

Project -Patrick Gorrington

Addition of monochamol improves trapping efficiency for *Monochamus* beetles in a northern United States forest

My overall goal is to publish this statistical analysis as a part of a paper on trapping methods for these beetles that are potentially invasive. It would be nice to include a statistical 'best practices' for other researchers doing similar studies. I would also like to push for a paradigm shift to GLM in statistical methods used for these type of trapping studies. The current default is to use general linear models and  $\log(x+1)$  transformed datasets. I want to use this dataset as an example of how to analyze the data and also eventually reanalyze other existing datasets.

Studies have shown the enhanced effect of multiple chemicals baited traps (Pajares et al. 2004; Miller 2006; Macias-Samano et al. 2012; Hanks and Millar 2013). The most successful of these multiple chemical lures in sawyer beetles has been one including host plant chemicals and bark beetle pheromones acting as kairomones (Allison et al. 2001; de Groot and Nott 2004; Pajares et al. 2004; Miller et al. 2011, 2012). A *Monochamus* aggregation pheromone, 2-undecloxy-1-ethanol ('monochamol') was discovered recently (Pajares et al. 2010) and has been shown to possibly synergize with other chemicals in traps (Teale et al. 2011; Hanks et al. 2012; Macias-Samano et al. 2012; Miller et al. 2012). In this study we plan to test how adding this new pheromone component increases the efficiency and total catch when compared to a combination lure without the pheromone.

Cerambycid species have varied feeding habits that involve different parts of a tree from root to branches in the crown. Beetles thought to be rare could just be under-collected due to a canopy lifestyle, and some longhorns have shown a preference for traps in the canopy or on the ground (Graham et al. 2012). Partitioning a resource may also be important for sympatric,

closely related, species. Considering these factors, trap height can be optimized for the habits of a preferred quarry. We will place traps at the ground and canopy level to test if height has an impact on the trapping of these *Monochamus* species.

**Methods:** Traps were arranged in a replicate block design with each block having a total of four traps: two on PVC pipe hangers at ground level (hanging at ~1.5m) and two at canopy level (hanging at ~7m). Lure sets are a proven blend of tree chemicals and kairomones (referred to as M-) or those components plus the aggregation pheromone monochamol (M+). Each block has each lure type at each height. A transect consists of six replicate blocks, 30 meters apart with 10m between traps. This spacing pattern has been proved independent for blocks and traps in prior studies (Graham et al. 2012; Miller et al. 2016). Insects were collected and traps were rotated in each block every four days (July 11, 15, 19) to account for any potential location effect and to increase the effective number of sampling events. In total 96 traps were sampled, 24 of each unique lure/height combination. Statistical evaluations were performed in the R environment v. 3.3.2 (R core team 2016).

### *Statistical methods*

A fake dataset was simulated with reasonable smaller coefficient values to test the model's ability to recover these values using large sample size. The data was simulated using a negative binomial model using Stan and the r packages rstan v. 2.16.2 and rstanarm v. 2.15.3. The model was constructed using two predictors (to represent lures and height) in the form  $y \sim b_0 + b_1 * x_1 + b_2 * x_2$ , phi. Phi is the 'theta' parameter of the negative binomial distribution. The mu parameter is formed by exponentiation of the model  $b_0 + b_1 * x_1 + b_2 * x_2$ . The simulated model was done with predictor values of  $b_0=1$  for the intercept,  $b_1=3$  for lures and  $b_2=1.5$  for

height,  $\phi$  was 1.2 to match the approximate shape of a negative binomial distribution skewed to the left, enhancing the zero count. Both  $x$  values were drawn from a binomial with a probability of .5 since each trap was dichotomous for lure type and trap height. The created neg. binomial model was fit to a dataset of 1000 sampled points and the true parameter values were compared with the model output.

### *Bayesian model comparisons*

The leave one out method of model comparison is one of the most stringent methods for comparing Bayesian models. With this method, the function is calculated as many times as data points in the set, each time leaving one of the data points out. The average error can then be used to evaluate the ability of the model. This value is then compared between models to determine if one better fits the data. We use the loo R package (v. 1.1.0) to compare Bayesian GLM poisson and negative binomial models run in rstan.

