# Hornfly Analysis Exploratory Model

*Joseph Blubaugh*

*03 April 2017*

**Overview**

My approach was to calculate the difference between the overall average of the control group vs the overall average of each treatment group in a given week. Any treatment observations that did not have a corresponding control group were dropped. Early models displayed non-constant variance in the residuals. To counter this, I am applied a transformation of $log(\text{diff} + 150)$ on the response variable.

**Additional Filters**

- Restricted Week range > 4 to eliminate observations that started before treatments were applied.
- Restricted Week range < 21. Many of the tags seemed to wear off towards the end of the season.
- Removed Anderson County due to the control herd actually being a treatment herd.
- Removed Camp and Palapinto Counties. Questionable that the control herds did not receive a treatment.
- Initially kept Pythonstrips and Vetgun.

Table 1: Sample of Dataset

| cty | trt | week | flies | ctrl_cnt | diff |
|-------|--------|------|-------|----------|-------|
| Bowie | Python | 7 | 22.5 | 195.8 | 173.2 |
| Bowie | Python | 8 | 28 | 223.5 | 195.5 |
| Bowie | Python | 9 | 29.5 | 223.8 | 194.2 |
| Bowie | Python | 10 | 22.67 | 161.8 | 139.1 |
| Bowie | Python | 11 | 28.67 | 204 | 175.3 |
| Bowie | Python | 12 | 33 | 218.8 | 185.8 |



Dotplot of Treatment vs Control
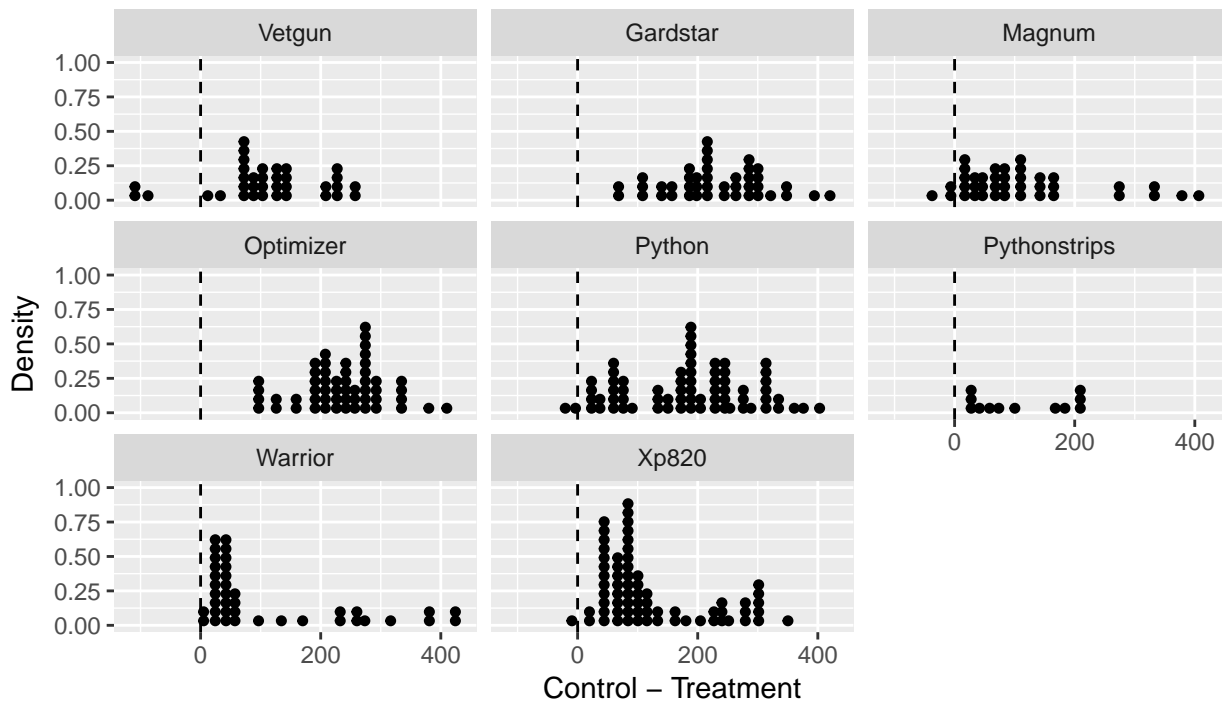
**Exploratory Model**

