

# Final models

Tim White

## Packages, data, and functions

```
library(car)
library(tidyverse)
library(knitr)

data_stems <- read_csv("data/data_stems.csv") %>%
  mutate(plot_id = as.factor(plot_id),
         size_class = as.factor(size_class),
         harvested = as.factor(harvested),
         vegetation_type = as.factor(vegetation_type),
         milpa = as.factor(milpa))
data_plots <- read_csv("data/data_plots.csv") %>%
  mutate(plot_id = as.factor(plot_id),
         harvested = as.factor(harvested),
         vegetation_type = as.factor(vegetation_type),
         milpa = as.factor(milpa))

summarize_glm <- function(mod) {
  summary <- summary(mod)
  table <- tibble(var = c("intercept", "harvestedYes", "vegetationKeelenche",
                         "vegetationNukuuchche", "milpaYes", "latitude", "longitude"),
                  coef = summary$coefficients[,1], se = summary$coefficients[,2]) %>%
    mutate_if(is.numeric, function(col) {round(col, 3)}) %>%
    kable(col.names = c("", "Coefficient", "SE"), align = "c")
  return(table)
}

analysis_of_deviance <- function(mod, p_adj) {
  summary <- Anova(mod, type = 2, test.statistic = "LR")
  table <- tibble(var = c("harvested", "vegetation", "milpa", "latitude", "longitude"),
                  chisq = summary$`LR Chisq`, df = summary$Df, p = summary$`Pr(>Chisq)`, p_adj) %>%
    mutate_if(is.numeric, function(col) {round(col, 3)}) %>%
    kable(col.names = c("", "Chisq", "df", "p-value", "adjusted p-value"), align = "c")
  return(table)
}
```

## Fit models and adjust p-values for analysis of deviance tables

```
stemden_mod <- MASS::glm.nb(stemden_totaltrees_count ~ harvested + vegetation_type +
                                milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

stemden_mod_seedlings <- MASS::glm.nb(stemden_seedlings_count ~ harvested + vegetation_type +
                                         milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

stemden_mod_saplings <- MASS::glm.nb(stemden_saplings_count ~ harvested + vegetation_type +
                                         milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

stemden_mod_trees05to09 <- MASS::glm.nb(stemden_trees05to09_count ~ harvested + vegetation_type +
                                         milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

stemden_mod_trees10to14 <- MASS::glm.nb(stemden_trees10to14_count ~ harvested + vegetation_type +
                                         milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

stemden_mod_trees15plus <- MASS::glm.nb(stemden_trees15plus_count ~ harvested + vegetation_type +
                                         milpa + latitude + longitude + offset(log(plots_per_ha)), data = data_plots)

ba_mod <- glm(ba_totaltrees ~ harvested + vegetation_type + milpa + latitude + longitude,
               family = Gamma(link="log"), data = data_plots)

ba_mod_saplings <- glm(ba_saplings ~ harvested + vegetation_type + milpa +
                           latitude + longitude, family = Gamma(link="log"), data = data_plots)

ba_mod_trees05to09 <- glm(ba_trees05to09 ~ harvested + vegetation_type + milpa +
                           latitude + longitude, family = Gamma(link="log"), data = data_plots)

ba_mod_trees10to14 <- glm(ba_trees10to14 ~ harvested + vegetation_type + milpa +
                           latitude + longitude, family = Gamma(link="log"), data = data_plots)

ba_mod_trees15plus <- glm(ba_trees15plus ~ harvested + vegetation_type + milpa +
                           latitude + longitude, family = Gamma(link="log"),
                           data = data_plots %>% filter(stemden_trees15plus_count > 0))

p_values <- c(Anova(stemden_mod, type = 2, test.statistic = "LR")[,3],
             Anova(stemden_mod_seedlings, type = 2, test.statistic = "LR")[,3],
             Anova(stemden_mod_saplings, type = 2, test.statistic = "LR")[,3],
             Anova(stemden_mod_trees05to09, type = 2, test.statistic = "LR")[,3],
             Anova(stemden_mod_trees10to14, type = 2, test.statistic = "LR")[,3],
             Anova(stemden_mod_trees15plus, type = 2, test.statistic = "LR")[,3],
             Anova(ba_mod, type = 2, test.statistic = "LR")[,3],
             Anova(ba_mod_saplings, type = 2, test.statistic = "LR")[,3],
             Anova(ba_mod_trees05to09, type = 2, test.statistic = "LR")[,3],
             Anova(ba_mod_trees10to14, type = 2, test.statistic = "LR")[,3],
             Anova(ba_mod_trees15plus, type = 2, test.statistic = "LR")[,3])

# Apply Benjamini-Hochberg procedure
adjusted_p_values <- p.adjust(p_values, method = "BH")
```

