How does type and number of volunteers affect data collected during animal surveys in Peru?

Tom Bowey Strange and Dr Sarah Papworth



Report produced for Operation Wallacea as part of the Royal Holloway University of London Microplacement Scheme

21st July 2017

Introduction

Citizen science, which utilises groups of volunteers to collect survey data, has been praised as an inexpensive method of acquiring large amounts of data (Delaney et al., 2008). Using groups of volunteers provides a larger manpower pool that can provide more data than could be obtained without them, and volunteers receive enjoyment from and education about the environment (Fore et al., 2001; Boudreau and Yan, 2003). Trained volunteers can gather field samples of a comparable quality to professional ecologists for a fraction of the cost (Fore et al., 2001). However, every volunteer is different and the ability to differentiate between animals of similar sizes and appearances may be affected by differences in volunteers, such as age and education. Delaney et al. (2008) found that education level of students was a strong predictor of a student's ability to identify the species and sex of crabs found in the field, with sex of crabs requiring a higher level of education to identify correctly than species of crab.

Despite the benefits of using trained volunteers to collect data, it is not widespread practice to used data obtained from volunteers in scientific literature because it is believed to be less reliable than professionally obtained data (Delaney et al., 2008). However, there is a growing number of studies that have tested the reliability of volunteer obtained data and have found that it is accurate and reliable for its cost (Delaney et al., 2008; Boudreau and Yan, 2003; McNicol et al., 1995).

Operation Wallacea is a conservation research organisation that utilises student volunteers to obtain data for real-world research in a variety of locations around the world (Operation Wallacea, 2017b). The data used in this experiment was obtained from study sites located in the Pacaya-Samiria Reserve in Peru (Operation Wallacea, 2017a). This study aimed to compare the data obtained by school students with university students who were volunteering with Operation Wallacea to assess whether there were any differences in the numbers of animals seen or caught in a session while considering other variables such as weather, time of day and location of sampling. The number of students participating in each survey was recorded to see if it had any effect on the data collected, although this is anticipated to not affect the identification of species found in the field, consistent with the findings of Delaney et al. (2008). The data obtained also provides indications of the biodiversity of the study site in Peru.

Methods

Study site

The Pacaya-Samiria National Reserve is the largest protected area in Peru, spanning over 20,000 km^2 of tropical rainforest and is a truly exceptional wilderness area. Situated deep in the rainforest of the western Amazon basin, the Pacaya-Samiria reserve teems with aquatic and terrestrial life. The two major rivers that bound the reserve are the Ucayali and Marañon, and they join to form the River Amazon right at the point where the reserve begins. Data was collected from June 13 to July 5, 2017 - in the middle of the low water season –at Tacshcocha guard camp.

Data collection

On each survey, the trained biologist and guide which led data collection were joined by groups of school and university students. Total group size varied between 5 and 16. Five surveys are analysed in this report: terrestrial transects, aquatic dolphin transects, aquatic bird transects, fishing bat transects, and fishing surveys.

Terrestrial transects

Standard distance sampling methods were used to census eight terrestrial transects. Each transect was slowly walked at 1 km/h, and the distance from the transect was recorded for any primates or ungulates found along the transect. Once the end of the transect was reached, there was a 20 minute resting period before data was collected on the return journey. Morning transects were started at around 7:30 AM, and afternoon transects were started at approximately 3 PM. For this analysis, the total number of individual seen, and the total number of groups seen were used as response variables.

Aquatic dolphin surveys

5 km aquatic transects were used to survey for the two dolphin species present, the pink Amazonian dolphin, *Inia geoffrensis,* and the gray dolphin, *Sotalia fluviatilis*. Morning transects were started at 9:30 AM, and afternoons transect started at 2:30 PM. The species, group size, and group composition were recorded for each dolphin sighted during the transect. In this analysis, total number of dolphins seen were used as the response variable.

Aquatic bird transects

3 km aquatic transects were used to survey for wading birds, starting at 6 AM and 4 PM daily. All individuals of 13 species of aquatic birds seen on the transect were recorded as part of the survey. In this analysis, total number of individuals seen was used as the response variable.

