BIOL 5500: Research Design & Data Analysis

I. General Information

Professors: Drs. Kenneth Schoenly, Michael Fleming
Office: N271, N269
Phone: 667-3949, 664-6923
Office Hrs: MW 9:00-10:30 (KS), W 2:30-3:30, Th 9:30-10:30 (MF)
Email: kschoenly@csustan.edu, mfleming1@csustan.edu

Semester: Spring 2020
Credits: Lec/Activity 4
Class Times: Tu/Th 2:00-4:50
Room: N334 (Tu), ____ (Th)

II. Course Description

(Catalog Description): Survey of experimental designs and investigative methods for graduate level research. Content will be tailored to research needs of enrolled students. Includes sampling schemes, modeling, ANOVA and multivariate analyses. Prerequisite: Graduate standing in M.S. program (Lecture, 3 hours; Activity 3 hours) (Spring).

III. Student Learning Outcomes:

1. The student will use established methods and models to address biological problems and answer questions related to study design, sampling, and investigative approaches.

2. The student will evaluate peer-reviewed literature in science, its cumulative, self-correcting, and hypothesis-testing features, and be able to critique the investigative methodology used, thereby refining his/her communication, quantitative, and critical thinking skills.

> ***"… if we want to test hypotheses, model data, and forecast future environmental conditions in the 21st century, we must move beyond basic statistical literacy and attain statistical fluency."***
> - Ellison & Dennis 2010, p. 362

3. The student will apply sampling schemes and study designs learned in the class to plan thesis experiments and analyze results using appropriate investigative tools and models.

4. This course addresses the University's graduate student learning goals numbered 1, 2, 5 and 7.

IV. Required Textbook/Readings:

1. Textbooks (2):*Havel, Hampton & Meiners. 2019. Introductory Biological Statistics, 4th edition. Mlodinow. 2008. The Drunkard's Walk: How Randomness Rules Our Lives*. Bring to every class period. Used copies available through on-line web sites.
2. USB of papers from the biological literature (supplied); web sites that offer free software (see "to do list" and below).
3. Handouts of spreadsheet exercises and accompanying questions to solve as in-class activities or take-home assignments. These will be given out in class or sent as attachments in future emails.

V. Pre-Requisites:

Two undergraduate math courses (statistics and calculus) are necessary (and enforced) pre-requisites for this course. Additional mathematical courses that ecologists take to bolster their quantitative understanding include probability, multilinear regression, matrix algebra, multivariate statistics, and graph theory, among others. With each new math course taken, your "comfort zone" with quantitative concepts increases. As a graduate student, we expect that you will perform at a higher level, take the initiative on class assignments, have excellent problem-solving skills, become a student member of one or more professional societies, and use quantitative methods in your thesis research (just to mention a few). It also affords you the opportunity to take an active role in shaping your future professional growth and for making contributions in your chosen field of study.

To maximize these benefits, both you and the university must make commitments. The university must make available the resources, both intellectual and physical, that are necessary for you to complete coursework, pursue your program of studies, and realize your potential. For your part, you must take advantage of the opportunities the program has to offer by being highly motivated and focused in the first place, and by directing that motivation in a productive manner. If you proceed appropriately, the lectures, homework, exams, presentations, and other course assignments will become exciting challenges and guidelines rather than hurdles and hindrances.

VI. Course Requirements:

This course is a required research design and analysis class for the M.S. degree in Biological Sciences. **Biology is an experimental, observational, mathematical and correlative science.** This course will expose you to many quantitative methods and experimental designs that are used by biologists to answer diverse questions. Because one semester is not enough time to cover all research methods and designs used by biologists, **we will focus on only some of the more common tools and case studies.** Mastering these tools is critical to successful completion of your thesis and to future employment and career building as a professional biologist.

The course's subject matter will be introduced through formal lectures, assigned journal readings, student presentations, and self-paced lab and field activities that use R programs, ecostatistical software and EXCEL spreadsheets. Written exams will require you to know the proper uses, assumptions, limitations, interpretations (but not intimate mathematical steps) of biostatistical methods & models used. Successful completion of EXCEL activities will demonstrate literacy of the individual calculation steps leading to the correct solution and interpretation.

Presentations: You will be required to deliver **two, 25-minute critiques of journal papers** (listed below) and **one 25-minute spreadsheet/software exercise/demonstration** featured in the topics list below. I will provided you all the papers (with few exceptions) as PDF's on a thumb drive on the first class day. Each **presentation grade** will be based on your **verbal performance** and a **1-page synthesis** (i.e., study guide for your classmates) that summarizes the paper's questions/hypotheses, methodology, results, and conclusions. Paper assignments and software demonstrations will be drawn at random (i.e., sampling without replacement!) and have a deadline a few weeks before the 1st presentations. It will then be your job to 'marry yourself' to each paper or exercise and hand out your summary page and spreadsheet/software output when your name is called, effectively 'teaching' the topic to your classmates and instructor.