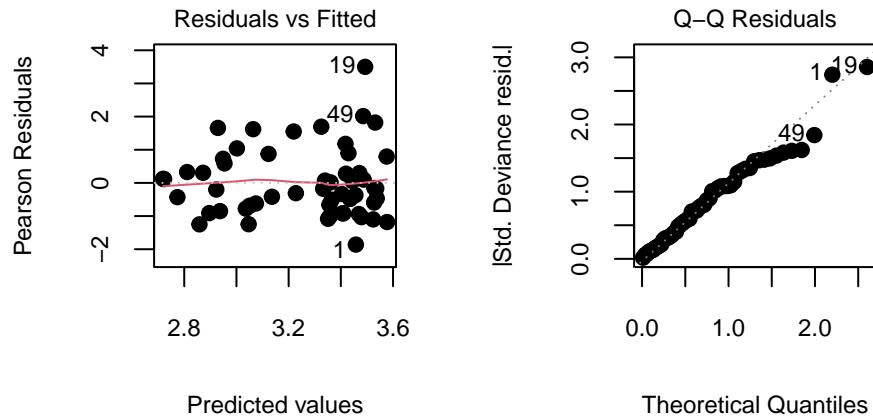
# Stem density

## Model

$$\log E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \log(\text{plots per ha})$$

where  $Y$  is the total number of stems per plot and  $X = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ , where  $X_1$  is a binary variable indicating whether a plot has been harvested,  $X_2$  and  $X_3$  are indicator variables for the Keelenche and Nukuuchche vegetation types,  $X_4$  is a binary variable indicating whether a plot has been exposed to milpa,  $X_5$  is latitude, and  $X_6$  is longitude. We set  $\text{Var}(Y|X) = E[Y|X](1+\theta E[Y|X])$ , where  $\theta$  is the negative binomial dispersion parameter.

## Diagnostic plots



## Model summary

	Coefficient	SE
intercept	6.738	0.244
harvestedYes	-0.116	0.112
vegetationKeelenche	-0.005	0.185
vegetationNukuuchche	-0.314	0.231
milpaYes	0.212	0.166
latitude	0.002	0.061
longitude	-0.084	0.073

## Analysis of deviance table

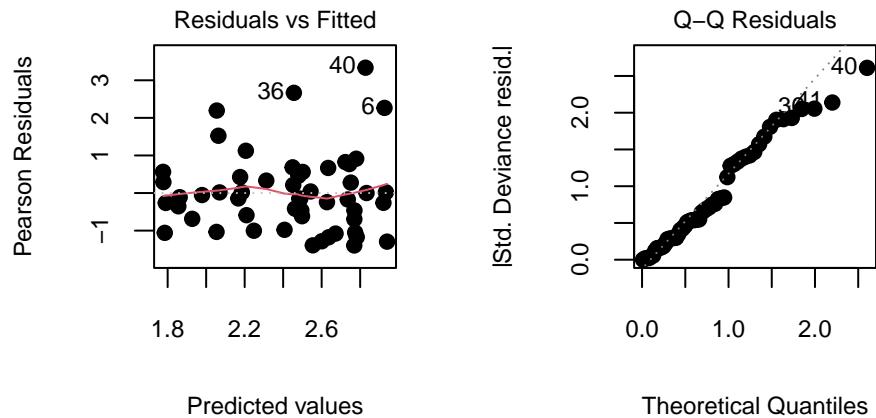
	Chisq	df	p-value	adjusted p-value
harvested	1.041	1	0.308	0.616
vegetation	3.451	2	0.178	0.466
milpa	1.644	1	0.200	0.492
latitude	0.001	1	0.979	0.979
longitude	1.322	1	0.250	0.551

## Stem density by size class

For each size class, the model is the same as the one written above except  $Y$  denotes the number of individuals (per plot) in each size class.

