SPECIAL ARTICLE

Which statistics should tropical biologists learn?

Natalia Loaiza Velásquez[1], María Isabel González Lutz[2] & Julián Monge-Nájera[3]

1. Biología Tropical, Universidad de Costa Rica, 2060 San José, Costa Rica; natalia.loaiza@ucr.ac.cr
2. Escuela de Estadística, Universidad de Costa Rica, 2060 San José, Costa Rica; mariaisabel.gonzalezlutz@ucr.ac.cr
3. Laboratorio de Ecología Urbana, UNED, 2050 San José, Costa Rica; Universidad de Costa Rica, 2060 San José, Costa Rica; julianmonge@gmail.com

Received 13-IX-2010. Corrected 03-II-2011. Accepted 01-III-2011.

Abstract: Tropical biologists study the richest and most endangered biodiversity in the planet, and in these times of climate change and mega-extinctions, the need for efficient, good quality research is more pressing than in the past. However, the statistical component in research published by tropical authors sometimes suffers from poor quality in data collection; mediocre or bad experimental design and a rigid and outdated view of data analysis. To suggest improvements in their statistical education, we listed all the statistical tests and other quantitative analyses used in two leading tropical journals, the *Revista de Biología Tropical* and *Biotropica,* during a year. The 12 most frequent tests in the articles were: Analysis of Variance (ANOVA), Chi-Square Test, Student's T Test, Linear Regression, Pearson's Correlation Coefficient, Mann-Whitney U Test, Kruskal-Wallis Test, Shannon's Diversity Index, Tukey's Test, Cluster Analysis, Spearman's Rank Correlation Test and Principal Component Analysis. We conclude that statistical education for tropical biologists must abandon the old syllabus based on the mathematical side of statistics and concentrate on the correct selection of these and other procedures and tests, on their biological interpretation and on the use of reliable and friendly freeware. We think that their time will be better spent understanding and protecting tropical ecosystems than trying to learn the mathematical foundations of statistics: in most cases, a well designed one-semester course should be enough for their basic requirements. Rev. Biol. Trop. 59 (3): 983-992. Epub 2011 September 01.

Key words: education of tropical biologists, statistical courses, free statistics software, tropical biology journals.

Tropical biologists study the richest and often the most endangered biodiversity in the planet and in these times of climate change and mega-extinctions, the need for efficient, good quality research is more pressing than in the past. Previous authors have mentioned the need to improve the statistical training that tropical biologists receive, in order to reach those goals of efficiency and quality (Monge-Nájera 2002).

However, the statistical component in research published by tropical authors has been qualified as suffering from three problems: poor quality in data collection; mediocre or bad experimental design and a rigid and outdated view of data analysis (Fielding & Lauckner 1992).

In comparison with the scientific leaders, tropical statisticians are few, isolated and not always properly trained, having for example a poor grasp of the concept of variability and its usefulness in biology. Some scientific journals that publish in this field accept articles that lack appropriate descriptions of the experimental design and in some cases, 70 % of the articles have incorrect statistical analyses (Camacho 1997).

Some problems detected by the authors of Camacho's (1997) compilation are not so serious today, thanks to the Internet, which provides acceptable statistical software for free and helps statisticians from tropical countries to keep abreast of new developments and in contact with colleagues. However, the key to the problem lies in education and progress is slow in this field. The results of statistics courses taught to biology students are often unsatisfactory, and students too often end up hating statistics and being unable to select and apply appropriate test and procedures (Camacho 1997).

To summarize the problem, tropical biologists are not being taught the statistics they really need. The objective of this article is to identify the statistical tests that are most frequently used in the current literature of tropical biology, and to use that information to recommend an updated syllabus to all institutions that educate tropical biologists.

MATERIALS AND METHODS

We listed all the statistical tests and other quantitative analyses used in two leading tropical journals, the *Revista de Biología Tropical* and *Biotropica*. For the *Revista* we reviewed all of the articles from editions 56-4 (December 2008) to 57-4 (December 2009), not including the supplements. For *Biotropica*, we used only the articles included in the "Papers" section from editions 40-4 (July 2008) to 41-4 (July 2009).

RESULTS

We found that the articles in both journals used a large number of unusual statistics, but that each of these highly specialized procedures and tests was used very few times (Appendix).

The 12 most frequent tests in the articles were: Analysis of Variance (ANOVA), Chi-Square Test, Student's T Test, Linear Regression, Pearson's Correlation Coefficient, Mann-Whitney U Test, Kruskal-Wallis Test, Shannon's Diversity Index, Tukey's Test, Cluster Analysis, Spearman's Rank Correlation Test and Principal Component Analysis (Table 1).