Fishing bat transects

5 km aquatic transects were used to survey for fishing bats, *Noctilio leporinus*, starting at 6:30 PM each evening. A spotlight was swept from side to side across the river as the boat moved at 14 km/h to spot fishing bats. In this analysis, total number of individual seen were used as the response variable.

Fishing survey

The fishing survey assessed a new location each morning at 9:30 AM and afternoon at 2:30 PM. Two capture methods were used: a 30m long, 3m deep and 3.5in-mesh net, and fishing rods. Both methods were used for one hour. In this analysis, the total number of fish caught were used as the initial response variable, with supplementary analysis on the number of fish caught in the net and number of fish caught by rod.

Analysis

In each analysis, the location of surveying, the time of surveying, the weather while surveying and the type and number of students who went on each survey were used as explanatory variables. In addition the interaction between the type of student and the number of students who went on a survey was analysed. The exception was the fishing bat survey which did not use time of surveying as an explanatory variable, because all surveying took place in the evening.

Results

Terrestrial transects

Individuals per km

Five data entries were removed from the analysis because they were missing data points, and one outlier was excluded from the analysis because it had a Cook's distance of 1, possibly distorting the accuracy of the analysis. These exclusions made the sample size n = 46. The fixed effects of the mixed effects linear model explained over half of the variance in animals spotted per kilometre ($R^2 = 0.522$). Transect used was included as a random effect and accounted for 32% of the variance. None of the fixed effects had a statistically significant effect on the number of animals spotted, however there was a trend between student type and number of students on the number of animals spotted. The summary of the fixed effects are shown in Table 1 and the interaction between student type and number of students on animals spotted per kilometre is shown in Figure 1.

Table 1: Summary of the effects of various fixed variables on animals spotted per kilometre in terrestrial transects.

Variable	Statistical test	χ^2	df	p
Time	Chi-squared	1.553	1	0.213
Weather	Chi-squared	6.916	4	0.140
Student type	Chi-squared	1.61	1	0.204
Number of students	Chi-squared	1.322	1	0.250
Student type: Number of students	Chi-squared	3.701	1	0.054 *

