1. (5 pts) Name 2 shortcomings of pie charts discussed in Week 7.

2. (5 pts) Finish the following sentence: A 90th percentile confidence interval is...

3. (10 pts) Null Hypothesis Significance Testing is based on the likelihood of the data conditional on the null hypothesis. For a one-sample t-test, this would be written as $P(data|\mu = c)$ where c is the value of the parameter μ under the null hypothesis H_0. How would you calculate the marginal probability of obtaining your data? (I am looking for a mathematical expression. Partial credit for defining "marginal probability".)

4. (5 pts) Sketch the probability density function for the $Beta(\alpha = 3, \beta = 3)$ distribution. (Your sketch is not going to be exact, but should contain some key features.)

5. (15 pts; 5 pts each) A researcher is measuring the volume of humpback whales (in m^3) in multiple ocean regions over several years, and is interested in the relationship between the change in volume over the course of the summer (Y_i) and habitat characteristics such as average sea surface temperature (SST; in Celsius), estimated density (in kg/hectare) of preferred prey in the region, the length of the individual at the start of the summer (in m), and the demographic group (adult male, adult female, juvenile male, juvenile female) to which the whale belongs.

The researcher's model looks like ($Y_i; i = 1, \dots, n$)

$$Y_i \sim N(\beta_0 + \beta_1 SST_i + \beta_2 PreyDensity_i + \beta_3 StartingLength_i + \beta_4 Group_i, \sigma^2)$$

(a) What is the Mean Squared Error (MSE) for this model?

(b) What are the units of β_2?

(c) What are the units of σ^2?

6. (10 pts) Assume that you have data

$$X = \{X_1, X_2, X_3, \dots, X_{100}\}$$

representing the petal length of a population of flowering plants, which you assume are drawn from a Normal distribution, i.e.

$$X \sim N(\mu, \sigma^2)$$

Using your data, you calculate that

$$\frac{1}{n}\sum_{i=1}^{100} X_i = 3.2\ cm$$

and

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{100}(X_i - \bar{X})^2} = 0.8\ cm$$

a) Approximately what proportion of the population would be expected to have a petal length greater than 4.5 cm? (5 pts)

b) What is the standard error on the statistic $\bar{X}$? (5 pts)

7. (20 pts)

a) Fill in the three empty boxes. (12 pts)

Test	Hypothesis (assuming two-tailed tests)	Test statistic T	f(T|H$_0$) (Distribution of T under H$_0$)	Assumptions
Two sample unpaired t-test (equal variance)	$H_0: \mu_A = \mu_B$ $H_A: \mu_A \neq \mu_B$			

b) Under what conditions is the unpaired two-sample t-test sensitive to violations of its assumptions? (10 pts)

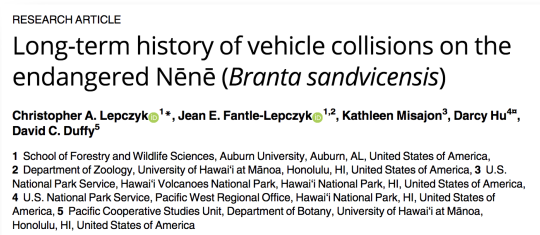
8. (20 pts)

a) Using a Bonferroni Correction for multiple comparison, what should the per-comparison rate α' be set to to maintain a family-wise error rate of α if we are doing k comparisons? What about if we use the Dunn-Šidák's correction? (10 pts)

b) What assumption does each correction make regarding the hypothesis tests involved? (5 pts)

c) Which test is more conservative and why? [be sure to define conservative] (5 pts)

9. (52 pts) Lepczyk et al. (2019) studied vehicle collisions on the endangered Nēnē in Hawaii.



