Hornfly Analysis

Joseph Blubaugh

08 April 2017

Problem Description

Horn flies are destructive insects that reduce the productivity and health of livestock. They are a type of biting fly, approximately a quarter inch in length, and are most active during the summer months. Horn flies survive on the blood of their hosts which in cattle inhibits milk production, weight gain, and the overall health and comfort of the animal. They reproduce quickly and an untreated cow can have hundreds of flies living on them at one time.

An experiment has been conducted to test the efficacy of ear tag pesticides on horn fly counts during the summer months. Research suggests that ear tags have been very successful in reducing the number of horn flies. The goal of this analysis is to determine which treatment is most effective at reducing the number of horn flies on cattle.

Data Description

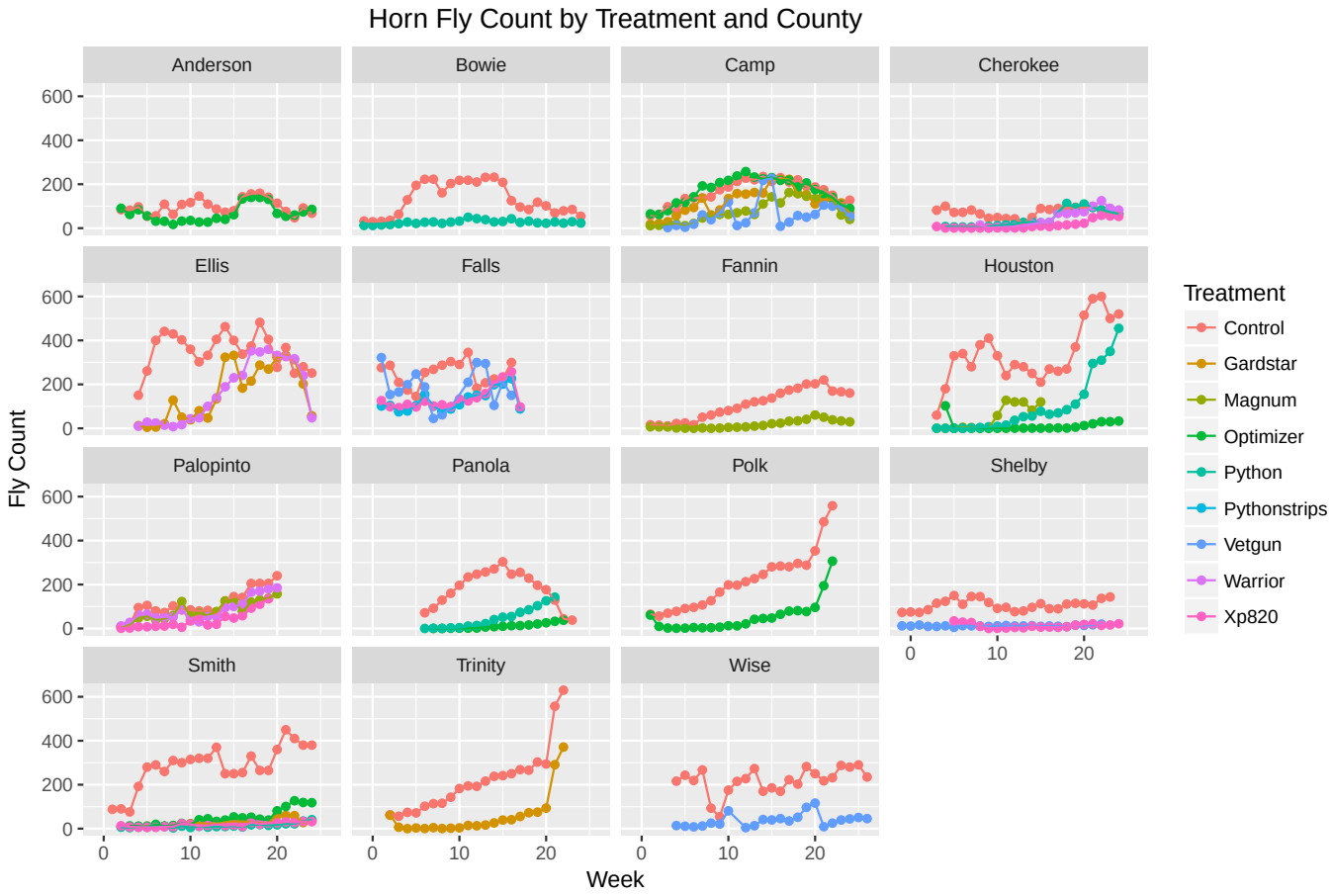
The experiment was conducted over a 26 week period between May and October in 15 Texas counties. The treatments consisted of 8 varieties of pesticides that were applied mostly using ear tags. Vetgun (injection) and Pythonstrips (pesticide strips attached to the ear tag) were the two treatments with an application method different from the rest of the treatments. In each county there were a minimum of two herds. One herd in each county served as the control and received no pesticide. All other herds in the county that were participating in the experiment received a randomly selected pesticide. Each week a sample of 10 cows were randomly selected and manually observed from one side. The number of flies observed on each cow were recorded. The observed number of flies was an approximation and not the actual count. The observations were also recorded between 7am-11am or after 5pm to avoid counting during the hottest part of the day when the flies may be on the underside of the cow.

