

# *Biology*

## **An Exploration of Biostatistics: A Comparison of Species and Biodiversity Within Two Small Creeks in Western Georgia**

Sponsoring Faculty Member: Professor Randall Colvin

Mary Hannah Robertson

### **Introduction/Background**

Biostatistics is a branch of statistics which allows for the comparison of biological data using methods primarily developed to account for the unique situations in biological data collection. Biostatisticians relate empirical data collected in the field or lab settings to a normal distribution or some methods allow them to non-parametrically relate biological data. Data can be used to compare groups non-parametrically by certain indices such as species diversity, evenness, or richness. This is helpful when determining correlations between data points or relating populations to a normal distribution. Many tools within biostatistics are used in conservation efforts. Statisticians can monitor success or failure of efforts for creek cleanup, water quality, among others. Statisticians can pair data to compare the whether the efforts allowed the creek to be more diverse or allowed for a greater density of species. Statisticians can assess water quality based on the species found within the environment. These species are called bio-indicators and their tolerance or intolerance for a harsh or clean environment provides a tool to determine the overall health of a habitat. "Understanding how site-specific habitat attributes affect macroinvertebrate assemblages and their calculations is crucial in bio-monitoring" (Stepenuk et al., 2008). Conservationists can also use biostatistics when examining the impact of humans and other disturbances on ecosystems. Medical professionals use biostatistics to relate diseases, measure growth, examine success or failure rates of certain types of medicines, as well as a wide variety of other uses. In parametric studies, statisticians can use samples and relate them to a normal distribution; this allows the statistician to reject or fail to reject hypotheses based on the results of the statistical methods and tests used. In summary, biostatistics is a way for biologists to take real data, organize, and analyze it to provide real and statistically significant results.

To examine properties of data collected in field or lab settings biostatisticians employ either parametric tools utilizing a normal distribution or non-parametric methods. Calculating species diversity, evenness, or richness metrics can be used to compare groups non-parametrically. Many of these tools are used in conservation efforts and statisticians can utilize these metrics to monitor the

impact of restoration efforts on biota and water quality. Making paired comparisons, using a reference stream of presumed high quality and a stream undergoing restoration, can be used to examine whether efforts enhanced species diversity or abundances.

In my study, I used biostatistics to investigate the biodiversity of two similarly sized small creeks in western Georgia. One creek, a small tributary of Park Branch Creek (PBC), runs through LaGrange College campus and parallels Forest Ave. The other creek is a small un-named tributary to the Chattahoochee River we named "Young Pine Creek" (YPC) it runs through managed forest property with little to no human disturbance.

### Methods

Sampling was conducted with help from students attending a General Ecology class at LaGrange College. Both Young Pine and Park Branch Creek were divided into ten transects. Samples of benthic macro invertebrates were gathered and isolated by transect using the kick net sampling technique. An inventory of the taxonomic families and species number of the species was taken. To gather an inventory of the amphibians and fish, backpack electro-fishing techniques were used following the transect system. Data was collected and organized. The following hypotheses were evaluated.

H <sub>0</sub> : Young Pine Creek, which has a very low anthropogenic impact, and Park Branch Creek, which has a high level of anthropogenic impact, will show no significant difference in the density of benthic macro invertebrates.	H <sub>a</sub> : Young Pine Creek, which has a very low anthropogenic impact, and Park Branch Creek, which has a high level of anthropogenic impact, will show a significant difference in the density of benthic macro invertebrates.
H <sub>0</sub> : Young Pine Creek and Park Branch Creek will show no significant difference in terms of species density of amphibians.	H <sub>a</sub> : Young Pine Creek and Park Branch Creek will show a significant difference in terms of species density of amphibians.
H <sub>0</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show a significant difference in terms of species abundance of fish.	H <sub>a</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show a significant difference in terms of species abundance of fish.
H <sub>0</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show no significant difference in terms of species diversity of benthic macroinvertebrates.	H <sub>a</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show a significant difference in terms of species diversity of benthic macroinvertebrates.
H <sub>0</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show no significant difference in terms of diversity of the community of benthic macroinvertebrates.	H <sub>a</sub> : Young Pine Creek, low impact, and Park Branch Creek, high impact, will show a significant difference in terms of diversity of the community of benthic macroinvertebrates.