### Seedlings

#### Diagnostic plots



#### Model summary

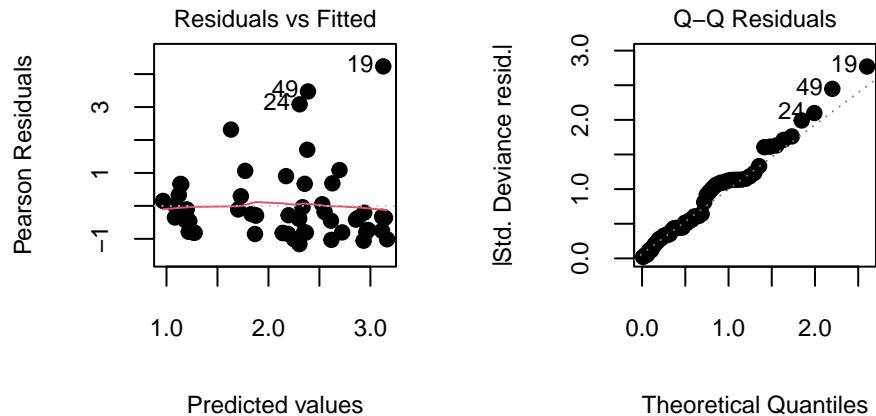
	Coefficient	SE
intercept	6.406	0.329
harvestedYes	-0.282	0.154
vegetationKeelenche	-0.162	0.247
vegetationNukuuchche	-0.669	0.313
milpaYes	-0.067	0.225
latitude	0.049	0.083
longitude	-0.091	0.100

#### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	3.315	1	0.069	0.235
vegetation	5.976	2	0.050	0.200
milpa	0.096	1	0.757	0.873
latitude	0.305	1	0.580	0.798
longitude	0.844	1	0.358	0.616

## Saplings (0-4 cm DBH)

### Diagnostic plots



### Model summary

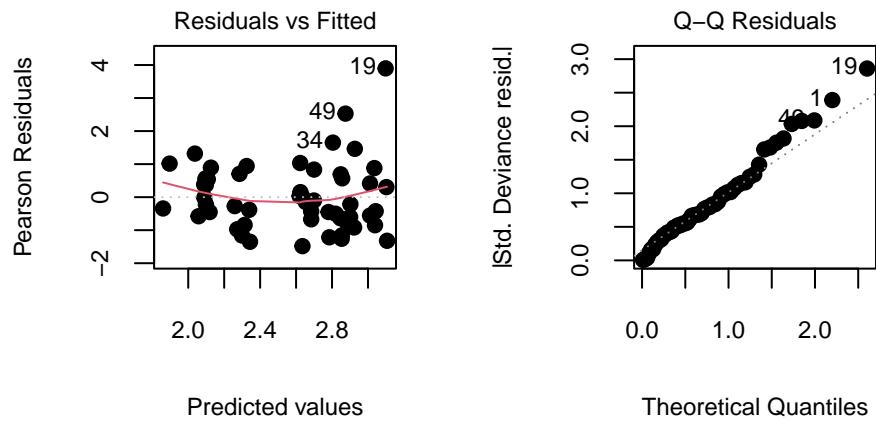
	Coefficient	SE
intercept	7.191	0.458
harvestedYes	-0.068	0.217
vegetationKeelenche	-0.575	0.343
vegetationNukuuchche	-2.160	0.451
milpaYes	-0.650	0.315
latitude	0.041	0.117
longitude	0.243	0.152

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.092	1	0.762	0.873
vegetation	23.667	2	0.000	0.000
milpa	3.694	1	0.055	0.200
latitude	0.148	1	0.701	0.873
longitude	2.398	1	0.122	0.371

## Trees (5-9 cm DBH)

### Diagnostic plots



### Model summary

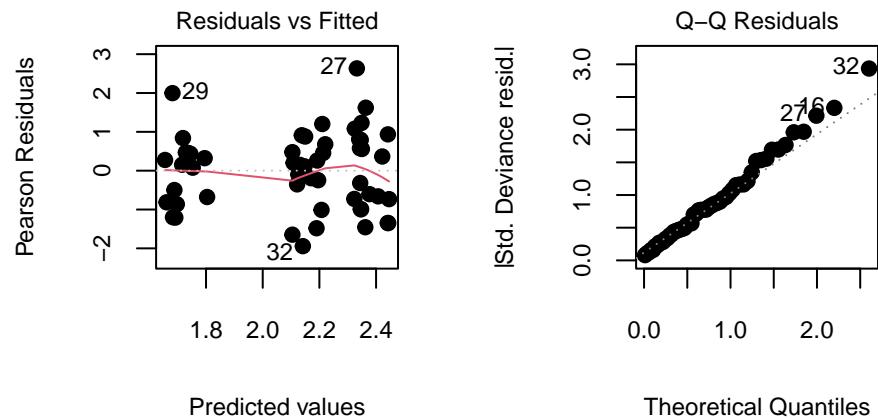
	Coefficient	SE
intercept	6.482	0.342
harvestedYes	-0.225	0.159
vegetationKeelenche	-0.109	0.257
vegetationNukuuchche	-0.803	0.329
milpaYes	0.041	0.233
latitude	0.030	0.086
longitude	0.062	0.108