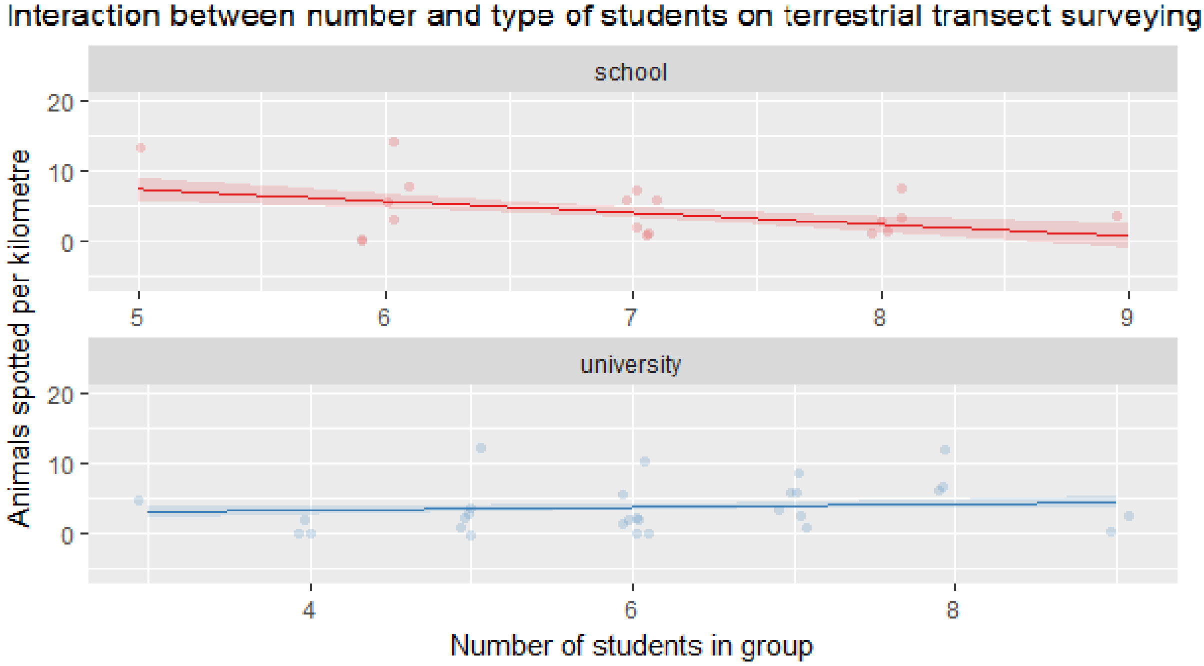

Figure 1: Effect of the interaction of student type and number on animals spotted per kilometre on terrestrial transects.

Groups per km

The fixed effects of the mixed effects linear model explained about 3% of the variance in groups per kilometre ($R^2 = 0.027$). Random effects (transect used) accounted for 19% of the variance in groups per kilometre. None of the fixed effects had a statistically significant effect on the number of animal groups spotted per kilometre. The summary of the fixed effects is shown in Table 2.

Table 2: Summary of statistical analysis of the effect of fixed variables on the total number of groups of animals spotted per kilometre.

Variable	Statistical test	χ^2	df	p
Time	Chi-squared	1.509	1	0.219
Weather	Chi-squared	2.61	4	0.625
Student type	Chi-squared	2.375	1	0.123
Number of students	Chi-squared	1.39	1	0.238
Student type: Number of students	Chi-squared	0.368	1	0.544

Dolphins

Two entries had missing data points so they were removed from the analysis and one outlier was removed because it had a Cook's distance value of over 1, thus skewing the results. This brought the sample size to n = 44. The fixed effects of the mixed effects linear model explained about 43% of the variance for total number of dolphins spotted (R^2 = 0.431). Random effects (location of sampling) accounted for 32% of the variance in the total number of dolphins spotted. None of the fixed effects had a statistically significant effect on the number of dolphins spotted. The results of the analysis are summarised in Table 3.

Table 3: Summary of statistical analysis of the effect of various variables on the total number of dolphins spotted.

Variable	Statistical test	χ^2	df	p
Time	Chi-squared	2.286	1	0.131
Weather	Chi-squared	3.878	3	0.275
Student type	Chi-squared	0.789	1	0.375
Number of students	Chi-squared	2.469	1	0.116
Student type: Number of students	Chi-squared	0.161	1	0.688

Wading birds

One entry had missing data points and two outliers had a Cook's distance of 1, so these were removed from the analysis, bringing the sample size to n=27. The fixed effects of the mixed effects linear model explained about 42% of the variance for birds spotted per kilometre (R^2 = 0.421). Random effects (location of sampling) accounted for 38% of the variance in the total number of birds per kilometre. None of the fixed effects had a statistically significant effect on the total number of wading birds spotted. The results of the analysis are summarised in Table 4.

Table 4: Summary of statistical analysis of the effect of various variables on the number of wading birds spotted per kilometre

Variable	Statistical test	χ^2	df	p
Time	Chi-squared	0.350	1	0.131
Weather	Chi-squared	2.110	2	0.275
Student type	Chi-squared	0.372	1	0.375
Number of students	Chi-squared	0.126	1	0.116
Student type: Number of students	Chi-squared	0.014	1	0.907

Fishing bats

One entry had missing data points so it was excluded, leaving the sample size at n=17. The fixed effects of the mixed effects linear model explained about 62% of the variance for bats spotted per kilometre (R^2 = 0.621). Random effects (location of surveying) accounted for about 13% of the variance in the total number of bats per kilometre. Number of students in the surveying group had a statistically significant effect on the number of bats spotted, as did the interaction between number of students and student type on the number of bats spotted. In addition, there was a trend between student type and the number of bats spotted. The results of the analysis are summarised in Table 5 and Figure 2 shows the effect of the interaction between student type and number of students on the bats spotted per kilometre.

Table 5: Summary of statistical analysis of the effect of various variables on the number of bat spotted per kilometre.