In this first model I am primarily interested in looking at the overall fit and testing some contrasts to determine if Pythonstrips and Vetgun should be removed from the analysis.
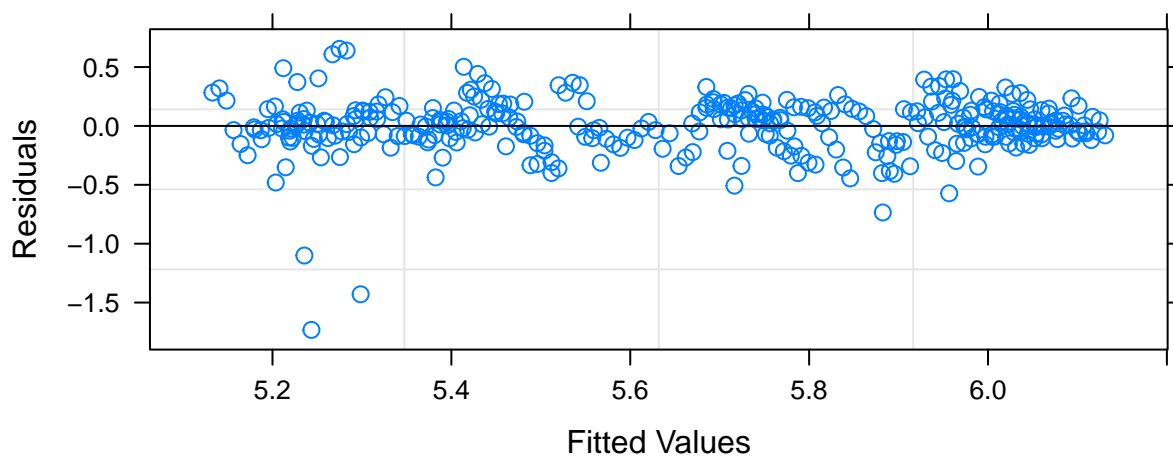
**Proposed Model**: $y_{ijk} = \mu + \beta_1(week) + \alpha_{(treatment)} + c_{(county)} + e_{ijk}$

```
Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
     Sum Sq Mean Sq NumDF  DenDF F.value   Pr(>F)
week 0.3357 0.33570     1 337.66  5.8597 0.016018 *
trt  1.2216 0.17452     7 345.85  3.0463 0.004013 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Fitted Values vs Residuals



The ANOVA table shows that both week and treatment are significant overall. The non-constant variance issue seems to be fixed by the transformation, but there are some notable outliers near the bottom right of the residuals. These are actually 3 Vetgun observations indicating weeks where the pesticide had worn off and was reapplied.

```
Linear mixed model fit by REML ['merModLmerTest']
Formula: linkfun(diff) ~ week + trt + (1 | cty)
   Data: modeling
REML criterion at convergence: 67.6645
Random effects:
 Groups   Name        Variance
 cty      (Intercept) 0.07095
 Residual             0.05729
Number of obs: 358, groups:  cty, 13
Fixed Effects:
   (Intercept)          week    trtGardstar      trtMagnum    trtOptimizer      trtPython trtPythonstrips
      5.687993     -0.007862       0.107235      -0.023076       0.141624       0.084412       0.268548
     trtWarrior        trtXp820
       0.048723        0.162475
```

The variance component County makes up approximately 55% of variance.

The treatment with the highest estimated mean count in hornfly reduction is the Python strips treatment. It is notable that Pythonstrips only shows up in one county. Pythonstrips also has the largest confidence interval of all of the treatments.

LSMeans for Treatments (Back Transformed)

