BIOL 54. Advanced methods in ecological research
Students will work in small groups to conduct original research projects addressing ecological questions that are developed by the students with inspiration from natural history observations, published research, and discourse within student working groups. Students will develop skills in exploring natural history, formulating interesting answerable research questions, deriving hypotheses from theory, developing research design, acquiring and analyzing data, making statistical and logical inferences, writing scientific papers, and presenting seminars.
Instructor: Ayres
Prerequisite: BIOL 16 and one additional course from among BIOL 21 – BIOL 32
Dist: SLA
Next offered: Spring 2021

BIOL 54: Advanced methods in ecological research

Course Objectives

Students will develop skills in exploring natural history, formulating interesting answerable research questions, deriving hypotheses from theory, developing research design, acquiring and analyzing data, making statistical and logical inferences, writing scientific papers, and presenting seminars.

Course structure

The curriculum is structured around three units of three weeks each. The core of each unit will be original research projects (ORPs) that are conceived and conducted by self-formed teams of 2-4 students. Research topics are not pre-determined and will be inspired by the student interests and the ecology of ecosystems that we are studying. Data will be drawn from online resources (see Textbook, references, and data sources), collected by student working groups, or a combination. At the end of each set of ORPs, students will present their studies in a research symposium. During the next week, while the next set of ORPs are being conceived and conducted, students will be writing their last ORP in the style and format of a peer-reviewed journal in ecology and evolutionary biology (see Instructions for Authors). The manuscripts will be reviewed, revised, and revised before being formatted and published on the Dartmouth server (similar to Dartmouth Studies in Tropical Ecology 2020).

In addition to working in small groups on ORPS, the curriculum includes lectures, labs, and structured modules on a variety of topics (details below). There are also regular student presentations of high quality papers from the primary literature that have been chosen for them by the staff and adapted to their research projects (see Student Presentations of Published Research).

Staff and office hours

Professor:	Matt Ayres (LSC 125); Office hours xxxx, xxxx, & by appt.
Graduate Assistant:	xxxx; Office hours xxxx

Crosscutting themes of the curriculum

Ecology and natural history of selected ecosystems and their biota including Hubbard Brook Experimental Forest, National Parks of eastern U.S., and the Arctic.

All aspects of conceiving and conducting original ecological research including question development, theoretical context, hypothesis testing, experimental design, analyses, writing, and presentations.

Teamwork: including on projects, in discussions, and in general.

Data management and statistics (see Data analysis reference guide)

The practice of science: including application of theory, strong inference, critical thinking, application of statistics, seeing parallels across systems, etc.

Classic and contemporary literature in ecology and evolutionary biology: diversity and coexistence, evolutionary ecology, animal behavior, conservation biology, ecosystem science.

Textbook, references, and data sources

Holmes, R. T., and G. E. Likens. 2016. Hubbard Brook: The Story of a Forest Ecosystem. Yale University Press, New Haven, CT. ISBN: 9780300203646.
Hubbard Brook data catalog. https://hubbardbrook.org/d/hubbard-brook-data-catalog
National Park Service Forest Vegetation Visualizer. https://npsforveg.shinyapps.io/netn/
National Park Service Land Bird Visualizer. https://netnaaronweed.shinyapps.io/BirdViz/
Ecological Data Wiki. https://ecologicaldata.org/find-data
USA National Phenology Network. https://www.usanpn.org/usa-national-phenology-network
Ecological Forecasting Initiative. https://ecoforecast.org/efi-rcn-forecast-challenges/
Ameriflux network of ecosystem studies measuring CO_2, water, and energy fluxes https://ameriflux.lbl.gov/
NOAA's National Centers for Environmental Information. https://www.ncdc.noaa.gov/
eBird, Cornell Lab of Ornithology. https://ebird.org/home
BirdNET, Cornell Lab of Ornithology. https://birdnet.cornell.edu/
Circumpolar Biodiversity Monitoring Program. https://www.caff.is/about-the-cbmp

Grading

Grades will be based upon the following criteria:
1. Quality of research projects: including design, analyses, writing, and presentations
2. Acquisition of knowledge regarding ecology and natural history of selected ecosystems
3. Teamwork skills: including on projects, in discussions, and in general
4. Paper presentations and associated discussions
5. The practice of science: including application of theory, strong inference, critical thinking, application of statistics, seeing parallels across systems, etc.

Element	Points	Percent
Original Research Project 1	40	14
Original Research Project 2	55	19
Original Research Project 3	70	25
Development and application of data management skills	20	7
Development and application of analytical skills	20	7
Paper presentations and associated discussions	20	7

Element	Points	Percent
Teamwork	20	7
Lightning presentations on study systems	15	5
Lightning presentations on data resources	15	5
Contributions to project development	10	4
Total	285	100

Special needs

We encourage students with disabilities, including invisible disabilities like chronic diseases, learning disabilities, and psychiatric disabilities to discuss appropriate accommodations with the professor after class or during office hours. You may also wish to talk with Craig Layne and your teaching assistant if laboratory accommodations would be appropriate.

Religious holidays

If you have a religious observance that conflicts with the course schedule, please speak with the professor and we will be happy to develop appropriate accommodations.

Health and well-being

If you or someone close to you becomes ill, please speak with the professor and we will develop appropriate accommodations. The academic environment at Dartmouth is challenging, our terms are intensive, and classes are not the only demanding part of your life. Dartmouth offers resources to support your wellness, including:

Your undergraduate dean (http://www.dartmouth.edu/~upperde/);
Counseling and Human Development (http://www.dartmouth.edu/~chd/); and
The Student Wellness Center (http://www.dartmouth.edu/~healthed/).

We encourage you to use these resources and to speak with the professor to take care of yourself throughout the term.

Academic Honesty

From the Dartmouth College Student Handbook: "Students who submit work that is not their own or who commit other acts of academic dishonesty forfeit the opportunity to continue at Dartmouth." The complete text of the Academic Honor Principle is available at: https://students.dartmouth.edu/community-standards/policy/academic-honor-principle.
Please read it carefully; you are responsible for abiding by the Dartmouth Honor Principles. Any violations of the Honor Principles within the context of Biology 54 will be referred to the Community Standards and Accountability Office and can result in a hearing before the Committee on Standards. Students found responsible for violating the honor principle can be suspended for multiple terms or, in the most extreme cases, separated from the College.
The activities of Bio 54 include team projects as well as individual work. Just ask if you are ever uncertain about applications of honor principles to work in Bio 54.

Consent to recording of course and group office hours

- I affirm my understanding that this course and any associated group meetings involving students and the instructor, including but not limited to scheduled and ad hoc office hours and other consultations, may be recorded within any digital platform used to offer remote instruction for this course;
- I further affirm that the instructor owns the copyright to their instructional materials, of which these recordings constitute a part, and distribution of any of these recordings in whole or in part without prior written consent of the instructor may be subject to discipline by Dartmouth up to and including expulsion;
- I authorize Dartmouth and anyone acting on behalf of Dartmouth to record my participation and appearance in any medium, and to use my name, likeness, and voice in connection with such recording;
- I authorize Dartmouth and anyone acting on behalf of Dartmouth to use, reproduce, or distribute such recording without restrictions or limitation for any educational purpose deemed appropriate by Dartmouth and anyone acting on behalf of Dartmouth.

Requirement of consent to one-on-one recordings

By enrolling in this course, I hereby affirm that I will not under any circumstance make a recording in any medium of any one-on-one meeting with the instructor without obtaining the prior written consent of all those participating, and I understand that if I violate this prohibition, I will be subject to discipline by Dartmouth up to and including expulsion, as well as any other civil or criminal penalties under applicable law.

Schedule

Biol 52. Schedule of activities. ORP = Original research project

Week	Lecture	Lecture topic	Lab activity
1	1	Avian ecology: natural history and prominent research questions	Avian ecology: development of group project comparing selected features of avifauna across student locations
	2	Avian ecology: data resources	
	3	Lightning presentations: posssible research questions. Research tactics: good questions	
2	1	National Parks of the northeastern U.S.: ecology and natural history	ORP 1: Conception and design of original research project.
	2	National Parks of the northeastern U.S.: data resources	
	3	Statistics I. Data frames, data management, and metadata	
3	1	Estimating ecosystem flux of CO2 and H20 with eddy-covariance systems	ORP 1: Research by working groups
	2	Lightning presentations on study systems, data resources, and potential research questions	
	3	Statistics II: Hypothesis testing	
4	1	The Hubbard Brook forest: ecology and natural history	ORP 1: Data analysis and interpretation. Writing Lab 1.
	2	The Hubbard Brook forest: data resources	
	3	Research symposium. Presentations of ORP 1	
5	1	Lightning presentations on study systems, data resources, and potential research questions	ORP 1: manuscript due. ORP 2: Conception and design of original research project.
	2	The Arctic: ecology and natural history	
	3	The Arctic: data resources	
6	1	Research tactics: connecting theories to hypotheses; deriving predictions from theory; strong inference.	ORP 2: Research by working groups. Writing lab 2. ORP 1 manuscript revisions.
	2	Lightning presentations on study systems, data resources, and potential research questions	
	3	Statistics III: special topics in ecological research*	
7	1	Research tactics: analysis and interpretation,	ORP 2: Data analysis and interpretation. ORP-1 revisions due.
	2	Student presentations of high impact research papers	
	3	Research symposium. Presentations of ORP 2	
8	1	Lightning presentations: posssible research questions	ORP 2: manuscript due. ORP 3: Conception and design of original research project.
	2	Student presentations of high impact research papers	
	3	Special topics in ecology and evolution*	
9	1	Student presentations of high impact research papers	ORP 3: Research by working groups. ORP 2 manuscript revisions.
	2	Special topics in ecology and evolution*	
	3	Presentations of avian ecology comparative project	
10	1	Research symposium. Presentations of ORP 3	No lab. ORP-2 revisions due. ORP 3 manuscript due.

* Selected to complement ongoing research projects

STUDENT PRESENTATIONS OF PUBLISHED RESEARCH

What are the properties of a high-impact research paper in ecology and evolutionary biology? How can one distill the most interesting and important parts of an intellectually rich research paper so that colleagues can appreciate the value with only a few minutes of explanation? As part of our activities each of you will give two presentations on a high quality research contribution from the primary literature that has been chosen for you. On each occasion, you will have about 20 minutes of undivided attention from your colleagues to rock their world with a scintillating synthesis from you, customized for them.

Plan on about 12 minutes for your presentation, leaving about 8 minutes for the rich and sophisticated discussion that you will inspire. You may use one visualization *of your creation* to help with the presentation. Your visualization should fit on one powerpoint slide or one panel of a whiteboard. We suggest that it be hand-drawn and acquired as a photo of the sketch you have created on a notepad. You may not use any visualizations reproduced from the paper.

We suggest that you take notes as you read your paper and bring questions ahead of time to the cognizant staff person regarding logic, technical approaches, vocabulary, relevant natural history, implicit theoretical models, etc. Also, talk with the staff person about what will be in the lecture that precedes your presentation so that you know what background your audience will and will not have previously received.

Take full advantage of knowing your audience. Help them connect your paper (beyond what the author could have possibly done) to your shared experiences in the course. Think broadly about how to start and end your presentation to grab the attention of your audience and later to leave them with new and lasting awareness. If you start and end your presentation as the authors have done with their paper then you have missed the opportunity to connect with your audience. We will shame you if you start your presentation with "The title of my paper was …"

We suggest the following structure for your presentation.

1. Provide the context that motivates the overarching research question. Frame the background in terms of general theory and big puzzles in ecology and evolutionary biology. Relate to the scholarship and experiences shared by you and your audience.
2. What is the research question? Seek to summarize the research objectives as a single thought-provoking sentence ending in a question mark. It is unlikely that you will find the perfect sentence for your purposes within the paper. Remember that all interesting questions have more than one plausible answer. Try to provide the context and identify the question without making reference to the particular study system in your paper.
3. Identify at least two possible answers to the research question that relate to general theory and big puzzles. Try to employ a simple graphic. Your goal is to make two possible answers sound so plausible and yet so incompatible that your audience can't wait to learn the outcome. This is not unlike how a good piece of writing in almost any genre captures the reader with suspense, intrigue, and interesting characters.
4. Describe the study system and technical approaches. Relate to things you and your colleagues have seen or will see on Bio FSP. Your aim is not to replicate the the paper by laboring through the introduction, methods, results, etc. Don't bore your audience. Filter to the information that is necessary for the story you are telling.
5. Present the results, including those that were not obtained but could have been. We suggest that you employ a very simple data figure, or schematic of the underlying theoretical model, and explain it. With a good visual aid, you can easily explain the different possible results, and having done that, finally tell them how it really came out. To continue the analogy with good writing, this is the climax. Make it meaningful and memorable. Note that the best visual aids for this purpose can rarely be taken directly from the paper (where figures and tables are designed to present detail and to be studied at length). You will need to develop a simplification from the package of figures, tables, and text in the results of the paper. Use your creativity, intuition, and intellect to imagine different possible visual aids and pick the ones that will work best for your audience.
6. Summarize interesting conclusions with respect to the study system and general theory. Remember to tell us what interesting possibilities we can now say are not true.
7. Identify important limitations of the study. Where are the edges of what we can and cannot say as a result of the study?
8. End broadly. Consider finishing with one or two thought provoking questions that might stimulate continuing thought and discussion. When you succeed well, there will be good questions at the end of your presentation, and you will hear people talking about your research paper days later.

Instructions for authors: Bio 54 (2021)

Conducting original ecological research is the focus of Bio 52. Scientific papers are the commodity of research. It is axiomatic that the science is not completed until the paper is published. Great scientists are those who write great papers. Writing is described as a rate-limiting step by almost every scientist you will ever meet. For all these reasons, writing the papers that describe your research is a crucial part of Bio 52. Alumni consistently note the improvement in their writing skills as among the most valuable lasting benefits of the course. Make the most of this opportunity to cultivate your growth as a writer. Read everything we give you about writing. Study the feedback from colleagues, TAs, and faculty. Watch how others do it. Different tactics work for different writers and advice will vary accordingly. Find what works.

Readings about writing.

Papers as pdfs within this electronic package

Mack, R. N. 1986. Writing with precision, clarity, and economy. *Bulletin of the Ecological Society of Americ*a 67:31-35.

Knight, J. 2003. Clear as mud. Nature 422: 376-378.

Goldwasser, L. 1998. A collection of grammatical points. *Bulletin of the Ecological Society of America* 79: 48-50.

Hammond, G. Thoughtful writing.

Books in our physical library

Karban, R., and M. Huntizinger. 2006. How to do ecology. Princeton University Press. 145 pp. [See Chapter 6 on communication].

McMillan, V.C. 2012. Writing papers in the biological sciences. 5th Edition. Bedford/St. Martin's, Boston. 241 pp.

Strunk, J.R., and E.B. White. 2000. The elements of style. 4th Edition. Longman, New York.

Some general suggestions for Bio 52 authors.

Make the team work. Research on Bio 52, like contemporary ecological research in general, is almost all conducted by work groups. Great research is a product of teams that work effectively. Learning to be a good team member is a prescription for professional success in science. Learn how to cultivate the power of human teams (e.g., creativity, problem solving, and division of labor) without succumbing to the hazards (e.g., documents that appear to be have been written by committee, duplication of effort, petty acrimony, violent discord, etc.). In your spare time, read Hong and Page (2004) and ask your colleagues what they think of it. Note the NY Times article appended to the original research paper.

Hong, L. and S. E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Nation. Academy of Sciences* 101:16385-16389.

Have a process for writing, revising, and publishing the paper. For example:

1. Work with your team to agree on the precise wording of the questions that the research addressed.
2. Agree on the precise answer to each question by locating the outcome in the matrix of possibilities (below).

Table 1. A matrix of possible research results. One likes to land in the upper left box or the lower right box, but the upper right and even lower left are also common. Authors should agree on where the results fall within this matrix and use vocabulary and tone in describing the results that conveys their judgment.

Is the relationship between *X* and *Y* biologically relevant if real?

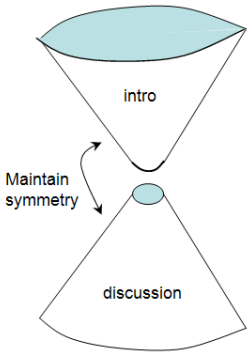
Do the statistics indicate a relationship between *X* and *Y*?	No	Maybe	Yes
No	"There were no differences." "Treatments were very similar." "Nectar addition had no effect." [This box is a good place to be when the question was interesting]		"Higher…" "tended to be higher…" "40% higher…" [a syndrome of too few data]
Maybe			
Yes	"Effects were small but significant" [a syndrome of too many data]		"Species richness was higher ( )." "was significantly higher…" "was much higher…" "was 2x higher…" [This box is a good place to be when the question was interesting]

3. Agree on the figures, tables, and results presentation in general.
4. Try out your results presentation on someone outside the work group (e.g., from course staff); revise presentation of questions and results as appropriate.
5. Outline the manuscript together (to about the level of topic sentences for each paragraph).
6. Write the manuscript. Revise the manuscript (with all authors providing input on the entire paper). Submit the manuscript.
7. When the manuscript comes back from review, go over all of the comments as a group. Talk to the reviewer for clarifications and further suggestions. Agree on the responses. Perform the revisions. Polish the revised manuscript (with all authors reading the entire revised paper before resubmission). Submit the revised manuscript.
8. Repeat as needed until the manuscript is accepted for publication and the galley proofs have been checked.

Some more detailed suggestions for Bio 52 authors.

Write with: Clarity -- Precise words; simple direct sentences; topic sentences; logical paragraph development.
Economy – Purge repetition with good organization; eliminate superfluous words, sentences, and paragraphs. See Table 1 in Mack (1986).
Style -- Active tense/ 1st person. Use descriptive precise vocabulary. Emphasize what you learned (not what you still wonder about); embrace the surprises that were revealed during the work. Employ creativity & logic in the presentation of results, and in developing the intellectual contents of intro & discussion.

Introduction. Funnel in.
Begin with broad framework; big question; relevant theory. A good paper is relevant beyond the species and place.
Provide enough background about the study system to understand hypotheses, predictions, and technical approach.
Articulate the research questions, alternative possible answers, and related biological hypotheses.
Foreshadow potential consequences of different possible answers.
Summarize the logic of hypothesis tests and the technical approach.
Do not overly emphasize and/or describe the methodology used in the research, unless your research is aimed to test/compare efficacy of certain methods.



Discussion. Funnel out.
Give clear direct answers to your research questions. Describe the fate of hypotheses that were tested. Discuss the causes and consequences of patterns that were revealed.
Note any unexpected results (not foreshadowed in introduction) and their causes and consequences.
Consider the rigor of inferences, potential caveats, and strengths and weaknesses of technical approach. However, keep in mind that negative results do not necessarily imply technical mistakes or caveats of the methodology. Consider biological explanations as to why a hypothesis is rejected.
Finish with the most cosmic conclusions: e.g., broad causes and consequences of what you have learned (relevance to other species, systems, contexts?). Connect the reader back to how the paper began.

More on Introductions. Distinguish among: theory, hypothesis, prediction, assumption. Embrace multiple working hypotheses (Platt 1964). Emphasize biological hypotheses not statistical null hypotheses. De-personalize theories, hypotheses, and predictions. Avoid the construction "We predicted that …" Justify your hypotheses and predictions with logic and reference to existing theory, not with how you thought it would turn out. Good papers are usually constructed so that readers don't know or care how the investigators thought it would turn out.

More on discussions. Develop connections to the literature as appropriate. Carefully distinguish in your thinking and writing between conclusions from the present research vs. new hypotheses suggested by the research. Characterize the strength of inference for conclusions as: (1) a valid deductive argument – and therefore unassailable; (2) a reasonably compelling inductive argument; or (3) informed speculation. When offering a new hypothesis suggested by the research, justify its logic and rigor by identifying a critical test that could be done (in another study). Seize the awesome power of work groups (APOWG) to brainstorm whenever diverse creative input is helpful: e.g., when considering possible causes of observation X, possible consequences of process Y, broader connections of the work to other systems, next generation research questions, etc.

Methods. No need to be chronological; organize the methods to be logical given the structure of the introduction and results. Consider people like yourself and your colleagues to be the target audience. Describe statistics only as needed; frequently simple statistics are sufficiently described in the course of presenting the results.

Results.

Try different data presentations. Be creative and stylistic. Try out different versions. Pick the presentations that work best. Not everything needs to be a figure or a table. Some results are best presented as sentences.

The legends for figures and tables should complement but not repeat that which is in the text of the results. See "Sample results materials" (p.4-5).

Organize the results around answering the questions identified in the introduction.

Use topic sentences that refer to the research questions.

Describe the biology with reference to the statistics, not vice versa. See "Sample results materials" (p.4-5).

Illustrate your points with specifics from the data.

Describe the clearest and most relevant patterns first.

Consider the important things you learned that were not part of the original study plan (and may have been a surprise) but which may belong in the "Results" even if they do not involve statistics (e.g. when expecting to sample males and females equally, "18 out of the 20 (90%) animals sampled were males")

Make your figures and tables outside of the manuscript document and then paste them into the end of your document (usually one per page) with their legends. See Figure and Table formatting guidelines for fonts, etc.