### *Log transformation tests*

To see how this trapping data responds to the traditional  $\log(x+1)$  data transformation with subsequent ANOVA, the base R environment was used. We compared the different multivariate linear models and their variance analysis. There were models run on the raw, untransformed as well as the log transformed data set. A likelihood ratio test of the ANOVA scores can also be done using the `anova()` command.

### *Posterior predictive checks*

Posterior predictive checks are one of the best ways to determine if a given distribution is representative of your dataset's variance. With several traps with zero beetles caught, and larger standard deviations, it is important to explore this in this dataset. In the Rstanarm

package, posterior predictive checks were done by simulating 4000 datasets using the lures + height model under the GLM poisson and negative binomial distributions. The model is then fit to the simulated dataset and the parameter values are compared to the original dataset.

**Results:** 536 *Monochamus* beetles were captured over 16 days: 453 *M. mutator* LeConte, 57 *M. s. scutellatus* (Say) and 26 *M. notatus* (Drury), both had a mix of sexes. Since these species have similar biology they are combined into one dataset (Fig.1). The combined species average beetles/trap are: high M+ 4.25, SD 3.7; high M- 2.08, SD 2.4; low M+ 12, SD 11.4; low M- 4, SD 3.6. The modeled predictors of lure type and height are supported by all model types (table 1). A low height and monochamol addition both increased beetle capture.

*Fake data-* The fake dataset created to test the negative binomial distribution on a two-predictor model worked fairly well to return the original values (Fig. 2). These predictors differed by a small value, representing a small effect size for the height and lures and the model performed well under the negative binomial. Some tweaks in differentiating the predictors may still be needed.

*Model choice-* The choice of the count based GLM model was done through a combination of the leave one out test and posterior predictive checks on the models. The leave one out test highly supported a negative binomial over poisson for two predictors (diff 106.4, se 30.8). It also supported having four predictor variables in the negative binomial. Posterior predictive checks (Fig. 3) also show strong support for the negative binomial when looking at the proportion of zeroes in the dataset and the ability to have extreme maximum values. When the model has all four predictors (fig 3, bottom row), the difference lessens but is still present. The

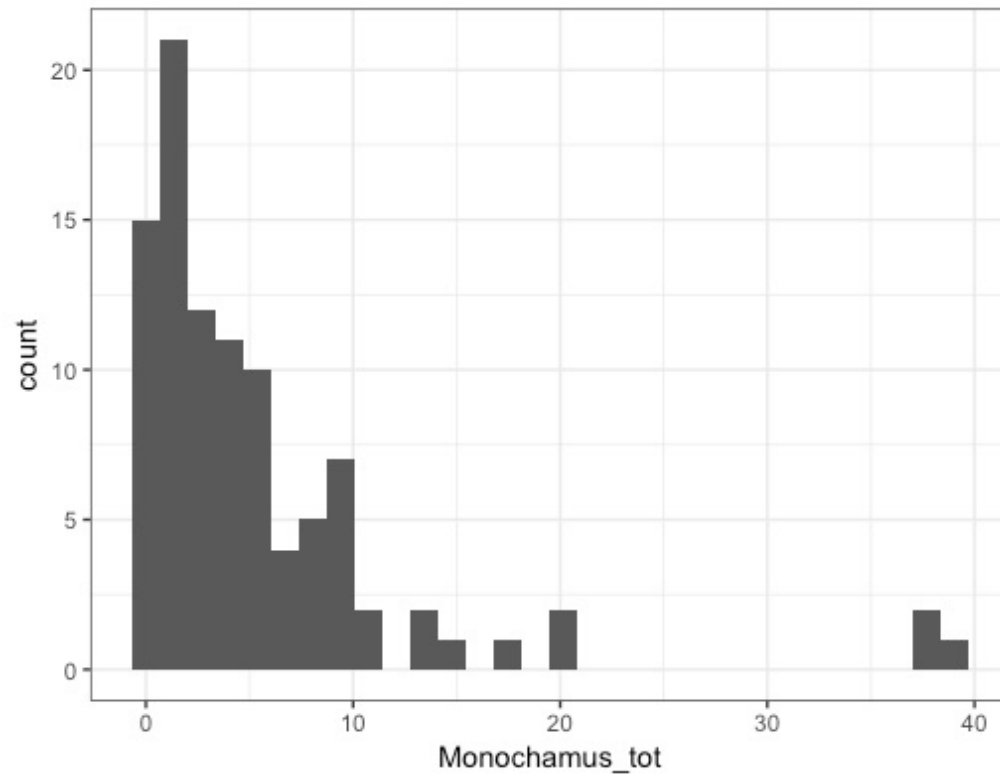
proportion of large trap values in the dataset also followed a similar trend (fig. 4). Leave one out does not prefer one model over the other for four variables (diff .2, se 1.5). The four predictor negative model is ultimately the best representation of the data including zeroes and overdispersion.

