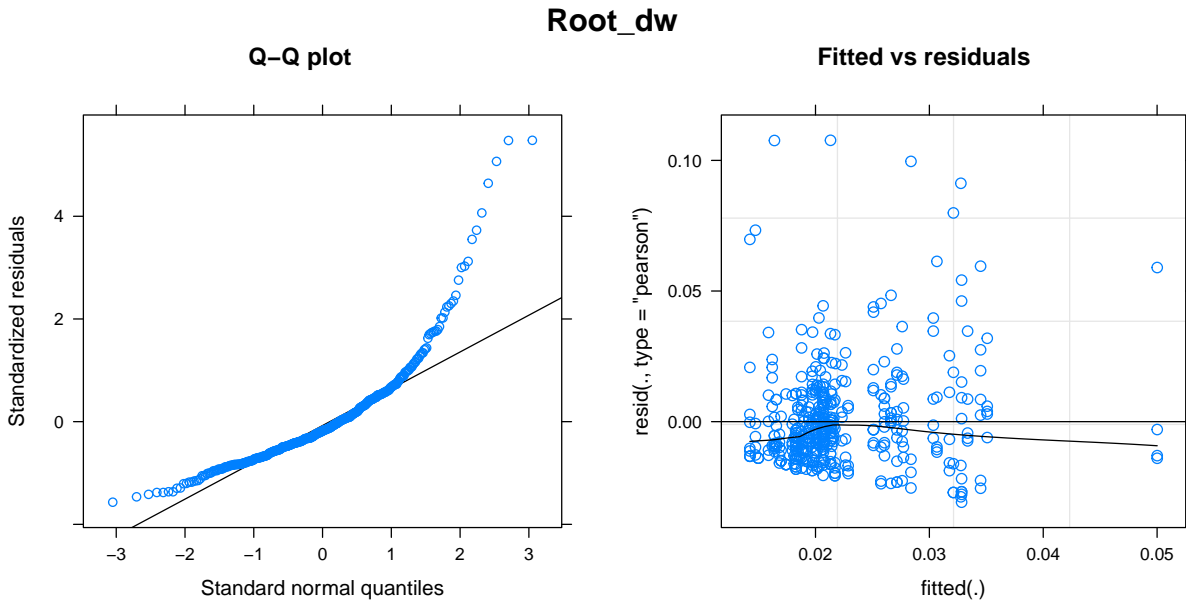


Fitted vs residuals

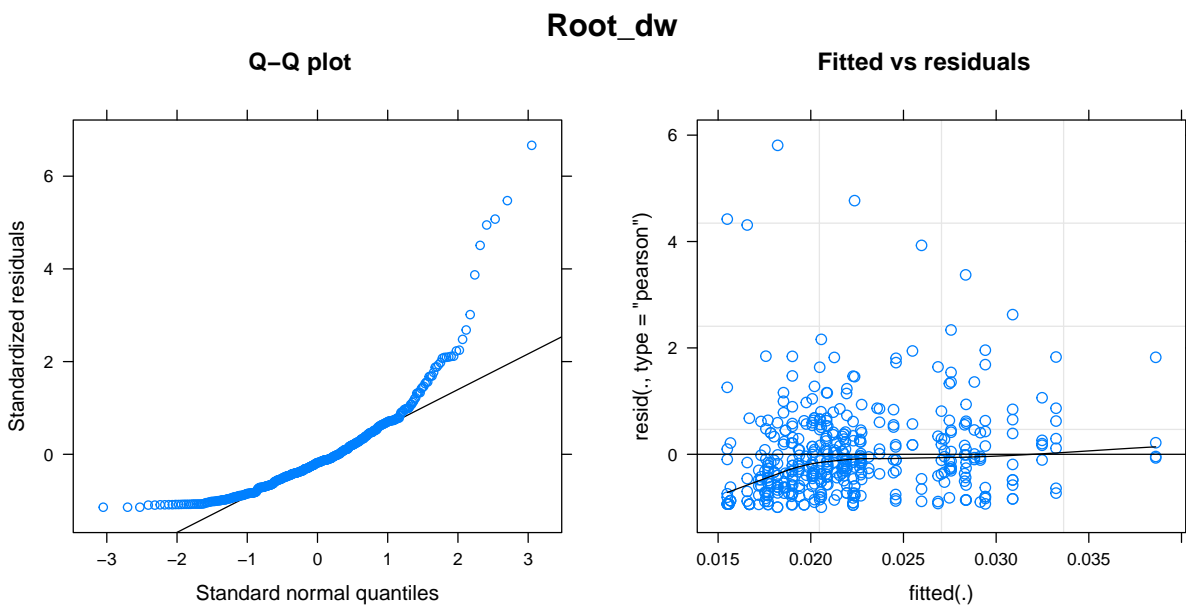


This doesn't necessarily look better, but it does make more sense (number of leaves is much more likely to be poisson distributed than gaussian, since it's a discrete trait) so for num leaves we'll use a poisson distribution going forward in place of a gaussian.