Variable	Statistical test	χ^2	df	p
Weather	Chi-squared	0.815	1	0.367
Student type	Chi-squared	3.261	1	0.071 *
Number of students	Chi-squared	4.036	1	0.045 **
Student type: Number of students	Chi-squared	5.736	1	0.017 **

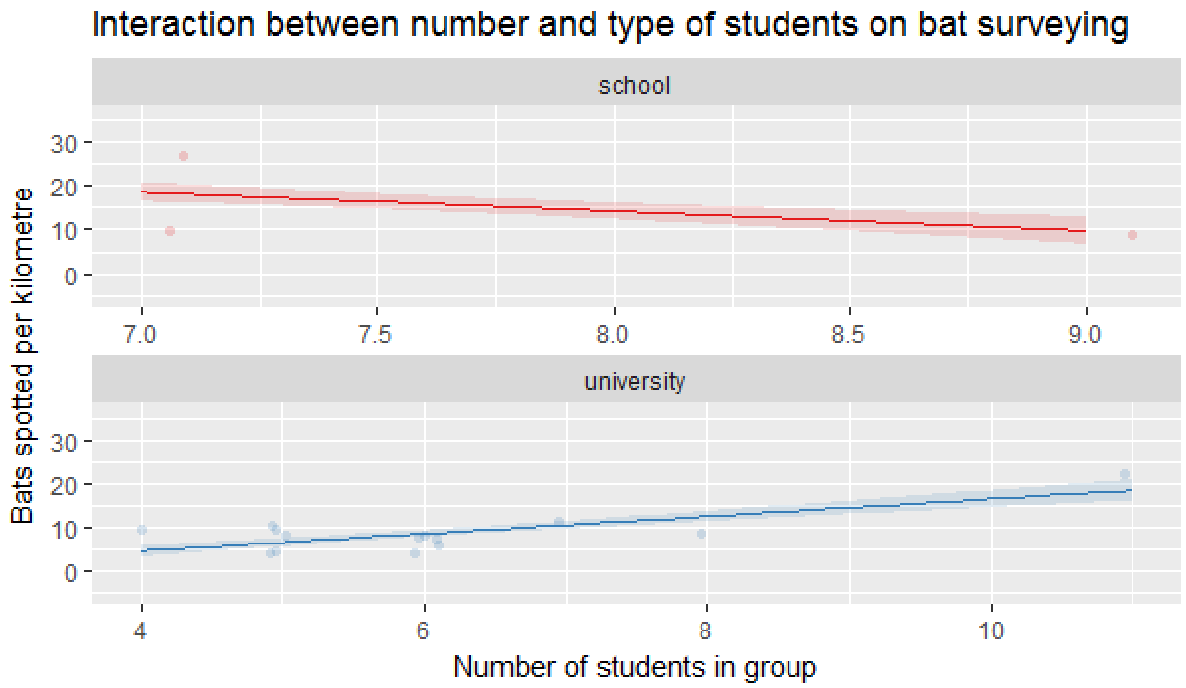

Figure 2: Effect of the interaction between student type and number of students on bats spotted per kilometre.

Fishing

No data was omitted so the sample size was n= 41. The fixed effects of the mixed effects linear model explained about 27% of the variance for total fish caught (R^2 = 0.265). Random effects (location of sampling) accounted for about 0.01% of the variance in the total number of fish caught. The number of students in a group and the interaction between student type and number of students had a statistically significant effect on the number of fish caught. The results of the analysis are summarised in Table 6 and the interaction between student type and number of students on total number of fish caught is summarised in Figure 3.

Table 6: Summary of statistical analysis of the effect of various variables on the number of total number of fish caught.