Pesticides:	Gardstar, Magnum, Optimizer, Python, PythonStrips Vetgun, Warrior, XP820
Texas Counties:	Anderson, Bowie, Camp, Cherokee, Ellis, Falls, Fannin, Houston, Palopinto, Panola, Polk, Shelby, Smith, Trinity

Exploratory Analysis

Since the experiment was originally conducted across several counties in Texas, it makes sense to observe the data stratified by county and note any differences. It is immediately obvious that there is a lot of variation between the counties. Focussing on the control herds for a momemnt, it is interesting to note that although there is a variety of control count patterns ocurring over the season, there is not a single county that displays a completely unique pattern. Some counties have smooth weekly averages that steadily increase, some of which have large uptick towards the end of the season. Other counties have larger swings, but still trend upwards. A few coutnies stay relatively constant all season, and some counties exibit a season pattern where they peak around the middle of the season. These similar patters suggests that some counties may share similar

environments that influence horn fly counts in different ways, therefore it will be important to retain the association between each herd and county.



One of the most interesting things about the treatments is that some perform very well against horn flies in some counties and in others, are no better than the control. Most of the treatments show a fairly small changes in horn fly counts from week to week with the exception of Vetgun which is the only treatment in the experiment which was not administered from an ear tag. In some of the counties, the treatment counts have a very similar pattern to the control counts, in other counties they are completely different. This suggests that the effectiveness of the treatments may be dependent on the surrounding environment.

Restricting Observations

County plays a significant role in this analysis and has been treated conservatively. Anderson county is confirmed to have not had a control herd and has been removed. Camp county has not been confirmed, but the difference between the reported control herd and the treatment herds are suspicious enough that it has been removed from this analysis as well. Vetgun has also been removed given that the treatment method was different then the rest of the treatments. Pythonstrips was removed for a few reasons: It has a slightly different treatment method (strips are applied to the tag manually), it only appears as a treatment in one county with no other treatments, and contrasts show that pythonstrips a python are not significantly different. Python is used in several counties so there is no need to combine to pythonstrips into the python treatment.

An additional filter was applied to weeks so that only data between weeks 5 through 18 were used. The reason for starting at week 5 was to ensure that treatment observations that were recorded

before the tag was applied were removed. Since the overall goal of the analysis is to determine the best or most effective pesticide, observations beyond week 18 were removed because some treatments appeared to be losing the effectiveness.

Before building a model the difference in average fly count between the control herd and treatment herds were computed. Any treatment observations that did not have a corresponding control observation were removed.

Table 2: Sample of Final Dataset

County	Treatment	Week	Treatment Count	Control Count	Difference
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

The following tables are computed from the final dataset. All remaining treatments are used in at least 3 counties. Optimizer has the highest mean difference vs the control. Magnum has the lowest average difference and Warrior is slightly higher.

Table 3: Mean Difference in Count vs Control

Treatment	Mean	Counties
Optimizer	235.8	4
Gardstar	229.6	3
Python	183.8	5
Xp820	123.8	5
Warrior	115.3	3
Magnum	111.9	3

