Writing An Analysis Report

Today we'll go through a sample analysis report. There's lots of potential guidelines to follow; the most important is to answer your client's questions! This format may or may not always be appropriate for doing that, but it's a good starting point.

For the Storm Damage Analysis Report, what do you notice for each of these sections? What's important to include? What's left out, possibly either on purpose or by mistake?

Executive Summary

Study Design and Data

Statistical Methods

Results

For the Running Case Study

With your group, draft what these sections might look like.

Statistical Methods

Results

Executive Summary

Storm Damage Analysis Report

Executive Summary: We analyzed the tree root failure data collected after the June 2013 Minneapolis wind storm, and find that street or sidewalk replacement is associated with an increase in root failure, approximately doubling the risk. This increase is larger for trees on narrow boulevards, again with approximately twice the risk for a wide (8ft) vs narrow (4ft) boulevard. The type of the tree was also significant, with Tilia at most risk, then Fraxinus, Acer, Ulnus, and others. Finally, tree size was found to have a smaller effect, when no work was performed. This analysis used a logistic mixed model, which accounted for the different conditions on each block.

Background and Goals: (skipped here, same info as in proposal)

Study Design and Data: Data was collected on every tree on each block where two or more root failures occurred. The final data set contained 3076 trees on 122 blocks. The primary response was root failure; this occurred in 367 of the 3076. Trees were grouped into five categories by Genus, using the four largest genuses, and then grouping the remaining trees in an "other category". There were 1365 Tilia, 940 Fraxinus, 289 Acer, 229 Ulnus, and 253 Other. DBH values were centered around the mean for each category separately. For analysis purposes, the boulevard width was log-transformed, as the distribution was skewed to the right. Boulevard width was calculated for each block using the average of the values for which it was collected. For street work, a combined variable denoting any replacement work was calculated; it denoted when the sidewalk, the curb, or the street was replaced, or utility repaving due to utility work was done. Of the 3021, 945 had sidewalk, 2 had curb, 14 had street, and 54 had utility repaving; a combined 979 had any replacement work done. Because of the relatively small numbers of curb, utility, and street replacement, these variables were not considered separately; only the combined variable was used. Variables of secondary interest were percent organic matter, sand, silt, and clay.

Statistical Methods: To investigate which variables were associated with root failure, a logistic mixed model was fit, with root failure as the response and block as a random effect. Several models were fit to explore the various predictors, and the most parsimonious model was chosen to explore further. First, the primary predictors, replacement work, boulevard width, DBH (diameter at breast height), and Genus group, were considered, along with all two-way interactions. Non-significant terms were excluded, and the resulting model compared using AIC with the full model. Secondary predictors of percent organic matter, sand, silt, and clay, were then added to this model, along with interactions with genus group and replacement work done. The final model included replacement work and Genus group, as well as boulevard width, DBH, and their interactions with replacement work. For each variable of interest, odds ratios with 95% confidence intervals and p-values are reported, and the corresponding proportions are calculated for a representative example.

Results: The major finding is that having replacement work done increased the odds of root failure by 2.24 times (95% CI: 1.77, 2.83; $p < 0.0001$). For illustration, when no replacement work was done, the average Tilia had a 10.6% chance of root failure; this increased to 21.0% when replacement work was done.

Running Study Analysis

```
##   ID Sex Age YearsRunning Smoking Treatment Before After
## 1  1   F  21            9     Yes       Trt    242   243
## 2  2   F  30            3      No       Trt    201   198
## 3  3   F  34            9      No       Trt    156   155
## 4  4   F  23            6      No   Control    286   290
## 5  5   F  19           10      No       Trt    275   272
## 6  6   F  30            6      No       Trt    247   237
```

```
d$Diff <- d$After - d$Before
m <- lm(Diff ~ Treatment * (Sex + Age + YearsRunning + Smoking), data=d)
m2 <- lm(Diff ~ Treatment + (Sex + Age + YearsRunning + Smoking), data=d)
anova(m2, m); summary(m2)
```

```
## Analysis of Variance Table
##
## Model 1: Diff ~ Treatment + (Sex + Age + YearsRunning + Smoking)
## Model 2: Diff ~ Treatment * (Sex + Age + YearsRunning + Smoking)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     34 624.29
## 2     30 606.80  4    17.492 0.2162 0.9273
##
## Call:
## lm(formula = Diff ~ Treatment + (Sex + Age + YearsRunning + Smoking),
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0676 -2.6903 -0.0395  2.7100  9.2434
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.7319     4.3106   2.026  0.05071 .
## TreatmentTrt  -2.9681     1.3693  -2.168  0.03729 *
## SexM          -2.0069     1.4311  -1.402  0.16988
## Age           -0.4049     0.1369  -2.958  0.00561 **
## YearsRunning   0.2250     0.2295   0.980  0.33386
## SmokingYes    -0.8820     2.3554  -0.374  0.71039
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.285 on 34 degrees of freedom
## Multiple R-squared:  0.3504,Adjusted R-squared:  0.2548
## F-statistic: 3.667 on 5 and 34 DF,  p-value: 0.009236
```