Variable	Statistical test	χ^2	df	p
Time	Chi-squared	0.389	1	0.533
Weather	Chi-squared	2.643	2	0.267
Student type	Chi-squared	1.138	1	0.286
Number of students	Chi-squared	5.937	1	0.015 **
Student type: Number of students	Chi-squared	4.068	1	0.044 **

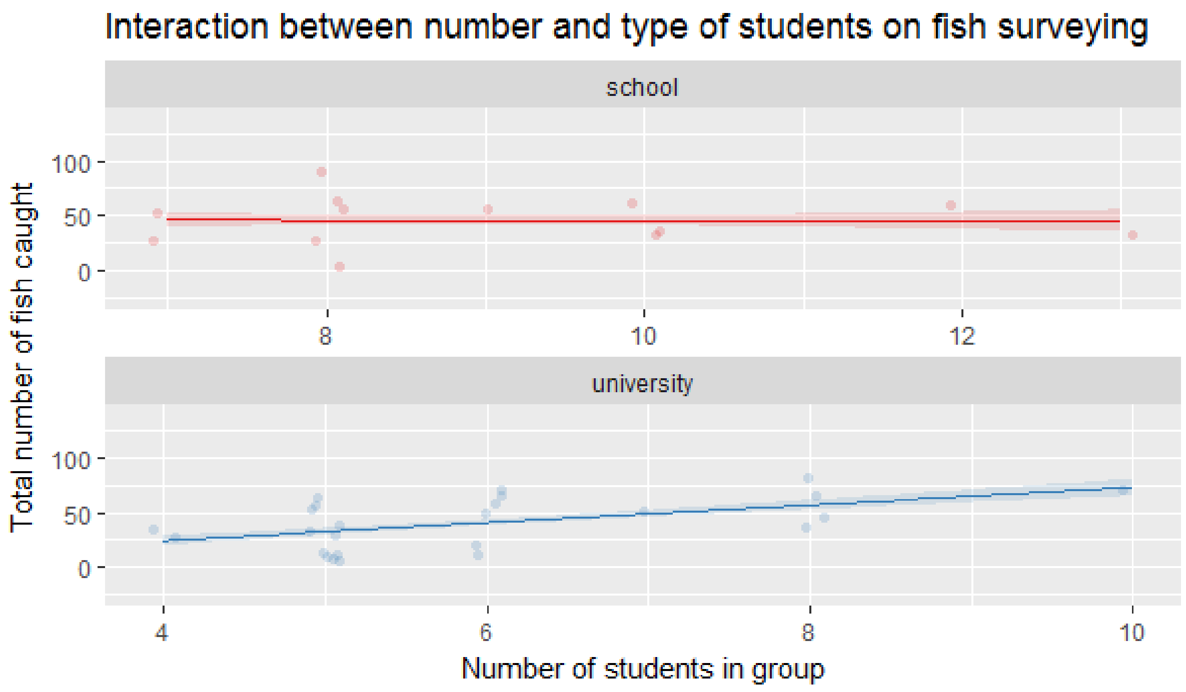

Figure 3: Effect of the interaction between number of students and student type on total number of fish caught.

When broken down to number of fish caught by rod and net, it was found that the number of students had a statistically significant effect on the number caught by rod (Chi-squared, $\chi^2 = 4.390$, p=0.036) and there was a trend between the interaction between student type and number of students for fish caught by rod (Chi-squared, $\chi^2 = 3.335$, p=0.0678). These fixed effects were not statistically significant for those caught by net, except for a statistically significant effect of the weather on fish caught by net (Chi-squared, $\chi^2 = 9.983$, p=0.007). The results for number of fish caught by rod and net are shown in Figures 4 and 5 respectively.