The other trait that looks odd in the above diagnostic plots is root dry-weight. We already know that this trait suffers from some skew, but we'd hoped to improve that by removing outliers earlier. Perhaps some kind of inverse link function could help with this?



That doesn't look any better. What about a gamma distribution?



Even worse! Okay, there doesn't seem to be any way to fit this trait nicely, so let's just model it with a gaussian distribution but take the results with a pinch of salt.

Parental effects testing

Before running any other models, we'd like to check if any traits exhibit strong differences between those accessions with parental effect standardization and those without. Note that for *Raphanus*, all species/subspecies have both standardised and unstandardised samples, so we can just compare all against all, whereas for *Brassica* standardisation is present only for some *B. rapa* (both wild and domesticated) so we need to subset before running the comparison.

First for the Brassicas:

```
## [1] "Parental effects p for Height_1820 in Brassica rapa: 0.00482"
## [1] "Parental effects p for Height_2326 in Brassica rapa: 0.00581"
## [1] "Parental effects p for Leaf_length_1820 in Brassica rapa: 0.0332"
## [1] "Parental effects p for Leaf_length_2326 in Brassica rapa: 0.00167"
## [1] "Parental effects p for Num_leaves_1820 in Brassica rapa: 0.268"
## [1] "Parental effects p for Num_leaves_2326 in Brassica rapa: 0.00483"
## [1] "Parental effects p for Sla_1820 in Brassica rapa: 0.0337"
## [1] "Parental effects p for Sla_2326 in Brassica rapa: 0.00257"
## [1] "Parental effects p for Sldw_1820 in Brassica rapa: 0.0129"
## [1] "Parental effects p for Sldw_2326 in Brassica rapa: 4.26e-05"
## [1] "Parental effects p for Aboveground_dw in Brassica rapa: 0.0149"
## [1] "Parental effects p for Root_dw in Brassica rapa: 0.00163"
## [1] "Parental effects p for Log_root_shoot_ratio in Brassica rapa: 0.0986"
```

And for Raphanus:

```
## [1] "Parental effects p for Height_1820 in Raphanus: 0.71"
## [1] "Parental effects p for Height_2326 in Raphanus: 0.0032"
## [1] "Parental effects p for Leaf_length_1820 in Raphanus: 0.0562"
## [1] "Parental effects p for Leaf_length_2326 in Raphanus: 0.00118"
## [1] "Parental effects p for Num_leaves_1820 in Raphanus: 0.56"
## [1] "Parental effects p for Num_leaves_2326 in Raphanus: 0.0812"
## [1] "Parental effects p for Sla_1820 in Raphanus: 0.00553"
## [1] "Parental effects p for Sla_2326 in Raphanus: 6.41e-05"
## [1] "Parental effects p for Sldw_1820 in Raphanus: 0.0458"
## [1] "Parental effects p for Sldw_2326 in Raphanus: 0.0211"
## [1] "Parental effects p for Aboveground_dw in Raphanus: 0.804"
## [1] "Parental effects p for Root_dw in Raphanus: 0.221"
## [1] "Parental effects p for Log_root_shoot_ratio in Raphanus: 0.0241"
```

The large majority of traits in both Raphanus and Brassica display some kind of parental effect, so we will have to include parental standardisation status as a random effect in our subsequent models.

Work package 1: Comparing domesticates to their wild progenitors

For this work package, we wish to compare wild progenitor species to modern domesticates. For Brassica, the comparison is between domesticated B rapa with wild B rapa, while for Raphanus the comparison is between R raphanistrum (wild) and all other Raphanus samples, all of which appear to be domesticated. Our analytic design is:

expression ~ *domestication*+*treatment*+*domestication*treatment*+(1|*population*)+(1|*parental standardization status*)

Intercept	0.79 (0.061)	0.51 (-0.27)	0.51 (-0.22)	0.5 (-0.45)	2.8e-216 (1.8)	2.2e-1 (1.9)
Cultivated_environment	0.49 (-0.19)	0.23 (0.35)	0.91 (0.018)	0.33 (0.24)	0.91 (0.015)	0.91 (0.025)
Domesticated_history	0.88 (-0.26)	0.92 (0.078)	0.88 (0.26)	0.88 (0.38)	0.91 (-0.037)	0.88 (0.052)
Interaction	0.62 (0.47)	0.96 (-0.21)	0.96 (0.19)	0.97 (0.019)	0.97 (0.054)	0.97 (-0.03)
	Height_1820	Height_2326	Leaf_length_1820	Leaf_length_2326	Num_leaves_1820	Num_leaves_2326

Our primary interest is in the bottom row, since significant terms here will indicate traits that exhibit differential plasticity in domesticated vs wild rapa. However, none of these interaction terms are significant. Nor, indeed, are any traits significantly different between wild and domesticated rapa overall, which is very surprising!

Let's try the same analysis for *Raphanus*, comparing *R. raphanistrum* to the various domesticated *Raphanus* using the same model.

