Blowdown Analysis

Ivan Li

2024-09-11

A logistic regression on Black Spruce Blowdown Data

```
# Import the Blowdown data
blowdown <- read.csv("C:/Users/ivani/OneDrive/Documents/Blowdown.csv")
head(blowdown)
```

```
##   X  d         s y        spp
## 1 1  9 0.0217509 0 balsam fir
## 2 2 14 0.0217509 0 balsam fir
## 3 3 18 0.0217509 0 balsam fir
## 4 4 23 0.0217509 0 balsam fir
## 5 5  9 0.0217509 0 balsam fir
## 6 6 16 0.0217509 0 balsam fir
```

We are only interested in the Black Spruce trees. We must subset the data so that all the trees we perform a regression on are Black Spruces.

```
# Subset the data to only include Black Spruce trees

blkspruce_blowdown <- subset(blowdown, spp=='black spruce')

head(blkspruce_blowdown)
```

```
##     X  d         s y          spp
## 18 18  9 0.0242120 0 black spruce
## 25 25 11 0.0305947 0 black spruce
## 26 26  9 0.0305947 0 black spruce
## 51 51  9 0.0341815 0 black spruce
## 52 52  5 0.0341815 0 black spruce
## 53 53  8 0.0341815 0 black spruce
```

```
# Logistic Regression with d and s
model <- glm(y ~ d + s, data=blkspruce_blowdown, family=binomial(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = y ~ d + s, family = binomial(link = "logit"), data = blkspruce_blowdown)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9460  -0.6563  -0.4055   0.5910   2.4684
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.2238     0.3902 -13.387   <2e-16 ***
## d             0.2850     0.0290   9.828   <2e-16 ***
## s             4.4242     0.5037   8.784   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 582.66  on 656  degrees of freedom
## AIC: 588.66
##
## Number of Fisher Scoring iterations: 5
```

Here, we notice that only d and s are valid predictors, as they represent the diameter of the tree, and local storm severity respectively. x is just the number identifier for the tree, and spp is the species of the tree in string format. First, let's run an analysis using only d as a predictor.

```
# Logistic regression with only d as a predictor
model <- glm(y ~ d, data=blkspruce_blowdown, family=binomial(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = y ~ d, family = binomial(link = "logit"), data = blkspruce_blowdown)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1749  -0.7357  -0.5637   0.8264   2.0857
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.53118    0.28019  -12.60   <2e-16 ***
## d            0.29533    0.02681   11.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 674.01  on 657  degrees of freedom
## AIC: 678.01
##
## Number of Fisher Scoring iterations: 4
```

Notice that d is the raw unit data of the diameter of the tree in cm. Using the current model, we would be assuming that each cm of diameter would contribute equally to whether or not the tree blows down. I believe that this would be an incorrect assumption. For example, if a tree is very small, then extra cm of diameter would affect it's odds of blowing down. However, a huge tree would be affected as much, as it would make sense for it to catch more "storm" than a small tree and blowdown anyways. So, taking the log of d would make sense here, as it would represent a proportional relationship that falls off as cm gets bigger. The bigger the tree gets, the less it's blowdown rate is affected from extra cm in diameter.