TABLE 1
Statistical tests and other quantitative procedures that were used more than once during the studied year

Revista de Biología Tropical		Biotropica	
Test or procedure	Frequency	Test or procedure	Frequency
ANOVA	50	ANOVA	31
Chi-Square Test	26	Chi-Square Test	26
Student's T Test	23	Student's T Test	25
Linear Regression	18	Linear Regression	17
Pearson's Correlation Coefficient	17	Pearson's Correlation Coefficient	16
Kruskal-Wallis Test	16	Mann-Whitney U Test	14
Shannon's Diversity Index	13	Generalized Linear Models	13
Abundance	11	Spearman's Rank Correlation Test	11
Richness	11	Akaike's Information Criterion	10
Tukey's Test	11	Bray-Curtis Coefficient	10
Cluster Analysis	10	Tukey's Test	9
Spearman's Rank Correlation Test	10	Kruskal-Wallis Test	8
Principal Components Analysis	9	Multiple Regresion Analysis	8
Bray-Curtis Coefficient	9	ANCOVA	7
Canonical Correspondence Analysis	8	Nonparametric Multidimensional Scaling	7
Duncan Test	8	Richness	7
Kolmogorov-Smirnov Test	8	Wilcoxon Test	7
ANCOVA	7	Jackknife Statistics	6
Chao Index	6	Kolmogorov-Smirnov Test	6

TABLE 1 (Continued)
Statistical tests and other quantitative procedures that were used more than once during the studied year

Revista de Biología Tropical		Biotropica	
Test or procedure	Frequency	Test or procedure	Frequency
Mann-Whitney Test	6	Monte Carlo Test	6
UPGMA	6	Abundance	5
Linear Correlation Analysis	5	ANOSIM	5
Jaccard Index	5	Cluster Analysis	5
Morisita Index	5	Least Squares	5
Variation Coefficient	5	MANOVA	5
Euclidean Distance	4	Morisita Index	5
Dunnett's T3 Test	4	Principal Components Analysis	5
Jackknife Statistics	4	Amova	4
Least Significant Difference	4	Bonferroni Correction	4
Margalef Index	4	Fst Statistics	4
SIMPER	4	Kaplan Meier Method	4
Student-Newman-Keuls Test	4	Mantel Test	4
Bonferroni Correction	4	Bootstrap	3
ANOSIM	3	Chao Index	3
Correlacion Matrix	3	Fisher's Exact Test	3
Frecuency	3	G Test	3
Laverne Test	3	Jaccard Index	3
Levene Test	3	Logistic Regresion	3
Modal Progression Analysis	3	Maximum Likihood	3
Simpson Index	3	Poisson Distribution	3
Discriminant Function Analysis	2	Regression Analysis	3
Hardy-Weimberg Equilibrium	2	Relative Growth Rate	3
Evenness Index	2	Shannon Evenness Index	3
Robertson and Hill' F Statistics	2	SIMPER	3
Wright's F Statistics	2	Simpson Index	3
Fisher's Alpha Diversity Index	2	Backward Elimination	2
F Statistic	2	Bayesian Analysis	2
Fulton's Correlation Test	2	Canonical Correspondence Analysis	2
Generalized Linear Models	2	Correlation Matrix	2
Wilks' Lambda	2	Dunn Test	2
Logistic Regression	2	Fixation Index	2
MANOVA	2	Hosmer and Lemeshow's Goodnes of Fit Test	2
Least Squares	2	Kendall Tau Rank Correaltion Test	2
Multiple Dismensional Scaling	3	Linear Correlation	2
Non-Parametric Multidimensional Scaling	2	Log Rank Test	2
Fisher's Exact Test	2	Least Significant Difference	2
Multiple Regression	2	Minimun Convex Polygon Method	2
Shapiro-Wilk Test	2	Monte Carlo Simulation	2
Wilcoxon Test	2	Multidimensional Scaling	2
		Multiple Linear Regresion Model	2
		Multivarate Ordinations	2
		Odds Ratio	2

TABLE 1 (Continued)
Statistical tests and other quantitative procedures that were used more than once during the studied year

Revista de Biología Tropical		Biotropica	
Test or procedure	Frequency	Test or procedure	Frequency
		Pairwise Comparisons Test	2
		Sørensen Similarity Index	2
		Wilks' Lambda	2
		UPGMA	2
		Univariate Linear Regression	2
		Chao-Sorensen Index	2
		Canonical Analysis of Principal Coordinates	2

DISCUSSION

In tropical ecosystems, a few species are very common, while most species are rare; and in our sample, the same applies to statistical procedures. There is no justification to burden students with a knowledge of procedures that they will seldom, if ever, use; if the need arises, such procedures can be applied by professional statisticians. On the other hand, the main mathematical procedures were almost the same in both journals, even though *Biotropica* is more limited to terrestrial ecology, while the *Revista* publishes articles from a variety of fields in marine and terrestrial biology, as well as on the conservation of tropical ecosystems.