```
 trt          response       SE    df  lower.CL upper.CL
 Vetgun       115.7498 25.89802 28.89  67.72016 174.3748
 Gardstar     145.8317 26.88936 23.66  95.19484 206.9259
 Magnum       109.6875 23.62744 23.93  65.22155 163.3404
 Optimizer    156.1818 26.42960 19.68 105.68044 216.6581
 Python       139.1564 24.00765 17.10  92.71149 194.4889
 Pythonstrips 197.6175 40.73122 59.18 124.96836 289.4611
 Warrior      129.0184 24.66982 21.96  82.26846 185.1779
 Xp820        162.6330 25.70063 16.59 112.76570 221.9640
```
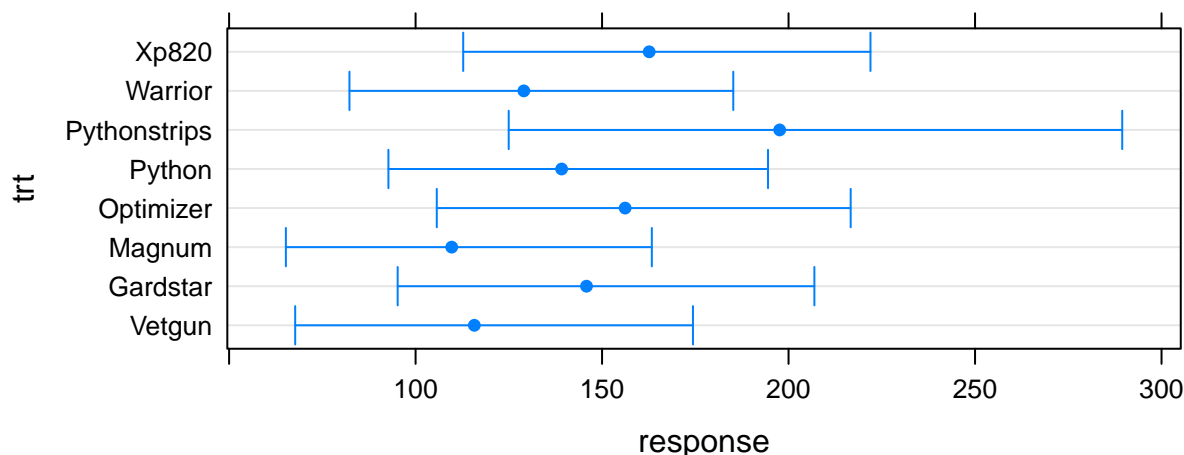
Degrees-of-freedom method: satterthwaite
Confidence level used: 0.95
Intervals are back-transformed from the log(mu + 150) scale

## LSMeans for Treatment



Since Vetgun has a different type of treatment application it is worth looking at a contrast to assess if Vetgun is different from the overall average of all other treatments. The following Vetgun contrast excludes Pythonstrips. Pythonstrips and Python are the same pesticide, but are applied differently. Since Pythonstrips is only tested in one county it is worth testing a contract against Python.

```
        Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: lme4::lmer(formula = linkfun(diff) ~ week + trt + (1 | cty),
    data = modeling)

Linear Hypotheses:
                         Estimate Std. Error z value Pr(>|z|)
Vetgun vs All Tags == 0    0.5214     0.4371   1.193    0.391
Python vs Pythonstrips == 0  0.1841   0.1039   1.772    0.139
(Adjusted p values reported -- single-step method)
```

**Refined Model**

- Pythonstrips have been removed because it is not significantly different than Python and it is only used in one county.
- Vetgun have been removed because its a different treatment method that needs to be reapplied. Observations where the pesticide had worn off were outliers in the original model.