Species diversity was measured using the Shannon's H diversity index for hypothesis 4 and Simpson's diversity index for hypothesis five. The relative abundance was measured using the Log- Normal Distribution, because of small sample size. Amphibians, benthic macro-invertebrates, and fish were sampled and parametric statistical analyses along with the Shannon's diversity index, Simpson's diversity index, percent EPT analysis, and analysis of the riparian vegetation of the creek. The Simpson's diversity index is a parametric test that allows for the examination of species evenness. Species evenness determines the health of the individual species populations within the ecosystem, based on their densities relating to the total number of species. It is important to analyze the evenness of a species to affirm that the species is stable and has established a niche within the ecosystem, and therefore is relevant to the study. The Shannon's H diversity index is used to determine species diversity within the ecosystem of the creek. It allows for the comparison of the two creeks in terms of number and variation of species. In the comparison of the two creeks with Shannon's H diversity index, diversity can be used to determine the overall health of the creek. The diversity of a creek shows the health by the variety for species and the availability of multiple niches, intra-, and inter-competition between species. The EPT analysis examines the water quality based on three specific families of macro-invertebrates. These taxa are: Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies). These three families of macro-invertebrates are pollution intolerant species, meaning they must have good water quality to survive. Lastly, the proportion of riparian cover with our environmental examinations were compared to the density of each species within each transect as well as between the two creeks. These comparisons allowed for the determination of the health of Young Pine and Park Branch Creeks.

## **Results**

The study suggests that the health, in terms of biodiversity, density, and habitat survey, of Young Pine creek supports a greater and wider variety of species. This is supported by the statistical analyses that were completed. An EPT taxa analysis, Shannon's H test of the community, Simpson's diversity index for each individual transect, along with physical habitat percentage measurements. For the EPT Analysis (Table 1 & Table 2), Young Pine has a percent EPT taxa of 51.33%. This evidence provides support for alternative hypothesis five. The EPT taxa index for Park Branch is 0.00%. This shows a detrimental water quality, allowing for no sampling of an EPT taxa.

Table 1. Proportion of EPT taxa within Young Pine Creek, sampled using kicknet sampling at ten transects

Young Pine: % EPT Analysis		
	Total Benthic Macroinvertebrates	300
	Total EPT	154
	% EPT	51.33%

Table 2. Proportion of EPT taxa within Park Branch Creek, sampled using kicknet sampling at ten transects

Park Branch: % EPT Analysis		
	Total Benthic Macroinvertebrates	83
	Total EPT	0
	% EPT	0.00%

Riffle beetle (Coleoptera) density of Young Pine is greater than that of Park Branch. This supports alternative hypothesis one. "Significant differences between macroinvertebrate communities inhabiting snag and riffle habitats existed in stream quality assessments" (Stepenuk et al., 2008). Young Pine creek has a greater population density of riffle beetles than Park Branch. This difference in species diversity suggests a greater water quality for Young Pine. The ability to make that assumption is supported by the evidence that riffle beetles require a stricter habitat. Shannon's diversity index was also completed. (Table 3 & 4) Shannon-Weiner Index results (1.89 (YPC), 0.799 (PBC)) indicates greater diversity of the invertebrate community in Young Pine Creek. The Shannon's diversity index shows a higher H value for greater diversity and less for less diversity (Park, 2007). (Table 3 and Table 4)

Table 3. Macroinvertebrate totals and Shannon's H Diversity Indices for Young Pine Creek.

Young Pine: Shannon's H Test						
	Benthic Macroinvertebrates		Totals	Pi	lnPi	PiInPi
	Plecoptera		36	0.120	-2.12026	-0.25443
	Chironomidae		90	0.300	-1.20397	-0.36119
	Decapoda		3	0.010	-4.60517	-0.04605
	Ephemeroptera		72	0.240	-1.42712	-0.34251
	Trichoptera		46	0.253	-1.37305	-0.34784
	Tipulidae		29	0.097	-2.33649	-0.22586
	Oligochaeta		12	0.040	-3.21888	-0.12876
	Coleoptera		4	0.013	-4.31749	-0.05757
	Plecoptera		1	0.003	-5.70378	-0.01901
	Simuliidae		3	0.010	-4.60517	-0.04605
	Amphipoda		3	0.010	-4.60517	-0.04605
	Chilopoda		1	0.003	-5.70378	-0.01901
	Total		300			1.894333