Look for opportunities to enhance your manuscript with one or two illustrative photographs of your study system. These can be included as "Figure X". We have Photoshop on the FSP notebook computer, which can be used for cropping and polishing photographs.

Consider including as electronic supplements tables, figures, photographs, mathematical derivations, or other details that do not need to be published as hard copy but which support the research. These files will be electronically accessible to subsequent readers. Refer to them in the text as "Electronic Supplement 1", etc. Electronic supplements might include photographs, sound files, large tables, R-code, an Excel spreadsheet, etc. When possible, include the electronic supplements within your metadata / data file.

See formatting guidelines toward end of this document regarding presentation of statistical results, etc.

Abstract

Write abstract last. Start with topic sentences from each section and then shorten and polish.

Manuscript preparation for submission, review, and subsequent revisions

1. We recommend that you begin with the template file, Bio52.template_ms-writing&editing.docx
1. Name your ms file as "ORP-#.ProjectName.v##" (e.g., ORP-2.Caecilians.v01.docx). Update version ##s during process.
2. Order of manuscript elements = Title, Authors, Abstract, Intro, Methods, Results, Discussion, Acknowledgement (if appropriate), Author Contributions, Literature Cited, Tables, Figures.
3. Tables and figures, with their captions, should appear at the end of the document, usually one per page.
4. No columns. We recommend single spacing when writing and editing.
5. Submit the ms for review by the staff.
7. Read and consider reviewer comments. Revise appropriately. Resubmit.
8. Repeat as needed.

If and when your manuscript is accepted for publication

File the following items with the course staff.

- Electronic manuscript as *.docx with all final revisions in place and no comments or track changes.
- Project data and metadata as a descriptively named Excel file (e.g., ORP-2.Caecilians.data.v##.xls) following the conventions within 03.Data&Metadata_example.xlsx.

See sample results material on next pages.

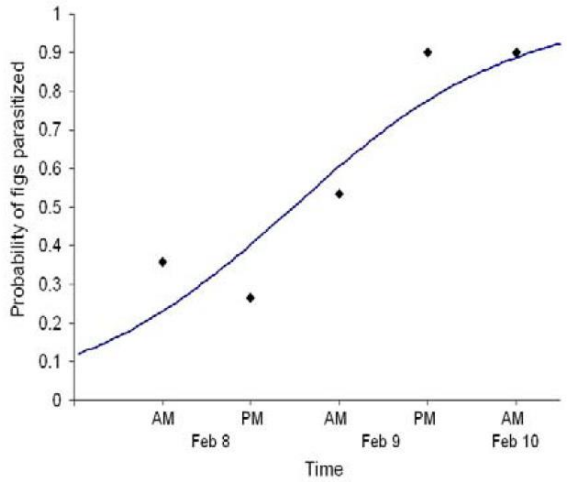


Figure 3. Proportion of figs that contained fig wasp parasitoids as a function of the time ripened fruit dropped to the ground. Data from a single *Ficus obtusifolia* tree near Sirena Bioloogical Station in Corcovado National Park, Costa Rica: $y = 1-1/(1+e^{-2.023+0.817x})$.

<u>Corresponding text from results for Fig. 3 above:</u>

Fruits that ripened later were about 3-times more likely to be parasitized than fruits that ripened earlier (Fig. 3). The probability of any fig fruit containing visible parasitoids increased over 48 hours from about 30% to about 90% (chi-square = 11.61, df = 1, P = 0.0007). …

Figure 3. Close up of *N. leporinus* flying next to roost – 12 Feb 2008. Photo by Alex Spinoso.

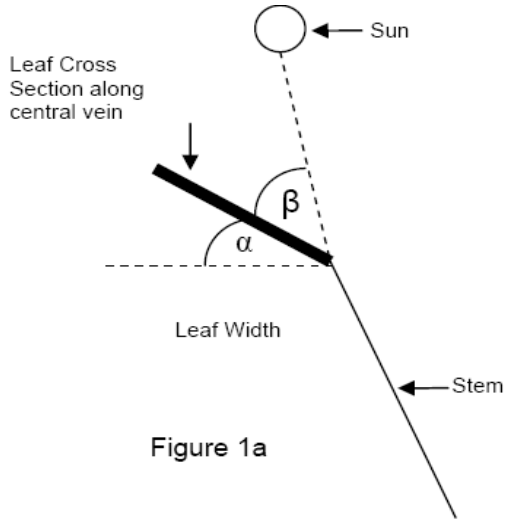


Figure 1a

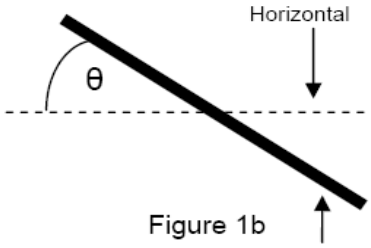


Figure 1b

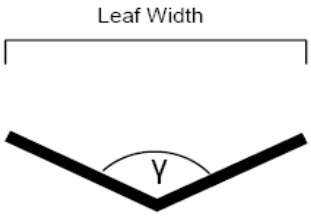


Figure 1c

Figure 1. A-C are diagrams of various angular measurements. A) Picture of a leaf cross-section taken along the central vein. The vertical angle between the leaf face and horizontal is pictured by alpha (α), while Beta (β) refers to the vertical angle between the leaf face and the sun. B) Graphical representation of leaf cross section (perpendicular to central vein). Theta (θ) refers to the rotational leaf angle relative to horizontal, in degrees. C) Gamma (γ) represents the angle of 'cupping' of each leaf. The smaller the γ value, the more 'cupped' a leaf is, and the less photosynthetic surface there is exposed.

<u>Corresponding text from methods:</u>

For each individual leaf we took five measure-ments; leaf face's vertical angle (α) from horizontal (Fig. 1a), leaf face's vertical angle to the sun (β) (Fig. 1a), leaf face's rotation relative to horizontal (angle θ) (Fig. 1b), and the 'cupping' angle of a leaf face (γ) (Fig. 1c).

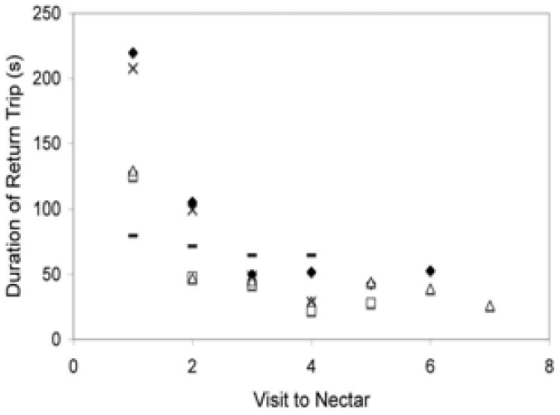


Figure 2. Relationship between the return trip duration of the first ant to find the nectary and its number of visits to it at La Selva, Costa Rica. Symbols differentiate first foragers from six separate experiments.

Corresponding text from results for above figure:

In all experiments, the first trip back to the nest required the longest time, and subsequent return times decreased non-linearly until reaching a minimum by an individual's 3rd or 4th trip to the nectar (Fig. 2).

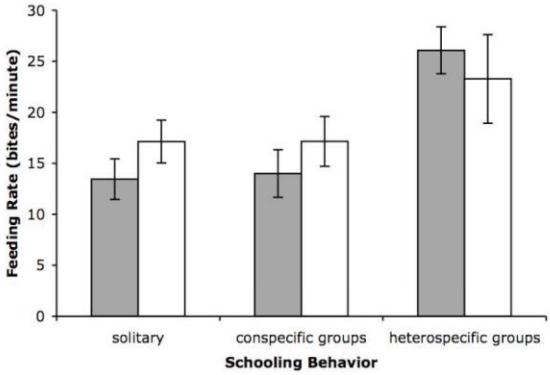


Figure 1. Feeding rate (mean ± SE) of adult and juvenile Ocean Surgeonfish (*Acanthurus bahianus*) with different schooling behaviors in a back reef of Grape Tree Bay, Little Cayman Island. Gray bars = juvenile (N=48), white bars = adults (N=38).

Corresponding text from results for above figure:

Schooling behavior affected feeding rate of juveniles and adults differently (Table 2, Fig. 1). Feeding rate of individuals in con- and heterospecific groups differed for juveniles (linear contrast $F_{1,44} = 10.11$, P = 0.003) but not adults (linear contrast $F_{1,33} = 1.33$, P = 0.23). When combined, feeding rate of individuals in conspecific and heterospecific groups did not differ from those of solitary foragers, for either adults (linear contrast $F_{1,33} = 1.17$, P = 0.29) or juveniles (linear contrast $F_{1,44} = 2.64$, P = 0.11). Juveniles in heterospecific groups fed the fastest (Fig. 1).

Table 1. Actual mean adult-adult distance, larva-larva distance, and minimum adult-larva distance (average distance from a larva to the nearest adult) compared to the median value for the same metric from randomization tests. Studies involved five passalid colonies found under logs at Cuerici Biological Station, Costa Rica. All colonies that had at least 2 adults and 1 larva were analyzed. Units are cm.

			Adult-adult distance		Larva-larva distance		Larva-nearest adult distance	
Log #	Adults	Larvae	Actual	Random	Actual	Random	Actual	Random
2	3	11	16.20	7.95	4.50	7.77	5.80	4.13
6	2	3	4.00	2.48	2.00	2.33	1.30	1.63
9	3	12	12.30	6.57	3.40	6.60	4.80	3.03
10	4	4	3.60	11.21	17.40	11.00	10.00	6.06
11	2	3	12.30	6.57	3.40	6.60	4.80	3.03
Median	2.80	6.6	9.68	6.96	6.14	6.86	5.34	3.57

Corresponding text from results for above table:

We found no evidence of spatial structure (Table 1). There was no aggregation in either larvae (t = 0.38, P = 0.72, df = 4) or adults (t = 0.97, P = 0.39, df = 4). The position of adults relative to larvae did not suggest care of larvae: the mean larva-nearest adult distance (3.58 cm) was 1.77 cm greater than the randomly predicted distribution (marginally significant, t = 2.61, P = 0.06, df = 4)

3. Guidelines for the presentation of statistics

Stats:

Mean with 1 SE	mean ± 1 SE = ### ± ## mg
Mean with 1 SD	mean ± 1 SD = ### ± ## µMoles
Confidence interval	95% CI = ### - ### ºC
simple t-test:	(t = #.##, P = #.##, df = ##)
paired t-test:	(paired-t = #.##, P = #.##, df = ##)
ANOVA:	($F_{df1, df2}$ = #.##, P = #.##)
chi-square:	(chi-square = #.##, P = #.##, df = ##)
correlation:	(r = 0.##, P = #.##, df = ##)
regression:	(slope = ### ± ##, P = #.##, r^2 =0.##)
comparing models	(delta AIC = ##.##); see also log-likelihood

Note the number of significant digits indicated in above table: *t*-statistics, *F*-statistics, chi-square statistics, correlation coefficients (*r*), and coefficients of determination (r^2) are almost always expressed to two decimal places. *P*-values are usually best expressed to two decimal places, except where $P < 0.01$; then use one significant figure using rounding, e.g. if $P = 0.0023$, use $P = 0.002$. With only rare exceptions, report exact p-values, rather than "$P < 0.05$". Use your judgment with the number of digits to report for parameter estimates (means, regression slopes, etc.).

Use your judgment in how you report regressions. Sometimes the most salient result is the slope ± SE, sometimes the r^2, sometimes the full regression model, and sometimes the *P*-value.

When reporting *F*-statistics from general linear models more complex than a one-way ANOVA, be sure it is clear to the readers what is being tested by each reported *F*-statistic (e.g., main effect of plant species, plant species x habitat interaction, etc.)

The style for reporting statistics in ecology is evolving. Favored styles vary among authors, journals and reviewers. The trend, which we support, is to emphasize parameter estimates and de-emphasize *P*-values more than has been done historically. Make liberal use of sentence constructions and table formats that highlight parameter estimates (e.g., mean ± SE or 95% confidence interval = ### - ###) and relegate *P*-values, if they are needed, to parentheses. In general, describe the data (with reference to *P*-values) and avoid flogging the reader with lots of *P*-values. The guiding principle is to clearly and efficiently communicate the biological results of your research to your readers.

Specify the correct SI units whenever reporting measurement values. Use a single space between the value and units (e.g., 7 cm)

4. Construction tips and formatting guidelines for Tables

- Create tables in Excel and then paste into Word. Avoid using the table editor in Microsoft Word because it is miserable. Organize all of your tables for each manuscript within one excel file and submit that file with the other files when the manuscript is accepted.
- OK to use 10-pt font and landscape format for contents of table when it allows the table to fit on one page and be more accessible to readers..
- Format the tables from top to bottom as follows: caption, a double line, the category headings with additional single lines as helpful, the data, a single line at the bottom, then any footnotes. Do not include any vertical lines in tables.
- Your table caption should appear above your table.

Table 1. Number of calls of each bat sonotype (putative species) recorded during early and late sampling times at each site at each of four locations in Costa Rica.

	Palo Verde		Corcovado		La Selva		
Sonotype	Early	Late	Early	Late	Early	Late	Original morpotype codes [a]
1	88	19	24	0	22	29	pva, pvd, pvm, lcb, coe, lsb, pvb, pvc
2	0	16	0	5			pve, coi
3	0	1	0	3			pvf, coj
4	6	0					pvg
5	0	2					pvh
6	0	1					pvi, cog
7	1	0					pvj
8	1	0					pvk
9	2	0					pvl
10	0	1					pvn
11			0	8			coa
12			4	5			cob
13			7	0	10	2	cod, lsg
14			1	0			cof
15			2	1			coh
16			1	0			cok
17			2	0			col
18			0	1			coo
19							lca
20							lcc
21					4	4	lsa, lse
22					3	18	lsd
23					0	8	lsf
Total:	98	40	41	23	39	61	

[a] site codes = pv, co, lc, ls

4. Construction tips and formatting guidelines for Figures

- Create graphs in the software of your choosing (Excel and R are frequent choices). Polish your figures as needed in powerpoint. Capture the figure as a high resolution jpg (easily done with a screenshot). Then paste the jpg into your manuscript document and add the figure caption.
- Consider inserting a visible break in the axes when it is not practical to have the axis scaling begin at 0 (e.g., when the range of the data displayed are far from 0 or when the axis is on a log scale.). Some of us do this very simply in powerpoint at the last step by creating a visual break on the axis with a white box.
- If you use color in figures, choose from a pallet that is accessible to those with alternative color vision (http://jfly.uni-koeln.de/color/#pallet).
- Your figure caption should appear below your figure. Position and edit your caption within the Word document independently of the figure itself (i.e., do not cut and paste the figure and its caption simultaneously).
- Styles vary among journals, authors, and editors for the style of figure captions. Some prefer a neutral description of what is on the X and Y axes, leaving the reader to judge the pattern: e.g., ‘‘Relationship between Ovenbird territory size and distance from roads.’’ Others prefer that the caption conveys the biological message to the reader: e.g., “Territory size of Ovenbirds decreased with distance from roads.” We encourage you to lean towards the latter style, but to stop short of interpretations that would be placed in the Discussion rather than in the Results. Figure captions are very important to papers. Try different tacts. Solicit editorial suggestions. Expect some variation in style preferences among reviewers and editors.

Example of a figure with caption:

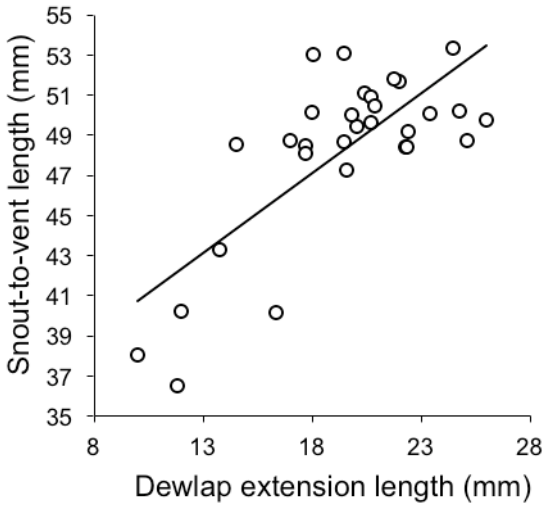


Figure 1. Snout-to-vent length was greater in lizards with larger dewlaps.

Appendix 1. Sample legends. Examples of figure and table legends. In some cases there is an early version of the legend along with the final (better) legend.

Early legend

Table 3. Shows the results of an ANOVA comparing the proportion of invertebrates in each functional group found above versus below each dam. There was no difference between functional group proportions above versus below each dam.

Better legend

Table 3. ANOVAs comparing the proportion of invertebrates in each functional group above versus below dams. Asterisks indicated significant effects. Error degrees of freedom = 16.

		Collector/Gatherer		Predator		Filter Feeder		Grazer/Scraper		Shredder	
Source	df	F	P	F	P	F	P	F	P	F	P
Position	1	0.00	0.98	1.30	0.27	3.62	0.08	1.80	0.20	2.50	0.13
Dam	7	0.58	0.76	0.77	0.62	4.61*	0.01	0.50	0.82	3.89*	0.01
P x D	7	0.38	0.90	1.23	0.33	1.10	0.41	0.56	0.78	0.34	0.93

Final legend

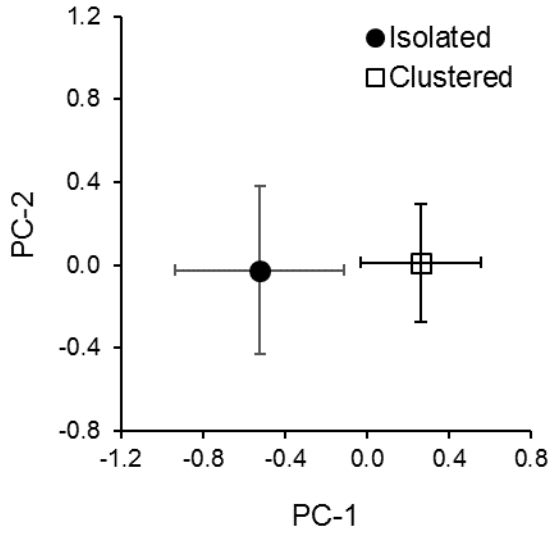
Table 1. Total number of each taxa captured above and below dams.

Taxa	Functional Feeding Group	Below	Above
Damselflies (Odonata)	Predator	4	14
Dragonflies (Odonata)	Predator	3	2
Non-segmented flatworms (Platyhelminthes)	Predator	5	9
Caddisflies (Trichoptera: Hydropsychidae)	Filter Feeder	11	28
Mayflies (Ephemeroptera: Heptageniidae)	Scraper	2	8
Beetles (Coleoptera: Ptilodactylidae)	Shredder	17	37
Crustaceans	Collector-Gatherer	3	1
Midges (Diptera: Chironomidae)	Collector-Gatherer	2	3
True bugs (Hemiptera)	Predator	3	1
Black flies (Diptera: Simuliidae)	Filter Feeder	6	18
Aquatic worm (Oligochaete)	Collector-Gatherer	3	4
Adult beetles (Coleoptera)	Collector-Gatherer	3	4

Final legend

Table 1. Characteristics of macroinvertebrate taxa found in bromeliad tanks near the Monteverde Biological Station.

Taxon	Dispersion coefficient	Functional group	Mean abundance	Variance of abundance
Chironomidae	1.09	filterer	0.83	1.94
Culicidae	2.14	filterer	3.90	98.23
Cyclopidae	2.29	filterer	3.90	42.51
Oribatida	2.59	gatherer	0.37	3.34
Psychodidae	1.30	gatherer	1.17	4.07
Ceratopogonidae	1.15	gatherer	0.20	0.37
Dityscidae	1.43	predator	0.40	1.21



Early legend

Figure 2. The isolated and clustered bromeliads shared similar taxonomic compositions. Positive PC-1 values indicate a closer relationship between culicidae, cyclopidae, and oribatida; positive PC-2 values indicate a closer relationship between culiciduae, chironomidae, and psychodidae.

Final legend

Figure 2. Isolated and clustered bromeliads did not differ in their taxonomic compositions as characterized by a principle components analysis.

Early legend

Table 3. Loadings from a principle components analysis are shown for each collected taxon, with higher values indicating a higher probability of finding that taxon in a given community. Percent variance indicates how much of the data PC-1 and PC-2 explain

Final legend

Table 3. Loadings from a principle components analysis of the abundance of 7 taxa that occurred in the phytotelmata of bromeliads near Estacion Biologica Monteverde.

Taxon	PC-1 Loadings	PC-2 Loadings
Culicidae	0.45	0.41
Cyclopidae	0.59	-0.21
Chironomidae	0.02	0.67
Psychodidae	-0.30	0.50
Dityscidae	0.04	0.14
Ceratopogonidae	-0.36	-0.26
Oribatida	0.47	-0.05
% Variance	25	22

Early legend

Table 1. The proportion of behavioral events performed by individuals with juveniles is compared to individuals without juveniles. The total number of events observed is listed in the last row. We conducted a chi-squared test on the numbers of behaviors each group performed, which indicated that the deviation between the two groups was significant ($\chi^2 = 22.39$, df = 8, p = 0.0042). The four largest differences in proportions between these groups indicate the specific behaviors whose difference resulted in the significant chi-squared value.