```
Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
      Sum Sq Mean Sq NumDF  DenDF F.value   Pr(>F)
week 0.22997 0.22997     1 293.15  7.2481 0.007506 **
trt  0.60588 0.12118     5 300.85  3.8192 0.002275 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Linear mixed model fit by REML
t-tests use  Satterthwaite approximations to degrees of freedom ['lmerMod']
Formula: linkfun(diff) ~ week + trt + (1 | cty)
   Data: modeling

REML criterion at convergence: -113.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1960 -0.4926  0.0843  0.6461  2.3590

Random effects:
 Groups   Name        Variance Std.Dev.
 cty      (Intercept) 0.07476  0.2734
 Residual             0.03173  0.1781
Number of obs: 311, groups:  cty, 12

Fixed effects:
             Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  5.647050   0.094136  20.730000  59.988  < 2e-16 ***
week        -0.006917   0.002569 293.150000  -2.692 0.007506 **
trtGardstar  0.121488   0.059796 303.630000   2.032 0.043053 *
trtOptimizer 0.155883   0.051973 302.600000   2.999 0.002931 **
trtPython    0.101545   0.049126 302.280000   2.067 0.039582 *
trtWarrior   0.070000   0.051757 300.940000   1.352 0.177243
trtXp820     0.187250   0.049227 302.640000   3.804 0.000172 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
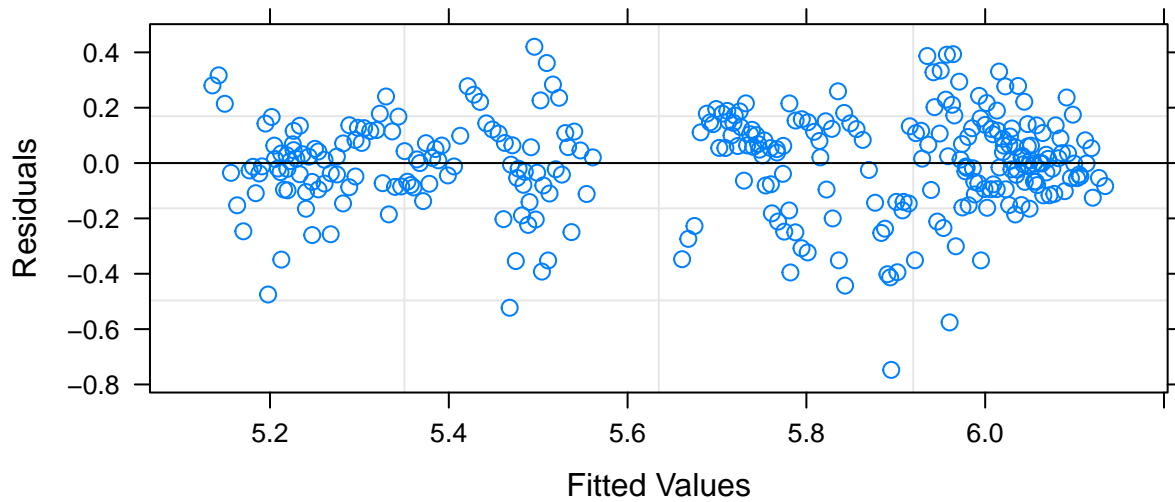
## Fitted Values vs Residuals



In the refined model, both week and treatment are still significant. The variance component County now makes up approximately 70% of the variance in the model. The model residuals have improved as well with the removal of the Vetgun treatment. The model summary shows the output in the transformed scale. All but one treatment appears to be significantly different than the base treatment, Magnum. The negative week estimate indicates that the effect of the pesticides are wearing off as the season progresses. A future model will test the interaction between treatment and week to see if some treatments wear off at different rates.