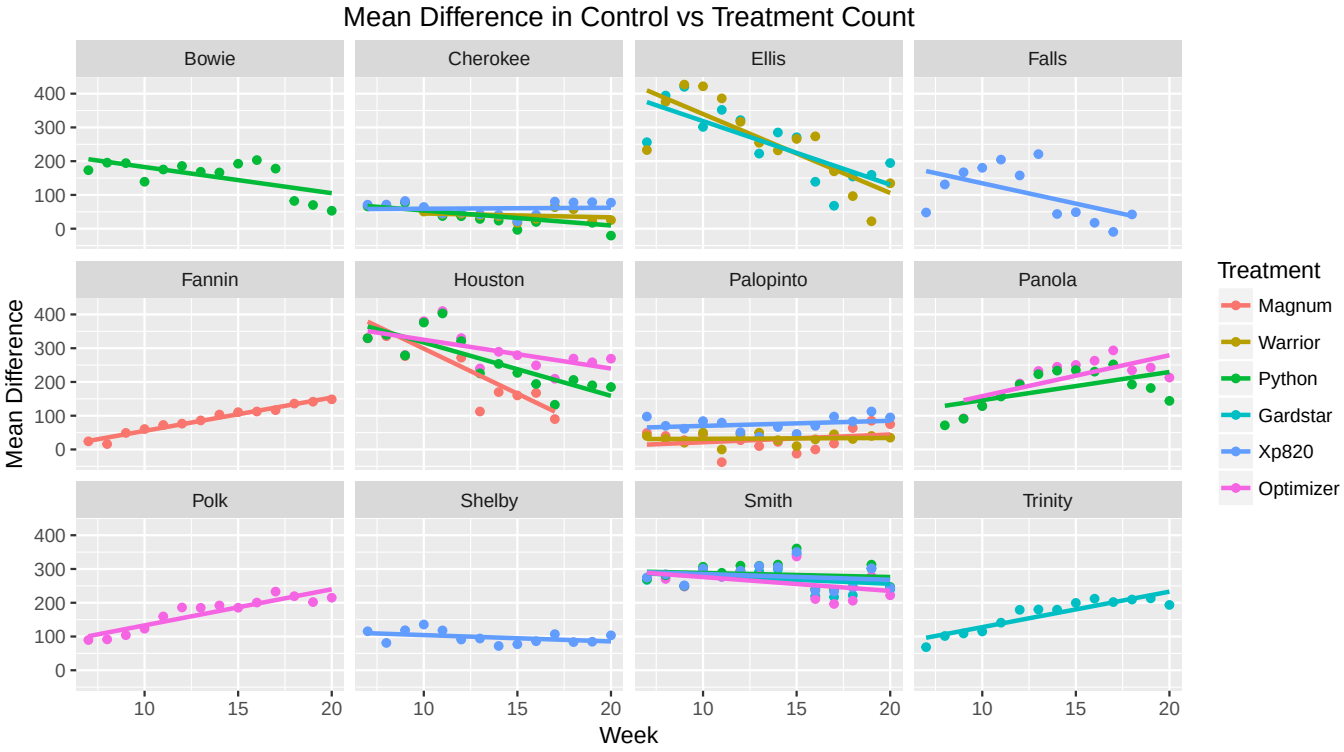
Table 4: Mean Difference in Count by Treatment and County

County	Magnum	Warrior	Python	Gardstar	Xp820	Optimizer
Bowie			155.5			
Cherokee		39.3	38.3		60.1	
Ellis		257.8		252.8		
Falls					104.4	
Fannin	89.5					
Houston	245.4		261.5			295.2
Palopinto	29.4	32.5			75.1	
Panola			179.5			212.5
Polk						170.4
Shelby					97.6	
Smith			284	271.4	279.1	261.7
Trinity				164.5		

Statistical Method

The response variable is the difference in average fly count between the control herds and the treatment herds. In early modeling attempts there were issues with non-constant variance in the residuals so in the final model a transformation of log(difference + 150) has been used. The +150 is added to ensure that none of the response observations have a negative value when applying the log transformation.

The original experiment was a repeated measures design and as a result the observations are correlated. The following plot illustrates the weekly change in average fly count by county and treatment. Six of the twelve remaining counties have multiple treatments applied.



Simple regression lines indicate that the intercept and slopes within county do not look very different, however they look quite different across counties. This could be an indication of each counties overall influence on the effectiveness of the treatments over time. One way to approach modeling this situation is with a random coefficient model.

Treatments are the primary interest and will be used as fixed effects. County will be the random effect since we want to use inference to make statements on the effectiveness of treatments in general. Since the slopes across counties appear to be different, week will be used to model the rate of change within each county. Finally, to account for the correlation among observations we will use an AR(1) correlation structure for the residuals.

