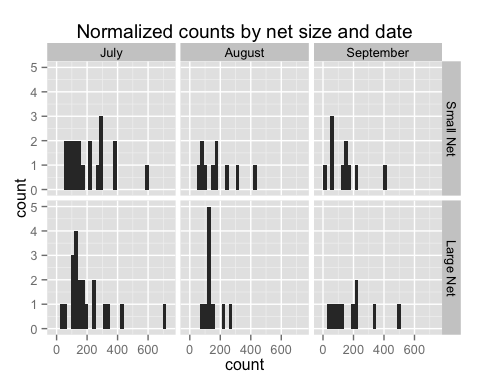
Effect of net size on estimates of abundance, size, age and sex ratio of Mysis diluviana Journal of Great Lakes Research - Statistical Analysis

John Tipton

January 3, 2016

To determine if there was any difference between the two sized nets, we have a few questions of interest. Our main questions of interest are 1) Is there a difference in density of *Mysids* caught between the two nets, 2) Is there a difference in mean length of *Mysids* caught between the two nets, and 3) Does the density when broken down by age and sex class differ between the two net sizes? Note that we will use count and density interchangeable in what follows because we normalize to a one area.

1. Is there a difference in density of *Mysids* caught between the two nets?



By looking at the distributions of counts across time, we see that they are similar between net sizes, but the normality of the counts is questionable. From the histogram, we see little differences between the two net sizes, except for perhaps a spike in catches of about 120 by the small net in August. Because counts cannot be less that zero and often are right skewed, the assumption of normality in the data is questionable. Hence, methods like linear regression, t-tests, and ANOVA that assume normal distributions are not ideal, but, for completeness we begin with a paired t-test.

From the paired t-test, we find that there is no significant difference ( = 0.069, = 39, -value = 0.945) in expected counts.

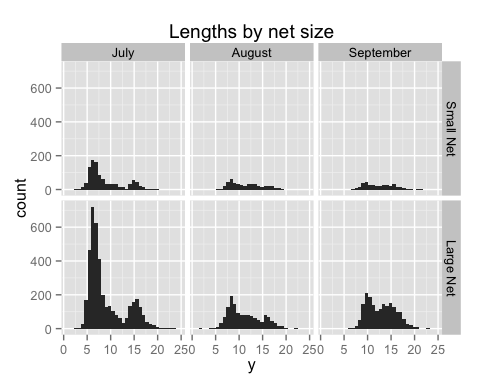
We test for a difference in density between the nets formally by fitting a model to compare if the count of *Mysids* caught by the large net is the same as the count of *Mysids* caught by the large net, after normalizing for the area of the two nets and controlling for covariates. The class of models known as generalized linear models (glms) offers a rigorous solution for modeling count data. These models allow for regression models that account for the effects of covariates like in linear regression but do not assume a normal distribution. For count data, there are two natural distributions to use: the Poisson distribution and the negative binomial distribution. The Poisson distribution has the strong assumption that the mean equals the variance, which is often not met in practical datasets. In our data, the count means are and the count variances are , for the small and large nets, respectively. Therefore we use a negative binomial model

where is the mean of the negative binomial distribution for sampling event and is an overdispersion parameter that allows for the mean and variance to be different. We model with covariates using the canonical log link function

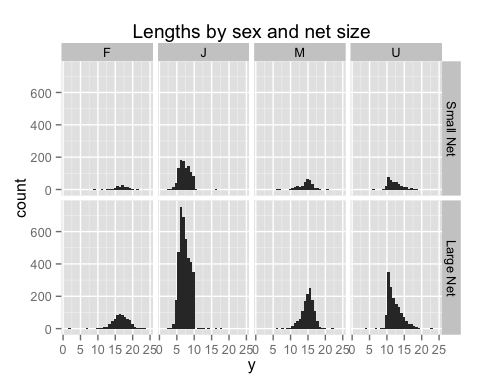
where is the set of covariates for observation . Then we perform inference on our coefficients , where the interpretation of is a percent change in count per unit change in .

To test for a change in total counts between the two net sizes, we construct a negative binomal regression model that examines the effect of net size while accounting for date of sampling and sampling location. The results shown below show that there is no evidence of an effect on the number of counts observed between the two net sizes (-value = 0.095, -value=0.924).