Storm Damage Analysis Report continued

The genus of the tree was also significant ($p < 0.0001$), even after adjusting for the average DBH of each group. Compared with Tilia, which was considered the baseline genus, Fraxinus had a decrease in odds of root failure of 0.94 times (95% CI: 0.73, 1.21); Acer of 0.47 (95% CI: 0.29, 0.77), Ulmus of 0.39 (95% CI: 0.22, 0.69) and other of 0.22 (95% CI: 0.11, 0.46). For illustration, Fraxinus had a 10.0% chance of root failure when no replacement work was done, compared with a 20.0% chance when it was done; for Acer these were 5.3% and 11.1% respectively, for Ulmus, 4.4% and 9.4%, and for other, 2.6% and 5.6%.

Boulevard width was found to have a significant interaction with replacement work (p=0.011). When work was done, an increase in boulevard width of 1.42 times (one standard deviation) reduced the odds of root failure by 0.64 (95% CI: 0.49, 0.84; p=0.001). For illustration, two otherwise average Tilia on streets with widths of 3.94 and 7.99 (one sd up and one down from the geometric mean) have a 29.4% and 14.6% chance of failure, respectively, when work is done. However, when no replacement work done, boulevard width was not significant (p=0.50).

DBH was also found to have a significant interaction with replacement work (p=0.008). In this case, when no replacement work done, an increase in DBH of 6.77 (one sd) increases the odds of root failure by 1.27 times (95% CI: 1.08, 1.51; p=0.005). So two otherwise average Tilia with DBH of 8.2 and 21.7 (one sd up and one down from mean) has 8.5% and 13.2% chance when no work was done. When replacement work was done, DBH was not significant (p=0.29).

Additional soil parameters were tested for by adding them to the logistic mixed model with tree genus, replacement work, DBH, log boulevard width, and interactions between replacement work and DBH and also log boulevard width. None were statistically significant.

- Percent organic matter was not found to be significant ($p = 0.94$). An increase of 1.80 (one sd) increased the odds of root failure by 1.00 times (95% CI: 0.90, 1.12). For example, two trees with values of 4.47 and 8.07 (one sd up and one down from the mean) have an estimated 10.33% and 10.41% chance of root damage.
- Percent sand was not found to be significant ($p = 0.55$). An increase of 11.02 (one sd) decreased the odds of root failure by 0.96 times (95% CI: 0.85, 1.09). For example, two trees with values of 39.31 and 61.35 (one sd up and one down from the mean) have an estimated 10.76% and 10.06% chance of root damage.
- Percent silt was not found to be significant ($p = 0.38$). An increase of 7.69 (one sd) increased the odds of root failure by 1.06 times (95% CI: 0.93, 1.20). For example, two trees with values of 10.43 and 25.81 (one sd up and one down from the mean) have an estimated 9.88% and 10.95% chance of root damage.
- Percent clay was not found to be significant ($p = 0.97$). An increase of 5.42 (one sd) increased the odds of root failure by 1.00 times (95% CI: 0.89, 1.13). For example, two trees with values of 26.15 and 36.98 (one sd up and one down from the mean) have an estimated 10.39% and 10.43% chance of root damage.

Homework 9: Due Mon April 3

Write an analysis report for your client. In addition to the sections here, include at least one table and at least one graphic. Your report will probably be longer; try to keep it to five total pages.

These are things I'll be looking for when I read them.

Executive Summary

- Does it have the take-home message?
- Is it written in everyday language and clear to a non-technical audience?

Background and Goals

- Why was the study done?
- What specific subject matter questions did it hope to answer?

Study Design and Data

- What was the experimental design (if an experiment)?
- How were the samples chosen?
- What were the variables?
- What data cleaning, manipulation, and transformations were done?
- Include some basic descriptive data, maybe in a table or graphic.

Statistical Methods

- What question it will answer
- What statistical method you used
- What variables were used in the method and how
- What output from the model you used to answer the question

Results

- Include estimated values and confidence intervals in addition to p-values
- Explain the meaning in the context of the original question.
- Make a table or graphic to demonstrate the results.
- Have the subject matter questions been answered?

STAT8801, March 22, 2017

Name: _____

Do you feel ready to perform the analyses and write an analysis report for your case?
If not, what questions do you have?

STAT8801, March 22, 2017

Name: _____

Do you feel ready to perform the analyses and write an analysis report for your case?
If not, what questions do you have?