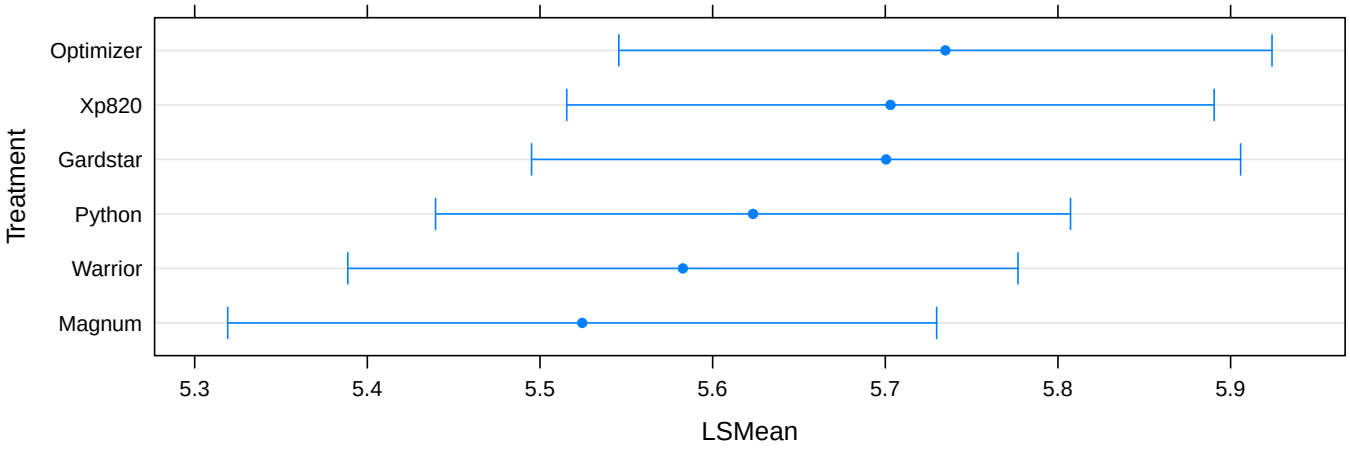
Model Results

ANOVA Summary

```
            numDF denDF  F-value p-value
(Intercept)     1   294 5683.888  <.0001
trt             5   294    2.700   0.021
```

Random Coefficient Model Estimates

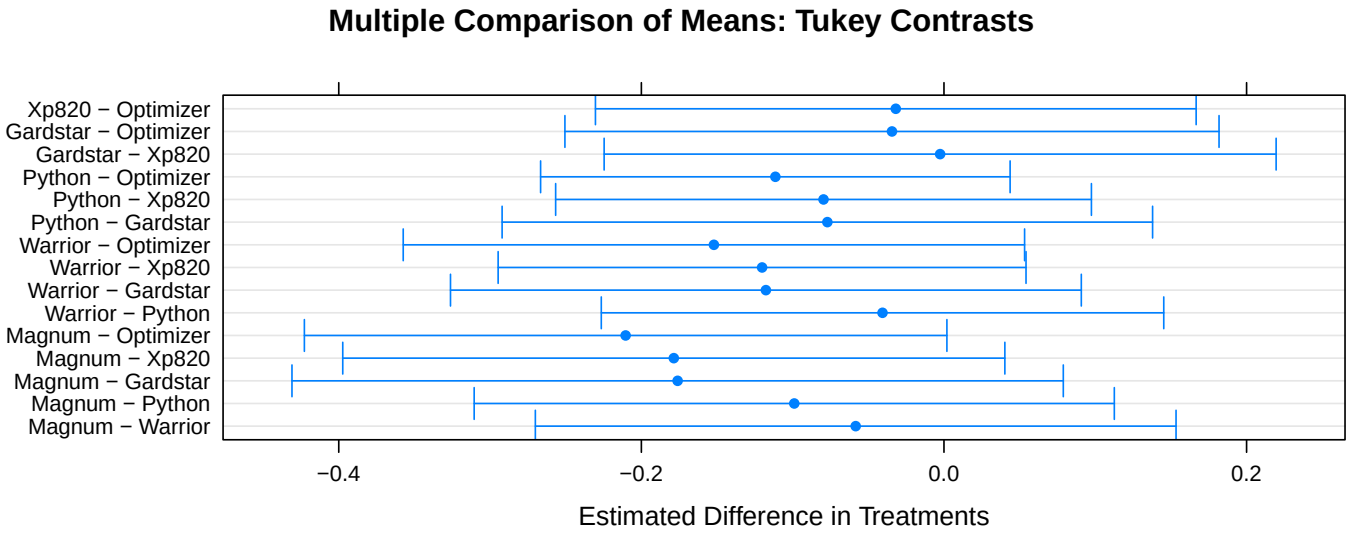
```
Linear mixed-effects model fit by REML
 Data: modeling
        AIC       BIC   logLik
  -299.3031 -258.3796 160.6515

Random effects:
 Formula: ~week | cty
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.44186507 (Intr)
week        0.02330313 -0.825
Residual    0.16269287

Correlation Structure: AR(1)
 Formula: ~1 | cty
 Parameter estimate(s):
      Phi
0.6189815
Fixed effects: log(diff + 150) ~ trt
                Value  Std.Error  DF  t-value p-value
(Intercept)  5.524433 0.09328343 294 59.22202  0.0000
trtWarrior   0.058338 0.07380857 294  0.79040  0.4299
trtPython    0.098955 0.07374355 294  1.34188  0.1807
trtGardstar  0.176028 0.08885931 294  1.98098  0.0485
trtXp820     0.178558 0.07627337 294  2.34103  0.0199
trtOptimizer 0.210394 0.07401910 294  2.84242  0.0048
 Correlation:
             (Intr) trtWrr trtPyt trtGrd trX820
trtWarrior   -0.463
trtPython    -0.520  0.614
trtGardstar  -0.476  0.615  0.589
trtXp820     -0.510  0.672  0.662  0.570
trtOptimizer -0.492  0.531  0.732  0.585  0.576

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.58023086 -0.48940343  0.06579332  0.57603416  2.74471174

Number of Observations: 311
Number of Groups: 12
```