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	1.929	1	0.165	0.460
vegetation	8.643	2	0.013	0.066
milpa	0.029	1	0.865	0.936
latitude	0.140	1	0.708	0.873
longitude	0.338	1	0.561	0.791

## Trees (10-14 cm DBH)

### Diagnostic plots



### Model summary

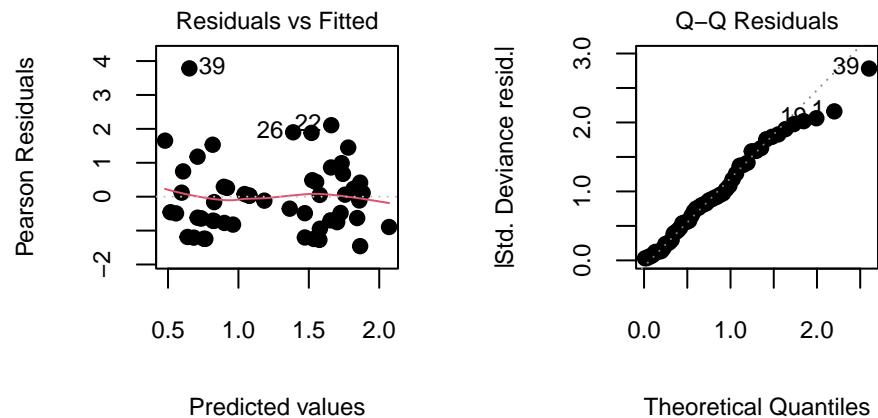
	Coefficient	SE
intercept	5.594	0.272
harvestedYes	0.007	0.125
vegetationKeelenche	0.098	0.207
vegetationNukuuchche	-0.258	0.258
milpaYes	0.025	0.185
latitude	-0.024	0.067
longitude	-0.147	0.081

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.003	1	0.956	0.974
vegetation	3.577	2	0.167	0.460
milpa	0.018	1	0.893	0.936
latitude	0.128	1	0.721	0.873
longitude	3.223	1	0.073	0.235

## Trees (15+ cm DBH)

### Diagnostic plots



### Model summary

	Coefficient	SE
intercept	3.339	0.455
harvestedYes	0.142	0.185
vegetationKeelenche	0.257	0.345
vegetationNukuuchche	1.188	0.401
milpaYes	1.044	0.304
latitude	-0.134	0.103
longitude	-0.395	0.114

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.587	1	0.444	0.718
vegetation	14.037	2	0.001	0.009
milpa	13.504	1	0.000	0.007
latitude	1.602	1	0.206	0.492
longitude	11.840	1	0.001	0.009