*Log(x+1) transformation*- The transformed dataset did get closer to a normal distribution for the data (Fig. 5). This was not an ideal accounting for variance in the dataset and future tests on other datasets will find the limits of where the transformation can fail to even return the correct predictors in simulation or with real data. The variance in the data seems to be loading certain parts of the model disproportionately, like the intercept of 27.2 in the two predictor model (Table 1). It is inappropriate to model count data using a continuous variable framework and the intricacies of what problems arise by doing this through transformation to continuous variables are still being explored.

### **Conclusion:**

This study shows a *Monochamus* beetle preference for monochamol baited traps at ground level. When modeling species such as these, it was found best to also account for the date of the transect and the block of traps. A negative binomial distribution for the model was found to have the best fit and seems to account for overdispersion caused by traps with zero catch and some traps with extreme numbers captured. A comparison of GLM models should be undertaken as a part of the statistical work flow for researchers dealing with trapping count data.

**Trapping result histogram**



*Figure 1*-raw number of traps for each count of Monochamus beetles

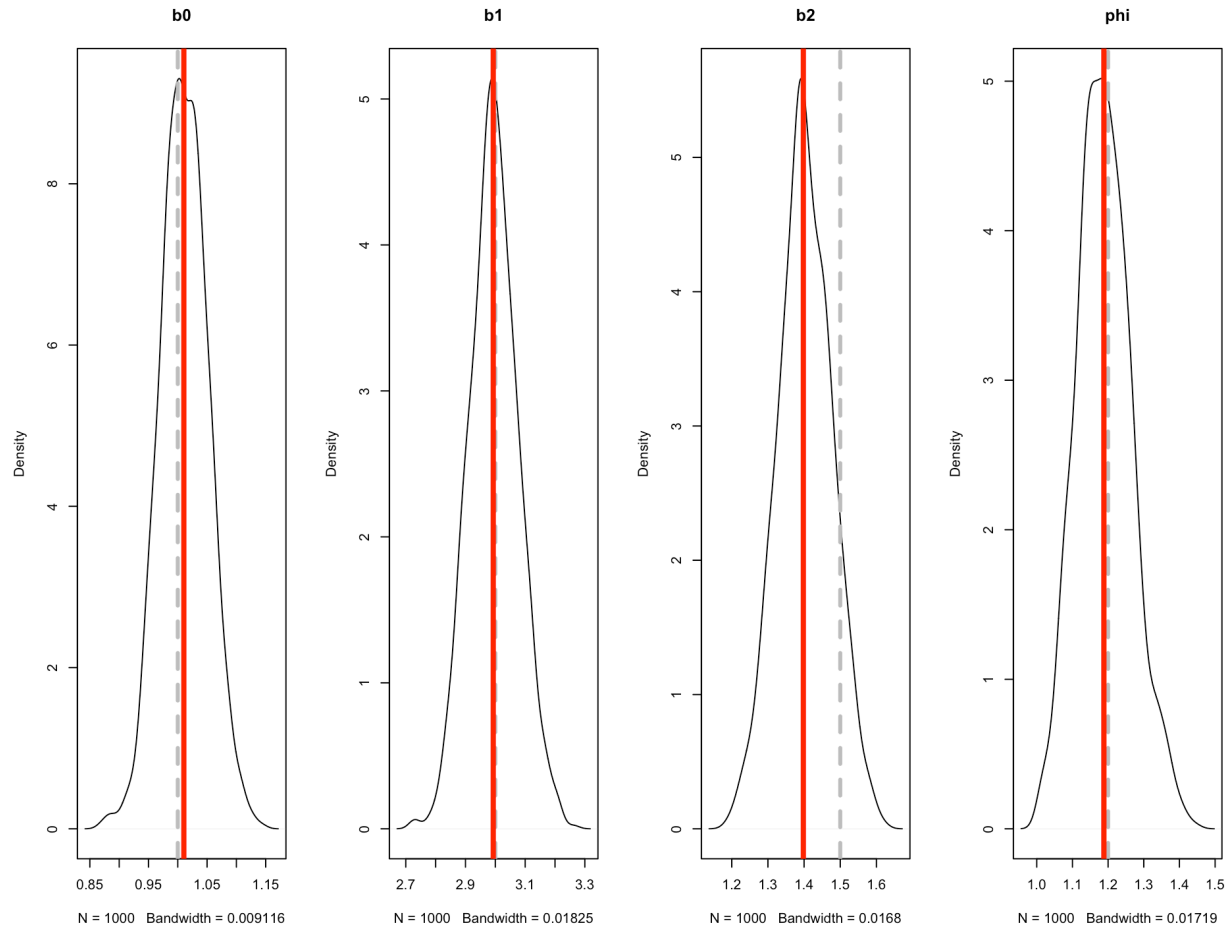


Figure 2-results comparison using fake data simulation. Gray line=true value, red line=estimated value, posterior distribution of each parameter based on fake dataset

Model	Intercept	SE	height low	SE	lures plain	SE
untransformed Anova 4 var	-0.04	1.72	4.83*	1.04	-5.08*	1.04
transformed Anova 2 var	27.2*	0.38	3.96*	0.45	-0.79*	0.45
transformed Anova 4 var	3.12*	0.58	3.96*	0.32	-0.79*	0.32
		SD		SD		SD
Neg Binomial 4 var	0.1	0.2	0.9	0.1	-0.9	0.1
Poisson 4 var	0.1	0.2	0.9	0.1	-1	0.1