Assigned homework will be due at the next class meeting (no exceptions) and handed in-person (as hard copies); no homework will be accepted as email attachments. **Persons may be called upon at random to present homework and answer questions**. Presentations need to be practiced (several times, if necessary) so they finish under/on time (**no longer than 20/5 min for talk/questions**). Your instructor will evaluate your verbal performance and 1-page syntheses of your papers plus your spreadsheets/other outputs using a standardized grading rubric. **In the event you miss (or**

are unready to present) your scheduled talk, you will receive an automatic '0' for that assignment. HAVE YOUR PRESENTATIONS/SPREADSHEETS/FREEWARE WITH YOU AT THE START OF CLASS (ON COMPUTER OR THUMB DRIVE) AND BE READY TO TEACH THE CLASS WHEN YOUR TIME COMES!!

VII. Personal Responsibility:

Behavior that interferes with the instructor's ability to teach or the ability of students to benefit from instruction will not be tolerated. Examples include: audible ring tones, repeated late arrivals or early departures, irrelevant conversation, and inappropriate use of phones or computers. At the start of class, turn off cell phones. Check your university email daily for updates of information items.

Cheating in any form is inappropriate conduct and will be dealt with swiftly and severely according to Sections 41301 through 41304 of Title 5 of the *California Code of Regulations*" which includes expulsion, suspension or probation.

VIII. Biological Software:

We will be using software from commercial and "freeware" sources, such as R and PAST. These programs run on your laptop (PC, Mac, or emulated versions). Commercial software is copyright protected and requires site licenses for individual student use. Unauthorized copying of commercial software is a serious crime, constituting copyright violation. See below for some examples:

Other Helpful Websites for EXCEL Users (tools and YouTube videos):

PopTools (add-in): http://sunsite.univie.ac.at/Spreadsite/poptools/index.htm.
Mr. EXCEL: http://www.mrexcel.com/.
EXCEL Tutorial- Advanced Functions: https://www.youtube.com/watch?v=bpQ3KCVHLc4

IX. Grading Procedure

As per university regulations, students with excessive absences will be dropped from the class. Two midterm exams will be 150 point mixed format. Exams will consist of long and short answer essays, graph/output interpretation, calculations, and questions that will come from the lecture notes, textbook, presentations, software, journal articles, and activities. Allow at least two weeks for exams and other assignments to be graded and returned.

It is your responsibility to contact us in the event you miss an assignment (within 24 hrs) documenting your absence (e.g., doctor's note, jury summons, funeral notice); otherwise, an unexcused absence for a gradable event will result in no score. All requests to take exams at other than scheduled times must be in writing to the instructor **at least 3 lecture days** prior to the scheduled exam date. Makeup exams, if approved by us, must be made up before the test day or within 2 days of the exam. The final decision to offer a makeup exam rests with us. Assigned homework or presentations will have a no-exceptions due date and time

In any professional endeavor, correct spelling and grammar, and good penmanship are necessary for effective communication. In this class, initiative, resourcefulness, or originality on homework and presentations will be rewarded with extra points! On the other hand, poor spelling, grammar, and penmanship will result in lost points and illegible answers on exams and homework will receive no credit (i.e., If I can't read it, it's wrong).

Grades and Weighting

Lecture Exams (2 @ 150 points each)	300 (30%)
Journal Presentations (2 @ 150 points each)	300 (30%)
Spreadsheet/Software Exercise	150 (15%)
Homework	200 (20%)
Miscellaneous (punctuality, attendance, participation, conduct)	50 (5%)
Total	1000 points

A = 900-1000, B = 800-899, C-F = 700 or fewer. No +/- grading will be applied to your final grade.