Multiple Comparisons of Means

contrast	estimate	SE	df	t.ratio	p.value
Warrior - Magnum	0.058337984	0.07380857	294	0.790	0.9690
Python - Magnum	0.098955198	0.07374355	294	1.342	0.7614
Python - Warrior	0.040617214	0.06479678	294	0.627	0.9890
Gardstar - Magnum	0.176028222	0.08885931	294	1.981	0.3557
Gardstar - Warrior	0.117690238	0.07267048	294	1.620	0.5863
Gardstar - Python	0.077073025	0.07493464	294	1.029	0.9081
Xp820 - Magnum	0.178558079	0.07627337	294	2.341	0.1811
Xp820 - Warrior	0.120220095	0.06081210	294	1.977	0.3581
Xp820 - Python	0.079602881	0.06171983	294	1.290	0.7905
Xp820 - Gardstar	0.002529857	0.07741441	294	0.033	1.0000
Optimizer - Magnum	0.210393705	0.07401910	294	2.842	0.0537
Optimizer - Warrior	0.152055721	0.07158083	294	2.124	0.2777
Optimizer - Python	0.111438507	0.05408732	294	2.060	0.3112
Optimizer - Gardstar	0.034365482	0.07535551	294	0.456	0.9975
Optimizer - Xp820	0.031835625	0.06920178	294	0.460	0.9974

P value adjustment: tukey method for comparing a family of 6 estimates



Statistical Summary

ANOVA Results

- Treatment has a significant effect on fly count vs the control (pvalue: .021).

Random Coefficient Model Summary

- Week is highly correlated within county (ρ:-.825).
- County accounts for most of the variance: County:87.9%, Week(County):.2%, Residual:11.9%.
- The error structure AR($\phi = .618$), indicates the correlation between observations at lag = 1.
- Optimizer has the highest estimated mean difference in fly count vs. the base treatment Magnum.

LSMeans and Tukey Contrasts

- The LSMeans plot indicates that the treatments with the highest mean difference in fly count are Optimizer (159.2), Xp820 (149.5), and Gardstar (148.9) *(Backtransformed)*.
- The LSMeans 95% Confidence Interval for treatments are all greater than 0 indicating that all treatments are effective against repelling horn flies vs the control (no treatment).
- Tukey's test of multiple comparisons indicates that none of the treatments are significantly different (pvalue >.05).