Table 1. summary of model outputs, \*= significant <.05



## Posterior predictive checks-Negative binomial versus poisson

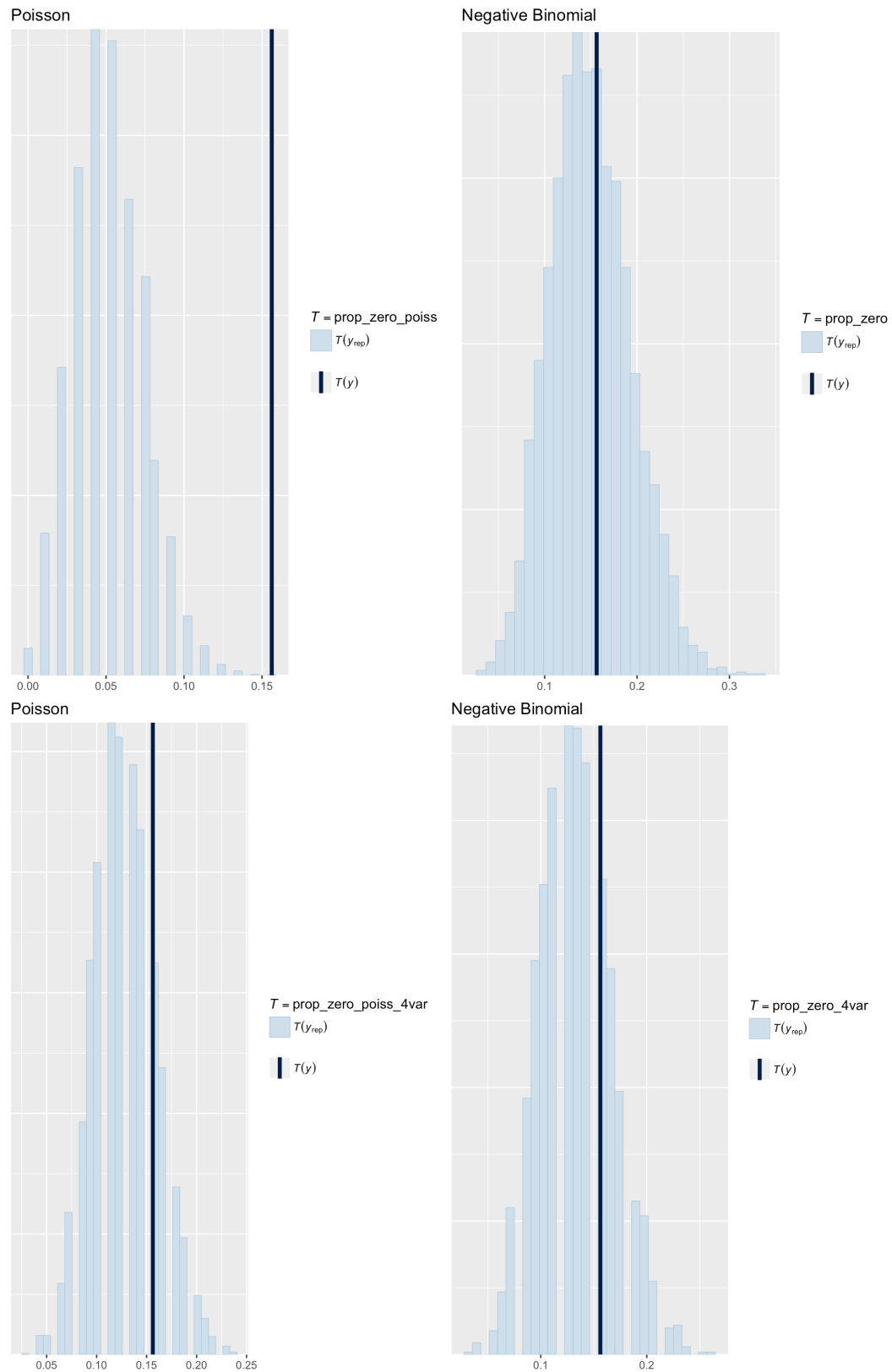


Figure 3- Results of posterior predictive checks for negative binomial and poisson stan GLMmodels. Top boxes represent a two predictor model and bottom row are from a four predictor model. black is dataset, blue is simulated data