Final legend

Table 1. Proportions of different behaviors in adult jacanas with and without young.

	Proportion of total activity	
Event	**No Young**	**Young**
Foraging	0.32	0.34
Flight	0.15	0.07
Vigilance	0.14	0.21
Wing Up	0.14	0.08
Vocalization	0.14	0.12
Preening	0.03	0.04
Antagonistic Behavior (conspecific)	0.03	0.01
Hopping	0.03	0.03
Walking	0.02	0.06
Snaking Head	0.00	0.01
Loafing	0.00	0.01
Antagonistic Behavior (heterospecific)	0.01	0.01
Total	**1.00**	**1.00**
Total events	**394**	**158**

BIO 52: formatting details

Use the "Oxford comma" in all instances for items in a series. For example: "a, b, and c"
Single space between each sentence.
Single space between number and units (e.g., 2 cm).
Single space around mathematical operators: e.g., mean ± SE = 2.82 ± 0.17 mg.

Following details are preset within the manuscript template file: Bio52.template_ms-writing&editing.docx
Font type: Calibri 11 point throughout
Margins: 1 inch on all sides
Title: only the first word and proper names should be capitalized.
Abstract: up to 250 words
Keywords: 3 – 5 keywords

References

For citations within body of text:
one author: Darwin 2008
two authors: Darwin and Wallace 2008
more than two authors: Darwin et al. 2008
For example. "Darwin (2008) reported that …" *or* "Gomphotheres are apparently no longer extant in the Neotropics (Darwin and Wallace 2008)."

In the literature cited section:

Narwani, A. and A. Mazumder. 2010. Community composition and consumer identity determine the effect of resource species diversity on rates of composition. Ecology 91: 3441-47. [**Article]**

Starsky, F. G. and A. F. Hutch. 1989. A guide to the birds of Costa Rica. Robinho Publishing Associates, Ithaca, NY. **[Book]**

Gómez, L. D., and J. M. Savage. 1983. Searchers on that rich coast: Costa Rican field biology, 1400-1980. Pages 1-11 in D.H. Janzen, editor. Costa Rica Natural History. University Press of Chicago, Chicago, IL. **[Chapter in an edited book]**

Baker, C. P. 2008. Mammals. CentralAmerica.com. Published online at http://centralamerica.com/cr/moon/momammal.htm, accessed 1/2/11 **[Website]**

Other features of style and format should follow author guidelines for Ecology
https://esajournals.onlinelibrary.wiley.com/hub/journal/19399170/resources/author-guidelines-ecy

Excerpts from Goldwasser 1998.

A COLLECTION OF GRAMMATICAL POINTS

Most of these points boil down to common sense. Very few of them are inviolable: your goals should be clarity and ease of expression. Read aloud what you have written, and make sure that it says what you mean to say. Be succinct.

1. **Data**: This word is plural. The singular is *datum,* a *piece* of information. Information is, but data are, data are, data are.

2. **Less/few**: Use *less* for quantities that aren't composed of identifiable units, as in "less water," "less light." Use *few* and *fewer* for quantities with identifiable units, as in "few animals," "fewer references."

3. **Comprise/compose**: Many people misuse the first of these words, which means "include." The whole *comprises* the parts. The parts *compose* the whole. "The study comprises five experiments." "The study is composed of five experiments." What does "is comprised of" mean?

4. **Avoid noun phrases**. Would you want to look something up in a Forest Rodent Age Distribution Survey Volume Index?

5. **This/these/that/those**: The "this" to which you're referring is probably clearer to you than it is to your readers. Say explicitly what you mean: "This conclusion," "These measurements," "Those species."

6. **Only** comes directly before the word that's "only." "He only died last week" means "That's all he did, he didn't do anything else, he only died." (Courtesy J. Thurber)

7. **To split** is an infinitive; in spite of being two words it is a single unit and should not be split. (You thought I was going to carelessly split it, didn't you?)

8. **Which/that**: This distinction is sometimes tricky; *which* is often wrongly used in writing that adopts a formal tone. Use *which* for a phrase that is set off by commas and simply gives further information about a group that is already fully specified. Use *that* for a phrase that restricts or defines the group in question. A *which* phrase may be optional, but a *that* phrase is usually essential for the meaning of the sentence. Say "The blue car, which is on the driveway, has blown a gasket" if there is only one blue car around, and you want to tell someone where to head with the tools; say "The blue car that is on the driveway has blown a gasket" if you don't want this person to waste time messing around with the perfectly intact blue car that is sitting on the street.

9. **Passive**: Avoid using the passive voice. Don't be bashful. Instead of saying, "The plants were measured weekly," say, "I measured the plants weekly," or, better, "I measured the plants daily."

10. **Subjunctive**: If I *were* to use the subjunctive incorrectly, then I would say, "If I *was* to use the subjunctive. . . ."

11. A **preposition** is not a word to end a sentence with.

12. "Chocolate **and/or** salsa" means that at least one of chocolate and salsa is included and that both of them might be—which is exactly what the shorter and less awkward "chocolate **or** salsa" also means. In contrast, the exclusive or, "chocolate **or** salsa **but not both**," must be specified explicitly.

13. One can **predate** a check, but a predator must **prey** on its prey.

14. **Affect/effect**: In their common meanings, the first is a verb ("to change something in some way") and the second is a noun ("the change itself").

15. The abbreviations **i.e./e.g.**: The first means "that is" (Latin *id est*) and is followed by an explanation. The second means "for example" (Latin *exempli gratia*) and is followed by an example. Consider using the English equivalents instead, since everyone already knows what they mean.

38. Beware of run-on sentences and be sure to use a short word when a polysyllabic one is unnecessary (*use* vs. *utilize, turkey* vs. *autoproctophragmotist*).

39. It may seem obvious that two or more subjects, even when conceptually linked, take a plural verb, yet phrases like "The abundance and distribution of animals constitutes the principal subject matter of ecology," crop up in embarrassingly conspicuous places. "There's my mom and dad." "Where's your shoes?"

40. Be consistent in your use of tense. The past tense is usually appropriate for descriptions of what you did or observed, the present tense for general statements about nature.

41. Capitalize all Latin taxonomic names except the species part of a binomial. Either italicize or underline all Latin names for genera and species, but treat adjectives made from taxonomic names as normal English words: "*Homo,*" but "hominid"; "Rosaceae," but "rosaceous."

42. **Is, are**, and other forms of **to be** provide only vague information about the relationship between the words on either side of them: this verb may, for instance, define, describe, equate, or identify members of a set. Any time you write one of these words (or forms of the only slightly less vague **to have**), ask yourself whether a more specific verb might convey your meaning more clearly.

43. **Position and emphasis**: Words at the beginning or the end of a phrase automatically receive extra emphasis, and placing important words there can help reinforce your point. In contrast, the wording of, "This is an exciting result" emphasizes not that what you have found is exciting, but that it is a result. "This result is exciting" is somewhat better; recasting this sentence as part of a more informative one, such as, "This exciting result shows that . . ." improves it even further.

44. Don't be afraid to repeat a word in **parallel constructions**. Although varying your choice of words can often make your writing more lively, using synonyms simply for the sake of variety can distract the reader by suggesting misleading distinctions. For example, "First we counted the papaw trees and then we tallied the hedgehogs," can leave the reader pondering the difference between *counting* and *tallying* rather than your exciting discovery about hedgehogs and papaw trees.

Acknowledgments

I thank R. Colwell and A. Johnson for their contributions to this collection. R. Colwell also invented the longest word here.

Lloyd Goldwasser
Marine Science Institute
University of California
Santa Barbara, CA 93106-6150
E-mail: goldwass@lifesci.ucsb.edu

Excerpt from Mack 1986

Table 1. Common expressions with superfluous words (left column) and suggested substitutes (right column).

The purpose of this study was to test the hypothesis	I (or We) hypothesized
In this study we assessed	We assessed
We demonstrated that there was a direct	We demonstrated a direct
were responsible for	caused
played the role of	were
On the basis of evidence available to date	Consequently
in order to provide a basis for comparing	to compare
as a result of	through; by
for the following reasons	because
during the course of this experiment	during the experiment
during the process of	during
during periods when	when
for the duration of the study	during the study
the nature of	(eliminate by rearrangement)
a large (or small or limited) number of	many (or few)
conspicuous numbers of	many
substantial quantities	much
a majority	most
a single	one
an individual taxon	a taxon
seedlings, irrespective of species	all seedlings
all of the species	all species
various lines of evidence	evidence
they do not themselves possess	they lack
were still present	persisted; survived
the analysis presented in this paper	our analysis
indicating the presence of	indicating
despite the presence of	despite
checked for the presence of	checked for
in the absence of	without
a series of observations	observations
may be the mechanism responsible for	may have caused
It is reasonable to assume that where light is not limiting	With light not limiting
in a single period of a few hours	in a few hours
occur in areas of North America	are in North America
adjacent transects were separated by at least 20 m	adjacent transects were at least 20 m apart
in the vicinity	nearby
separated by a maximum distance of 10 m and a minimum distance of 3 m	3–10 m apart
the present day population	the current population; the population
their subsequent fate	their fate
whether or not	whether
summer months	summer
are not uncommon	may be
due to the fact that	(eliminate by rearrangement)
showed a tendency toward higher survival	had higher survival
devastated with drought-induced desiccation	killed by drought

Data Analysis Reference Guide

This guide was designed as a reference manual/cheatsheet for students taking ecology courses with no statistics prerequisite. The assumption is that students will learn an operational understanding of statistics in class, and that they will learn the ins and outs of analysis under the guidance of teaching staff. The information contained here is meant to merely serve as a reference to jog your memory when minor details are forgotten. See further N.B. on page 4.

The JMP "how-to" materials were pulled largely from the JMP support web pages, and the sources are noted in the headers for each section. Some of the other material comes from power points or worksheets used in Bio 22. I've tried to amend these so they will make sense to someone who hasn't taken it, but there might still be a reference or two leftover to something we did "in class."

Table of Contents

How to make inference from data

What to Do	Where to go in this guide
1. Ask a fascinating research question.	
2. Formulate a (biological) hypothesis from your question. (More than one, if appropriate.)	
3. If you have not done any statistical analysis before, read the example starting on p. 3.	
4. Devise a study that will help you test your hypothesis(-es). a. What is/are your response variable(s)? b. What are your predictor/treatment variables? c. How will you measure these in the field or lab? d. Ensure independence of samples and/or randomization as appropriate. e. Draw one or more figures of your possible outcome(s).	
5. Figure out how your variables will be organized into a spreadsheet, so that you collect the data in a way that will make sense later for analysis (i.e., figure out what your rows and columns will be, and ensure everyone on your team is doing it the same way! Or assign one person to record the data).	See pp. 7-10 for a data organization example.
6. Perform the study / collect the data.	
7. Transfer the data from your notebook to a spreadsheet; import the data into statistical software.	See p. 11 for importing instructions.
8. Ensure normality and homoscedasticity of the data. Recall that parametric statistics requires a) independent samples; b) normality; c) homogeneity of variance. (Note to Bio 22 '17: We didn't emphasize checking the variances as much, but you should perform that test as well as testing for normality. See instructions.) While you're at it, check for extreme values, and decide what to do with them if you have them. Ask teaching staff for help.	See bottom of p. 24 (Distribution of Continuous Variables) for testing for normality, and p. 28 for testing for equal variances (among groups/treatments).

Rev. Sept 2017, Aoki

9. If assumptions have not been met, perform appropriate transformations. (But be sure to keep your original data! See p. 7-10.)	For common transformations (and how to do them in JMP), see p. 29; for additional info, see http://www.biostathandbook.com/transformation.html[1]
10. If transformations don't work, consider other methods of analysis.	See pp. 35-38.
11. Make sure that your data are not pseudo-replicated. Summarize (sum or average) across groups as necessary. Again, practice good data hygiene (keep your original data separate from transformed or summarized data).	For data summary instructions, see p. 18. For data hygiene examples, see pp. 7-10.
12. Perform statistical analyses. Interpret the outcome in terms of both your statistical and biological hypotheses. Ask teaching staff for help as needed.	Start with methods cheatsheets, p. 42. If your data are more complicated, see additional cheatsheet, p. 44. JMP how-to tutorials for the basic methods can be found beginning on p. 30 (see table of contents for details).
13. Save statistical output along with your data.	Data hygiene, pp. 7-10.

[1] McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

Basics: A Motivating Example

Snakes, Part I: An Example Test of a Hypothesis

Adapted from a document created for Dartmouth Bio FSP
by Peter Schulze, David Peart, and Scot Zens

"There are the same number of fer de lance as there are bushmasters at La Selva." In statistics, this statement is a null hypothesis, a hypothesis that no difference exists between two populations, and in this case, no difference between the abundance of two species. (Note that a statistical hypothesis is different from a research/scientific/logical hypothesis. We will discuss this distinction further in class.) One alternative hypothesis is that there is a difference between the number of fer de lance and the number of bushmasters. Can you think of any other alternative hypotheses?

If we could count all the fer de lance and bushmaster individuals at the site, we would know whether the two populations were the same size. There would be no need to bother with statistics. Unfortunately, ecologists are rarely in a position to census the whole population. Instead, we are forced to examine **samples** of the population, and make **inferences** about the complete populations from the **samples**.

The design and execution of the sampling procedure is the point where randomness is critical. One could sample the snake densities by walking transects through the forest and catching snakes as they strike. But the fer de lance is more aggressive than the bushmaster, so one would expect to find more fer de lance than bushmasters sampling in this way, even if both populations are the same size. This is an example of how population size estimates can be biased. If we use the biased samples, a conclusion that the fer de lance is more abundant than the bushmaster could be erroneous; the error would be a result of **failing to sample randomly**.

Assume that we are able to randomly sample the two snake populations. In a single sample we find 12 bushmasters and 9 fer de lance. The question remains, are there the same number of bushmasters as fer de lance in the park? Can you answer the question yet? How confident are you that your guess is correct? Rather than relying on the results of a single sample, it would help to know the averages from a series of samples. Why?

After measuring several samples we find that there are, **on average**, 11 bushmasters and 15 fer de lance per sample (i.e., the **mean** abundance of bushmasters is 11 per sample). Which population is larger? If I roll a die 36 times and get 10 sixes and 2 fives, would you conclude that there are more sixes than fives on the die? Knowing what you know about dice, you would not draw the conclusion that there are more sixes than fives on the die. Instead, you would probably conclude that the numbers of sixes and fives found after the 36 rolls were a result of chance. If I rolled the die another 36 times I might get 10 fives and 2 sixes. The point is this: even the mean numbers of each species in a large number of samples cannot, by themselves, tell us whether there are more fer de lance or more bushmasters. It appears that there are more fer de lance than bushmasters, but we cannot yet be confident in this conclusion. We need to know something about the **variability, or standard deviation,** in the number of snakes per sample. We might also be interested in whether the means of 15 and 11 are the results of examining 3 samples or 30 samples (sample size). If the standard deviation is low, and the sample size is high, then we will feel more confident about concluding that there are more fer de lance than bushmasters at La Selva.

Statistically speaking, our question is, "If the population sizes of fer de lance and bushmaster are the same size (null hypothesis), how likely is it that our sampling procedure arrived at means of 15 and 11 per sample strictly as a result of chance?" **This is the fundamental statistical question in hypothesis testing.** Recall that we got 10 sixes in 36 rolls as a result of chance in the previous example.

N.B.: Because a course in statistics is not required for either Biology 22 or Biology FSP, we won't try to duplicate a stats course within our class. A standard statistics course would step methodically through a wide variety of procedures used to analyze data, and it would take up to three full semesters of coursework to get through just the standard methods. This term, we will, in a sense, work backwards, presenting you with a full palette of analytical methods, and try to teach you to recognize the broad strokes of which methods to use under which circumstances. We'll further give you as much help as you need to perform these analyses in JMP. If you plan to go on in ecology, for a senior thesis or perhaps even for graduate school, we highly recommend that you take Biol 29 (Biostatistics) as your quantitative course.

If you find the quantitative material interesting and want to know more of the details as we proceed, we recommend two books:

Baldi, Brigitte, and David S. Moore. **The practice of statistics in the life sciences**. WH Freeman, New York, 2013 (also 2011 and 2009). [This book has been used as the textbook for Biol 29 in the past. It is now in its 3rd edition, but the 1st edition is still perfectly relevant and available very cheaply online.]

Ellison, Aaron M., and Nicholas J. Gotelli. **A primer of ecological statistics**. Sinauer, Sunderland, Massachusetts, 2004 (2nd edition 2012). [This book is on reserve at the library, and the first edition is also available quite cheaply online.]

PSLS will give you the details on all the basics. Ellison and Gotelli gives a more detailed exploration of statistical methods particular to ecology, and is more narrative in format (less like a textbook).

Snakes (and pine needles), Part II: Descriptive Statistics

Adapted from a document created for Dartmouth Bio FSP
by Peter Schulze, David Peart, and Scot Zens
with material from Baldi and Moore, PSLS

The bushmaster/fer de lance example relies upon 3 concepts: mean, standard deviation, and sample size.

Sample size	= n = the number of samples taken
Mean	= $\bar{x}$ = the sum of the counts in the samples divided by the number of samples = $\frac{x_1 + x_2 + \cdots + x_n}{n}$ or more succinctly, = $\frac{1}{n}\sum x_i$ (Σ is the capital Greek letter sigma, and means "the sum of." The subscript *i* stands for the individual observations, so x_i here simply means all the observations x_1 through x_n.) For example, we might go out and count bushmasters on 9 occasions, resulting in the following counts: 9, 12, 11, 8, 8, 12, 11, 13, and 15. Our mean, or $\bar{x}$, = $\frac{9+12+11+8+8+12+11+13+15}{9}$ = 11
Variance	= s^2 = the average of the squares of the deviations of the observations from their mean = $\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1}$ or = $\frac{1}{n-1}\Sigma(x_i - \bar{x})^2$
Standard deviation	= s = the square root of the variance The standard deviation is thus in the same units as the observations. (The variance is in squared units.)

Can you calculate the variance and standard deviation for the bushmaster counts?

The mean is a **measure of center**, and the variance and standard deviation are closely related **measures of spread**, or how far the observations deviate from the mean.
Another measure of center is the **median**, or the midpoint of a distribution of observations. To find the median:

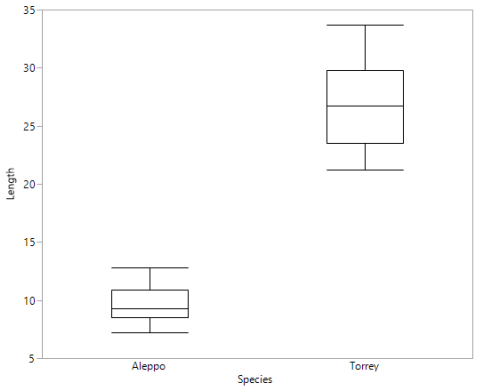
- Arrange the observations from smallest to largest.
- If there are an odd number of observations, the median is the center observation in the ordered list.
- If there are an even number of observations, the median is the mean of the two center observations in the ordered list.

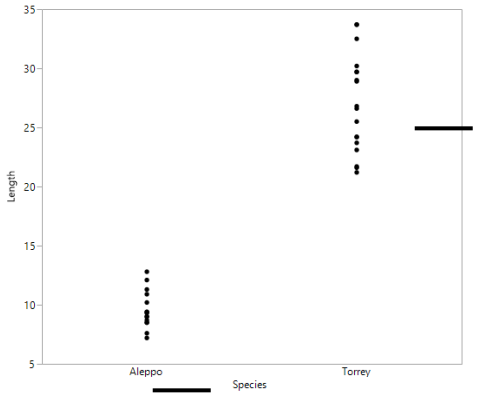
To measure spread around the median, we can use quartiles. To calculate the quartiles:

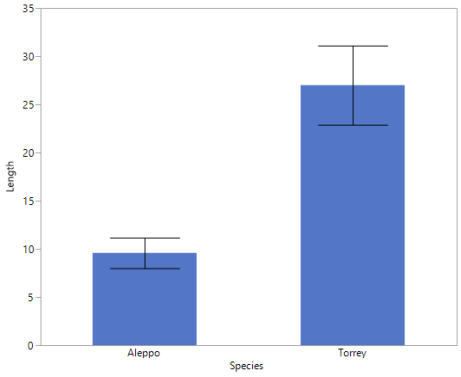
- Arrange the observations in increasing order and locate the median in the ordered list.
- The **first quartile Q_1** is the median of the observations whose position in the ordered list is to the left of the location of the overall median.
- The **third quartile Q_3** is the median of the observations whose position in the ordered list is to the right of the location of the overall median.