1. Is there a difference in mean length of *Mysids* caught between the two nets? First we examine some histograms and get an idea of what the length data look like.



When looking at the distribution of lengths between the two net sizes, we see that the histograms look quite similar across the different dates, with a general increasing trend through time. What stands out is that the small net catches about a quarter of the number of *Mysids*, as expected by the difference in net sizes, whereas the general shapes of the histograms appear quite similar across net sizes, suggesting there is not a difference in *Mysid* length distribution between nets across time.



When we plot the distribution of lengths between the net sizes with respect to sex class, the distributions look quite similar as well. This suggests that there is not a lot of difference in length distribution of *Mysids* between the two net sizes when broken down by sex class.

From the histogram of paired differences, it looks like there is some light evidence that the smaller net catches slightly smaller *Mysids*, although the distribution does seems to be centered near 0. From the paired t-test, we find that there is a significant difference ( = -2.158, = 39, -value = 0.037 ) in expected counts, although the -value is close to the 0.05 level. Despite the statistically significant difference, the practical difference is quite small as judged by the histograms and small difference in mean *Mysid* size difference (-0.293 mm).

One could argue that because the larger net catches four times the number of *Mysids*, the large net is more likely to catch *Mysids* at the extremes of the size classes (either really small or really large *Mysids*). To account for this, we perform a paired t-test using a trimmed mean, removing the smallest and largest 5% of the lengths before calculating the sample mean. From the paired t-test using the trimmed means, we find that there is not a significant difference ( = -1.754, = 39, -value = 0.087 ) in expected counts, although the -value is close to the 0.05 level. As with the untrimmed means, the practical difference is quite small as judged by the histograms and small difference in mean *Mysid* size difference (-0.247 mm).

To better account for the different sample sizes (the group means are from unbalanced sample sizes), we fit a linear model and a robust linear model (that assumes overdispersion in the data). We start by checking for heterogeniety in the group variances.

When looking at the distribution of sampling variance by sampling occasion (10 locations, 2 nets, and 4 different sampling dates gives 80 total sampling ocassions), we see that there is some difference in variance of lengths by sampling occasion and Bartlett's test (-value = 7.4310^{-175}) and Levene's test (-values = 5.36110^{-145}).

To test if there is a difference in length of net sizes while accounting for heterogeneity, we use a weighted least squares regression where the weights are the inverse of the group variances.

From the weighted least squares regression, we find that there is a statistically signficant difference in mean *Mysid* length caught between the two nets ( = 2.134, -value = 0.033). Now, after accounting for the location, date, sex distribution, and heterogeneity of variance we find that the difference in mean length caught between the two net sizes is 0.086 mm.

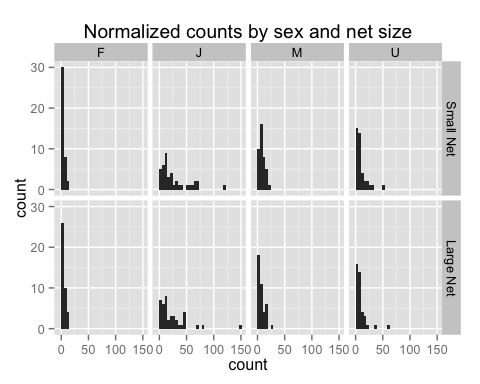
As a final analysis, we apply a robust regression, using an M-estimator model that accounts for heterogeniety in variance and presence of outlying observations.

For the robust linear model, we still find statistically significant differences in mean length between the net sizes (-value = 2.09, -value = 0.037 on 9111 degrees of freedom), with the effect size 0.078 about the same as from the weighted linear model

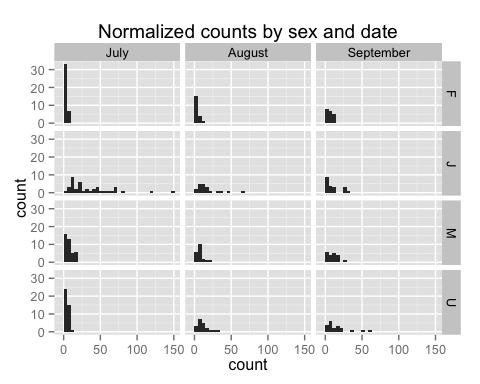
Although the statistical tests using both the weighted linear model and the robust linear model show statistically significant differences in mean length caught between the net sizes, this is not unexpected because have very large sample sizes and and the regression coefficients are very small relative to the other sources of variation in the model. Of greater interest is whether the observed effect is of practical significance.