Posterior predictive checks-Negative binomial versus poisson

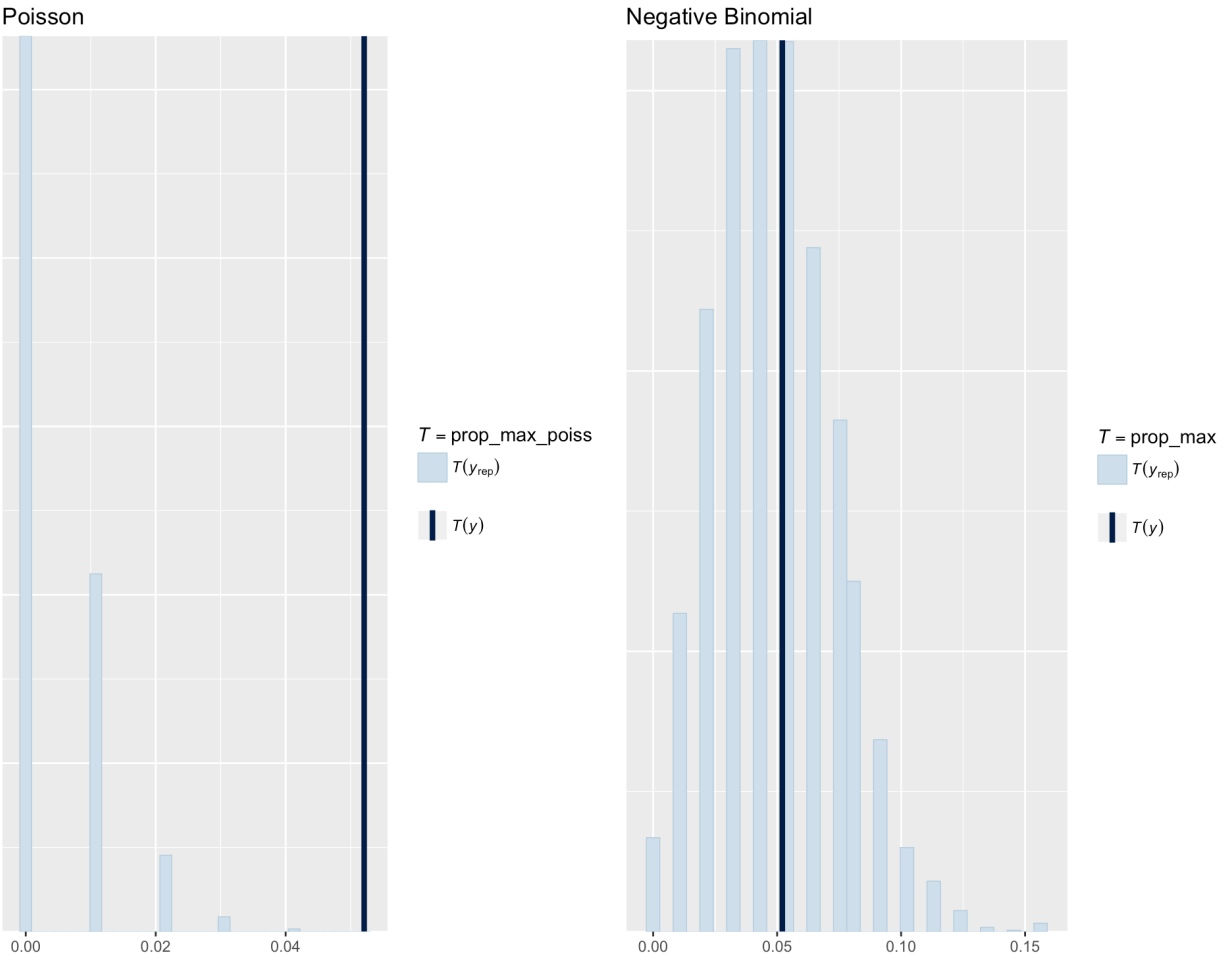


Figure 4- posterior predictive check of large trap catch values, black is dataset, blue is simulated data