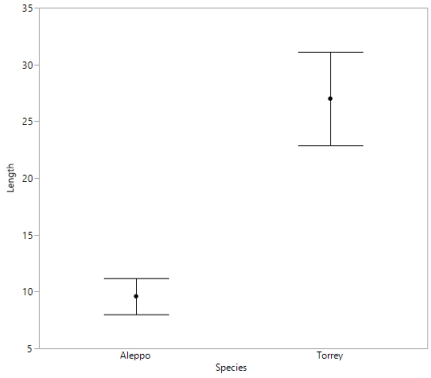
The **five-number summary** includes the smallest observation, Q_1, the median, Q_3, and the largest observation. A **boxplot** shows these five numbers with a box spanning the distance from Q_1 to Q_3, a line across the middle of the box showing the median, and vertical lines from the box to the smallest and largest observations.

Below are four different methods of showing center and spread for lengths of pine needles in two different species, using the same data in each figure. Which do you think gives the most accurate picture of the data?









Data and Metadata Example

(contributed by Matt Ayres)

MPA (28 December 2012):
This spreadsheet provides a general example of how to organize and manage data and analyses. This is not the only approach, but it is a good one when working with Excel and JMP as we will be doing. Please save this file on to your computer and refer to it (before beginning data entry) as a model for structuring your own data management.

Note that it will be convenient if your computer has some sort of "screen grab" software for saving the results of JMP analyses, which can then be permanently included within your master Excel data file, and annotated to record your interpretations.

Following these guidelines will be a great time saver for you on FSP. Furthermore, this is part of your broader training because developing good habits for managing data is fundamental to conducting good science. As is becoming the custom in refereed scientific journals, we will ask that you deposit your data for each project in this form, with metadata, to the Bio FSP data archive.

For further reading on data management see:
Chapter 8, "Managing and curating data", in Gotelli and Ellison (2004), A Primer of Ecological Statistics. In Bio FSP library.

Borer et al. (2009). "Some simple guidelines for effective data management." Attached here as a pdf (click on the icon to open).

Metadata Tab:

	A	B
1	The study was conducted in Aug 2005 in the Chickasawhay National Forest, MS, by Deepa Pureswaran (then at Dartmouth) and Brian Sullivan (US Forest Service).	
2		
3	Each row represents 1 southern pine beetle. Ntotal = 124	
4		
5	**Treatment descriptors**	
6	AttackTime	Putative **Pioneers** or **Scroungers** = early attacking beetles (2-3 after attacks began on that tree) or late attacking beetles (10-13 d after attacks began on that tree), respectively
7	MeasureTime	Time following landing on tree when beetle was sacrificed for quantitative extraction of pheromones from hindgut: **Landing** = captured within minutes after landing on tree; **Boring** = extracted one day after the beetle had begun boring into the phloem.
8	Sex	Sex of beetle: **Male** or **female**
9		
10	**Potential covariates**	
11	GalleryLength	length of the gallery that had been excavated by the beetle (cm); a measure of how long the beetle had been within the tree
12	BodyLength	size of the beetle (mm length)
13	HeadCapWidth	size of the beetle (head capsule width in mm)
14		
15	**Response variables**	
16	frontalin	pheromone amount in hindgut: ng of frontalin
17	endo	pheromone amount in hindgut: ng of end-brevicomin
18	trans-v	pheromone amount in hindgut: ng of trans-verbenol
19	verbenone	pheromone amount in hindgut: ng of verbenon
20	myrtenol	pheromone amount in hindgut: ng of myrtenol
21		

MPA:
In metadata, give:
1. brief explanation of study;
2. unambiguous description of each column in data file.

Original Data Tab:

	A	B	C	D	E	F	G	H	I	J	K
1	AttackTime	MeasureTime	Sex	GalleryLength	BodyLength	HeadCapWidth	frontalin	endo	trans-v	verbenone	myrtenol
2	Pioneer	Landing	Female	0	3.7	1.1	0	0	24.5	0	1
3	Pioneer	Landing	Female	0	4.1	1.1	0	0	14	0	2.25
4	Pioneer	Landing	Female	0	4.1	1.2	0	0	32.75	0	1.25
5	Pioneer	Landing	Female	0	3.3	1	0	0	2.25	0	0
6	Pioneer	Landing	Female	0	4	1.3	3.25	0	222	0	1.5
7	Pioneer	Landing	Female	0	4	1.3	10.25	0	1102.25	0	23.25
8	Pioneer	Landing	Female	0	4.1	1.2	0	0	5.25	0	1
9	Pioneer	Landing	Female	0	3.6	0.9	0	0	0	0	0
10	Pioneer	Landing	Female	0	4.2	0.8	0.75	0	173.75	0	3
11	Pioneer	Landing	Female	0	3	1.1	4.75	0	79.75	0	3.75
12	Pioneer	Landing	Female	0	3.9	1.1	2.75	0	259.75	0	3
13	Pioneer	Landing	Female	0	3.9	1.2	0	0	69.25	0	3.5
14	Pioneer	Landing	Female	0	4	0.9	0.5	0	57.75	0	7
15	Pioneer	Landing	Female	0	3.2	1	0	0	100.75	0	2
16	Pioneer	Landing	Female	0	3.3	0.9	1.25	0	320	0	17.75
17	Pioneer	Landing	Female	0	3	0.9	4	0	484.75	0	7.75
18	Pioneer	Landing	Female	0	3.9	1.1	0.75	0	58	0	19
19	Pioneer	Landing	Female	0	3.8	1.1	0.5	0	52.5	0	1
20	Pioneer	Landing	Female	0	3.9	1	0.25	0	74.25	0	3.75
21	Pioneer	Landing	Female	0	2.9	0.8	3.75	0	91.5	0	4.75
22	Pioneer	Landing	Female	0	3.9	1.1	1.5	0	213	0.5	2.25
23	Scrounger	Landing	Female	0	3.8	1.2	0.75	0	78.75	0	4.25
24	Scrounger	Landing	Female	0	3.1	1	0	0	52.5	0	4.25
25	Scrounger	Landing	Female	0	3.2	0.9	0.5	0	58.25	0	7.75
26	Scrounger	Landing	Female	0	3.4	1	0.5	0	134	0	2.5
27	Scrounger	Landing	Female	0	3.7	1.2	0.75	0	50	0	3
28	Scrounger	Landing	Female	0	3.3	0.9	13.25	0	119	0	4.5
29	Scrounger	Landing	Female	0	3.5	1.1	1.5	0	278	0	13.75
30	Scrounger	Landing	Female	0	3.4	1	0	0	15.75	0	1.5
31	Scrounger	Landing	Female	0	4	1.1	0	0	26	0	0
32	Scrounger	Landing	Female	0	4	1.2	0	0	125.5	0	0
33	Scrounger	Landing	Female	0	3.3	0.9	66	0	532.5	0	4.75

Notes | Metadata | **OrigData** | preJMP1 | preJMP2 | analysis1 | analysis2

MPA:
Original full data. Each row = one sample. Each column = a descriptor (and/or independent variable), or a response variable. Each column described in previous sheet (Metadata).

Descriptor variables to the left (from most coarse to most fine, left to right). Response variables to the right. Potential covariates in between.

1st row = column descriptors. No spaces or weird characters.

MPA:
Highlighted records are suspect because of problems with the gas chromatograph. These were excluded from subsequent analyses, which began with data file in subsequent sheet, PreJMP1

PreJMP1 tab:

	A	B	C	D	E	F	G	H	I	J	K
1	AttackTime	MeasureTime	Sex	GalleryLength	BodyLength	HeadCapWidth	frontalin	endo	trans-v	verbenone	myrtenol
2	Pioneer	Landing	Female	0	3.7	1.1	0	0	24.5	0	1
3	Pioneer	Landing	Female	0	4.1	1.1	0	0	14	0	2.25
4	Pioneer	Landing	Female	0	4.1	1.2	0	0	32.75	0	1.25
5	Pioneer	Landing	Female	0	3.3	1	0	0	2.25	0	0
6	Pioneer	Landing	Female	0	4					0	1.5
7	Pioneer	Landing	Female	0	4					0	23.25
8	Pioneer	Landing	Female	0	4.1					0	1
9	Pioneer	Landing	Female	0	3.6					0	0
10	Pioneer	Landing	Female	0	4.2					0	3
11	Pioneer	Landing	Female	0	3					0	3.75
12	Pioneer	Landing	Female	0	3.9					0	3
13	Pioneer	Landing	Female	0	3.9					0	3.5
14	Pioneer	Landing	Female	0	4					0	7
15	Pioneer	Landing	Female	0	3.2					0	2
16	Pioneer	Landing	Female	0	3.3					0	17.75
17	Pioneer	Landing	Female	0	3					0	7.75
18	Pioneer	Landing	Female	0	3.9	1.1	0.75	0	58	0	19
19	Pioneer	Landing	Female	0	3.8	1.1	0.5	0	52.5	0	1
20	Pioneer	Landing	Female	0	3.9	1	0.25	0	74.25	0	3.75
21	Pioneer	Landing	Female	0	2.9	0.8	3.75	0	91.5	0	4.75
22	Pioneer	Landing	Female	0	3.9	1.1	1.5	0	213	0.5	2.25
23	Scrounger	Landing	Female	0	3.8	1.2	0.75	0	78.75	0	4.25
24	Scrounger	Landing	Female	0	3.1	1	0	0	52.5	0	4.25
25	Scrounger	Landing	Female	0	3.2	0.9	0.5	0	58.25	0	7.75
26	Scrounger	Landing	Female	0	3.4	1	0.5	0	134	0	2.5
27	Scrounger	Landing	Female	0	3.7	1.2	0.75	0	50	0	3
28	Scrounger	Landing	Female	0	3.4	1	0	0	15.75	0	1.5
29	Scrounger	Landing	Female	0	4	1.1	0	0	26	0	0
30	Scrounger	Landing	Female	0	4	1.2	0	0	125.5	0	0

MPA:
Data in form suitable for import to JMP. Each row = one sample. Each column = a descriptor (and/or independent variable), or a response variable. Each column described in previous sheet (Metadata).

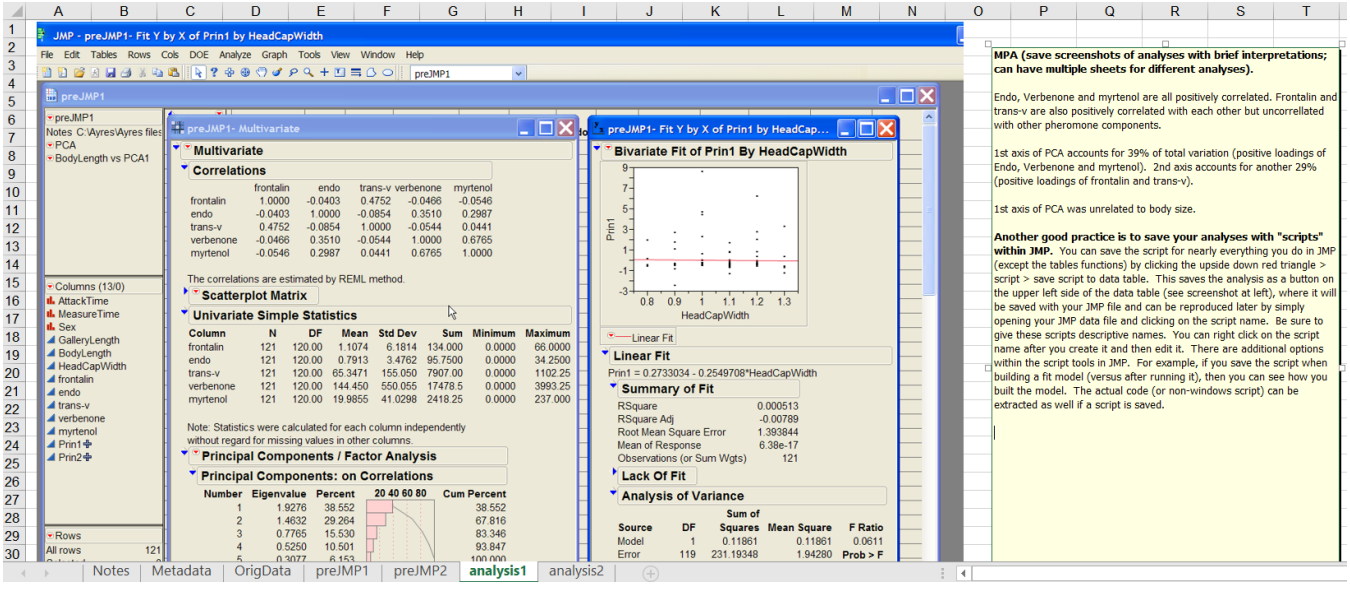
Descriptor variables to the left (from most coarse to most fine, left to right). Response variables to the right. Potential covariates in between.

1st row = column descriptors. No spaces or weird characters.

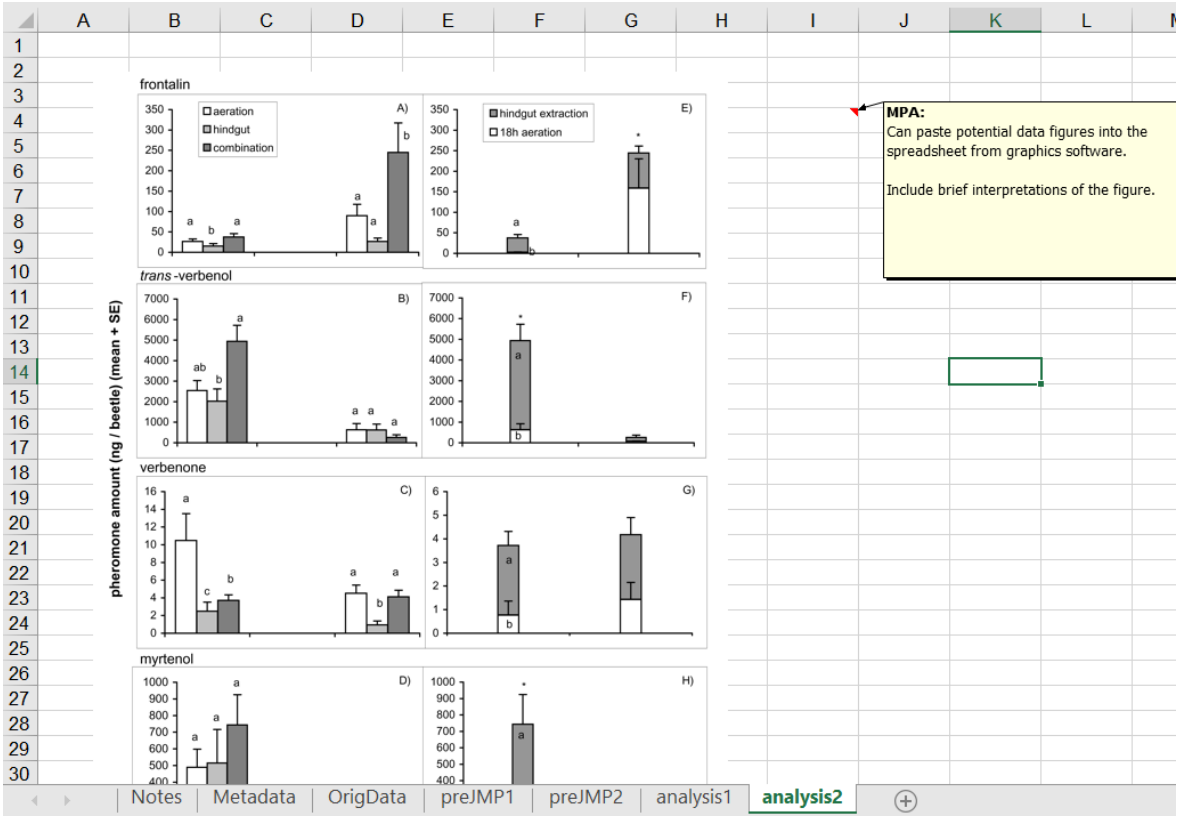
If data have been culled for some reason, the culled data are retained and identified within the sheet "OrigData".

Notes | Metadata | OrigData | **preJMP1** | preJMP2 | analysis1 | analysis2

Analysis1 tab:



Analysis2 tab:



Getting your data into JMP

There are three primary methods to get your data from your spreadsheet into JMP: copying and pasting, importing an Excel file directly, and importing an Excel file using the wizard. These three methods are described below. The methods vary in how well they retain your data type information (for example, you may have a categorical variable that JMP sees as a continuous variable). Don't forget to check all the variables in your JMP table once you've completed the data transfer. Select the method that works best for your computer platform. Your new data table in JMP is not automatically saved. Be sure to save it with a useful title if you don't want to re-import the data every time you use it.

Copying and Pasting Data

You can move data into JMP by copying and pasting from another application, such as Excel or a text file.

1. Open your spreadsheet file.

2. Select all of the rows and columns, including the column names.

3. Copy the selected data.

4. In JMP, select File > New > Data Table to create an empty table.

5. Select Edit > Paste with Column Names to paste the data and column headings.

If the data that you are pasting into JMP does not have column names, then you can use Edit > Paste.

Import a Microsoft Excel File Directly

Microsoft Excel files open in the Excel Import Wizard by default. This option is helpful when the structure of data in the worksheet is irregular. For example, you might want to exclude hidden columns or convert text in the third row to column headings.

To **always** open Microsoft Excel files outside the wizard, change the Excel Open Method preference.

The Excel Open Method preferences are in File > Preferences > General (Windows) and JMP > Preferences > General (Macintosh). From the Use Excel Labels for Headings list, select Always or Never (always means that JMP will always see the first row of your spreadsheet as column headers). Choose to open all worksheets at once or select them from a list.

To open a Microsoft Excel file (Windows)

1. After you set the Excel Open Method as described above, select File > Open.

2. Select the Excel Files file type, select the file, or enter the URL.

3. To convert text in the first row to column headings, select Always next to Always enforce Excel Row 1 as labels.

4. To import all worksheets, click Open.

or

To select the worksheets that you want to open, click the Open button arrow, and then select Open Selected Worksheets. Select one or more worksheets and click OK.

To open a Microsoft Excel file (Macintosh)

1. Select File > Open and select the file.

2. (.xls only) To convert text in the first row to column headings, select Use Excel Labels as Headings.

3. (.xls only) To open specific worksheets, select Select Individual Excel Worksheets.

4. Click Open.

If you chose to open specific worksheets, select those worksheets from the list, and then click OK. You can also click Select All if you change your mind and want to import all worksheets.

If you chose to open specific worksheets, select those worksheets from the list, and then click OK. You can also click Select All if you change your mind and want to import all worksheets.

If you selected an .xlsx file, a preview of the data appears in the Excel Wizard. See Preview and Import the Microsoft Excel Data (below) for details.

Import Microsoft Excel Files Using the Import Wizard

Microsoft Excel files open in the Excel Import Wizard by default. The wizard shows a preview of the data. You can then modify the settings before importing the data. For example, you might indicate which row the data begin on and whether the worksheet contains column headers or hidden rows or columns. Microsoft Excel .xls, .xlsm (on Windows), and .xlsx file formats are supported.
Notes:

- Password-protected Microsoft Excel .xlsx files cannot be opened in JMP.
- Between Windows and Macintosh, the number of digits after a decimal point and the date format of imported data might differ. For example, “10/25/2012” might be formatted as “25Oct2012” on Macintosh. Columns might be imported as character columns on Macintosh but not on Windows. For more details about formatting dates, see Pasting Dates from Another Application.

Preview and Import the Microsoft Excel Data

Before you import a worksheet, open the spreadsheet in Excel and decide how you want the data to be structured in the final data table. For example, you need to know whether the worksheet includes hidden or merged cells. In the wizard, you can then exclude hidden columns or rows.
To import a Microsoft Excel file that contains several worksheets, follow these steps:
1. Open the worksheet in Microsoft Excel.
For the figures in this example, we used the Team Results.xlsx file located in the JMP Samples/Import Data folder. The file has the following characteristics:

– the data begin on row 4, column 2 and end on row 9, column 5
– two worksheets
– the second worksheet has two sets of merged cells
– no hidden rows or columns

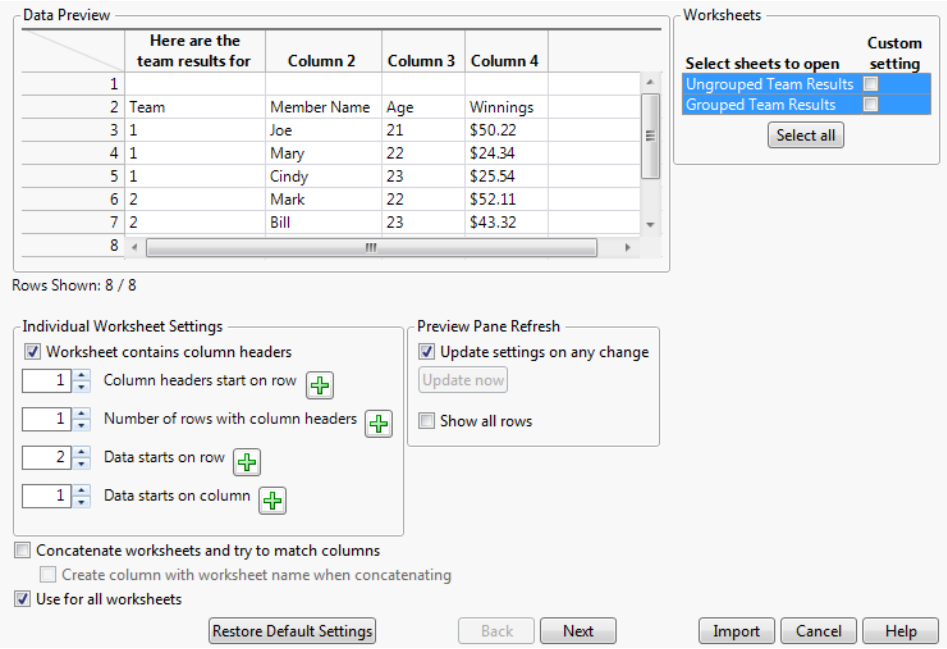
Team Results.xlsx Worksheet

	A	B	C	D	E	F
1		Here are the team results for this year's game.				
2						
3		**Team**	**Member Name**	**Age**	**Winnings**	
4		1	Joe	21	$50.22	
5		1	Mary	22	$24.34	
6		1	Cindy	23	$25.54	
7		2	Mark	22	$52.11	
8		2	Bill	23	$43.32	
9		2	Jennifer	24	$11.23	

2. To open an Excel file in JMP, select File > Open.
The Open Data File window appears.

3. Select the Excel file.
The worksheet opens in the Excel Import Wizard, where a preview of the data appears along with import options (Example Initial Data Preview).