Model Diagnostics

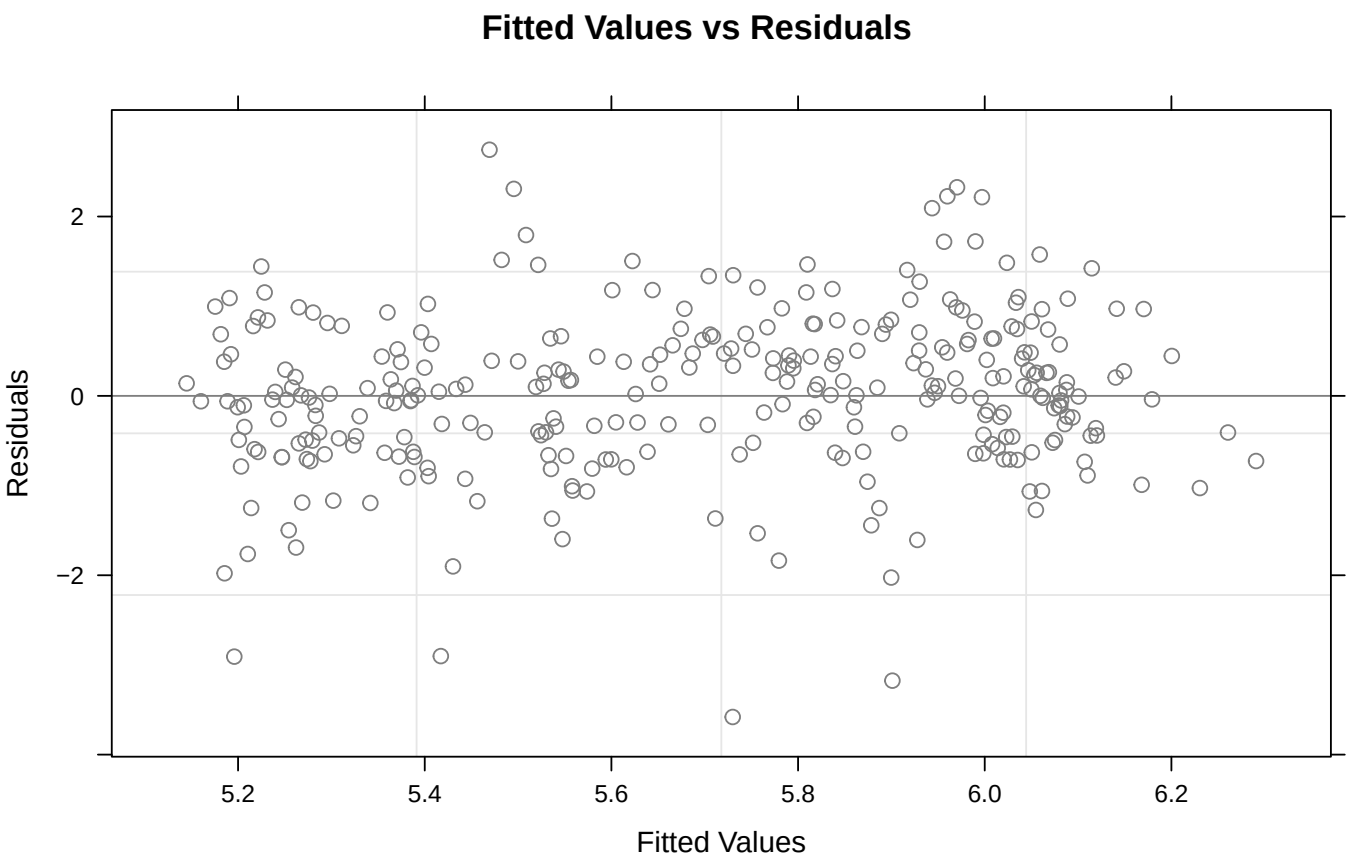
- Nice spread of residuals, but not exactly centered on 0 (median = .06).
- Predicted vs Actuals look okay, Warrior and Xp820 do not have great fits
- Box Plots of residuals by treatment show some notable outliers
- ACF plots of residuals by county show some slightly significant correlation at lag = 1, but no other significant correlation for the other lags