This gives us some confidence to recommend that statistical education of tropical biologists be based on learning when to use the dozen tests that they are more likely to need in their professional life (top of Table 1). The mathematical procedures involved in those tests, as well as any mathematical proof to justify those procedures, must be eliminated from the courses and left to professional statisticians. Highly useful tests such as Chi-Squared and the G test are simple, but more complex ones –such as multivariate tests– can also be understood from a practical perspective without need for the subjacent mathematics.

Descriptive statistics is also something everyone needs, but again only to understand what averages, standard deviation and the like mean. Learning the algorithms to calculate them is not necessary because computers already "know" how to do it for us.

A frequent weakness in most articles, not only in tropical biology but in all fields and nations, is the poor selection of graphics that has become omnipresent after the introduction of computer spreadsheets. But this problem is easy to identify (a simple guide to good statistical graphics is here: www.biologiatropical.ucr.ac.cr).

Students should also learn that the statistician must be consulted *before* collecting the data, and not afterwards, when it may be too late. They should not finish their statistical courses hating statistics, as we know is the case in some institutions, but rather feeling confident that they can select the correct procedure and apply it, as also found by previous authors (Garfield and Ben-Zvi 2007, Metz 2008).

Software is another field in which current education clearly fails. Excellent free software is available to everyone in the Internet, yet most universities pay huge yearly amounts in licenses of professional software for students. Very good programs that are fit for tropical biologists are available in http://faculty.vassar.edu/lowry/VassarStats.html and in other addresses (e.g. http://statpages.org/, www.freestatistics.info and www.macstats.org). Professional programs should only be bought for professional statisticians.

Students should also learn that statistics has space for humor, as shown by Gary C. Ramseyer's *First Internet Gallery of Statistics*

Jokes (http://my.ilstu.edu/~gcramsey/Gallery.html), and that the motto of the guild is "Rubbish in, rubbish out". They must always put mathematics at the service of biology and not the opposite; they should understand that only controlled experiments can identify cause and effect relationships; if this requirement is not met, no statistical procedure will (James & McCulloch 1990).

Additionally, it is of the greatest importance that:

1. Statistics be taught within the career's subject courses (for example, as part of biology or medicine courses, instead of being a separate course) or at least, that all examples used in courses be from the career's field (Metz 2008).
2. Related subjects, such as distribution, center and spread of data, be learned as a single concept (Garfield & Ben-Zvi 2007).

Currently, students can pass statistics courses with good marks even when they cannot understand statistics (Garfield & Ben-Zvi 2007, Metz 2008). Furthermore, they cannot apply statistical procedures outside the contexts in which they learned them and cannot understand the meaning of statistical tests and graphics (Garfield & Ben-Zvi 2007).

Every year unnecessary statistics textbooks continue to be written and published everywhere, despite the fact that all the information needed by students is freely available in Internet. Apart from that, almost all statistics textbooks available today are basically the same, a situation that would not be so bad if these they had the information that the students need, but a comparison of contents in books being published now and those from half a century ago will show little difference except for the addition of multivariate tests. Not surprisingly, student learning also is poor practically everywhere (Garfield & Ben-Zvi 2007, Metz 2008).

In conclusion, statistical education for tropical biologists must abandon the old syllabus based on the mathematical side of statistics and concentrate on the correct selection of procedures and tests, on their biological interpretation and on the use of reliable and friendly freeware. In most cases, a well designed one-semester course should be enough for them. It is a matter of common sense: their time will be better spent understanding and protecting tropical ecosystems than trying to learn the mathematical foundations of statistics.