### Monochamus transformed data histogram with overlain normal

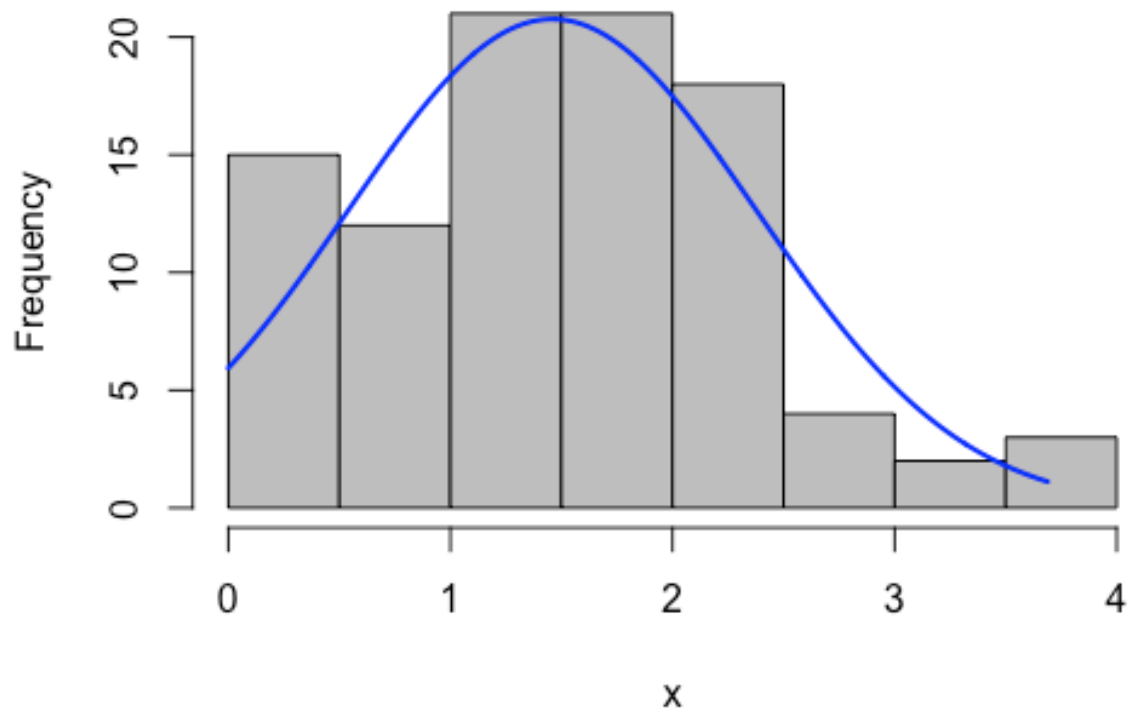


Figure 5- number of traps by  $\log(x+1)$  beetle count for the transformed all *Monochamus* dataset.