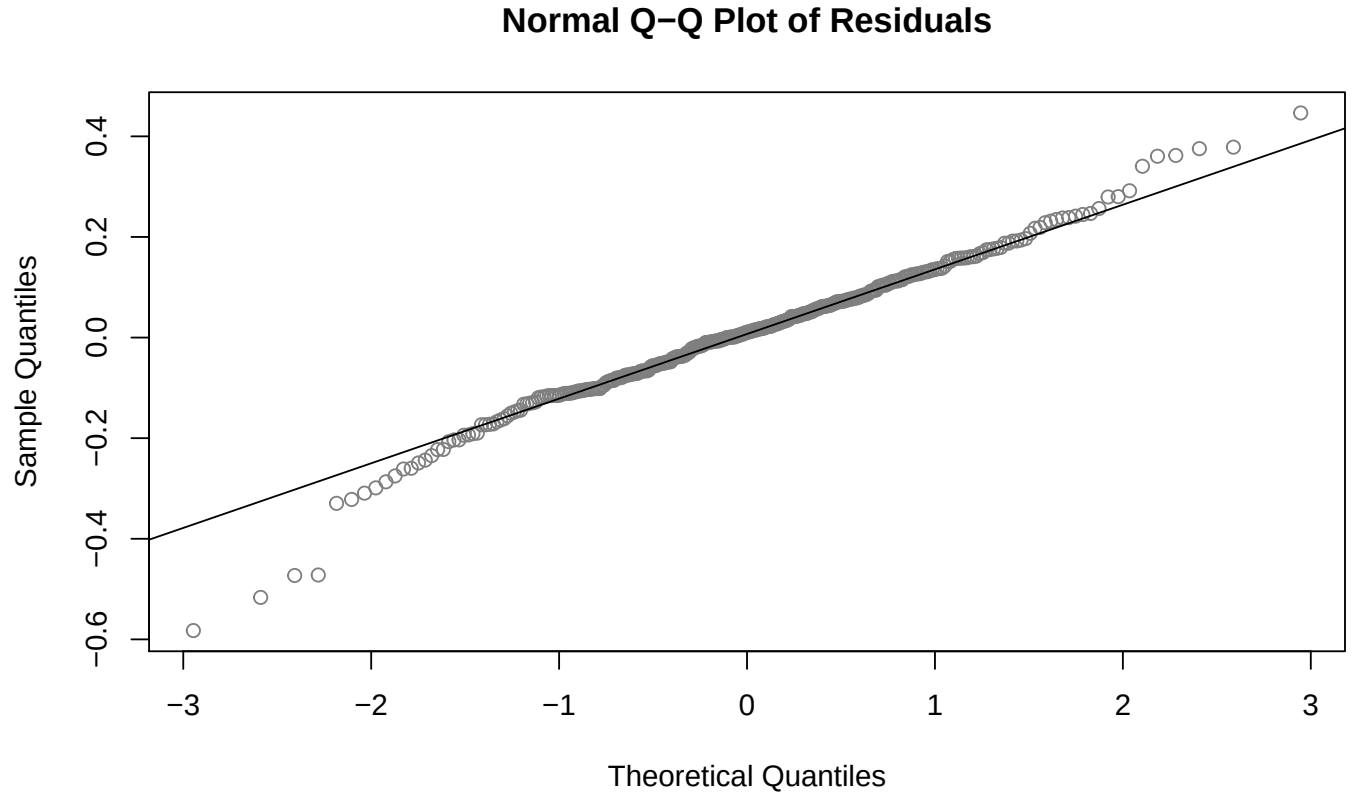
Conclusion

In this experiment, county appears to have an influence on the effectiveness of the pesticides. Many of the treatments had a large range in fly counts when compared across counties. Only six of the twelve counties were able to compare multiple treatments directly. As a result, many of the treatments were never tested against each other within the same county. It would be beneficial to do a follow up study with a balanced experimental design so see if one of the treatments with a higher predicted effectiveness proves to be statistically different than the other treatments.

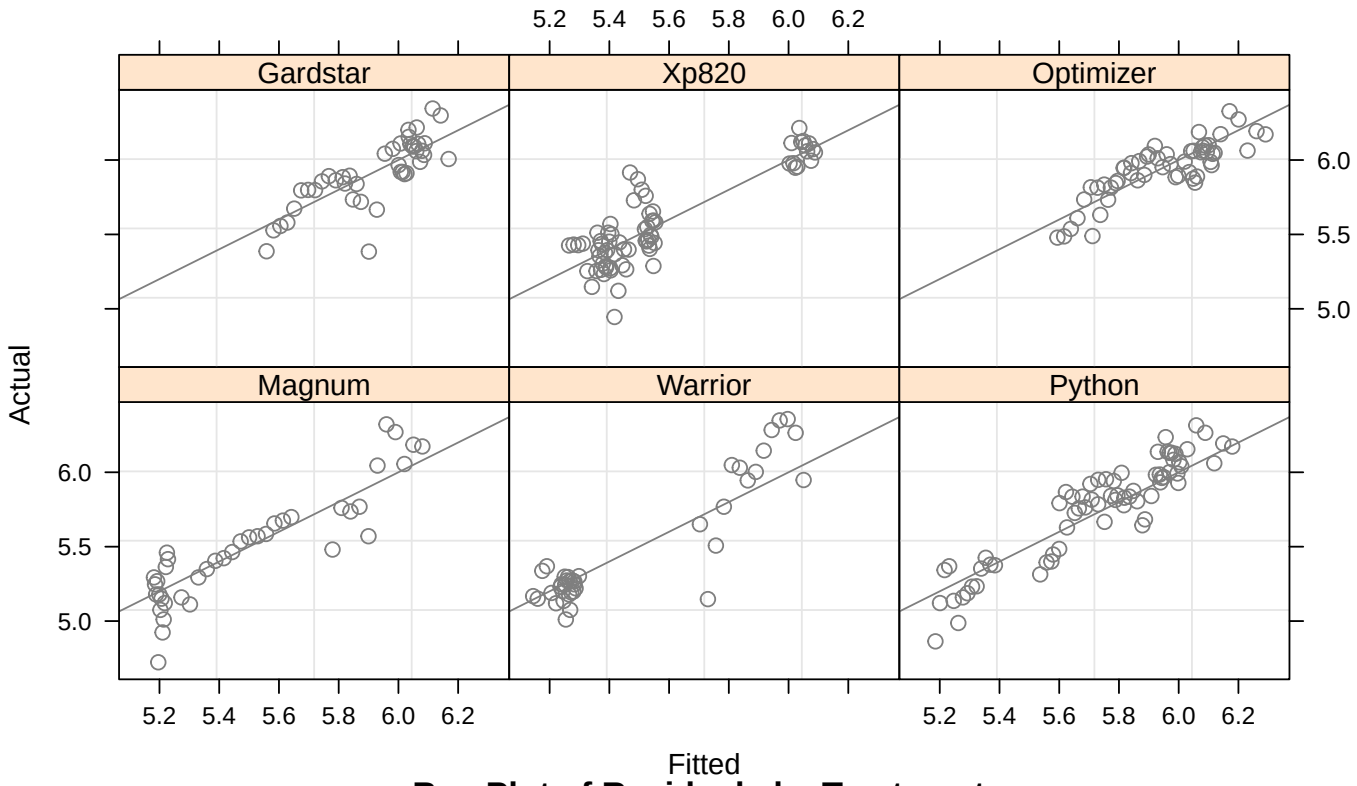
The six pesticides used in the final model all had a significant effect on horn fly counts, meaning they were all more effective than the herds that received no pesticides. However, a multiple comparison test showed that the treatments are not stastically different enough to declare any single treatment the best or most effective.