Intercept	0.99 (0.004)	0.99 (−0.038)	0.99 (−0.039)	0.99 (0.0042)	0.99 (−0.0026)	0.99 (0.12)
Cultivated_environment	0.96 (0.044)	0.96 (−0.0082)	0.96 (−0.08)	0.96 (−0.036)	0.96 (−0.2)	0.96 (−0.36)
Domesticated_history	0.93 (0.23)	0.93 (0.29)	0.99 (−0.0042)	0.93 (−0.15)	0.93 (−0.2)	0.93 (−0.05)
Interaction	0.9 (−0.036)	0.87 (0.15)	0.69 (0.39)	0.76 (0.35)	0.13 (0.89)	0.13 (0.78)
	Height_1820	Height_2326	Leaf_length_1820	Leaf_length_2326	Num_leaves_1820	Num_leaves_2326

Again, we reach the unfortunate conclusion that there seem to be no significant terms in any of the models. Might we find some if we subset by the separate *Raphanus* domesticates, rather than combining them all into one? Below we plot the interaction terms only for each comparison.

Raphanus_sativus_var_caudatus	0.98 (-0.0062)	0.95 (0.17)	0.95 (0.39)	0.95 (0.1)	0.95 (0.068)
Raphanus_sativus	0.95 (-0.036)	0.95 (0.15)	0.95 (0.39)	0.95 (0.35)	0.95 (0.14)
Raphanus_raphanistrum_munra	0.95 (0.042)	0.95 (0.27)	0.95 (0.36)	0.95 (-0.067)	0.95 (0.072)

To the contrary, subsetting by domesticate further erases any traces of interaction significance that we might have found for *Raphanus* (not least because we have to apply FDR correction over $3 \times 13 = 39$ comparisons here).

We must therefore conclude that domestication hasn't selected for greater phenotypic plasticity in domesticates relative to their wild progenitors.

Work package 2: Comparing wild progenitors to never-domesticated wild species

We now wish to ask a different question: do ‘progenitor species’ (i.e. those that were the original targets of human selection) differ systematically in their patterns of phenotypic expression from wild species that were never domesticated? We are particularly interested in knowing whether progenitor species express greater phenotypic plasticity than their never-domesticated relatives, since this could suggest that early human farmers either intentionally or inadvertently targeted species that were particularly receptive to the benefits of cultivation.

Unfortunately, for *Raphanus* we have no data for non-domesticated wild species, so this analysis will be restricted to *Brassica*, comparing wild *Brassica rapa* to a number of never-domesticated *Brassic*as. We’ll begin by running a lmer/glmer model as before, except that instead of comparing wild to domesticated, we’ll compare progenitors to non-progenitors.

Intercept	0.59 (-0.24)	0.88 (-0.054)	0.7 (0.17)	0.59 (0.41)	0.041 (-0.4)	0.59 (-0.16)
Cultivated_environment	0.58 (0.12)	0.99 (-0.0012)	0.58 (0.16)	0.67 (-0.053)	0.6 (0.067)	0.58 (-0.11)
Wild_progenitor	0.019 (0.88)	0.16 (0.47)	0.25 (-0.41)	0.021 (-1.1)	0.00046 (1.2)	0.17 (0.49)
Interaction	0.11 (-0.34)	0.085 (0.38)	0.54 (-0.13)	0.11 (0.34)	0.85 (0.028)	0.25 (0.23)
	Height_1820	Height_2326	Leaf_length_1820	Leaf_length_2326	Num_leaves_1820	Num_leaves_2326

This is somewhat more promising: many traits differ in their expression between *B. rapa* and the other wild species, and several of the interaction terms are near-significant. However, to be sure we'll again need to subset by species, since we once again may have lost power by combining distinct species. Below we plot only the interaction terms for each never-domesticated wild:

Brassica_villosa	0.64 (-0.21)	0.5 (0.36)	0.96 (0.013)	0.42 (0.4)	0.84 (0.077)	0.63 (0.36)
Brassica_rupestris	0.76 (-0.14)	0.5 (0.35)	0.96 (-0.083)	0.61 (0.37)	0.84 (-0.31)	0.74 (0.13)
Brassica_montana	0.47 (-0.4)	0.28 (0.5)	0.96 (-0.2)	0.14 (0.62)	0.84 (-0.11)	0.5 (0.4)
Brassica_macrocarpa	0.61 (-0.43)	0.9 (0.052)	0.96 (-0.31)	0.72 (0.14)	0.84 (0.3)	0.74 (0.18)
Brassica_incana	0.64 (-0.18)	0.5 (0.22)	0.96 (0.12)	0.72 (0.094)	0.84 (0.1)	0.74 (0.16)
Brassica_cretica	0.47 (-0.46)	0.5 (0.36)	0.96 (-0.44)	0.61 (0.3)	0.84 (0.061)	0.74 (-0.16)
	Height_1820	Height_2326	Leaf_length_1820	Leaf_length_2326	Num_leaves_1820	Num_leaves_2326

Unfortunately, as with the *Raphanus domesticates*, subsetting by wild species in *Brassica* results in a total loss of interesting interaction effects. We must therefore conclude that *Brassica rapa* does not display greater phenotypic plasticity than other wild *Brassica* species that were never domesticated.