Example Initial Data Preview



Note the following characteristics in the Data Preview:
– Both worksheets are selected for import in the upper right corner.
– The first column has been automatically been removed.
– Text from the first row of the worksheet appears as the column headings. However, you want the text in row 3 of the worksheet to be used as the column headings.
– The first data row is empty.

Note: JMP remembers your previous changes each time you import a worksheet, even after closing and reopening JMP. This feature is helpful when you want to reimport the same worksheet several times and experiment with options. To clear those changes when you import a different worksheet, click Restore Default Settings.

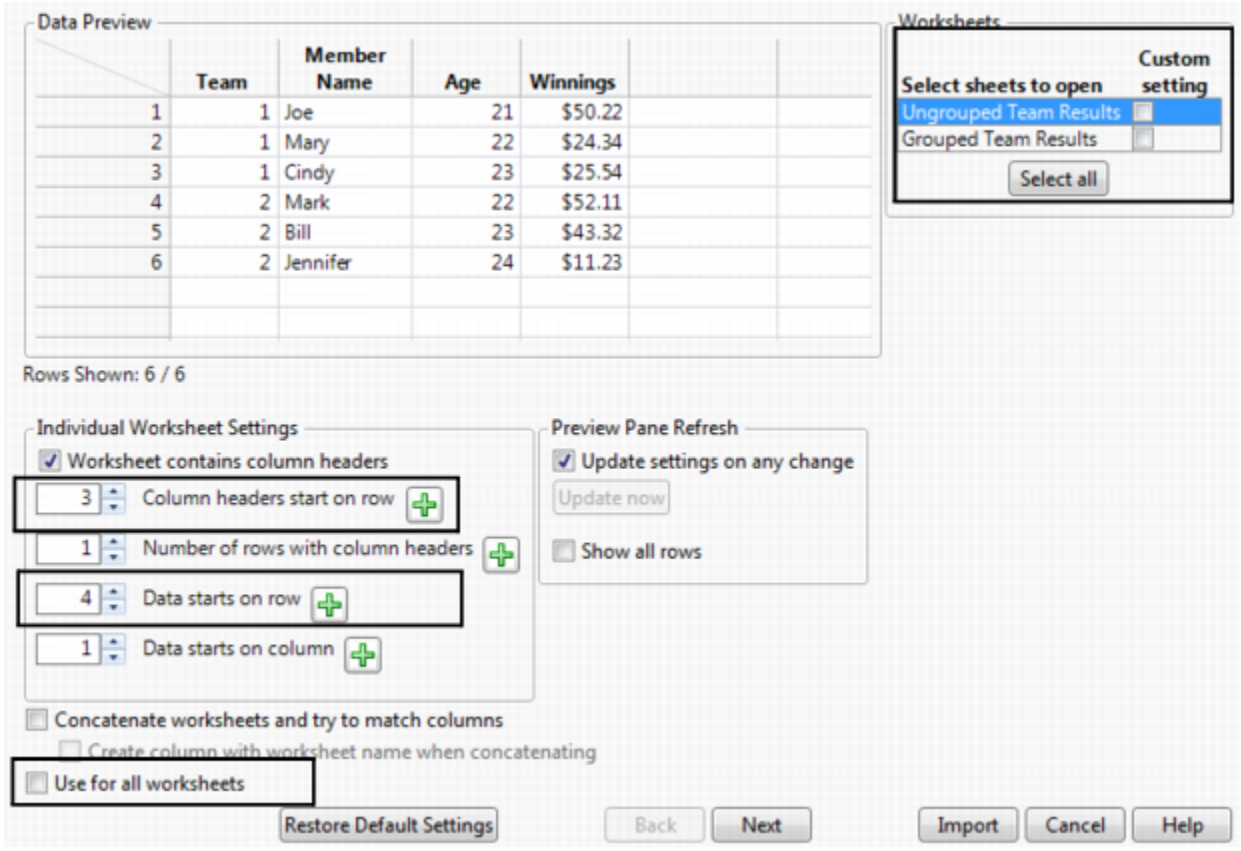
4. Type 3 for Column headers start on row

5. Type 4 for Data starts on row.

6. Select Ungrouped Team Results in the Worksheets pane.
Only this worksheet will be imported.

7. Deselect Use for all worksheets.
These settings apply only to Ungrouped Team Results.

Selecting the Column Header Row



See Individual Worksheet Settings (below) for details about all options.

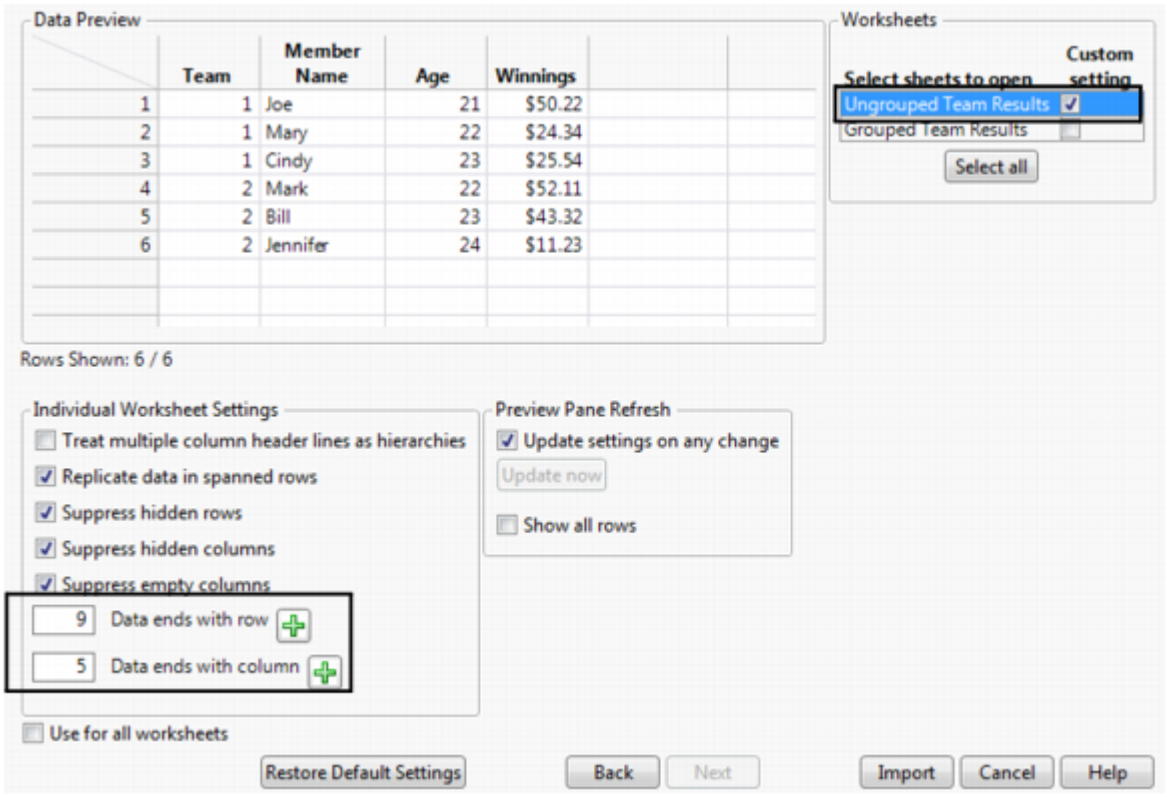
8. Click Next to configure other import settings.

The window displays additional import settings.

9. For Data ends with row, type 9.

10. For Data ends with column, type 5.

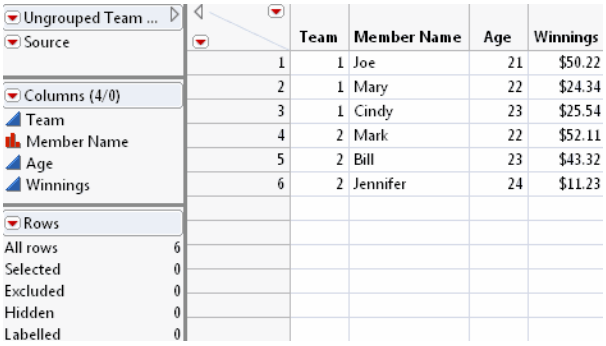
Specifying the Last Column



Custom setting is automatically selected when you modify the import options.

11. Click Import to convert the worksheet as you specified.

Final Data Table



The following sections describe options in the Excel Import Wizard.

Individual Worksheet Settings

Worksheet contains column headers
: Select if the worksheet contains rows with column headers.

Column headers start on row
: Indicates which row the column headers begin on in the worksheet. Click the up arrow until the headers begin on the correct row, or enter the row number and press Enter.

Number of rows with column headers
: Indicates whether the worksheet has multiple rows as column headers. Click the up arrow until the header rows appear correctly, or enter the number of rows and press Enter.

Data starts on row
: Indicates which row the data start on in the worksheet.

Data starts on column
: Indicates which column the data start on in the worksheet.

Concatenate worksheets and try to match columns
: Merges all worksheets into one data table. JMP matches columns that have the same header.

Create column with worksheet name when concatenating
: Adds a new Source Table column that lists the worksheet name for each imported table. This option is available after you select the preceding concatenate option.

Use for all worksheets
: Applies the current import settings to all worksheets that are selected in the upper right corner.

Additional Individual Worksheet Settings

Treat multiple column header lines as hierarchies
: Indicates that the worksheet contains multiple rows as column headers and you want these headers to be hierarchies

Replicate data in spanned rows
: Indicates cells are merged in the worksheet across rows. JMP unspan the cells and copy the cell contents into all of the resulting cells. The option is selected by default.
: If you deselect Replicate data in spanned rows, JMP unspans the cells and copies the cell contents into the topmost cell. The remaining unspanned cells are left empty.

Suppress hidden rows
: Prevents hidden rows from appearing in the data table. The option is selected by default.

Suppress hidden columns

Prevents hidden columns from appearing in the data table. The option is selected by default.

Suppress empty columns

Indicates whether an empty column that has a column header is imported. Deselect the option to import the column. The option is selected by default.

Data ends with row

Indicates the last row in the worksheet that contains data.

Data ends with column

Indicates the last column in the worksheet that contains data.

Tips:

- JMP remembers your previous changes each time you import a worksheet, even after closing and reopening JMP. This feature is very helpful when you want to reimport the same worksheet several times and experiment with options. To clear those changes when you import a different worksheet, click Restore Default Settings.
- Your import settings are saved in a data table script named Source. To reimport the worksheet using the same settings, run the script. The script includes the path to the worksheet, so make sure that other users have access to that location.
- As you experiment with settings for a large worksheet, the data preview might be slightly delayed. To speed up the preview, deselect Update settings on any change on the first wizard window. Modify the settings and then click Update now to refresh the data preview.
- To view all rows in the Data Preview pane, select Show all rows. The preview might be slightly delayed depending on the size of the spreadsheet.
- You can combine two worksheets from the same workbook into one data table. The column names are matched on import, so the order of the columns is irrelevant.

JMP Basics: Aggregating (summarizing) and Graphing Data

Aggregating the data spatially: Table Summary

(Note: This example originally included an exercise with data. The data are described, but not included here.) You will want to think carefully about the levels of spatial aggregation of the data: the lowest level is the *quadrat*, with 5 @ 1 m^2 quadrats nested within two *transects*, which are in turn nested within a 5 m x 5 m plot.

For most of you, our recommendation is to aggregate to the plot scale by SUMMING across the 10 quadrats, and if you need to convert to density, divide by 10 to get per m^2. This means that we sampled 20% of the total area of the plot, which should be a reasonable measure of the understory plants.

To aggregate, use the Table Summary option in JMP:

- Choose Table from the Main Menu
- Choose Summary from the pull-down menu
- Click on the response variable(s) of interest (use the command button on a Mac or the Alt button on a PC to select more than one) and then choose Statistics – Sum from that pull down menu. (You can also use this feature to get summary statistics like means and SE, but we recommend keeping the raw data as long as possible and looking at summary statistics using Graph Builder instead.)
- Choose the grouping variables: for most of you this will be something like year, site, plot, and treatment – so that site and treatment are available for use in the summarized data
- Click OK.

For practice, try getting the total number of plants per 25 m^2 plot, ignoring their taxonomic assignment, using the Total Plants column.

To make graphs: use Graph Builder

We strongly recommend using Graph Builder to make preliminary graphs. It's intuitive to some folks and clunky for others, but a nice first step especially for those of you who want to make plots both by year and by treatment (control/exclosure) and maybe even site. You can make prettier graphs later in Excel.

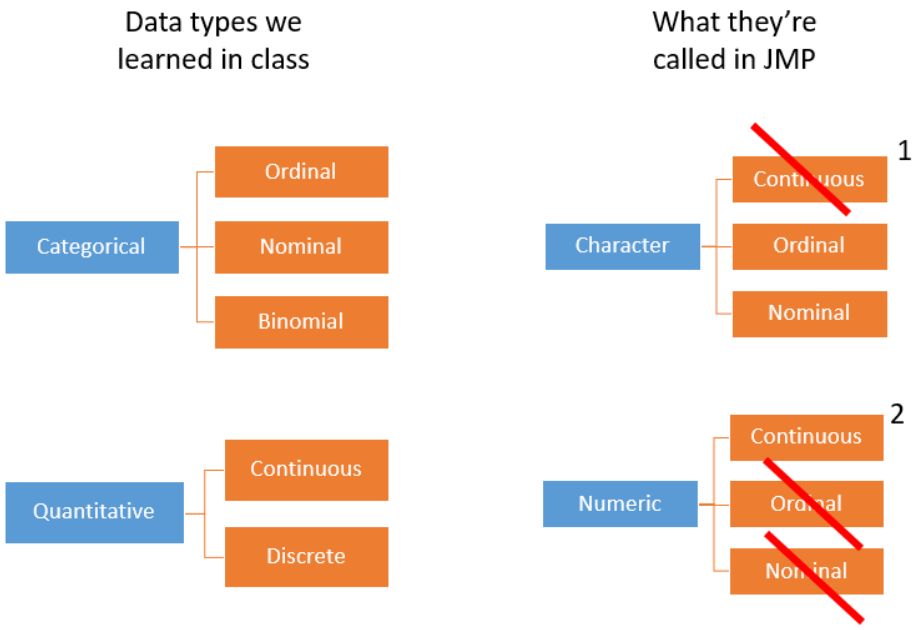
To plot total plants vs. treatment, within each year, from your Table Summary output:

- Click on Graph and choose Graph Builder
- Drag the Sum(Total Plants) column to the gray area that says "Y", which should make that variable pop up on the Y-axis and show some points in the graph itself.
- Drag the Year variable to the top of the graph panel where the gray area says "Group X" – you should now have data split out by year
- Drag the Site variable to the bottom gray area that says "X"
- Drag the Treatment variable to the area where the Site variable is, and wiggle a bit until you see Fenced and Open, by Site, for each year.

If you want bar charts, you can also Shift-Click onto the bar chart graph to add bars behind the symbols, and use the menu at least to change from mean to median summary statistics. (Boxplots aren't ok here because there are just 4 points per treatment combination.)

Types of Data, Analyzing Distributions, One sample t-test

Types of Data (pages 20-22): As we mentioned in the section on importing your data, the import method you choose may or may not retain the correct data types. The information below ("Understanding Modeling Types") demonstrates how data type differences can change the output in JMP, and shows you how to change the data type to the one you want. JMP has a relatively small number of data type (or as they call it, "modeling type") categories. We discussed a wider range of data types in class. The figure below shows how the categories we learned in class map onto the JMP categories. If you are brand new to data analysis, you may focus for now on simply learning how your data fit into the JMP data type system. The other categories on the left will be of interest later as we learn to identify the different types of analyses. (In JMP, you can perform many different analyses using just the "fit Y by X" and "fit model" platforms. You don't need to know the names of the analyses in order to perform them.)



[1]For some reason the "column info" window doesn't grey out the "character-continuous" choice. Never select this option. In practice, most categorical data you're likely to encounter in either Bio 22 or on FSP will be "character-nominal."
[2]For most of our purposes, *all* quantitative data should be designated numeric-continuous in JMP. See teaching staff if you think you have one of the rare exceptions to this rule.

To repeat, the JMP categories you will need on the vast majority of occasions are: numeric-continuous (for *all* numeric variables, including discrete count variables); and character-nominal (for almost all categorical data you will analyze in class/for projects).

Analyzing Distributions (pages 23-26): When you watched the audio powerpoint about univariate analysis, you learned how to assess a histogram for its various features. The section below on "Analyzing Distributions" describes how to obtain these histograms from your JMP data.

One-sample t-test (page 27): The instructions for a one-sample t-test are included here because they follow sensibly from the point and click functions used to analyze histograms.

Understanding Modeling (Data) Types

In JMP, data can be of different types. JMP refers to this as the modeling type of the data.

Modeling Types

Modeling Type	Description	Examples	Specific Example
Continuous	Numeric data only. Used in operations like sums and means.	Height Temperature Time	The time to complete a test might be 2 hours, or 2.13 hours.
Ordinal	Numeric or character data. Values belong to ordered categories.	Month (1,2,...,12) Letter grade (A, B,...F) Size (small, medium, large)	The month of the year can be 2 (February) or 3 (March), but not 2.13. February comes before March.
Nominal[2]	Numeric or character data. Values belong to categories, but the order is not important.	Gender (M or F) Color Test result (pass or fail)	The gender can be M or F, with no order. Gender categories can also be represented by a number (M=1 and F=2).

Different modeling types produce different results in JMP. To see an example of the differences, follow these steps:

1. Select Help > Sample Data Library and open Linnerud.jmp.

2. Select Analyze > Distribution.

3. Select Age and Weight and click Y, Columns.

4. Click OK.

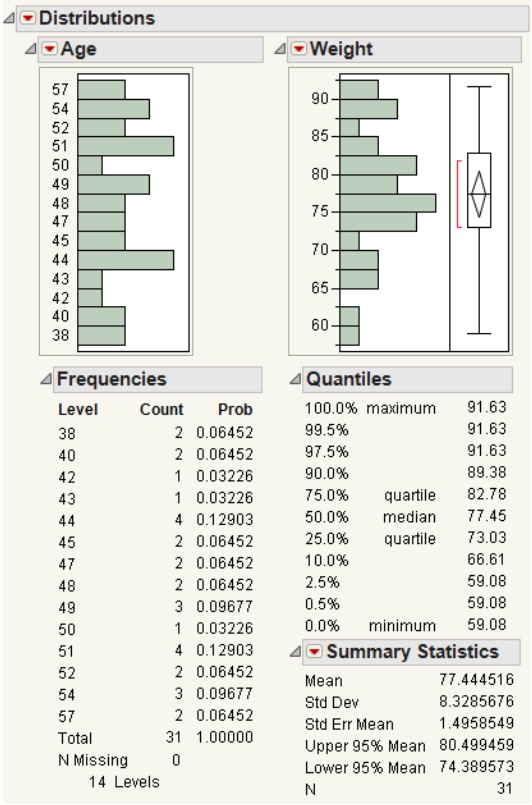
[2] Note to teaching staff: When a binomial GLM is required, the data type MUST be nominal. So although ordinal and nominal logistic regressions are both possible, I suggest we always (whether for simple logistic regression or for binomial GLM) specify binary variables as "nominal." From JMP's GLM help pages:

When you select Binomial as the distribution, the response variable must be specified in one of the following ways:

• If your data is not summarized as frequencies of events, specify a single binary column as the response. The response column must be nominal. If your data is summarized as frequencies of events, specify a single binary column as the response, along with a frequency variable in the Freq role. The response column must be nominal, and the frequency variable gives the count of each response level.