RESUMEN

Los biólogos tropicales estudian la biodiversidad más rica y amenazada del planeta, y en estos tiempos de cambio climático y mega-extinción, la necesidad de investigación de buena calidad es más acuciante que en el pasado. Sin embargo, el componente estadístico en la investigación publicada por los autores tropicales adolece a veces de baja calidad en la toma de datos, mal diseño experimental y una visión anticuada del análisis de datos. Para sugerir mejoras en la enseñanza de la estadística, hicimos una lista de todas las pruebas estadísticas y otros análisis cuantitativos aplicados en dos de las principales revistas tropicales, la Revista de Biología Tropical y Biotropica, durante un año. Las 12 pruebas más frecuentes en los artículos fueron: Análisis de Varianza (ANDEVA), Chi-cuadrado, t de Student, Regresión lineal, Coeficiente de Correlación de Pearson, U de Mann-Whitney, Kruskal-Wallis, Índice de diversidad de Shannon, Prueba de Tukey, Análisis de Conclomerados, Correlación de Spearman y Análisis de Componentes Principales. Concluimos que la enseñanza de la estadística para los biólogos tropicales debe abandonar el viejo plan de estudios basado en el lado matemático de la estadística y concentrarse en (1) la correcta selección de estos y otros procedimientos y pruebas, (2) su interpretación biológica y (3) la utilización de programas de fácil uso. En la mayoría de los casos, un curso bien diseñado de un semestre bastaría para sus necesidades básicas.

Palabras clave: educación de biólogos tropicales, cursos de estadística, *software* libre para estadística, revistas de biología tropical.

REFERENCES

Camacho, J. (ed.). 1997. Biometrical education: problems, experiences and solutions. Universidad Nacional and Centro Agrícola Tropical de Investigación y Enseñanza, Heredia, Costa Rica.

Fielding, W.J. & F.B. Lauckner. 1992. Biometric training in developing countries: Caribbean experiences. Statistician 41: 105-111.

Garfield, J. & D. Ben-Zvi. 2007. How students learn statistics revisited: a current review of research on teaching and learning statistics. Intern. Stat. Rev. 75: 372-396.

James, F.C. & C.E. McCulloch. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? Annu. Rev. Eciol. Syst. 21: 129-166.

Monge-Nájera, J. 2002. How to be a tropical scientist (50th anniversary editorial). Rev. Biol. Trop. 50: 19-23.

Metz, A.M. 2008. Teaching statistics in biology: using inquiry-based learning to strengthen understanding of statistical analysis in biology laboratory courses. CBE—Life Sci. Educ. 7: 317-326.

APPENDIX

Quantitative procedures used only once in the examined volumes

Revista de Biología Tropical	Biotropica
Andrews Graph	Accelerated Bias Corrected Confidence Limits
Autocorrelacion Coefficient	Acumulation Curve Analysis
Autocorrelacion Function	Additive Main Effect and Multiplicative Interactions Model
Binomial Probabilistic Model	Ancillary Variable Correlation
Binomial Proportion	Anderson-Darling Test
Bioplots Graph	ANODEV
Bootstrap	Assignment Test
Box Cox's Better Transformation	Association Matrix of Kulczynsky
Canonical Variate Analysis	Autocorrelation Function
Cochran Test	Average Effect Size
Complete Independence Test	Average Variance Effective Size
Cophenetic Correlation	B-Error Distribution With Log Link Function
Correlation Test	Best Subset Multiple Regression
Partitioning of Chi-Square, Method of	Binomial Error Distribution
Diametric Distribution	Binomial Test
Dice Index of Disimilarity	Chao-Jaccard Similarity Index
Disperssion Index	Chao-Jaccard-Raw- Abundance Based Index
Dunn Test	Circular Statistic
Factorial Correspondence Analysis	Compositional Analysis
Fit of Data For Log-Series, Lognormal, Geometric and Broken Strick Models	Contiguity Index
Form Index SHAPE	Contingency Table Analysis
Friedman Test	Contrast Analysis
Gamma Probabilistic Model With Log Link Funcion	Cormak-Jolly-Seber Models
Gauch and Whittakernon's Linear Regresion Model	Correlated Mating Model
Generalized Discriminant Analysis	Dendograms
Geometric Distribucion Model	Density
K Parameter of Negative Binomial Distribution	Detrended Correspondence Analysis
Kendrall's Tau B Test	Disimilarity Index
Likelihood Ratio Test	Disimilarity Matrix
Logaritmic Distribution Model	Distance Dissimilarity Coefficient
Log-Linear Analysis	Disturbance Index
Log-Normal Distribution Model	Dunnet T3 Test
Macarthur and Levene Index Modified by Pianka	Dunn's Test
Manhattan Distance Matrix	Edge Contrast
Mantel Test	Elasticity Analysis
Meyer I Distribution	Euclidean Distances
Meyer II Distribution	Expectation Maximization Numerical Method
Monte Carlo Test	Exponential Model
Morisita-Horn Index	F Statistic
Multiple Discriminant Analysis	F Watson-Williams Test
Multiple Regression Analysis	Fisher's F Statistic
Multivariate Analysis	Fisher's Least Singnificant Diference Test
Multivariate Analysis of Covariance	Fisher's Omnibus Test