Table 4. Macroinvertebrate totals and Shannon's H Diversity Indices for Park Branch Creek

Park Branch: Shannon's H Test						
	Benthic Macroinvertebrates		Totals	Pi	lnPi	PiInPi
	Oligochaeta		67	0.807	-0.21415	-0.17287
	Decapoda		3	0.036	-3.32023	-0.12001
	Gastropoda		1	0.012	-4.41884	-0.05324
	Chironomidae		6	0.072	-2.62708	-0.18991
	Formicidae		1	0.012	-4.41884	-0.05324
	Tipulidae		3	0.036	-3.32023	-0.12001
	Bivalvia		2	0.024	-3.72569	-0.08978
	Total		83			0.799046

Individual Simpson's D tests showed a diversity index greater for Young Pine Creek. An example of a striking difference in the Simpson's diversity index is transect three. Transect three for Young Pine creek has a Simpson's index of 0.92 and an index of 0.5 for Park Branch. This shows a stark difference for the two transects. (Tables 5 & 6)

Young Pine: Simpson's D Transect		Species	Transects	1	2	3	4	5	6	7	8	9
		Plecoptera		1	4	0	4	9	0	6		12
		Chironomidae		2	5	12	4	9	4	19		35
		Decapoda		1	0	1	0	1	0	0		0
		Ephemeroptera		8	10	0	11	13	3	27		0
		Trichoptera		5	2	0	8	2	6	20		3
		Tipulidae		2	5	0	2	1	4	11		4
		Oligochaeta		0	2	1	2	0	0	4		3
		Coleoptera		2	0	0	0	1	0	1		0
		Plecoptera		1	0	0	0	0	0	0		0
		Simuliidae		0	0	0	1	0	2	0		0
		Amphipoda		0	0	0	1	0	0	0		2
		Chilopoda		0	0	0	1	0	0	0		0
		Total		22	28	14	34	36	19	88		59
		n(n-1)		462	756	182	1122	1260	342	7656		3422
		D		0.952381	0.962963	0.923077	0.969697	0.971429	0.944444	0.988506		0.982759

Park Branch: Simpson's D Transect		Species	Transects	1	2	3	4	5	6	7	8	9	10
		Oligochaeta		9	0	3	7	10	23	8	3	2	2
		Decapoda		0	0	0	1	1	0	0	1	0	0
		Gastropoda		0	0	0	1	0	0	0	0	0	0
		Chironomidae		0	0	0	1	0	3	1	0	1	0
		Formicidae		0	0	0	0	0	0	0	0	1	0
		Tipulidae		0	0	0	0	0	0	0	0	1	2
		Bivalvia		0	0	0	0	0	0	0	0	0	2
		Total		9	0	3	10	11	26	9	4	5	6
		n(n-1)		72	0	6	90	110	650	72	12	20	30
		D		0.875	#DIV/0!	0.5	0.888889	0.9	0.96	0.875	0.666667	0.75	0.8

Species diversity and abundance also relates to the difference in the physical habitat measurements. For transect three, the fish cover, understory, and over story are all the same with a measurement of error; however, the ground cover for Young Pine is much less than for Park Branch. Young Pine has a 5% ground cover, while Park Branch has 45%. This difference in habitat could contribute to the diversity within those transects. These abiotic factors contribute to macroinvertebrate distribution (Suriano et al. 2011). These tests suggest a significant difference in the health, in terms of biodiversity and density, of the two creeks. (Table 7 & 8)

Young Pine Creek		Transects	1	2	3	4	5	6	7	8	9	10
	Ground Coverage		5%	2%	5%	7%	5%	10%	40%	10%	15%	20%
	Understory		55%	60%	35%	20%	45%	20%	2%	60%	70%	80%
	Overstory		90%	70%	55%	50%	45%	60%	80%	75%	80%	60%
	FC		15%	15%	40%	70%	75%	43%	40%	50%	30%	50%

Table 8. Riparian coverage of individual transects at Park Branch Creek.