The effect size of the mean length difference between the two net sizes is small (0.078mm) and the sample size is large (n = 9127). Given a large sample size, a hypothesis test will show statistical significance unless the population effect size is exactly zero (this explains why all of the p-values in the table above are less than 0.05). Therefore, the practical effect of a difference in mean length of 0.078mm on a species with a mean length of 10.351mm is small (this is the smallest effect of all the effects estimated by almost a factor of two) and a difference in means of this size is not of practical interest. Another measure of effect size is Cohen's which measures the difference in means relative to a pooled standard deviation. For our data, Cohen's =0.075, which implies that the effect of net size is quite small in terms of practical significance.

1. Does the density when broken down by age and sex class differ between the two net sizes? The next question we wish to explore is whether the total counts when grouped into sex classes of male, female, juvenile, and unknown vary with net size or other covariates. A visual inspection of counts by sex class shows no difference in counts by net size, but does show a change in counts over time for each sex class.



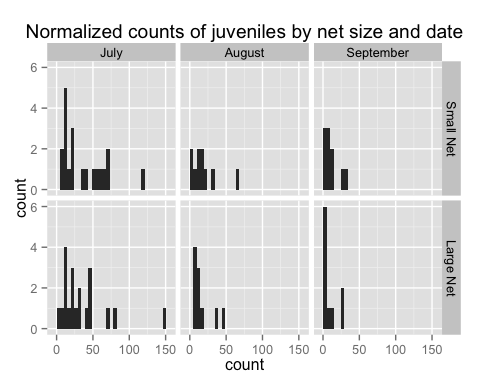
From the histogram of counts broken down by sex class and net size, we see no outstanding differences of counts between net size.



From the histogram of sex class by date, we see a pattern of juveniles maturing to males and females as time progresses and an increase in unknowns as juveniles grow in size but are not sexually differentiated.

To test for a change in counts broken down by sex category between the two net sizes, we construct a negative binomal regression model that examines the effects of net size, date of sampling, sampling location, and sex class. The results show that there is no effect of net size on counts when controlling for sampling date, sampling location, and sex class (-value = 0.185, -value = 0.853).

We can also examine if the counts within a sex class are different between the net sizes. To begin, we compare the counts of juveniles between the two nets.



The histogram shows no visual differences in the juvenile count between net sizes but shows a decrease in juvenile count through time, as expected. To test this formally, we construct a negative binomial model that controls for sampling date and station.

From our model, we see that time of year and location are important in influencing the juvenile counts, but net size is not (-value = -0.731, -value = 0.465). Similar negative binomial regression models for counts of males and females show similar effects (-value = -0.571, -value = 0.568 and -value = 1.607, -value = 0.108 for males and females, respectively). For further details and R code of the analysis, see the attached appendix and the online supplement at <http://jtipton25.github.io/mysis/>.

## Another issue to consider is the catch of a few large and small individuals in the larger nets. You should be able to test if that is simply due to the larger sample size obtained in the larger net. You suggest that may be the case.

Actually, this is quite a complicated statistical test as you are now testing the extreme observations of each sample. I prefer to look at the histograms and see that the distributions between net sizes for all of the different subsets of the data are consistent rather than do a formal statistical test (which is much more complicated). A quick and dirty extreme value analysis suggests that the model parameters for the generalized extreme value distributions are quite similar for the two net sizes, suggesting that both nets sample the large *Mysids* equally.

### Finally, the length analysis is using every measured mysids as an independent sampling unit. I think it would be more appropriate to use a paired t test on the mean lengths for that comparison as well, using each net instead of each mysid as the sampling unit. There is always the possibility that different size groups occur in different areas of the reservoir, and therefore that each mysids measured are not independent from the other mysids measured in a given net haul. Please consider that in your response.

It is true that different sized groups occur at different areas in the reservoir, but that is accounted for in the linear regression model (a paired t-test is a special case of a linear regression model). When performing a paired t-test, it requires collapsing each site's length distribution to a mean, which necessarily loses information about the distribution of lengths at each site. The linear model above properly accounts for that extra variability in the lengths at each site. See the comments above.