APPENDIX (Continued)
Quantitative procedures used only once in the examined volumes

Revista de Biología Tropical	Biotropica
Negative Binomial Distribution	Fixed Kernel Home Rage
Newman-Keuls Test	Flexible Beta Linkage Method
Normal Probabilistic Model With Indentity Link Function	Foundress Number
Normalized Difference Vegetation Index	Friedman Test
Probability Transitions Matrix	Gamma Error Distribution With Log Link Funcion
Randomization Analysis	Gehan's Generalized Wilcoxon Survival Analysis
Randomized Block Analysis	Geometric Mean Weigth of Prey
Rank-Abundance Plot	Greenhouse Geisser's E
Redundance Analysis	Hardy-Weimberg Equilibrium
Residual Analysis	HOF Model
Ripley's K Test	Hybrid Multidimensional Scaling
Sba K -Weiner Index	Individual Based Rarefaction Curves
Scheffe Test	Interclass Correlation Coefficient
Schoener Overlar Index	K Hat Estimate of Conditional Kapa Coefficient
Sign Test	Kernel Density
Similarity Matrix	Kishino-Hasegawa Test
Single-Factor Anlaysis of Variance	Kruskal's Measures of Stress
Square Cannonical Correlation	Kullback-Leibler Information
Trend Analysis	Levine's Index
Variance Test	Levine's Niche-Breadth Index
Wald Test	Linear Autocorrelation Coefficient
Warner Dendograma	Log Likelihood Test
Weibull Distribution	Logistic-Link Error Distribution
Z Test	Log-Likelihood Ratio
	Lowess Regression
	Manly's Selection Index
	Mantel Correlation Coefficients
	Mantel Correlation Comparisons Mantel's R
	Mantel Correlograms
	Mao Tau Estimator
	Mao Tau Function
	Mao-Tau Simple Based Rarefaction Curves
	Mardia-Watson-Wheeler Test
	Markov Chain Monte Carlo MCMC
	Matrix Randomizations
	Maximun Parsimony
	Mayfield Method
	Mean Czechanowski Index
	Mean Pianka Index
	Mean Square Error
	Minimazing Yield Parameter Estimates
	Minimum Model
	Minimun Area Probabilities MAP
	Mlmm PQL Function

APPENDIX (Continued)
Quantitative procedures used only once in the examined volumes

Revista de Biología Tropical	Biotropica
	Mode Clustering Index
	Moment Core Analysis
	Monte Carlo Based Kruskal Wallis H Test
	Monte Carlo Permutations
	Monte Carlo Randomization
	Multiple Post Hoc Comparison
	Multivariate Mantel Correlograms
	Nei's Diversity Index
	Newton-Rapson Numerical Method
	Non Linear Regresion Models
	Overall Population Structure Analysis
	Overdispersion Parameter
	Paetkau Method
	Pairwise Correlations
	Pairwise Population Differentiation Test
	Path Coefficient
	Pearson Partial Correlation
	PERMANOVA
	Permutation Method
	Quasi-Binomial Distribution
	Quasilikelihood Akaike's Information Criterion
	Randomization Test
	Rarefaction Analysis
	Rayleigh Test
	Regression Coefficient
	Relatedness Coefficient
	Renkonnen Index
	Ricker Function
	Rosario Randomization Algorithm
	Runs Test
	Shapiro Wilk Test
	Shared Species Analysis
	Similarity Matrix
	Simple/ Partial Mantel Test
	Sorensen Index
	Standardized Ripley's K Function
	Steinhays Similarity Coefficient
	Stepwise Regression Model
	Structural Equation Modeling
	Sum Test
	SUNIM Index
	Suvival Function
	T2 Hotelling Test
	Templeton Test
	Time Since Making Model

APPENDIX (Continued)
Quantitative procedures used only once in the examined volumes

Revista de Biología Tropical	Biotropica
	Triangular Similarity Matrix
	Two Generator Model
	Type III Analysis
	Variance Ratio Test
	Variance-Covariance Matrix
	Variation Coefficient
	Vulnerability Index
	Wald Statistic
	Wald Statistic Log-Rank Test
	Wilson-Shimida Index
	Wishart's Objective Function
	Wright's Fixation Index