## References:

- Allison, J. D., J. H. Borden, R. L. McIntosh, P. de Groot, and R. Gries. 2001. Kairomonal response by four *Monochamus* species (Coleoptera: Cerambycidae) to bark beetle pheromones. *J. Chem. Ecol.* 27:633–46.
- de Groot, P., and R. W. Nott. 2004. Response of the whitespotted sawyer beetle, *Monochamus s. scutellatus*, and associated woodborers to pheromones of some *Ips* and *Dendroctonus* bark beetles. *JEN* 128:483–487.
- Graham, E. E., T. M. Poland, D. G. McCullough, and J. G. Millar. 2012. A Comparison of Trap Type and Height for Capturing Cerambycid Beetles (Coleoptera). *J. Econ. Entomol.* 105:837–846.
- Hanks, L. M., and J. G. Millar. 2013. Field bioassays of cerambycid pheromones reveal widespread parsimony of pheromone structures, enhancement by host plant volatiles, and antagonism by components from heterospecifics. *Chemoecology* 23:21–44.
- Hanks, L. M., J. G. Millar, J. a. Mongold-Diers, J. C. H. Wong, L. R. Meier, P. F. Reagel, and R. F. Mitchell. 2012. Using blends of cerambycid beetle pheromones and host plant volatiles to simultaneously attract a diversity of cerambycid species. *Can. J. For. Res.* 42:1050–1059.
- Macias-Samano, J. E., D. Wakarchuk, J. G. Millar, and L. M. Hanks. 2012. 2-Undecyloxy-1-ethanol in combination with other semiochemicals attracts three *Monochamus* species (Coleoptera: Cerambycidae) in British Columbia, Canada. *Can. Entomol.* 144:764–768.
- Miller, D. R. 2006. Ethanol and (-)-alpha-pinene: attractant kairomones for some large wood-boring beetles in southeastern USA. *J. Chem. Ecol.* 32:779–94.
- Miller, D. R., J. D. Allison, C. M. Crowe, D. M. Dickinson, A. Eglitis, R. W. Hofstetter, A. S. Munson, T. M. Poland, L. S. Reid, B. E. Steed, and J. D. Sweeney. 2016. Pine Sawyers (Coleoptera: Cerambycidae) Attracted to  $\alpha$ -Pinene, Monochamol, and Ipsenol in North America. *J. Econ. Entomol.* 109:1205–1214.
- Miller, D. R., C. Asaro, C. M. Crowe, and D. a. Duerr. 2011. Bark Beetle Pheromones and Pine Volatiles: Attractant Kairomone Lure Blend for Longhorn Beetles (Cerambycidae) in Pine Stands of the Southeastern United States. *J. Econ. Entomol.* 104:1245–1257.
- Miller, D. R., K. J. Dodds, A. Eglitis, C. J. Fettig, R. W. Hofstetter, D. W. Langor, A. E. M. III, A. S. Munson, T. M. Poland, K. F. Raffa, and A. E. Mayfield. 2012. Trap Lure Blend of Pine Volatiles and Bark Beetle Pheromones for *Monochamus* spp. (Coleoptera: Cerambycidae) in Pine Forests of Canada and the United States. *J. Econ. Entomol.* 106:1684–1692.
- Pajares, J. A., G. Alvarez, F. Ibeas, D. Gallego, D. R. Hall, and D. I. Farman. 2010. Identification and field activity of a male-produced aggregation pheromone in the pine sawyer beetle, *Monochamus galloprovincialis*. *J. Chem. Ecol.* 36:570–83.
- Pajares, J. A., F. Ibeas, J. J. Diez, and D. Gallego. 2004. Attractive responses by *Monochamus galloprovincialis* (Col., Cerambycidae) to host and bark beetle semiochemicals. *JEN* 128:633–638.
- R-core-team. 2016. R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria.
- Teale, S. a, J. D. Wickham, F. Zhang, J. Su, Y. Chen, W. Xiao, L. M. Hanks, and J. G. Millar. 2011. A male-produced aggregation pheromone of *Monochamus alternatus* (Coleoptera: Cerambycidae), a major vector of pine wood nematode. *J. Econ. Entomol.* 104:1592–8.