• If your data is summarized as frequencies of events and trials, specify two continuous columns in this order: a count of the number of successes, and a count of the number of trials. Alternatively, you can specify the number of failures instead of successes.

Distribution Results for Age and Weight



Although Age and Weight are both numeric variables, they are not treated the same.

Results for weight and age

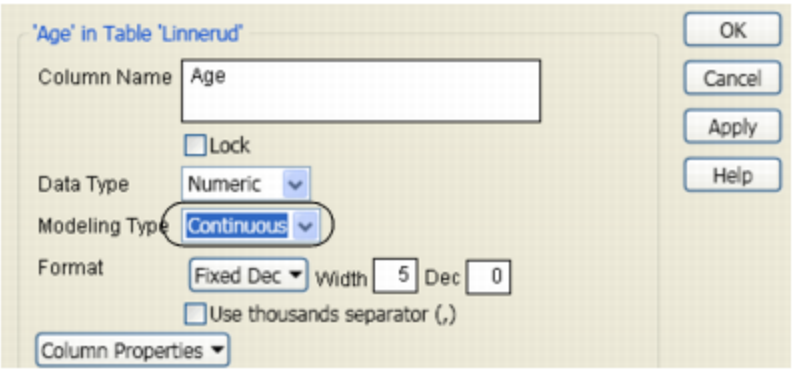
Variable	Modeling Type	Results
Weight	Continuous	Histogram, Quantiles, and Summary Statistics
Age	Ordinal	Bar chart and Frequencies

Changing the Modeling Type

To treat a variable differently, change the modeling type. For example, in Distribution Results for Age and Weight, the modeling type for Age is ordinal. Remember that for an ordinal variable, JMP calculates frequency counts. Suppose that you wanted to find the average age instead of frequency counts. Change the modeling type to continuous, which shows the mean age.

1. Double-click the Age column heading. The Column Info window appears.

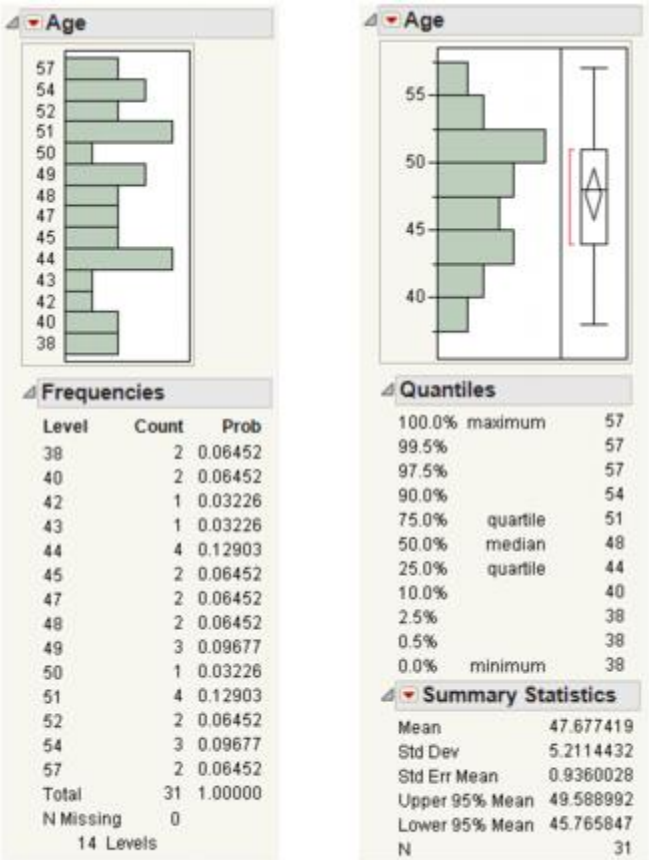
2. Change the Modeling Type to Continuous.



3. Click OK.

4. Repeat the steps in the example to create the distribution.

Different Modeling Types for age



When age is ordinal, you can see the frequency counts for each age. For example, age 48 appears 2 times. When age is continuous, you can find the mean age, which is nearly 48 (47.677)

Analyzing Distributions (and one-sample t-test)

To analyze a single variable, you can examine the distribution of the variable, using the Distribution platform. Report content for each variable varies, depending on whether the variable is categorical (nominal or ordinal) or continuous.
Note: For complete details about the Distribution platform, see the Basic Analysis book.

Distributions of Categorical Variables

Analyzing a categorical (ordinal or nominal) variable might include questions such as the following:

- How many levels does the variable have?
- How many data points does each level have?
- Is the data uniformly distributed?
- What proportions of the total do each level represent?

See the scenario in Distributions of Continuous Variables (below).
Now that the railroad company has determined that the average weight of the cars is not significantly different from the target weight, there are more questions to address.
The planning specialist wants to answer these questions for the railroad company:

- What are the types of cars?
- What are the countries of origin?

To answer these questions, look at the distribution for Type and Country.

Creating the Distribution

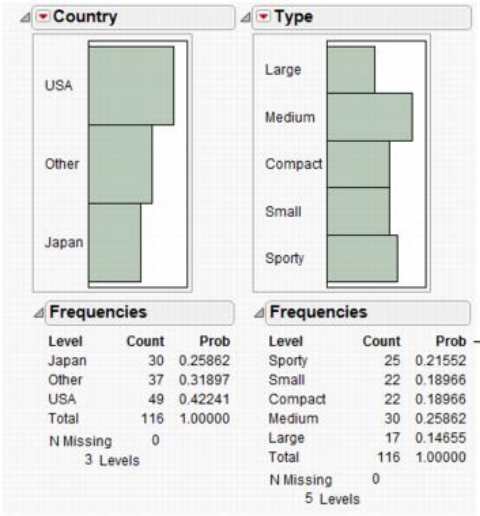
1. Select Help > Sample Data Library and open Car Physical Data.jmp.

2. Select Analyze > Distribution.

3. Select Country and Type and click Y, Columns.

4. Click OK.

Distribution for Country and Type



Interpreting the Distribution Results

The report window includes a bar chart and a Frequencies report for country and type. The bar chart is a graphical representation of the frequency information provided in the Frequencies report. The Frequencies report contains the following:

- Categories of data. For example, Japan is a category of Country, and Sporty is a category of Type.
- Total counts for each category.
- Proportion of the total each category represents.

For example, there are 22 compact cars, or about 19% of the 116 observations.

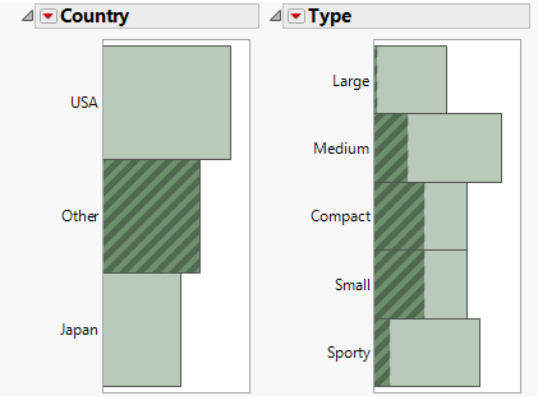
Interacting with the Distribution Results

Selecting a bar in one chart also selects the corresponding data in the other chart. For example, select the Japan bar in the Country bar chart to see that a large number of Japanese cars are sporty.

Japanese Cars

Select the Other category to see that a majority of these cars are small or compact, and almost none are large.

Other Cars



Distributions of Continuous Variables

Analyzing a continuous variable might include questions such as the following:

- Does the shape of the data match any known distributions?
- Are there any outliers in the data?
- What is the average of the data?

- Is the average statistically different from a target or historical value?
- How spread out are the data? In other words, what is the standard deviation?
- What are the minimum and maximum values?

You can answer these and other questions with graphs, summary statistics, and simple statistical tests.

Scenario

This example uses the Car Physical Data.jmp data table, which contains information about 116 different car models.

A planning specialist has been asked by a railroad company to determine the possible issues involved in transporting cars by train. Using the data, the planning specialist wants to explore the following questions:

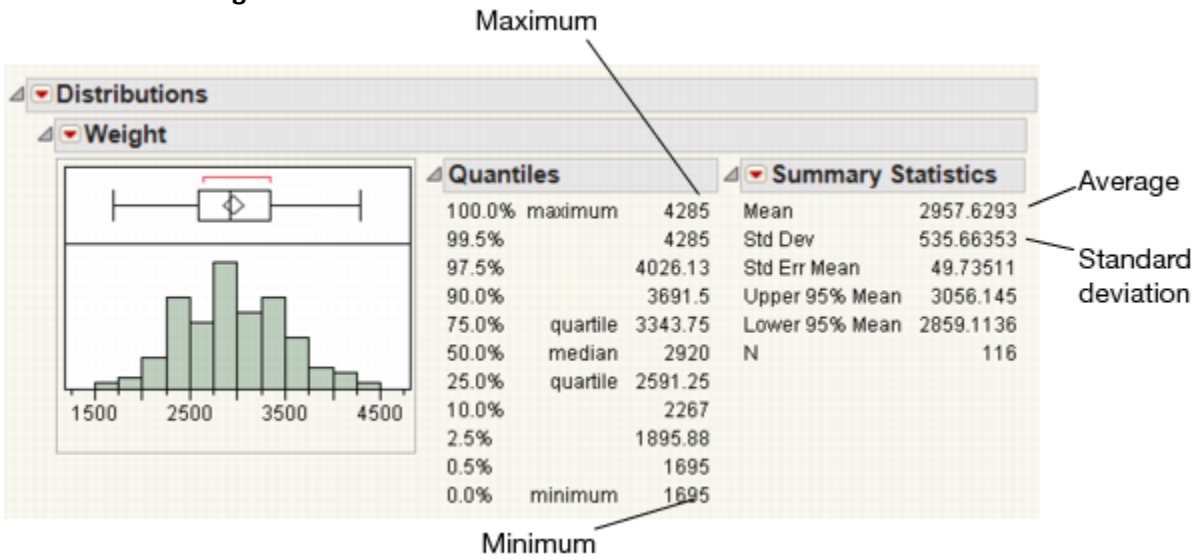
- What is the average car weight?
- How spread out are the cars' weights (standard deviation)?
- What are the minimum and maximum weights of cars?
- Are there any outliers in the data?

Use a histogram of weight to answer these questions.

Creating the Histogram

1. Select Help > Sample Data Library and open Car Physical Data.jmp.
2. Select Analyze > Distribution.
3. Select Weight and click Y, Columns.
4. Click OK.
5. To rotate the report window, select Display Options > Horizontal Layout from the red triangle menu next to Weight.

Distribution of Weight



The report window contains three sections:

- A histogram and a box plot to visualize the data.
- A Quantiles report that shows the percentiles of the distribution.
- A Summary Statistics report that shows the mean, standard deviation, and other statistics.

Interpreting the Distribution Results

Questions Answered Using the Report Window Results

Question	Histogram	Quantiles	Summary Statistics
What is the average car weight?	about 3000 lbs		2958 lbs
How spread out are the weights (standard deviation)?			536 lbs
What are the minimum and maximum weights?	about 1500 lbs and 4500 lbs	1695 lbs and 4285 lbs	
Are there any outliers?	No		

The default report window in Distribution of Weight provides a minimal set of graphs and statistics. Additional graphs and statistics are available on the red triangle menu.

Drawing Conclusions
Based on other research, the railroad company has determined that an average weight of 3000 pounds is the most efficient to transport. Now, the planning specialist needs to find out whether the average car weight in the general population of cars that they might transport is 3000 pounds. Use a t-test to draw inferences about the broader population based on this sample of the population.

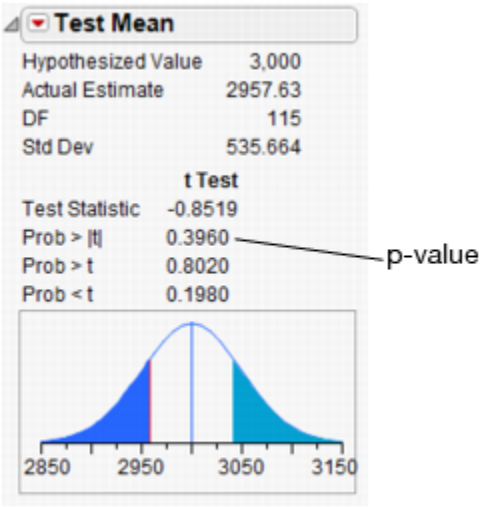
Testing Conclusions (One-sample t-test)

1. Using the "Distribution of Weight" results above, select Test Mean from the red triangle menu for Weight.

2. In the window that appears, type 3000 in the Specify Hypothesized Mean box.

3. Click OK.

Test Mean Results



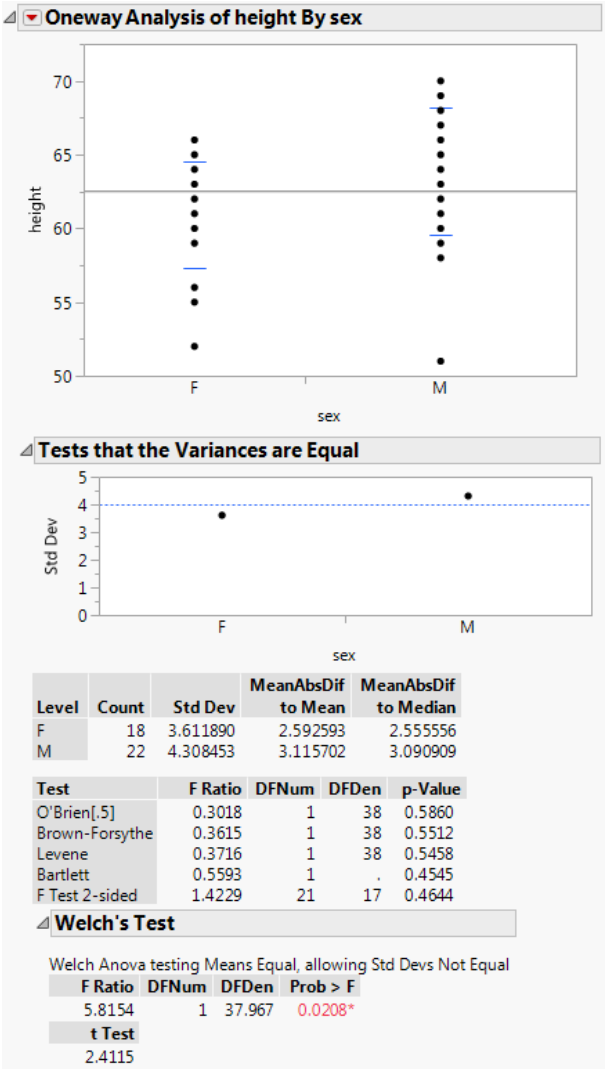
Interpreting the t-Test
The primary result of a t-test is the p-value. In this example, the p-value is 0.396 and the analyst is using a significance level of 0.05. Since 0.396 is greater than 0.05, you cannot conclude that the average weight of car models in the broader population is significantly different from 3000 pounds. Had the p-value been lower than the significance level, the planning specialist would have concluded that the average car weight in the broader population is significantly different from 3000 pounds.

For an example of a **one-sample t-test** used to test a **matched pairs design** (e.g., test subjects measured before and after a treatment), listen to the audio powerpoint **"One-sample t_matched pairs_compr.pptx."**

Testing for Unequal Variances

Suppose you want to test whether two variances (males and females) are equal, instead of two means.

1. Select Help > Sample Data Library and open Big Class.jmp.
2. Select Analyze > Fit Y by X.
3. Select height and click Y, Response.
4. Select sex and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Unequal Variances.



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
F	18	3.611890	2.592593	2.555556
M	22	4.308453	3.115702	3.090909

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0.3018	1	38	0.5860
Brown-Forsythe	0.3615	1	38	0.5512
Levene	0.3716	1	38	0.5458
Bartlett	0.5593	1	.	0.4545
F Test 2-sided	1.4229	21	17	0.4644

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
5.8154	1	37.967	0.0208*

t Test
2.4115

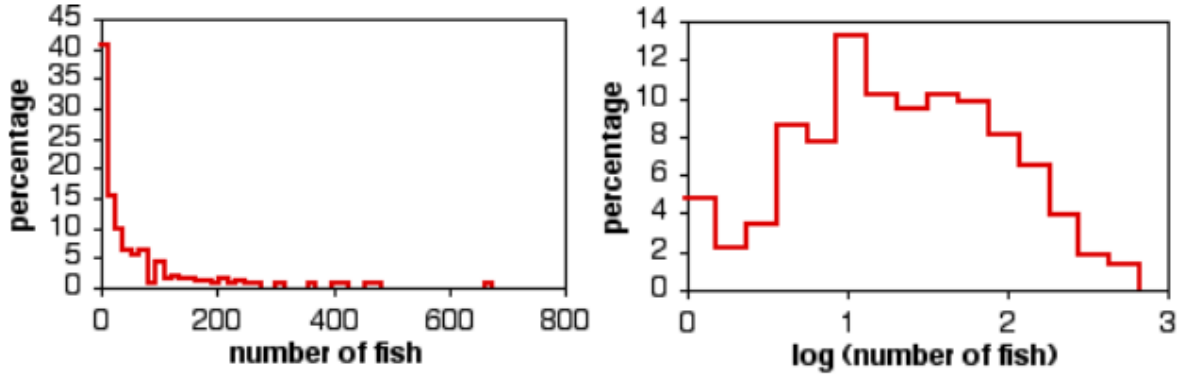
Since the p-value from the 2-sided F-Test is large, you cannot conclude that the variances are unequal.

Common transformations

For instructions on how to look at distributions to judge normality, see p. 24. For instructions on testing for unequal variances, see p. 28. Once you have determined that your data are not normal, or their variances are unequal, you can try the following transformations:

- For **positive continuous or count data**, try the log (either natural or base10) or square root transformations. You can do this directly in JMP by right-clicking on the column head, selecting "new formula column," then "transform," then your transformation of choice. Alternatively, you can right-click on an empty column and select "formula." Then you can input your formula by hand (for example, if you need to add 1 or .01 before log transformation).
- For **proportion data** (numbers between 0 and 1), you may use an arcsine square root transformation. Alternatively (and often better), you may use a generalized linear model with a binomial distribution, which doesn't require transformation (ask teaching staff for help). If your samples are actually binomial (i.e., yes/no, presence/absence, etc.) that became proportion data because you averaged across your sampling plots, and you're trying to do a regression (your x variables are continuous), you can perform a logistic regression (see p. X for instructions), without going into the generalized linear model menu. This won't seem different than performing a standard parametric ANOVA or regression. JMP will fill in the appropriate items for you. **For both logistic regression and binomial GLMs, be sure to specify your response variable as "nominal."**

For more details, including how to back transform, see http://www.biostathandbook.com/transformation.html (the figure below is from this page).



Histograms of number of Eastern mudminnows per 75 m section of stream (samples with 0 mudminnows excluded). Untransformed data on left, log-transformed data on right.

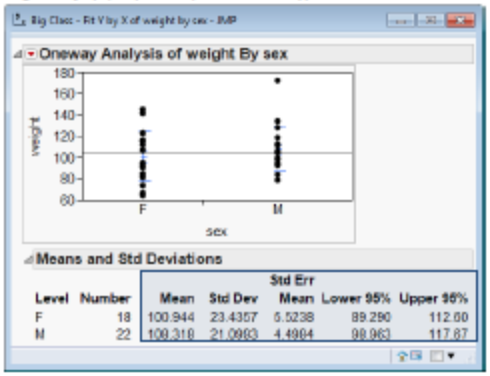
Two Sample t-Test and CIs

Estimate two population means (confidence intervals) or perform a hypothesis test for the difference between two independent means (two sample t-test) using the Fit Y by X platform. Note: If more than two means (more than two levels of the categorical X variable), refer to the page **One-Way ANOVA**.

Confidence Intervals

1. From an open JMP® data table, select **Analyze > Fit Y by X**.
2. Click on a continuous variable from **Select Columns**, and click **Y, Response** (continuous variables have blue triangles).
3. Click on a two-level categorical variable and click **X, Factor** (categorical variables have red or green bars).
4. Click **OK**. The Oneway Analysis output window will display.
5. Click on the **red triangle**, and select **Means and Std Dev** to produce summary statistics and individual confidence intervals for each mean (Lower 95% and Upper 95%).

Big Class.jmp (Help > Sample Data Library)

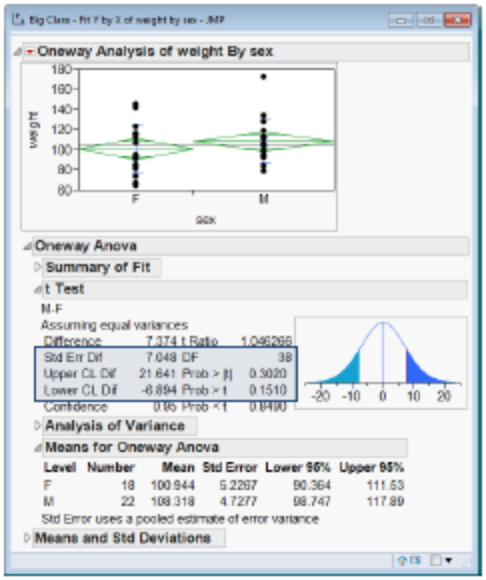


Two Sample t-Test

From the Oneway Analysis output window (shown above), click on the **red triangle** and select **Means/Anova/Pooled t**.