# Reviewer #1: Review of GLR-D-15-00174

### 8) Line 145, 150 I'm not sure you controlled for these factors, rather it seems like they were factors in the analysis?

By including these factors in the regression model, it controls for their influence.

### 9) Line 149 I don't think the mean length is the correct variable of interest. The size distributions could be very different and provide similar mean lengths. Rather, I think an analysis of the size structure is appropriate, perhaps using some sort of contingency table to evaluate the homogeneity of proportions. I also think the analysis of population structure (male-female, juvenile) and male:female ratio would be more appropriately done by testing for homegenity of proportions and then looking at standardized deviation to see where the nets differed (if at all).

While a contingency table is a possible analysis, the beta regression model is a more flexible model. A contingency table to evaluate a homogeneity of proportions is a special case of the beta regression model (just as the t-test is a special case of a regression model). I prefer to analyze the data using a regression framework as it is a more general framework that can account for multiple covariates at once (jointly estimating all the effects) instead of having to construct a table and perform analyses one at a time (marginally estimate effects).

### 10) The analysis of abundance (and other factors) was done basically to see if this study was done with one net or the other, would the results over the reservoir differ. This is an interesting question and an important one to address. But the net tows were done in paired combinations, so there is more information here. I would like to see something like a paired t-test (using each individual set of net tows as a sample), or a regression of values from one net plotted against the other from each sample to see if there was any bias based on the paired tows.

### 11) Line 163 and 166 I think you mean density and not number?

### 12) Please provide F values, df so the stats can be evaluated

We provide the data and the code in the appendix so the reader can evaluate the statistical analysis in its entirety.

### 13) Line 177 also caught more of the smallest individuals

# Reviewer #2: This manuscript has one straightforward message that will be useful to researchers interested in the collection of mysids, in that it supports the use of smaller diameter nets, which will be convenient from small vessels often used in lakes these animals have spread into. Overall this manuscript is acceptable for publication with some revisions. The details of the statistical tests need to be expanded upon and some language that downplays some of the size-distribution differences with different net sizes should be discussed in greater detail.

### 2) The manuscript is short so you have plenty of space to expand upon your statistical methodology which isn't made clear, and your presentation of only p values is insufficient to assess your analysis.

### 4) Downplaying the differences in size-distribution associated with the net sizes doesn't make the paper more persuasive. There are plenty of reasons where researchers may target the tiny fraction of smallest or largest individuals which are being missed by the smaller nets. Discuss this difference and end with the point you make that if possible, use the larger nets, but for most work, the smaller nets are an acceptable substitute.

Detailed comments: L38 Only 3 dates were sampled.

### L40 Size structure was slightly different (more at tails of size-distribution for large nets). Without this fact the last sentence of the abstract may not be clear.

### L122 use ind or N rather than lowercase n

### L130 use s or seconds

### L141 Be consistent N/m2, number/m2

### L142-151. You can be more concise here for some content and have to expand on other. You always have the same effects so you don't have to relist them. State the data isn't normal so a negative binomial transformation was used for regression on abundance, sex ratio…etc. See next note regarding statistical methods. You would be better off explaining the rationale for methods and the format in more detail rather than just the factors. For instance there is no detail given on the negative binomial model which can be constructed and assessed in a range of ways (likelihood format, AIC etc). Be descriptive here. If you used a specific stat software package, this would be the place to note it.

### L145-148 Why wouldn't you use a paired test? You collected each at the same time, so you could control much of the variability in catch (patchiness) by comparing them to each other (you comment L153-155 on this). Your size data shows multiple cohorts so it isn't surprising that you have non-normal data but the distributions are similar so should be easily discriminated using pairwise testing.

### L162-170. You give p values but no specifics as to method. The negative binomial model requires a range of output to assess the fit. The p values by themselves are not necessarily relevant so F (or z) values or r-mean squares should be included as well.

### L156. It is unusual for an author to cite unpublished data on their own paper. Ideally, you could easily add an additional graph that shows the series given the few number of tables and figures.