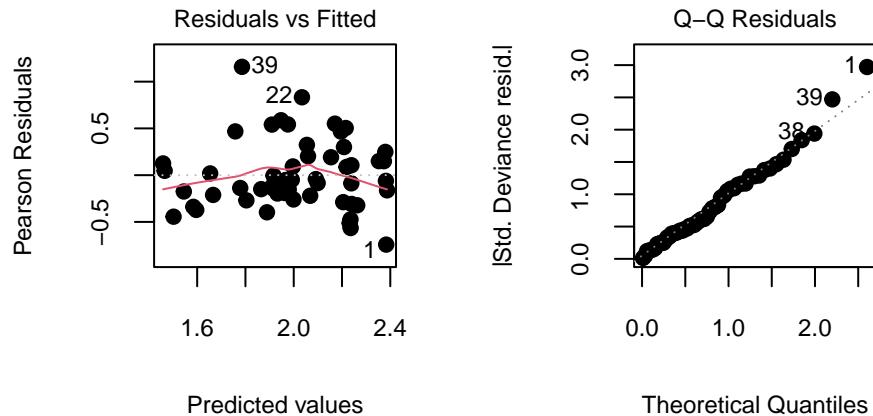
## Basal area

### Model

$$\log E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

where  $Y$  is the total basal area per hectare and  $X = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ , where  $X_1$  is a binary variable indicating whether a plot has been harvested,  $X_2$  and  $X_3$  are indicator variables for the Keelenche and Nukuuchche vegetation types,  $X_4$  is a binary variable indicating whether a plot has been exposed to milpa,  $X_5$  is latitude, and  $X_6$  is longitude. We set  $\text{Var}(Y|X) = \phi E[Y|X]^2$ , where  $\phi$  is the gamma dispersion parameter.

### Diagnostic plots



### Model summary

	Coefficient	SE
intercept	1.500	0.232
harvestedYes	0.013	0.106
vegetationKeelenche	0.079	0.179
vegetationNukuuchche	0.278	0.220
milpaYes	0.511	0.154
latitude	-0.055	0.059
longitude	-0.215	0.068

### Analysis of deviance table

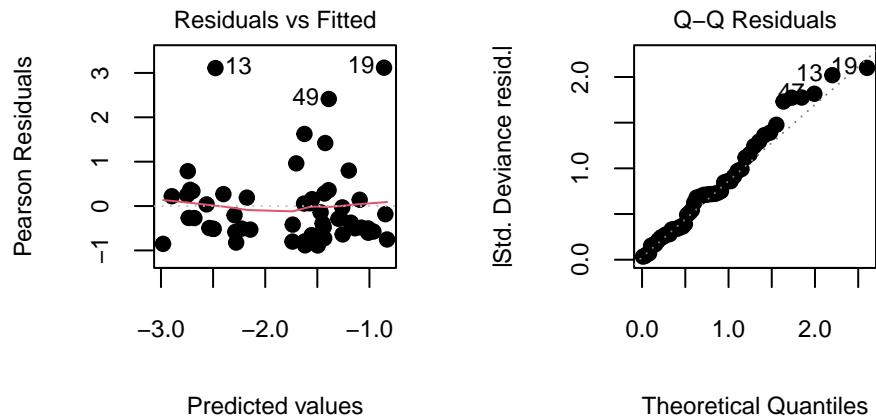
	Chisq	df	p-value	adjusted p-value
harvested	0.015	1	0.902	0.936
vegetation	2.159	2	0.340	0.616
milpa	11.147	1	0.001	0.009
latitude	0.852	1	0.356	0.616
longitude	10.315	1	0.001	0.009

## Basal area by size class

For each size class, the model is the same as the one written above except  $Y$  denotes the total basal area per hectare in each size class.

### Saplings (0-4 cm DBH)

#### Diagnostic plots



#### Model summary

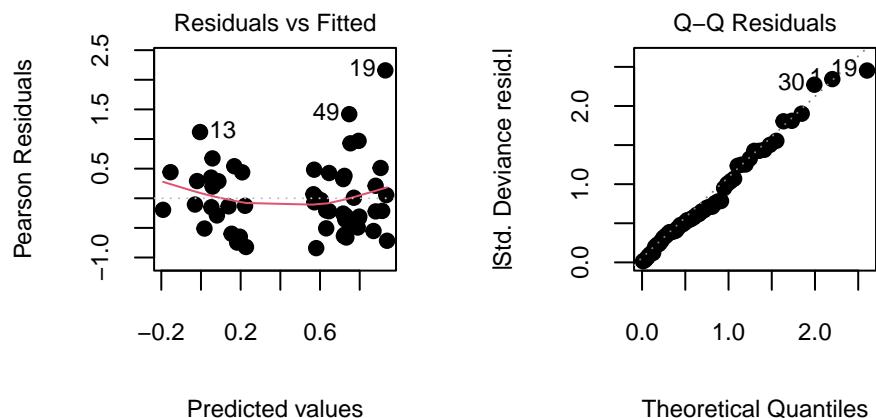
	Coefficient	SE
intercept	-0.490	0.570
harvestedYes	-0.195	0.261
vegetationKeelenche	-0.444	0.439
vegetationNukuuchche	-1.886	0.541
milpaYes	-0.392	0.378
latitude	0.021	0.145
longitude	0.184	0.167

#### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.508	1	0.476	0.748
vegetation	13.163	2	0.001	0.009
milpa	0.955	1	0.328	0.616
latitude	0.026	1	0.872	0.936
longitude	1.003	1	0.316	0.616