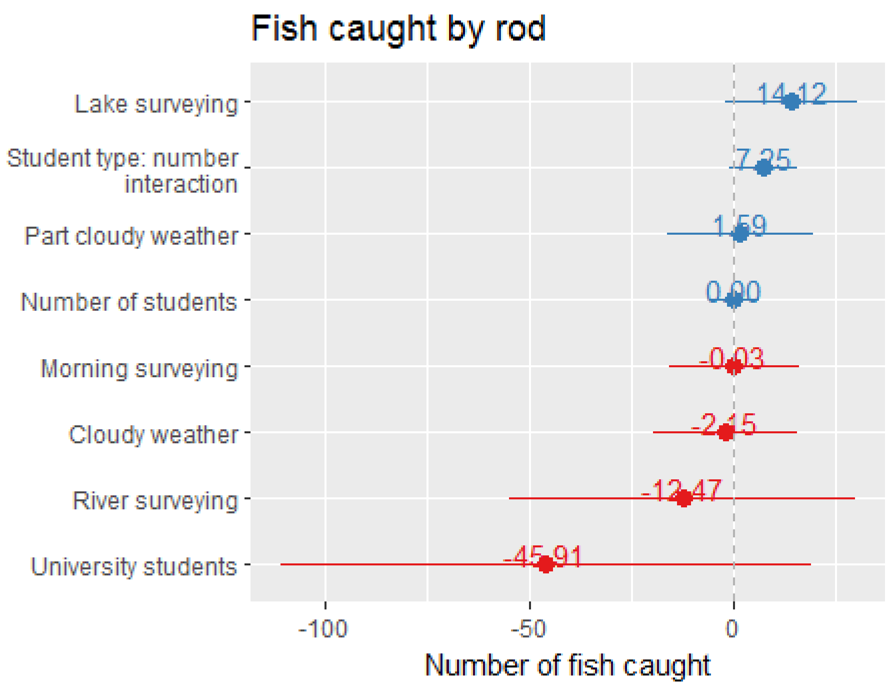

Figure 4: Graph summarising linear model of effects of various variables on number of fish caught by rod.

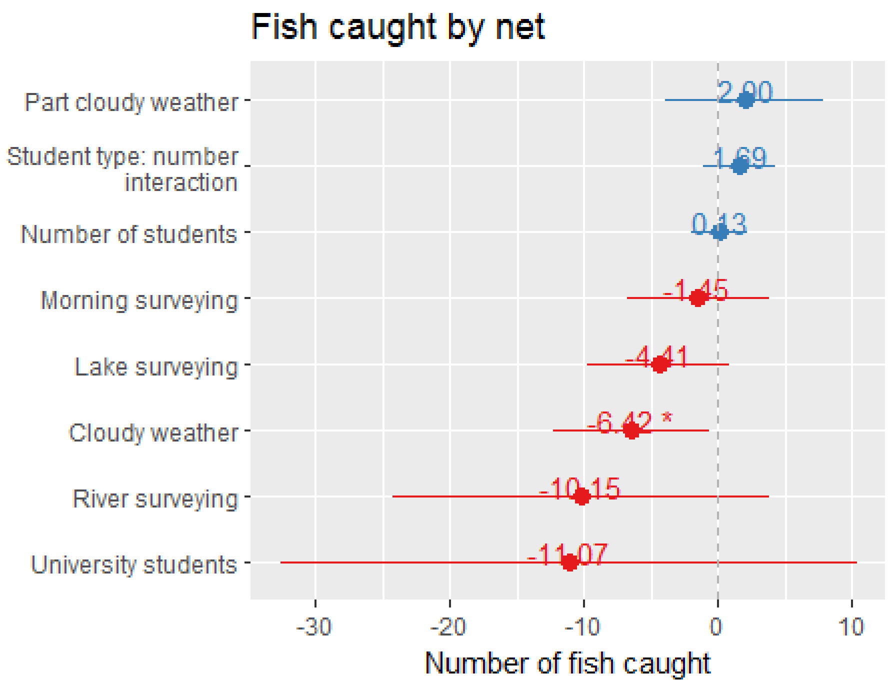

Figure 5: Graph summarising linear model of effects of various variables on number of fish caught by net.

Discussion

For the terrestrial transects, none of the fixed effects had any statistically significant effect on the total number of animals spotted per kilometre. There was, however a trend for the interaction between number of students and student type for the total number of animals spotted on the terrestrial transects. This may be due to larger groups of school students being louder and thus scaring away the wild animals, or they could be more distracted and thus less focussed on looking up to see monkeys in the tree tops. The interaction between student group size and student type on number of animals spotted on the terrestrial transects could be investigated further as it also has a statistically significant effect on number of bats spotted in the fishing bat survey and the total number of fish in the fishing survey. This interaction was not found to be statistically significant for the number of animal groups spotted per kilometre on the terrestrial transects, which could be explained by part of the group being scared away by the students, meaning the animal groups could be smaller. Investigating the effects of the fixed effects on animals per group might yield statistically significant results.