L161. Are you alluding to a drawback in your methods here given the lack of identification of sex increasing during the study? ### L167 Don't repeat yourself. "between net size and sampling date" redundant. Instead, "effect or interactions". ### L168 As above, you can include all of these as one statement: "density or juvenile proportion". L173 "likely not of practical interest". ### L175. So did you check for homogeneity of variance? A larger range but similar mode suggests that the larger nets had increased variance. This doesn't necessarily affect your results but should be noted. This might be relevant for any attempt at bioenergetics modelling for instance. The quartile plots in Fig. 2 suggest that the smaller net mean lengths are skewed downward (lower quartile larger than upper) which is reversed in the larger nets (except in July, which is more likely driven by the bimodal distribution in this month and suggests (there is no testing?) that the distributions are not normal and means are unlikely to be relevant for this month). You might be better off testing the bimodal distribution as two distributions (mode to mode). This may be a result of the different efficiencies of the two net sizes (you make no note of flowmeter readings and in your next section you calculate efficiency using areal ratio). It also might be worth more text describing the different distributions of %juveniles since the august and September distributions are very different even though their modes are similar (the distribution tail).

### L183. "probably not biologically relevant". Since you do not give these distributions (beyond the graphs), this isn't necessarily your call since some researchers may be focused on the very tails of the distribution that you dismiss. Instead, note that they are very small differences in numbers and that researchers interested in these larger individuals may wish to utilize the larger nets. Instead, move the sentence that is "Thus, these two nets…" up to be the second sentence (you state that there is essentially no difference in their distribution), then follow up with the caveat with respect to the larger size distribution afterwards.

### L187. The slight bias where the smaller nets are missing the largest individuals may by more relevant in late October or November when many programs sample to get egg production rates. Larger individuals appear to be able to avoid the smaller nets (they are certainly more powerful swimmers) and may also detect a difference in the net-mouth "head-wave" if your efficiencies are different.

### L203. As noted previously, be careful with your wording. Your equipment is similar to certain studies, but it can just as easily be said that your mesh size (for instance) is smaller than the majority that specifically are used to target mysids. Your comment on the size preventing loss of small mysids may be spurious. Given your distributions (and mean size of ~10mm) did you even see any mysids less than 1 mm in size (or even up to 3 mm)? Fig. 3 suggests you didn't, so you shouldn't expect net extrusion at a 3:1 ratio. If the mesh size was used to be consistent with other monitoring methods, then just state that, which is a perfectly fine reasoning and supported by your results.

L214 "Our findings…" ### L217 "While the larger net captured more…" - you have waited until here to note that there are two possible reasons (increased sampling area or reduced net avoidance). This should be in your first sentence of your discussion since it is the one factor which is clearly different between the sampling methods. Your second point about the large net also catching larger numbers of smaller individuals only suggests a sampling area effect (wording should be "did not support"). Both factors can be happening with the different size animals - larger mature adults are significantly more powerful swimmers than the smaller juveniles. Your last point in this paragraph should be expanded upon to note the comments listed above (that some researchers may be targeting the largest individuals to estimate population growth rates). L229. Remove "Conclusions" - it is the last paragraph anyway. L230. The term "biologically relevant" is loaded and likely to create disagreement even if there isn't any in the findings. You don't know what is biologically relevant to mysids or to other trophic levels that are affected by them (e.g. large individuals are known to have a significant impact on trophic dynamics beyond just their biomass (a la Fry and Quinones 1994, Ellis et al. 2011) and the presence of mysids will shift size structure of the entire plankton when present. Size is important. L236. This sentence essentially repeats the last in the previous paragraph. Combine them and expand on as noted in the L217 comment.

Table 1: You can save a huge amount of space in your table by just using "Conical" instead of "Conical plankton" and "Pyramid" instead of "Inverted pyramid" on two lines and note in the caption (you should also explain "framed net"). This table can easily be portrait with this change thus freeing up more space. This table is a major source of your extremely large reference section. This would be better suited for a supplementary appendix since it is not directly essential to the finding of this study.

Fig. 1. If you are going to note lat-long lines on your map, you should use a different line marking (e.g. dashed). Fig. 2 and Fig. 3 captions don't match your graphs because the next two figures pages are reversed. Fig. 2. The dots are means and the lines are modes? This isn't made clear. Axes titles may have to be larger. Fig. 3 The 1.0 m diameter net bars should be white to be consistent with the other graph. Font may have to be larger for the axes and the inset pie-graphs.