## Trees (5-9 cm DBH)

### Diagnostic plots



### Model summary

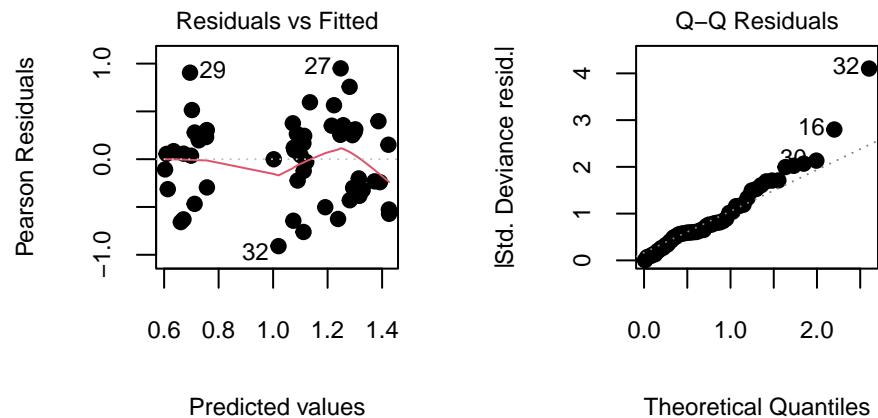
	Coefficient	SE
intercept	0.855	0.375
harvestedYes	-0.146	0.172
vegetationKeelenche	-0.071	0.290
vegetationNukuuchche	-0.758	0.356
milpaYes	0.039	0.249
latitude	0.029	0.095
longitude	0.066	0.110

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.691	1	0.406	0.677
vegetation	7.161	2	0.028	0.128
milpa	0.023	1	0.879	0.936
latitude	0.110	1	0.741	0.873
longitude	0.351	1	0.554	0.791

## Trees (10-14 cm DBH)

### Diagnostic plots



### Model summary

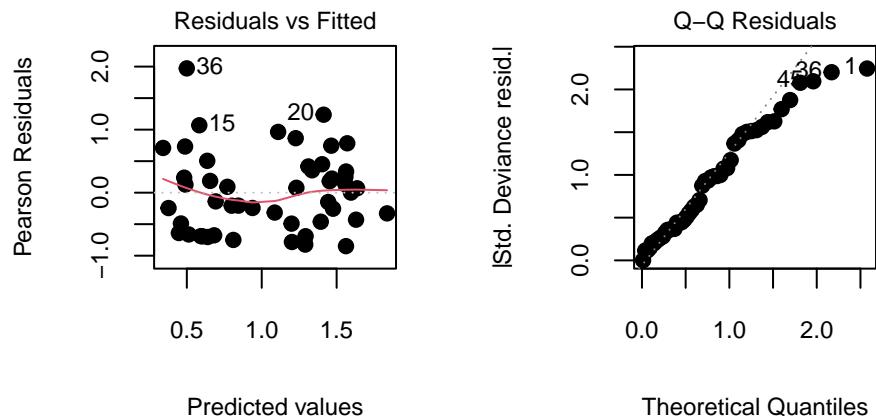
	Coefficient	SE
intercept	1.071	0.270
harvestedYes	0.046	0.124
vegetationKeelenche	0.020	0.209
vegetationNukuuchche	-0.267	0.257
milpaYes	0.081	0.180
latitude	-0.044	0.069
longitude	-0.156	0.079

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.137	1	0.711	0.873
vegetation	2.320	2	0.314	0.616
milpa	0.211	1	0.646	0.867
latitude	0.391	1	0.532	0.790
longitude	3.747	1	0.053	0.200

## Trees (15+ cm DBH)

### Diagnostic plots



### Model summary

	Coefficient	SE
intercept	-0.200	0.430
harvestedYes	0.128	0.191
vegetationKeelenche	0.218	0.329
vegetationNukuuchche	1.118	0.399
milpaYes	0.902	0.287
latitude	-0.139	0.106
longitude	-0.364	0.120

### Analysis of deviance table

	Chisq	df	p-value	adjusted p-value
harvested	0.437	1	0.509	0.777
vegetation	13.514	2	0.001	0.009
milpa	10.147	1	0.001	0.009
latitude	1.494	1	0.222	0.508
longitude	9.517	1	0.002	0.011