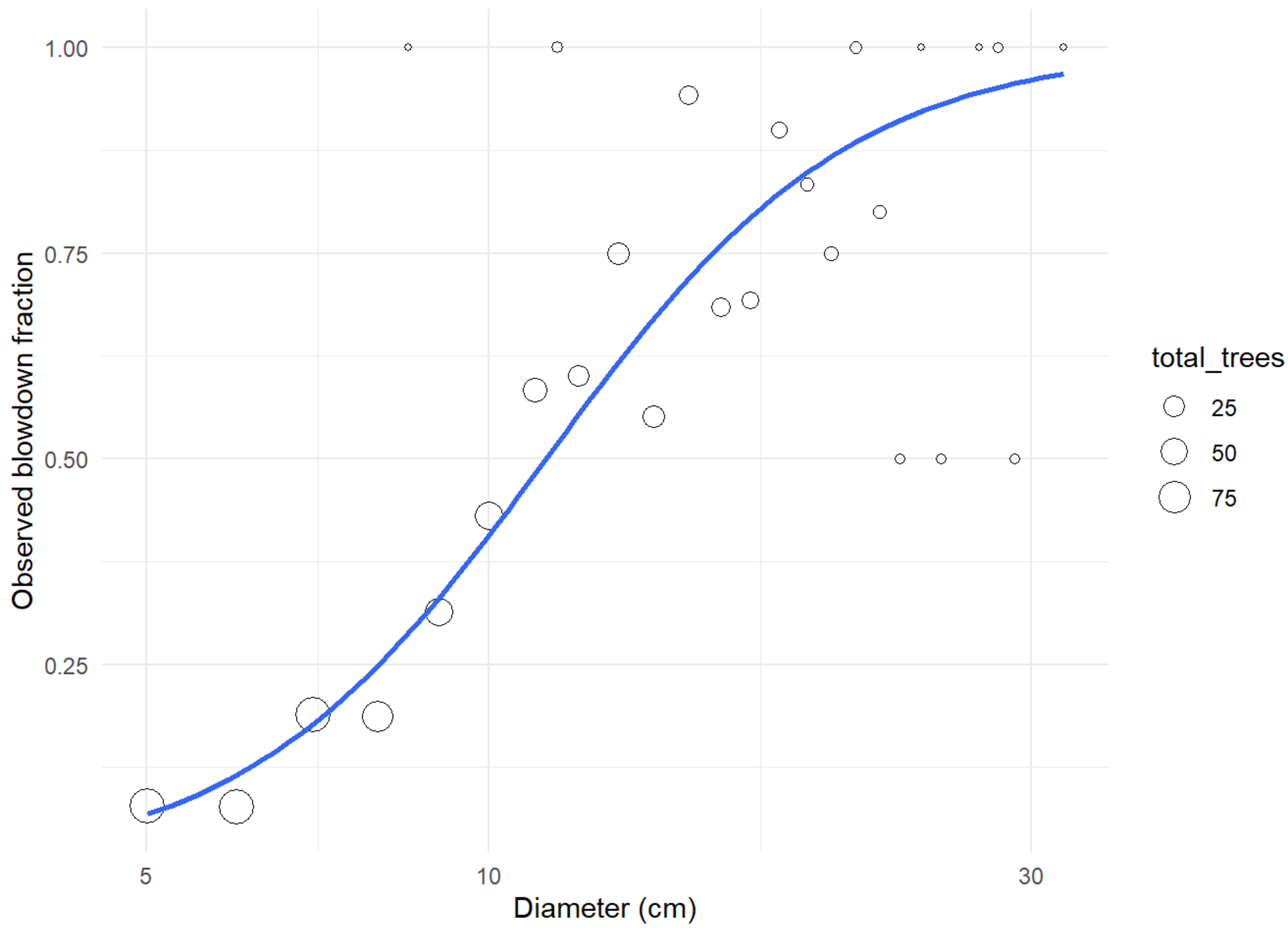
```
# Logistic Regression with Log(d)
log_model <- glm(y ~ log(d), data=blkspruce_blowdown, family=binomial(link="logit"))
summary(log_model)
```

```
##
## Call:
## glm(formula = y ~ log(d), family = binomial(link = "logit"),
##     data = blkspruce_blowdown)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5073  -0.7565  -0.4936   0.8096   2.3272
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.8925     0.6325  -12.48   <2e-16 ***
## log(d)        3.2643     0.2761   11.82   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 655.24  on 657  degrees of freedom
## AIC: 659.24
##
## Number of Fisher Scoring iterations: 4
```