JMP will plot means diamonds (95% confidence intervals for each mean), and will generate:

- The Summary of Fit (not shown).
- The t-test report, with a graph to aid in interpreting the results.
- The Analysis of Variance table.
- Means for Oneway Anova (not shown), which includes confidence intervals based on the pooled estimate of the standard error.



Interpretation of the results (using a significance level of 0.05 - click **the red triangle, Set α Level** to change significance level):

1. **Upper CL Dif** and **Lower CL Dif** give the 95% CI for the true difference. Since the **95% CI contains zero, conclude that there is not a significant difference** between the means.
2. **Prob > |t|** is the p-value for the two-tailed test. The null hypothesis is that means are equal (the mean difference is zero). Since the **Prob > |t| is greater than 0.05, cannot reject the null hypothesis** (i.e., we cannot conclude that there is a significant difference).

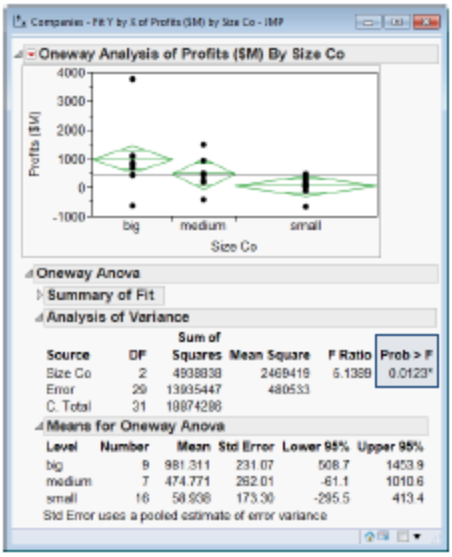
Notes: **Means/Anova/Pooled t** is the test under the assumption of equal variances. For a test without the assumption of equal variances, select **t Test** instead. See the ***Basic Analysis*** book (under **Help > Books**) for more details.

One-Way ANOVA

Test for differences between three or more population means using the Fit Y by X platform. Note: If testing two means (two levels of the categorical X variable), refer to the page **Two Sample t-Test and CIs**.

One-Way Analysis of Variance

Companies.jmp (Help > Sample Data Library)



1. From an open JMP® data table, select **Analyze > Fit Y by X**.
2. Click on a continuous variable from **Select Columns**, and Click **Y, Response** (continuous variables have blue triangles).
3. Click on a categorical variable and click **X, Factor** (categorical variables have red or green bars).
4. Click **OK**. The Oneway Analysis output window will display.
5. Click on the **red triangle**, and select **Means/Anova**.

   JMP will plot means diamonds (95% confidence intervals for each mean), and will generate:

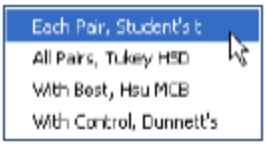
   - The Summary of Fit.
   - The Analysis of Variance (Anova) table.
   - Means for Oneway Anova, containing summary statistics and confidence intervals for each mean (based on the pooled estimate of the standard error).

Interpretation of the results in the Anova table (using a significance level of 0.05 – click **the red triangle, Set α Level** to change significance level):

- The null hypothesis is that there are no differences between the population means (i.e., all means are equal).
- **Prob > F** is the p-value for the whole model test. Since the **Prob > F is less than 0.05, reject the null hypothesis**. Conclude that there are differences between at least two of the means.
- To determine which means are different, a post hoc multiple comparison technique can be used.

Multiple Comparison Procedures

From the Oneway Analysis output window (shown above), click on the **red triangle**, select **Compare Means**, and select one of the four methods (described in JMP Help).

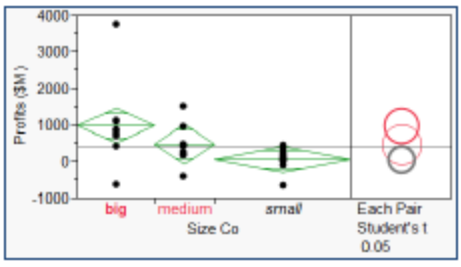


Each Pair, Student's t has been selected. This produces comparison circles (shown), along with statistical output (not shown).

Click on a circle for a mean to test for paired differences.

- The **selected mean** will have a **bold, red circle and variable label**.
- Means that **are not significantly different** from the selected mean will have **unbolded, red circles and variable labels**.
- Means that **are significantly different** from the selected mean will have **gray circles** and **gray italicized variable labels**.

In this example, the mean for **big** is significantly different from the mean for **small**, but is not significantly different from the mean for **medium**.

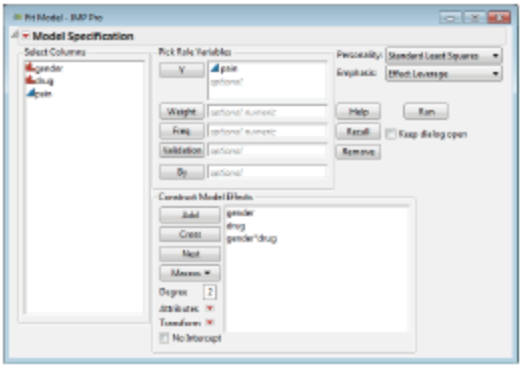


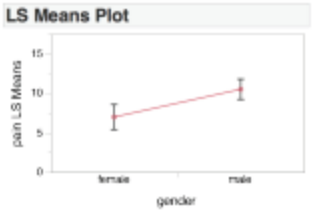
Two-Way (Factorial) ANOVA

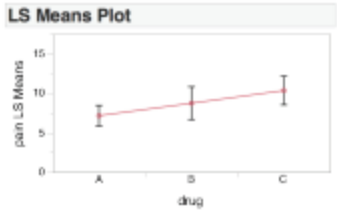
A two-way (factorial) analysis of variance tests the effects of two categorical variables (factors) and their interaction on one continuous (response) variable.

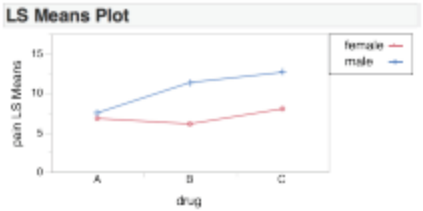
1. From an open JMP® data table, select **Analyze > Fit Model**.
2. Click on a continuous variable from **Select Columns**, and click **Y, Response** (continuous variables have blue triangles).
3. Click on two categorical variables from **Select Columns**, and click **Macros, Full Factorial** (categorical variables have red or green bars). This adds each factor and the interaction between the two factors as model effects.
4. Click **OK**. The Fit Model output window will display.
5. Above the leverage plots select **LS Means Plot** from the **red triangles** to display least square means plots.

Analgesics.jmp (Help > Sample Data Library)









Interpretation of the results in the ANOVA table under Effects Tests:

- The null hypothesis for a main effect is that there are no differences between the population means (i.e., all means are equal) in that factor, averaging over all other factors.
- The null hypothesis for the interaction between two effects is that the pattern of effects for one of the factors does not depend on the level of the second factor.
- **Prob > F** is the p-value useful for testing whether a particular source explains more variation in the data than would be expected by chance. Using alpha = 0.05:
  - Both main effects are significant, indicating that the mean for males differs from the mean for females, and that not all the means for the three drugs are the same.
  - We do not have evidence that the effect of drug depends on the gender of an individual, and equivalently, that the "effect" of gender does not depend on what drug someone is taking.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
gender	1	1	73.808295	12.6378	0.0014*
drug	2	2	51.058196	4.3713	0.0227*
gender*drug	2	2	30.542763	2.6148	0.0916

Tips:

- To determine which means are different (simple effects), a post hoc multiple comparison technique can be used (for details see the page **One-Way ANOVA**).
- The **Parameter Estimates** table provides results from tests of the parameterized (dummy) variables accounting for each source of variation (factors and interactions).

Notes: For more information on two-way analysis of variance, search for **Two-Way** in the book ***Fitting Linear Models*** (under **Help > Books**).

Simple Linear Regression

Simple linear regression is used to model the relationship between two continuous variables.

Simple Linear Regression Using Fit Y by X

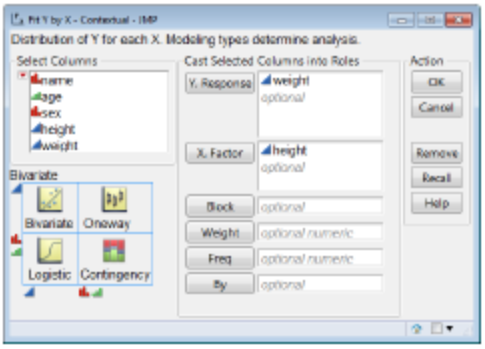
1. From an open JMP® data table, select **Analyze > Fit Y by X**.
2. Click on a continuous variable from **Select Columns**, and click **Y, Response** (continuous variables have blue triangles).
3. Select a second continuous variable, and click **X, Factor**.
4. Click **OK** to generate a scatterplot.
5. To fit a regression line, click on the **red triangle** and select **Fit Line**.

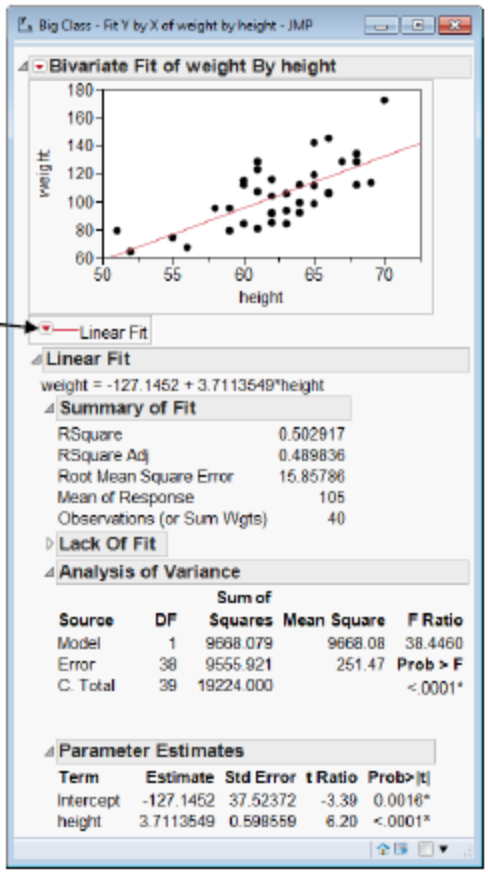
   By default, JMP will provide the following results:

   - The regression equation (under Linear Fit).
   - The Summary of Fit.
   - Lack of Fit (if the data table includes replicates of X values).
   - The ANOVA table.
   - The parameter estimates.

   Additional options, such as **residual plots** and **confidence curves**, are available from the **red triangle** next to **Linear Fit** (directly under the graph).

Big Class.jmp (Help > Sample Data Library)





Tips:

- For other fit options, such as **polynomial**, **transformation** (fit special) and **spline** (under flexible), use the **top red triangle**.
- To add a legend, change markers, or make other changes to the graphical display, right-click on the graph.
- To fit separate lines for categories of a grouping variable, click on the **top red triangle**, select **Group By**, and choose a grouping variable. Then, click on the **top red triangle** and select **Fit Line**.

  JMP will fit separate lines and provide results for each level of the grouping variable.

Notes: Simple linear regression can also be performed from **Analyze > Fit Model**. For more details on regression analysis, see the book ***Basic Analysis*** (under **Help > Books**) or search for "regression" in the JMP Help.

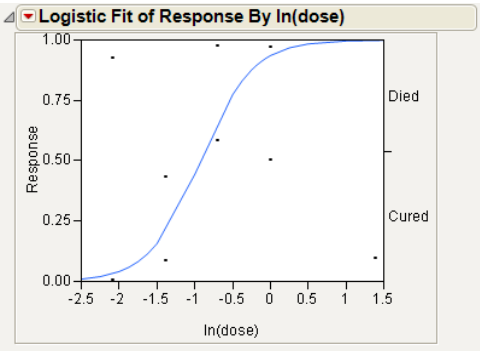
Example of Nominal Logistic Regression

This example uses the Penicillin.jmp sample data table. The data in this example comes from an experiment where 5 groups, each containing 12 rabbits, were injected with streptococcus bacteria. Once the rabbits were confirmed to have the bacteria in their system, they were given different doses of penicillin. You want to find out whether the natural log (ln(dose)) of dosage amounts has any effect on whether the rabbits are cured.

1. Select Help > Sample Data Library and open Penicillin.jmp.
2. Select Analyze > Fit Y by X.
3. Select Response and click Y, Response.
4. Select ln(Dose) and click X, Factor.

Notice that JMP automatically fills in Count for Freq. Count was previously assigned the role of Freq.

5. Click OK.



Freq: Count

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	18.968295	1	37.93659	<.0001*
Full	18.313369			
Reduced	37.281664			

RSquare (U)	0.5088
AICc	40.862
BIC	44.6047
Observations (or Sum Wgts)	54

Measure	Training	Definition
Entropy RSquare	0.5088	1-Loglike(model)/LogLike(null)
Generalized R-Square	0.5047	1-(Like(null)/Like(model))^(2/n)
Mean -Log p	0.3391	∑ -Log(p[j])/n
RMSE	0.3298	√∑(y[j]-p[j])²/n
Mean Abs Dev	0.2149	∑ |y[j]-p[j]|/n
Misclassification Rate	0.1481	∑ (p[j]≠pMax)/n
N	54	n

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	2.68517103	0.8253384	10.58	0.0011*
ln(dose)	2.9096072	0.7682096	14.35	0.0002*

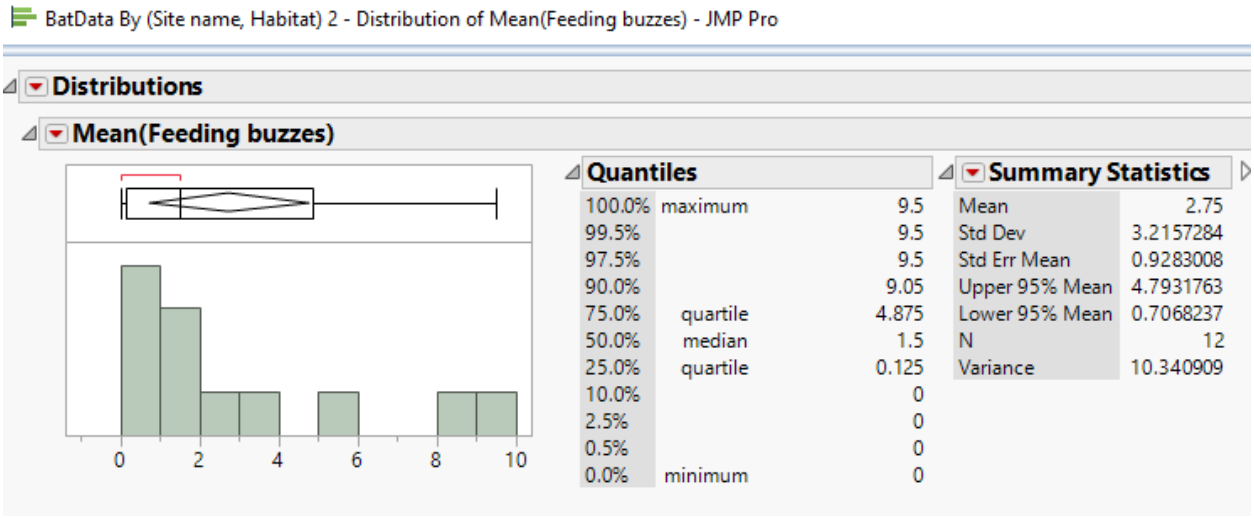
For log odds of Cured/Died

Covariance of Estimates

The plot shows the fitted model, which is the predicted probability of being cured, as a function of ln(dose). The p-value is significant, indicating that the dosage amounts have a significant effect on whether the rabbits are cured.

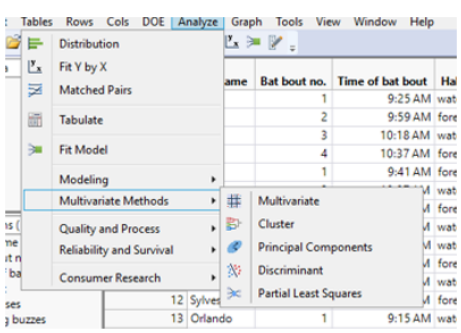
Non-parametric Tests and Generalized Linear Models

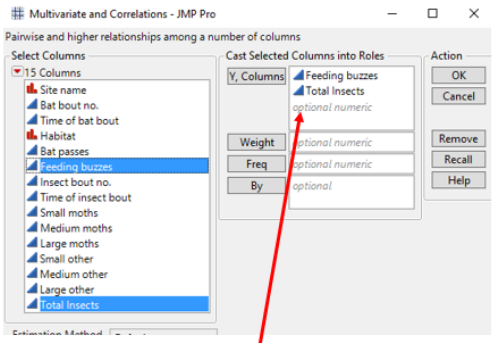
The following are data from Bio 22. Here the feeding buzz data are not normal, and summarizing the data over sites has not changed that feature of the data. Various transformations were tried, none of which improved normality.



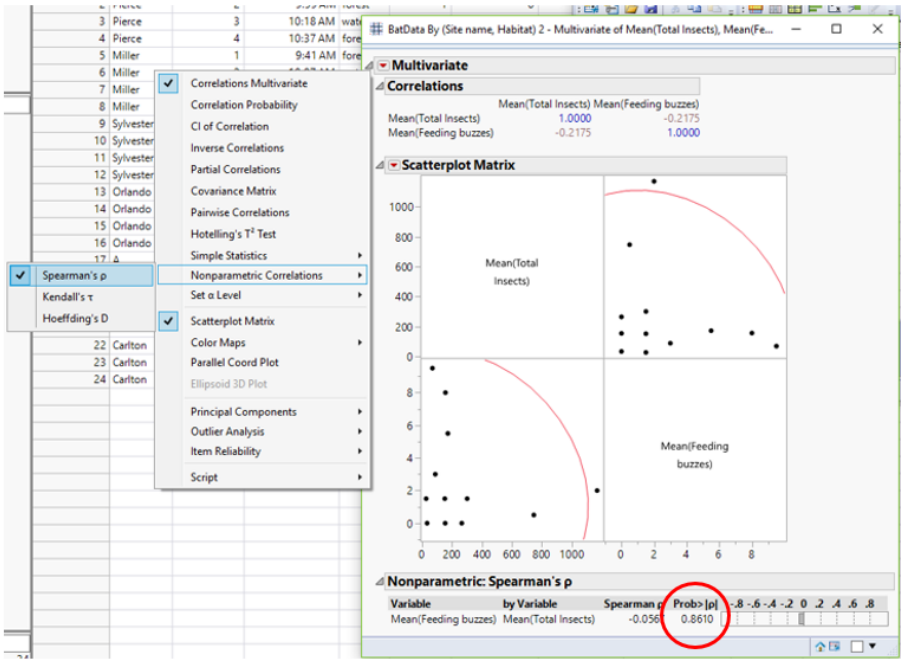
Note two things about these data: 1) there are a relatively large number of zeros; and 2) the variance is larger than the mean. When the latter condition is true, the data are "overdispersed."

There are two ways to deal with this problem. The first is to use a **non-parametric test**, such as Spearman's rank correlation (for continuous Y, continuous X). Get there in JMP using Analyze / Multivariate Methods / Multivariate.



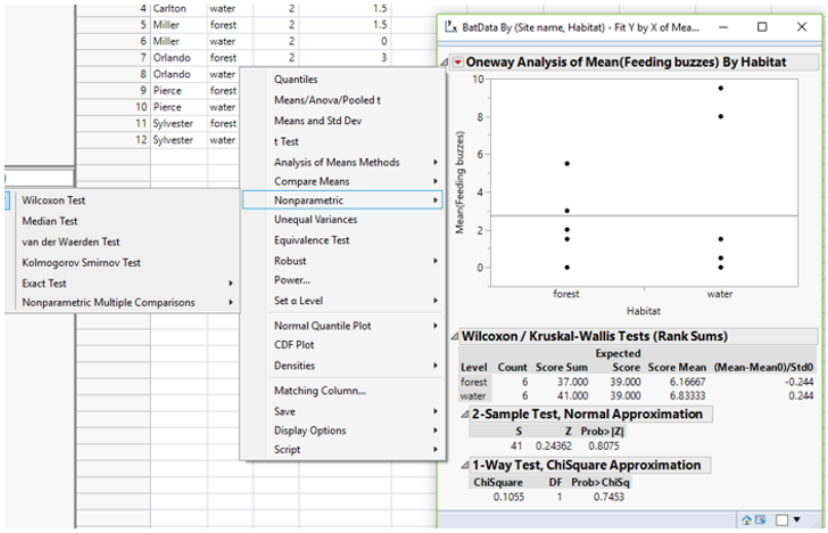


Let's say we want to see if feeding buzzes is related to total number of insects. In JMP, you now treat them as though they are both "y variables," moving both to the box on the right.