X. Topics*

TOPIC	BOOK READINGS**	ARTICLE READINGS**
Introduction: ❑ spreadsheet hints/tips ❑ math functions/macros/add-ins pack ❑ freeware programs (R, PAST, etc.) ❑ preparing data files ❑ paper discussions	M1 HHM:Appendix C & p. 68 M1	DW2 Curran-Everett 2008 Platt 1964, Gleick 1990
Properties/Challenges of Biological Data: ❑ visualizing biological information ❑ scales of measurement ❑ metadata, data archiving & sharing ❑ positional accuracy ❑ significant figures ❑ measurement repeatability ❑ handling & visitation effects	HHM3 HHM2, M7 M10	O'Donoghue 2010; Metcalf et al. 2016 Cook et al. 2001; Whitlock 2011 Dodd 2011 Wolak et al. 2012; Hayes et al. 1998 Ramsay et al. 2009; Klaring 1999
Statistical Foundations: ❑ use of (pseudo)random numbers ❑ standards for statistical inference ❑ statistical significance and power ❑ parametric vs. non-parametric methods ❑ central limit theorem, normal distribution ❑ accuracy, precision, bias ❑ data transformation ❑ one-tailed vs. two-tailed tests ❑ estimating uncertainty (confidence intervals) ❑ descriptive statistics, exploratory data analysis ❑ statistical modeling (bootstrap, jackknife)	M7 M5 HHM:1,2,4 DW5 M7-8, DW3-4 HHM9	Michaud et al. 2012 Halsey et al. 2015; Thomas & Juanes 1996 Wang & Riffel 2011 O'Hara & Kotze 2010 Ruxton & Neuhauser 2010 Curran-Everett 2009 Meyer et al. 1986
Study Designs: ❑ definitions, types (pilot studies, observational, experimental, quasi-	HHM:15,16	

experimental) ❑ true replication, pseudoreplication ❑ laboratory & field designs for ANOVA -- Completely Randomized Design -- Randomized Complete Block Design -- Repeated Measures Design -- Latin Square Design -- Split-Plot Design ❑ studying environmental impacts (BACI) ❑ critiquing study designs	HHM:11,12	Ramirez et al. 2001 Gill & Haps 1971 Steinbeck et al 2005
Proportions & Frequencies: ❑ binomial distribution ❑ Poisson distribution ❑ Chi-square goodness of fit test ❑ Contingency analysis (odds ratio, relative risk) ❑ Chi-square contingency test ❑ Fisher's exact test	HHM:5,7,8 DW3 M3-4	Bove et al. 2014 Shan & Gursteberger 2017
Minimum Sample Size Estimation ❑ single & two sample tests, statistical power ❑ optimal quadrat shape & size ❑ quadrats & sequential sampling ❑ collectors' curves, asymptotic sampling ❑ optimized sampling & stopping rules	HHM:6,9,10 Cobra	Marvier 2002, Button et al. 2013 Pringle 1984 Colwell & Coddington 1994 Cardoso 2009
Correlation & Regression: ❑ Correlation coefficients (parametric, non-parametric) ❑ Linear regression, confidence in predictions ❑ Non-linear regression ❑ Logistic regression ❑ Model fitting	HHM:13,14	Kaplan & Gurcan 2018
Meta-Analysis: ❑ theory & milestones ❑ methodology & critique ❑ applications	HHM:p.206	Gurevitch et al. 2018 Kotaiho & Tomkins 2002 Antman et al. 1992; Letourneau et al. 2011
Analysis of Multivariate Data: ❑ approaching multivariate data ❑ binary vs. quant. similarity measures ❑ multivariate distance ❑ ordination ❑ classification ❑ canonical correspondence analysis	HHM:p.202-203	Huhta 1979; Wolda 1981 Townsend et al. 1983 Stancampiano & Schnell 2004 Wiens et al. 2001
Review of Student Thesis Projects: presentation on research design & proposed data analysis		

* Topic content may be changed due to extenuating circumstances.
**HHM = Havel et al. 2019; M = Mlodinow; DW = Donovan & Welden 2002 chapters (EXCEL exercises)

Readings

Curran-Everett, D. 2008. Explorations in statistics: standard deviations and standard errors. Advances in Physiol. Educ. 32: 203-208.

Platt, J.R. 1964. Strong inference. Science 146: 347-353.

Gleick, J. 1990. The census: Why we can't count. NY Times Magazine (July 13): 22-26.

O'Donoghue, S. et al. 2010. Visualizing biological data—now and in the future. Nature Methods Supplement 5: S2-S4.

Metcalf, J.L. et al. 2016. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science 351: 158-162.

Cook, R.B. et al. 2001. Best practices for preparing ecological data sets to share and archive. Bulletin of the Ecological Society of America. 82: 138-141.

Whitlock, M.C. 2011. Data archiving in ecology and evolution: best practices. Trends in Ecology and Evolution 26: 61-65.

Dodd, M. 2011. Where are my quadrats? Positional accuracy in fieldwork. Methods in Ecology and Evolution 2: 576-784.

Wolak, M.E. et al. 2012. Guidelines for estimating repeatability. Methods in Ecology & Evolution 3: 129-137.