None of the fixed effects had a statistically significant effect on the total number of dolphins spotted. This may be because only the dolphins that came to the surface to breathe or near enough to be visible were recorded. This suggests that these dolphins were not afraid of humans, as it is taboo for local populations to intentionally hunt or kill dolphins (Bodmer et al., 2014), therefore they would appear regardless of the type or number of students who were in the motor boat observing them. The number of wading birds spotted was also not significantly affected by any of the fixed effects. This may be because they were disturbed by the noisy motor boat, rather than different student types or numbers in the boat.

Student type, number of students and the interaction between these two variables all appear to have effects on the number of bats spotted, but due to the low number of surveys with school children (n=3), these results should be interpreted cautiously. Further investigation with more school student data could be done to investigate this relationship. However, an increase in the number of bats seen with the number of students could be explained by the sampling taking place at night, when visibility was low and extra eyes would see more bats.

Finally, the number of students fishing affected the total number of fish caught. This is to be expected, as more fishing rods in the water would likely lead to more fish being caught. This is reflected in the number of fish caught by rod being significantly affected by number of students in the boat but not for those caught by net, as regardless of the number of students in the boat, there was only one net in the water. Furthermore, the interaction between student type and student numbers also had a statistically significant effect on the total number of fish caught and a trend on the number of fish caught by rod. The number of students in the boat has a greater impact on the total number of fish caught by university students, whereas larger school groups did not catch much more than smaller school groups. However, it should be noted that the school groups were mostly larger than the university groups, so repeating the surveys with more similar sized groups may affect the interaction between student type and number of total fish caught.

Perhaps the most unusual result obtained was the statistically significant effect of weather on the number of fish caught by net. This may be due to the fish having reduced visibility on cloudy days, or because of another variable that wasn't accounted for. Further research should be done to investigate this effect.

In conclusion, most of the fixed effects tested did not have any statistically significant effect on the number of animals seen or caught. Sometimes student numbers affected the number of animals seen or caught, but often these effects would be expected (e.g. more fish caught when more rods are used). The interactions between student number and student type on animals seen or caught may be worth further investigation as it is statistically significant or trending for some surveys but not others. Finally, the effect of weather on fish caught by net may be a random finding due to the limitations of the data, but it might be worth investigating further to see if more data yields the same results.

The following statistical findings that are worth reviewing with more data to determine whether the patterns hold:

1. The interaction between student type and number of students on the total number of animals spotted on the terrestrial transects per kilometre.

2. The interaction between student type and number of students on the total number of fishing bats spotted per kilometre.

3. The interaction between student type and number of students on the total number of fish caught and the number of fish caught by rod.

4. The effect of weather on the number of fish caught by net.

References

Bodmer, R., Fang, T., Puertas, P., Antunez, M., Chota, K. & Bodmer, W. 2014. Impacts of climate change on wildlife conservation in the Samiria River Basin of the Pacaya-Samiria National Reserve, Peru.

Boudreau, S. & Yan, N. 2004, "Auditing the accuracy of a volunteer-based surveillance program for an aquatic invader Bythotrephes", *Environmental monitoring and assessment*, vol. 91, no. 1-3, pp. 17-26.

Delaney, D.G., Sperling, C.D., Adams, C.S. & Leung, B. 2008, "Marine invasive species: validation of citizen science and implications for national monitoring networks", *Biological Invasions*, vol. 10, no. 1, pp. 117-128.

Fore, L., Paulsen, K. & O'Laughlin, K. 2001, "Assessing the performance of volunteers in monitoring streams", *Freshwater Biology*, vol. 46, no. 1, pp. 109-123.

McNicol, D., Mallory, M. & Vogel, H. 1995, "Using volunteers to monitor the effects of acid precipitation on common loon (Gavia immer) reproduction in Canada: The Canadian lakes loon survey", *Water Air and Soil Pollution*, vol. 85, no. 2, pp. 463-468.

Operation Wallacea, *Peru*. [online] Available from: http://opwall.com/undergrad-research-assistants/locations/peru/. [Accessed 18th July 2017a].

Operation Wallacea, *What is Operation Wallacea?* [online]. Available from: http://opwall.com/about-us/what-is-opwall/. [Accessed 18th July 2017b].