Model Diagnostic Plots

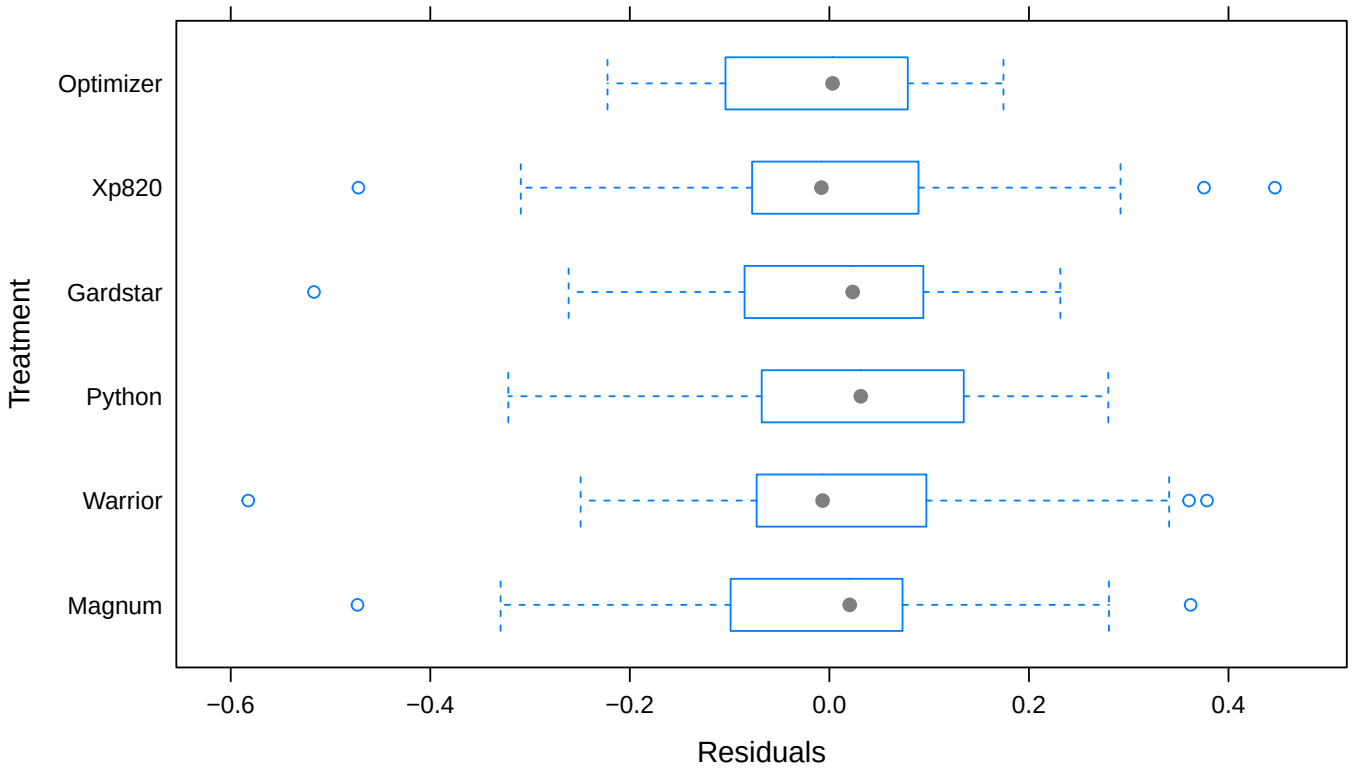




Fitted Values vs Actual Values



Box Plot of Residuals by Treatment



ACF Plot of Residuals by County