Park Branch Creek	Transects	1	2	3	4	5	6	7	8	9	10
	Ground Coverage		2%	45%	45%	10%	25%	31%	50%	22%	42%
Understory		0%	50%	20%	5%	5%	22%	25%	5%	10%	7%
Overstory		65%	80%	35%	27%	47%	18%	15%	81%	87%	93%
FC		40%	65%	50%	75%	62%	51%	50%	2%	60%	2%

The T-tests for both creeks resulted in insignificant results. We were unable to reject the null hypothesis for the densities of macroinvertebrates, fish, and amphibians. The alpha values were not consistent with a p value of less than or equal to the general accepted .01. The t-tests resulted in values of 0.32 for benthic macroinvertebrates of both creeks, while the result for amphibians and fish was 0.63. These values indicate no difference for the two creeks. The lack of statistical significance may be due to the small sample sizes in both cases. Small sample sizes result in a non-parametric distribution, making the results insignificant. In summary, the metric tests used to determine the health of each creek through the outlets of species diversity, density, and distribution resulted in significant differences between the two creeks. EPT taxa, individual Simpson's D, and Shannon's H all suggested Young Pine as the more diverse, healthier creek. The t-test showed no significant difference between the two creeks, but may be flawed due to small sample size. The results from all of the tests can then be examined for practical implications as well as repeatability of the study.

### Discussion

Our analyses helped determine that Young Pine Creek is an overall much healthier creek than Park Branch. The distinct effect on biodiversity and density of Park Branch creek is multi-causal. It is difficult to determine exactly what is causing the effect, but there could be a number of reasons. These reasons most likely stem from anthropogenic impact. The impact may be stemming from proximity to the roadway, grounds keeping, and/or allochthonous contributions that affect the pH and sedimentation of the creek. Sedimentation may contribute to low levels of EPT taxa in Park Branch (Wagenhoff et al. 2012). These effects can be minimized with proper monitoring and maintenance of the creek and the area within proximity of Park Branch creek. Efforts to minimize further anthropogenic impact from nearby roadway and grounds keeping practices will help further restore Park Branch. Continued study of Park Branch will help determine if the creek is regressing or progressing, and aid in identifying further restoration efforts.

Continued monitoring of allochthonous inputs, or materials from outside the creek such as leaves that contribute energy, may vary between creeks in part due to differences in riparian density (Wagenhoff et al. 2012). Monitoring these energy inputs between creeks may identify additional factors that influence species diversity. These allochthonous inputs could stem from the natural

riparian zone surrounding Park Branch, or they could result from grounds keeping deposits. Regular stream clean-up could also contribute to boosting creek health by ridding potential pH imbalance threats as well as material that could obstruct the creek and change the nature of the creek. These improvements will allow Park Branch to progress.

In conclusion, biostatistics provides results that express meaningful biological differences between sites of varying levels of ecosystem health. These results can be applied to practical uses that help biologists determine further actions. These methods and analyses can be repeated for further investigation. This begins a perpetual cycle for the improvement of biological systems and allows for a better understanding of those systems.

### References

- Park, C. (2007). Shannon–Wiener index. Oxford University Press. doi:10.1093/acref/9780198609957.013.7340
- Stepenuck, K. F., Crunkilton, R. L., Bozek, M. A., & Lizhu, W. (2008). Comparison of Macroinvertebrate-Derived Stream Quality Metrics Between Snag and Riffle Habitats. *Journal Of The American Water Resources Association*, 44(3), 670-678. doi:10.1111/j.1752-1688.2008.00197.x
- Suriano, M., Fonseca-Gessner, A. A., Roque, F. O., & Froehlich, C. G. (2011). Choice of macroinvertebrate metrics to evaluate stream conditions in Atlantic Forest, Brazil. *Environmental Monitoring & Assessment*, 175(1-4), 87-101. doi:10.1007/s10661-010-1495-3
- Wagenhoff, A., Townsend, C. R., & Matthaei, C. D. (2012). Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. *Journal Of Applied Ecology*, 49(4), 892-902. doi:10.1111/j.1365-2664.2012.02162.x