LSMeans for Treatments (Back Transformed)

```
trt        response       SE    df  lower.CL upper.CL
Magnum     108.1458 22.74009 15.95  64.16208 161.1627
Gardstar   141.4921 25.66096 15.88  91.83958 201.3389
Optimizer  151.6923 25.73951 14.10 101.27427 212.2266
Python     135.7364 23.85863 12.98  88.56929 192.2289
Warrior    126.8635 23.98505 15.03  80.18916 183.0017
Xp820      161.3056 26.06512 13.11 109.83548 222.9712
```

Degrees-of-freedom method: satterthwaite
Confidence level used: 0.95
Intervals are back-transformed from the log(mu + 150) scale

## LSMeans for Treatment
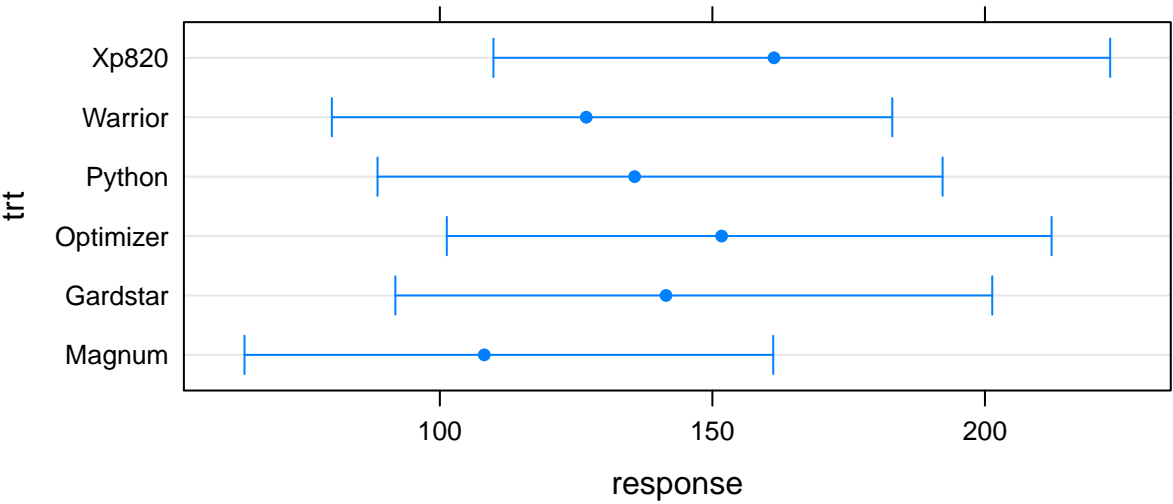
Tukey's test for multiple comparisons shows significant differences between Xp820 and Magnum, and Optimizer and Magnum. Xp820 and Warrior are borderline different.


     Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lme4::lmer(formula = linkfun(diff) ~ week + trt + (1 | cty),
    data = modeling)

Linear Hypotheses:
                         Estimate Std. Error z value Pr(>|z|)
Gardstar - Magnum == 0    0.12149    0.05980   2.032   0.3169
Optimizer - Magnum == 0   0.15588    0.05197   2.999   0.0312 *
Python - Magnum == 0      0.10155    0.04913   2.067   0.2978
Warrior - Magnum == 0     0.07000    0.05176   1.352   0.7488
Xp820 - Magnum == 0       0.18725    0.04923   3.804   0.0019 **
Optimizer - Gardstar == 0 0.03439    0.05303   0.649   0.9866
Python - Gardstar == 0   -0.01994    0.04993  -0.399   0.9986
Warrior - Gardstar == 0  -0.05149    0.04884  -1.054   0.8956
Xp820 - Gardstar == 0     0.06576    0.05014   1.312   0.7725
Python - Optimizer == 0  -0.05434    0.03807  -1.427   0.7031
Warrior - Optimizer == 0 -0.08588    0.05208  -1.649   0.5576
Xp820 - Optimizer == 0    0.03137    0.04644   0.675   0.9840
Warrior - Python == 0    -0.03155    0.04646  -0.679   0.9836
Xp820 - Python == 0       0.08571    0.04075   2.103   0.2787
Xp820 - Warrior == 0      0.11725    0.04306   2.723   0.0686 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)