We plot this logistic regression taking into account the values of diameter and the fraction of blowdown of these black spruce trees.

```
# There are 35 distinct values of diameter in the dataset. We group these into bins.
blkspruce_blowdown$diameter_bin <- cut(blkspruce_blowdown$d, breaks = 35)

# For each bin, we compute the blowdown fraction that we will plot using the tidyverse pipe func
tion
binned_data <- blkspruce_blowdown %>% group_by(diameter_bin) %>%
  summarise(diameter_center = mean(d), num_blowdown = sum(y),total_trees = n(), blowdown_fractio
n = num_blowdown/total_trees)

# Plot the bins and the logistic regression
# We ensure that the glm is binomial by setting the weight to the total number of trees
# These settings include hollow circles, no confidence interval, a logged x axis, and a minimal
layout
ggplot(binned_data, aes(x=diameter_center, y=blowdown_fraction)) + geom_point(aes(size=total_tre
es), shape=1) + stat_smooth(method="glm", method.args=list(family="binomial"), aes(weight=total_
trees), se=FALSE) + scale_x_log10() + labs(x="Diameter (cm)", y="Observed blowdown fraction") +
  theme_minimal()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



From this graph, we can observe that generally, when a tree has more diameter, its blowdown fraction is greater that of the smaller diameter trees. From this alone, it seems that the smaller diameter a tree is, the more likely it is to survive a blowdown. This makes sense, as bigger trees possibly catch more “storm” and as a result, blow over. However, this does not tell the whole story. The predictor ‘s’ must still be considered, because this is the measure of local storm severity, which will definitely affect the blowover outcome. If there is low storm severity, a tree will intuitively not blow over.

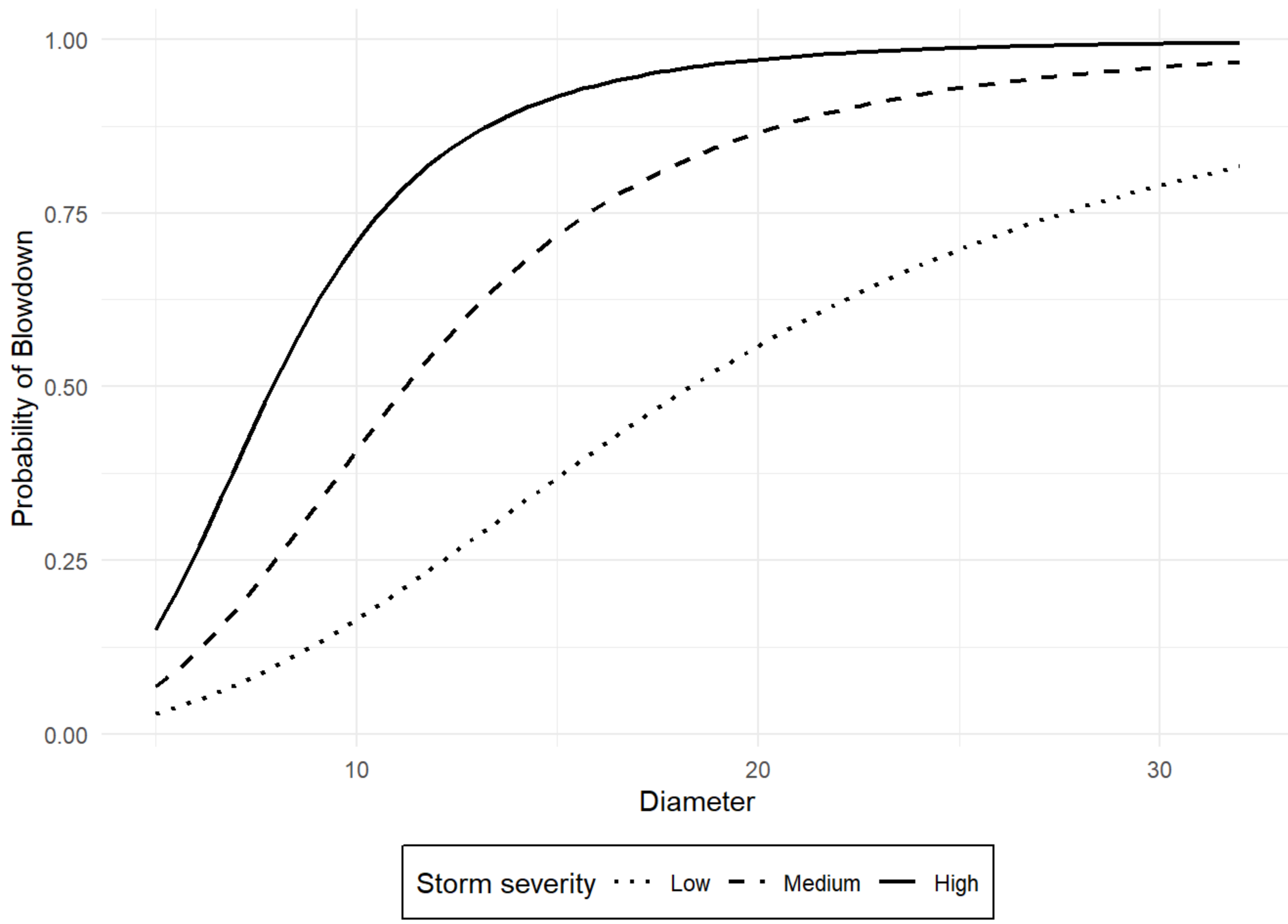
```
# Run a logistic regression with log(d) and s
ds_model <- glm(y ~ log(d) + s + log(d):s, data=blkspruce_blowdown, family=binomial(link="logi
t"))
summary(ds_model)
```

