Prof. David Draper
Department of
Applied Mathematics and Statistics
University of California, Santa Cruz

AMS 7: Homework 3

Due date: Thu 22 Feb 2007 in class *[100 total points]*

1. (biology *[35 points]*) In a study of mating calls in the tree toad *Hyla ewingi*, Littlejohn (1965) found the note duration of the call in a representative sample of 39 observations (one from each of 39 toads) from Tasmania to have a mean of 189 ms and a standard deviation of 32 ms; the sample histogram followed the normal curve pretty well.

(a) Set up a statistical model for this situation, being explicit about the population, sample and imaginary data sets, and use your model (including the usual inferential summary) to build a 95% confidence interval for the population mean μ (what *is* the population here, precisely?). *[15 points]*

(b) Other scientists working in south-eastern Australia had previously found that the mean mating call duration for this species was 210 ms.

(i) Does the difference between the Tasmania and Australia findings seem large to you in practical terms? Explain briefly. *[5 points]*

(ii) Is this difference large in statistical terms? Answer this question both by reference to your confidence interval in (a) and by calculating the P value for testing the hypothesis that the mean in Tasmania is 210 ms against the alternative that it's not. *[10 points]*

(c) When gathering data of this type, which of the following designs do you think would give the most information about mating calls for Tasmanian tree toads, or would they be equally informative, or would one design be better for some kinds of information and another design for other kinds (specify briefly in each case)? Explain briefly. *[5 points]*

(i) One observation from each of 39 toads;

(ii) 39 observations from a single toad;

(iii) 5 observations from each of 7 toads and 4 from an eighth toad.

2. (biology *[20 points]*) Continuing problem 1, run the clock back in your mind to the point where Dr. Littlejohn is planning his data-gathering. No one had ever taken data on the mating call duration of this particular species of tree toad in Tasmania before, so he doesn't know what to expect for the mean, but he thinks that either it's around 210 ms (the published Australian value) or it's smaller than that (because he theorizes that the Tasmanian animals may have found an evolutionary advantage in making themselves less noticeable to their predators by having shorter mating calls). The published value for the standard deviation of mating calls for this tree frog in Australia was 36 ms.

(a) Using the confidence interval approach to sample size determination, if his alternative theory is that the mean duration in Tasmania will be 195 ms, how many observations should he plan to take at the 95% confidence level to show a statistically significant difference? Explain briefly. *[10 points]*

(b) If instead he thinks the Tasmanian mean will be 190 ms, and if instead he uses the hypothesis-testing approach to sample size determination, what value of n should he plan to achieve to have 90% power with a 5% significance level? Explain briefly. *[10 points]*

Note: We'll soon learn how to get `JMP` to do some of these calculations, so it's obviously not enough just for you to do them in `JMP` for the homework; you need to show your work in correctly using the proper formula.

3. (environmental sciences *[30 points]*) Sokal and Rohlf (1995) report on results of the following experiment with acacia trees and ants. All but 28 trees of two species (let's call them A and B) of acacia were cleared from an area in Central America. There were 14 trees of each species, all of which had been freed from ants by insecticide. Next, 8 separate colonies of a particular species of ant (X, say), obtained from cut-down acacia trees of species A in an area nearby, were situated roughly equidistant from the 28 trees and allowed to invade them. In 6 of the 8 cases the ants invaded trees from species A, with the other 2 invading trees from species B.

(a) Set up a statistical model for this situation, being explicit about the population, sample and imaginary data sets, and use your model (including the usual inferential summary) to build an approximate 95% confidence interval for the population proportion p of invasions of species A (what *is* the population here, precisely?). *[10 points]*

(b) If the ants had been indifferent between species A and B, you would have expected 50% invasion for each species.

(i) Does the difference between the observed invasion percentage for species A and what you'd expect under indifference seem large to you in practical terms? Explain briefly. *[5 points]*

(ii) Is this difference large in statistical terms? Answer this question both by reference to your confidence interval in (a) and by calculating the P value for testing the hypothesis that the population proportion is 0.5 against the alternative that it's not. *[10 points]*

(c) Your results in (b) should have indicated that the researchers who ran this experiment didn't plan ahead correctly to ensure that the notions of practical and statistical significance were in rough correspondence. Using the confidence interval approach to sample size determination, if (ahead of time) you were expecting 50% of the ant colonies to prefer species A, approximately how large should n have been to demonstrate that an observed proportion of 75% was statistically significant? Explain briefly. *[5 points]*

4. (environmental sciences *[15 points]*) Zar (1999) reports on the results of a study comparing concentrations of nitrogen oxides and hydrocarbons (in $\mu g/m^3$) in an urban area that shall remain nameless (well, Zar forgot to name it, anyway). Both pollutants were measured on a representative sample of 11 days, with the following results:

Day	Nitrogen Oxides (NO)	Hydrocarbons (H)	Difference (H – NO)
1	104	108	+4
2	116	118	+2
3	84	89	+5
4	77	71	–6
5	61	66	+5
6	84	83	–1
7	81	88	+7
8	72	76	+4
9	61	68	+7
10	97	96	–1
11	84	81	–3
Mean	83.7	85.8	+2.1
SD	16.9	16.4	4.3

(a) Does the mean difference in pollutant concentration seem large to you in practical terms? Explain briefly. *[5 points]*

(b) Set up a statistical model for this situation, being explicit about the population, sample and imaginary data sets, and use your model (including the usual inferential summary) to build a 95% confidence interval for the population mean difference in pollutants (what *is* the population here, precisely?). Is this difference large in statistical terms? Explain briefly. *[10 points]*