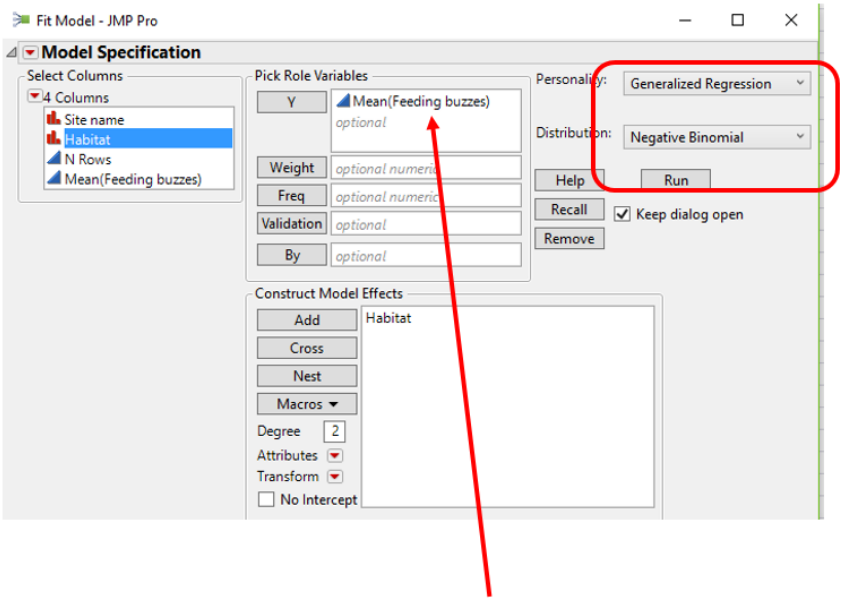


Then from the red triangle at the top of the resulting window, go to "Nonparametric Correlations" and choose "Spearman's ρ" to get the p-value.

To get the nonparametric version of a oneway ANOVA, do "Fit Y by X" as usual, then select "Nonparametric / Wilcoxon" instead of the usual "Means/Anova/Pooled t" that you've been using so far.

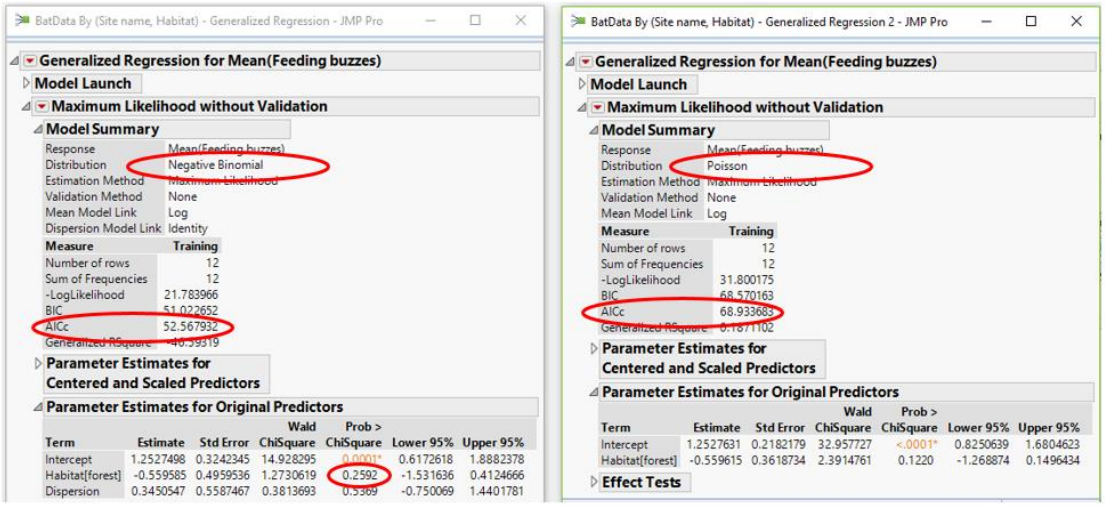


The second way to deal with non-normal data that are overdispersed is to try using a different distribution. This is called a "**generalized linear model**." (GLM) Opinions vary, but GLMs are becoming the norm for this kind of data. Advances in statistical software make the use of alternative distributions much more accessible for those who are not professional statisticians!



Untransformed feeding buzzes

In the next window, under "Estimation Method," select "Maximum Likelihood," then click "Go."



The two possible distributions for overdispersed data are "Poisson" and "negative binomial." A value called "AICc" tells you which model fits better (smaller AICc is better). In any event, habitat is not significant.

How do you know which distribution to choose?

You should not try this method without first trying to transform your non-normal data! If transformation does not work, ***then*** proceed to these other distribution methods. In this class, you probably won't need more than the following distributions (which are all for *discrete* data; we hope that most continuous data [in the general sense, not the JMP sense] you encounter in this class can be transformed successfully):

Counts of things (numbers of fish, flowers, etc.) → **Poisson**

Two-way outcomes (yes/no, presence/absence, etc.) → **Binomial**

Counts of things that are overdispersed → **Negative binomial**

More details, for those who are interested, are on the next page, but if you are new to stats, the information on this page more than suffices.

Note that the presence of a large number of zeros (as we saw earlier) is a special case. Ask for help if this comes up. Sometimes it can be taken care of just by using the negative binomial distribution, as we did above, but otherwise you'll need other methods.

Two-Factor Nested Random Effects Model

Consider a model with two factors, A and B, but where B is nested within A. Although there are situations where a nested effect is treated as a fixed effect, in most situations a nested effect is treated as a random effect. For this reason, in the model described below, the nested effect is entered as a random effect.

Effects to be entered: A, B[A]&Random

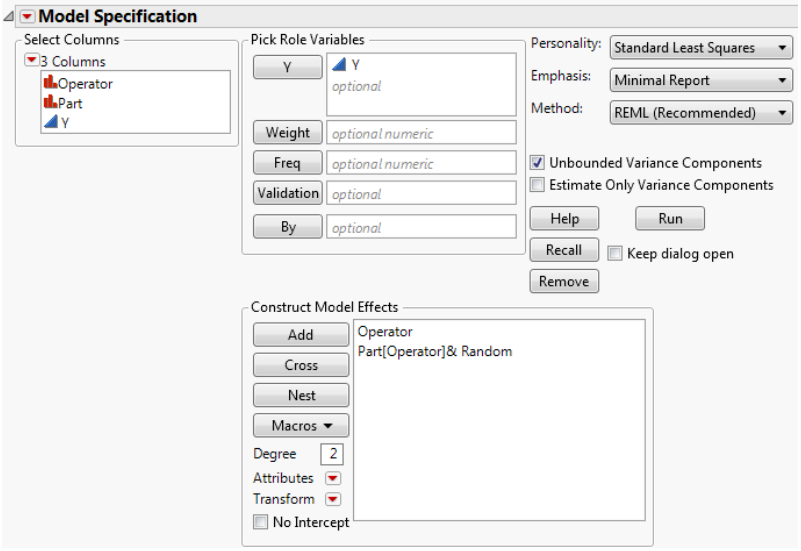
1. In the Select Columns list, select two nominal or ordinal effects, A and B.
2. Click Add.
3. To nest B within A: In the Construct Model Effects list, select B. In the Select Columns list, select A. The two effects should be highlighted.
4. Click Nest.
5. With B[A] highlighted in the Construct Model Effects list, select Attributes > Random Effect.

Example of Two-Factor Nested Random Effects Model

Open the 2 Factors Nested.jmp sample data table located in the Variability Data subfolder. As part of a measurement systems analysis study, 24 randomly chosen parts are measured. These parts are evenly divided among the six operators who typically measure these parts. Each operator make three independent measurements of each of the four assigned parts.

Since the parts measured by one operator are measured only by that specific operator, Part is nested within Operator. Since the parts are a random sample of production, Part is considered a random effect. Since these specific six operators are of interest, Operator is treated as a fixed effect. The appropriate model is specified as follows.

1. Select Analyze > Fit Model.
2. In the Select Columns list, select Y and click Y.
3. In the Select Columns list, select Operator and Part.
4. Click Add.
5. To nest Part within Operator: In the Construct Model Effects list, select Part. In the Select Columns list, select Operator. The two effects should be highlighted.
6. Click Nest.
7. With Part[Operator] highlighted in the Construct Model Effects list, select Attributes > Random Effect.



8. Click Run to obtain the Fit Least Squares report.

Model Fit for Two-Factor Nested Random Effects Model (below) shows two plots. The first is a Variability Chart showing the three measurements by each Operator on each of the four parts. Horizontal line segments show the mean measurement for each Operator.

To construct the Variability Chart, in the 2 Factors Nested.jmp sample data table, run the data table script Variability Chart - Nested. From the report's red triangle menu, deselect Show Range Bars and select Show Group Means.

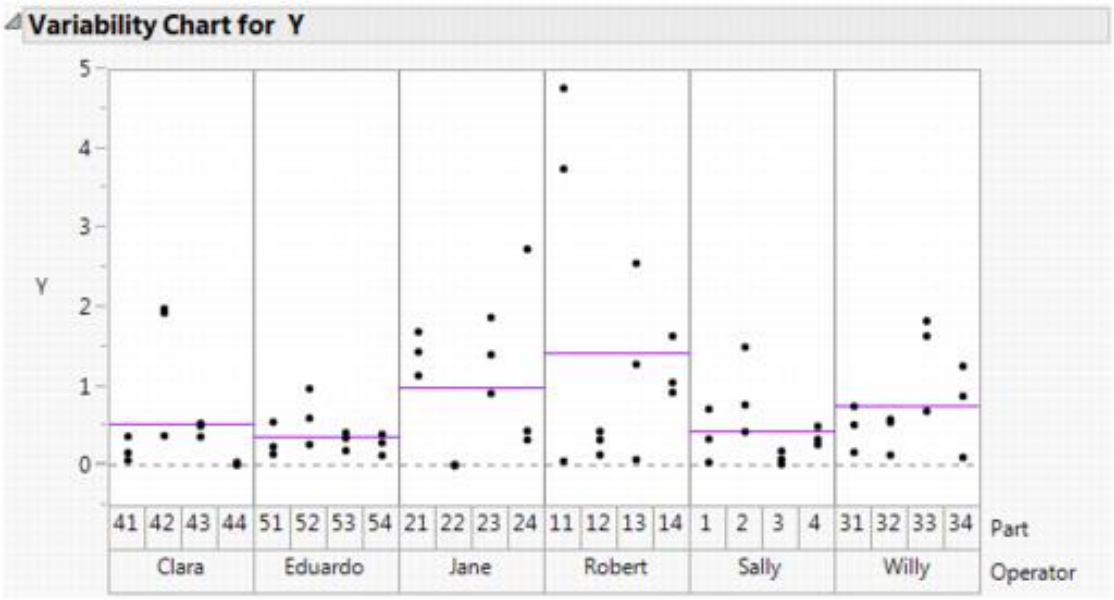
The second plot is the Fit Least Squares report Prediction Profiler plot for Operator. This plot shows the predicted response for each operator. The vertical dashed red line set at Jane indicates that Jane's predicted response is 0.997. You can see the correspondence between the model predictions given in the Prediction Profiler plot and the raw data in the Variability Chart.

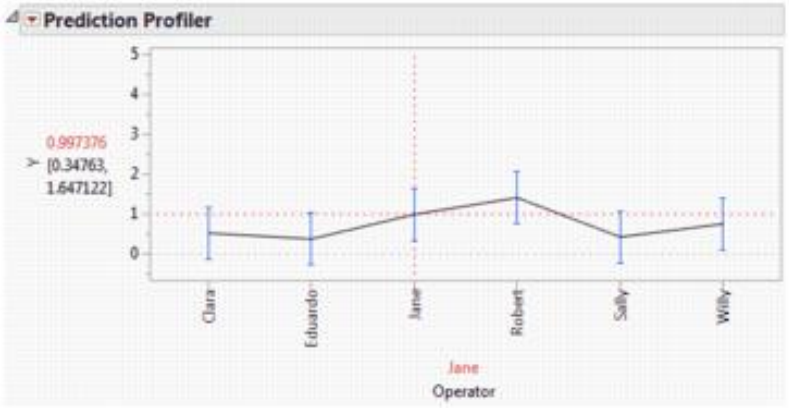
To obtain the Prediction Profiler plot, from the Fit Least Squares report red triangle menu, select Factor Profiling > Profiler.

These plots show how the predicted measurements for each Operator are modeled. However, keep in mind that you are not only interested in whether the operators differ in how they measure parts. You are also interested in the variability of the part measurements themselves, which requires estimation of the variance component associated with Part.

Model Fit for Two-Factor Nested Random Effects Model

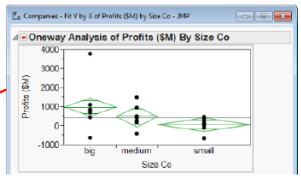
https://www.jmp.com/support/help/13-1/Two-Factor_Nested_Random_Effects_Model.shtml#228166



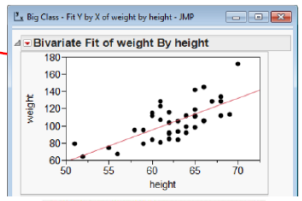


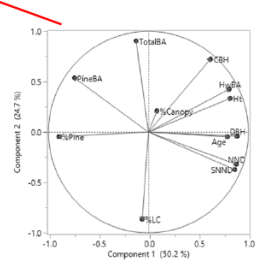
Analysis Methods Cheatsheets

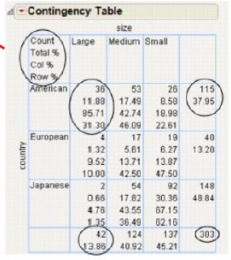
Most of the time, your data will fall into one of the categories below:



- Single continuous y, categorical x ➡ ANOVA
- Single continuous y, continuous x ➡ regression
- Single categorical y, categorical x ➡ contingency tables (chi-square analysis)
- Many y ➡ multivariate analysis (e.g., species occurrence across plots)







A brief cheatsheet: Analysis methods for categorical X, continuous Y and how to get there via JMP menus

Comparing a single mean (e.g., to a known quantity) OR
Comparing matched pairs in a single group
- One-sample t-test

➡ Analyze → Distribution

Comparing two means (e.g., fer de lance vs. bushmasters)
- Two-sample t-test

➡ Analyze → Fit Y by X

Comparing three or more means and a single X
- One-way ANOVA

➡ Analyze → Fit Y by X
(Fit Model also works)

Comparing three or more means and two X variables
- Two-way ANOVA (watch out for interactions)

➡ Analyze → Fit Model

For the most part, these menus in JMP are all you will ever need. These are the same menus you use to perform other forms of analysis, including simple linear regression, generalized linear models, multiple regression, mixed effects models, and more!

Rev. Sept 2017, Aoki

And as it happens, you use the same menus for continuous x, continuous y

Comparing one y against one x

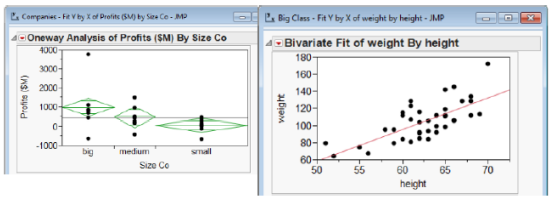
- Simple linear regression

→ Analyze → Fit Y by X

Comparing one y against two or more x variables

- Multiple regression

→ Analyze → Fit model



Again, note that the same menus are used for ANOVA (categorical X) and regression (continuous X) models. Whenever you have one X and one Y (regardless of the type of variable), you use Fit Y by X. Whenever you have two or more Xs and one Y, you use Fit Model. JMP can tell from the variable type whether you want an ANOVA or a regression.

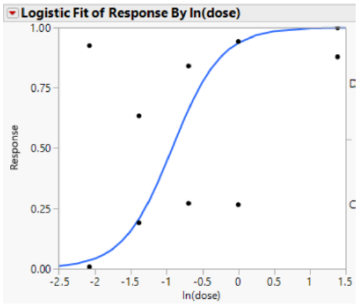
There is also a special case of regression called logistic regression

Comparing one **categorical y with only two possible levels** (yes/no; presence absence) against one continuous x

- Logistic regression

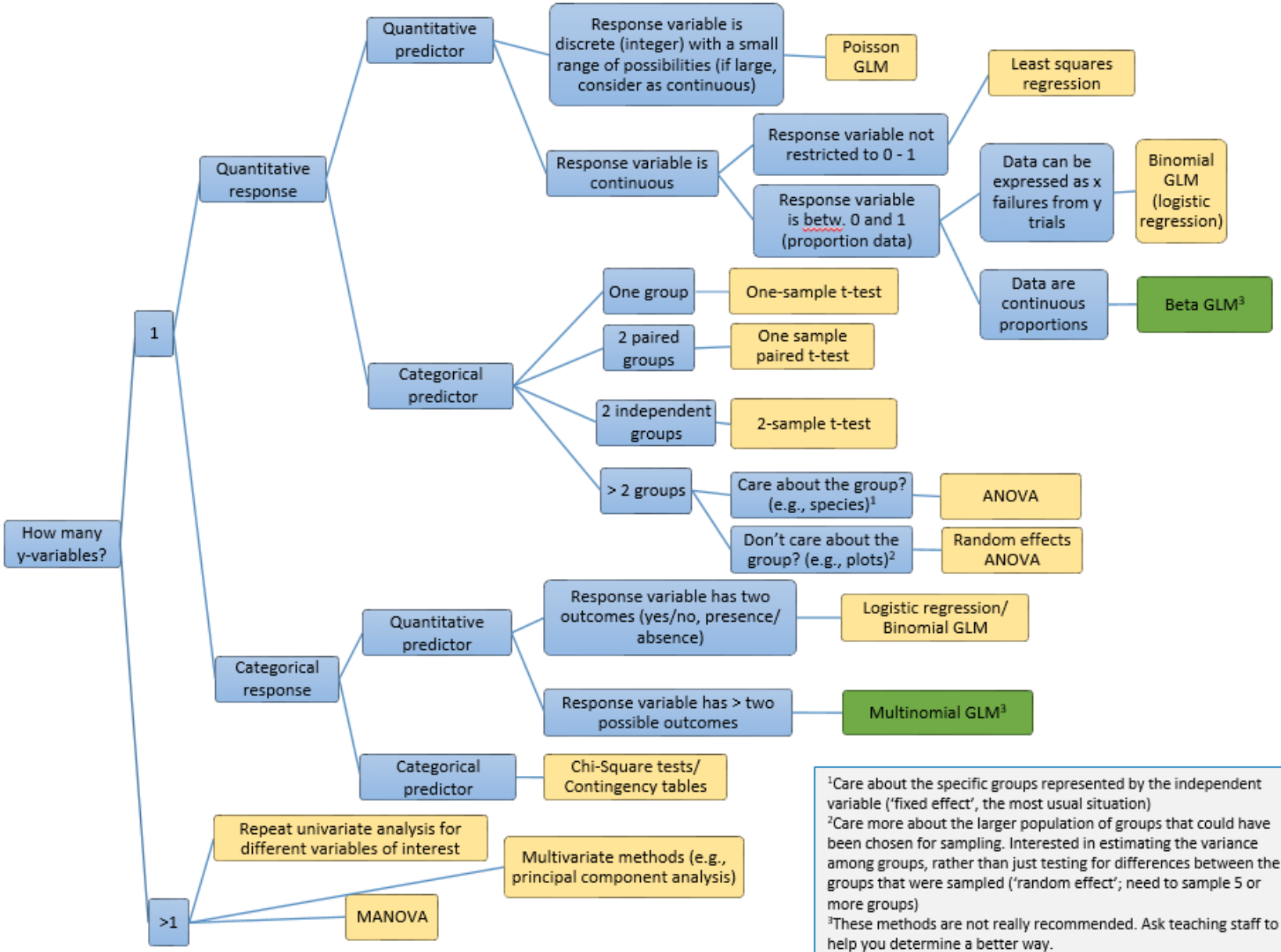
→ Analyze → Fit Y by X
(Fit model also works)

- NOTE: Logistic regression can also be used on proportional response data (between 0 and 1).



In addition to two-way ANOVA and multiple regression, there are other ways to perform analyses involving two or more x variables at a time, and even for these variables to be mixed (some continuous, some categorical). For the moment, it's less important that you know the names of these analyses by heart. In almost all cases, you can just use Fit Model as usual. Ask teaching staff for help interpreting your output.

Analysis Methods Cheatsheet (Advanced)



[1]Care about the specific groups represented by the independent variable ('fixed effect', the most usual situation)
[2]Care more about the larger population of groups that could have been chosen for sampling. Interested in estimating the variance among groups, rather than just testing for differences between the groups that were sampled ('random effect'; need to sample 5 or more groups)
[3]These methods are not really recommended. Ask teaching staff to help you determine a better way.

Rev. Sept 2017, Aoki