```
##
## Call:
## glm(formula = y ~ log(d) + s + log(d):s, family = binomial(link = "logit"),
##     data = blkspruce_blowdown)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3441  -0.6140  -0.4116   0.3824   2.3564
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.6782     1.4252  -2.581  0.00986 **
## log(d)        0.5781     0.6333   0.913  0.36136
## s           -11.2054     3.6384  -3.080  0.00207 **
## log(d):s      7.0847     1.6450   4.307 1.66e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 541.75  on 655  degrees of freedom
## AIC: 549.75
##
## Number of Fisher Scoring iterations: 5
```

Here, we do not use the log of s, because intuitively, it would make sense for a stronger storm affect a tree more. The assumption of a linear relationship is valid here. This model takes into account the effect of log(d) and s separately on the blowover outcome, as well as log(d) and s when they interact together. Specifically, this is how s (storm strength) affect's log(d)'s (proportional diameter) effect on the response (blowover outcome). We do this, as it makes sense for storm strength to affect how a tree's diameter may perform, and it doesn't make sense for a tree's diameter to affect storm strength. Finally, we observe that the model with the relationship log(d):s has the lowest p value by a large margin, which indicates that it is the best to use in this case.

We can plot the diameter with different storm strengths, and how they affect blowdown probability when interacting together.

```r
# Sort the storm severity into bins, and categorize them as low, medium, and high respectively
blkspruce_blowdown$storm_bin <- cut(blkspruce_blowdown$s, breaks = c(0, 0.25, 0.55, 1), labels = c("Low", "Medium", "High"))

# Logistic model with relationship log(d):storm_bin as a predictor
logds_model <- glm(y ~ log(d):storm_bin, data = blkspruce_blowdown, family = binomial(link = 'logit'))

# Create a data frame to store the new predictions
diameter_seq <- seq(min(blkspruce_blowdown$d), max(blkspruce_blowdown$d), length.out = 100)
new_data <- expand.grid(d = diameter_seq, storm_bin = c("Low", "Medium", "High"))

# Add log(d) column for predictions
new_data$log_diameter <- log(new_data$d)

# Predict the probabilities with the new model and match with the number of observations
new_data$blowdown_prob <- predict(logds_model, newdata = new_data, type = "response")

# Plot the logistic regression with log(d) and s bins as predictors
# The settings include a boxed legend, dotted/dashed/solid lines and a minimal theme
ggplot(new_data, aes(x = d, y = blowdown_prob, linetype = storm_bin)) +
  geom_line(size = 0.8) +  # Adjust the thickness of the lines
  labs(x = "Diameter", y = "Probability of Blowdown", linetype = "Storm severity") +
  scale_linetype_manual(values = c("dotted", "dashed", "solid")) +
  theme_minimal() +
  theme(
    legend.position = "bottom",
    legend.box.background = element_rect(fill = "white", color = "black")  # Box around the legend
  )
```



We observe that the blowdown probabilities are higher when storm severity is higher. It appears that the bigger the diameter of the tree, the more likely it is to blow down across all storm intensities. We also confirm that the relationship between diameter and blowdown is logarithmic, where the effect of diameter on blowdown probability tops off as diameter increases. The intuition from the beginning of the analysis is reinforced by these regression results

In conclusion, diameter and storm severity seem to be good predictors for the probability of the blowdown of a tree, and that the blowdown probability is likely dependent on these factors.