*Hayes, J.P. et al. 1998. Repeatability of mammalian physiology: evaporative water loss and oxygen consumption of *Dipodomys merriami*. Journal of Mammalogy 79: 475-485.

*Ramsay, J.M. 2009. Whole-body cortisol response of zebrafish to acute net handling stress. Aquaculture 297: 157–162

Klaring, H-P. 1999. Effects of non-destructive mechanical measurements on plant growth: a study with sweet pepper (*Capsicum annuum* L.). Scientia Horticulturae 81: 369-375.

Michaud, J-P, K. Schoenly, and G. Moreau. 2012. Sampling flies or sampling flaws? Experimental design and inference in forensic entomology. Journal of Medical Entomology 49: 1-10.

Halsey, L.G. 2015. The fickle P value generates irreproducible results. Nature Methods 12: 179-185.

Thomas, L. and F. Juanes. 1996. The importance of statistical power analysis: an example from animal behavior. Animal Behavior 52: 856-859.

Wang, M. & M. Riffle. 2011. Making the right conclusions based on wrong results and small sample sizes: interpretation of statistical tests in ecotoxicology. Ecotoxicology and Environmental Safety 74: 684-692.

O'Hara, R.B. and J. Kotze. 2010. Do not log-transform count data. Methods in Ecology and Evolution 1: 118-122.

Ruxton, G.D. & M. Neuhauser. 2010. When should we use one-tailed hypothesis testing? Methods in Ecology and Evolution. 1: 114-117.

Curran-Everett, D. 2009b. Explorations in statistics: confidence intervals. Advances in Physiol. Educ. 33: 87-90.

*Meyer, J.S. et al. 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. Ecology 67: 1156-1166.

*Ramirez, C.C. et al. 2000. Pseudoreplication and its frequency in olfactometric laboratory studies. Journal of Chemical Ecology 26: 1423-1431.

Gill, J.L. and H.D. Haps. 1971. Analysis of repeated measurements of animals. Journal of Animal Science 33: 331-336.

*Steinbeck, J.R. et al. 2005. Detecting long-term change in complex communities: a case study from the rocky intertidal zone. Ecological Applications 15: 1813-1832.

Bove, M.C. et al. 2014. Effect of El Niño on U.S. landfalling hurricanes, revisited. Bull. Am. Meteorol. Soc. 79(11): 2477-2482.

Shan, G. and S. Gerstenberger. 2017. Fisher's exact approach for post hoc analysis of a chi-squared test. PLOS One 12(12): e0188709.

*Marvier, M. 2002. Improving risk assessment for nontarget safety of transgenic crops. Ecological Applications 12: 1119-1124.

Button, K.S. et al. 2013. Power failure: why sample size undermines the reliability of neuroscience. Nature Reviews/Neuroscience 14: 1-13.

Pringle, J.D. 1984. Efficiency estimates for various quadrat sizes used in benthic sampling. Canadian Journal of Fisheries and Aquatic Science 41: 1485-1489.

Colwell, R. and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B 345:101-118.

*Cardoso, P. 2009. Standardization and optimization of arthropod inventories – the case of Iberian spiders. Biodiversity and Conservation 18: 3949-3962 (add software demonstration).

*Kaplan, S and Gürcan. 2018. Comparison of growth curves using non-linear regression function in Japanese quail. J. App. Anim. Res. 46(1) 112-117.

Gurevitch, J. et al. 2018. Meta-analysis and the science of research synthesis. Nature 555: 175-182.

Kotiaho, J.S. and J.L. Tomkins. 2002. Meta-analysis, can it ever fail? Oikos 96: 551-553.

*Antman, E.M. et al. 1992. A comparison of results of meta-analysis of randomized control trials and recommendations of clinical experts. Journal of the American Medical Association. 268: 240-248.

*Letourneau, D.K. et al. 2011. Does plant diversity benefit agroecosystems? A synthetic review. Ecological Applications 21: 9-21.

*Huhta, V. 1979. Evaluation of different similarity indices as measures of succession in arthropod communities of the forest floor after clear-cutting. Oecologia 41: 11-23.

Wolda, H. 1981. Similarity indices, sample size and diversity. Oecologia 50: 296-302.

*Townsend, C.R. et al. 1983. Community structure in some southern English streams: the influence of physicochemical factors. Freshwater Biology 13: 521-544.

*Stancampiano, A.J. and G.D. Schnell. 2004. Microhabitat affinities of small mammals in southwestern Oklahoma. Journal of Mammalogy 85: 948-958.

*Wiens, J.A. et al. 2001. A canonical correspondence analysis of the effects of the Exxon Valdez oil spill on marine birds. Ecological Applications 11: 828-839.

*student presentations