RESEARCH ARTICLE

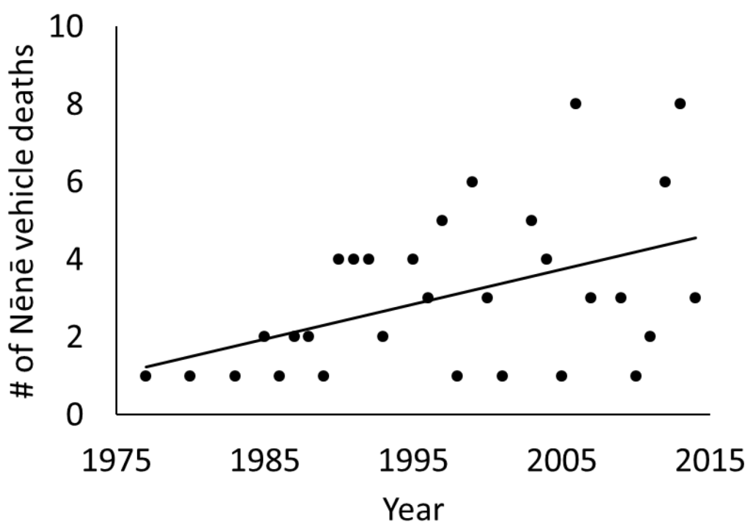
Long-term history of vehicle collisions on the endangered Nēnē (*Branta sandvicensis*)

Christopher A. Lepczyk[1]*, Jean E. Fantle-Lepczyk[1,2], Kathleen Misajon[3], Darcy Hu[4¤], David C. Duffy[5]

1 School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL, United States of America, **2** Department of Zoology, University of Hawai'i at Mānoa, Honolulu, HI, United States of America, **3** U.S. National Park Service, Hawai'i Volcanoes National Park, Hawai'i National Park, HI, United States of America, **4** U.S. National Park Service, Pacific West Regional Office, Hawai'i National Park, HI, United States of America, **5** Pacific Cooperative Studies Unit, Department of Botany, University of Hawai'i at Mānoa, Honolulu, HI, United States of America

Exhibit A: Behold, the endangered Nēnē.

a) In their analysis, Lepczyk et al. used ordinary least squares regression to model the data in the following figure. What assumption of regression did they violate? (5 pts)

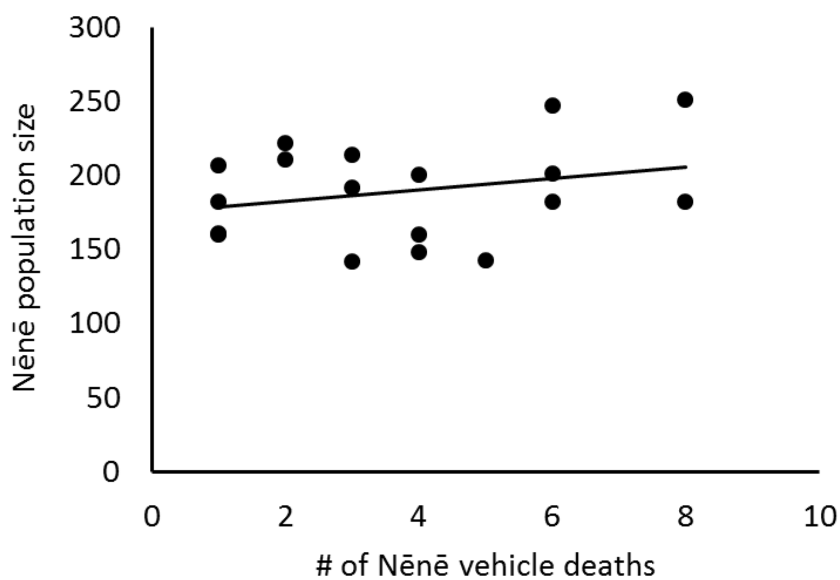


b) What would be a better method of modelling these data? Name two reasons why this approach would be better than the one the authors used. (10 pts)

c) In the figure above, circle one data point that has high leverage. (2 pts)

d) Does the point you circled have high influence? Why or why not? (5 pts)

e) Elsewhere in the manuscript, Lepczyk et al. examine data on Nēnē population size and the number of vehicle deaths. Indicate on the figure below <u>two</u> aspects that violate the best practices of data visualization discussed in Week 7. (5 pts)



f) For the data plotted above, is correlation or regression a more appropriate analysis? Why? (10 pts)

g) If the authors were to calculate Pearson's product moment correlation for the data on population size and vehicle deaths, how would they calculate the correlation coefficient r? (In other words, what is the formula.) (5 pts)

h) What would be the distribution of the test statistic used to determine if r was statistically significantly different from 0? Here I am looking for the distribution of the test statistic under the null hypothesis, not the test statistic itself. Full credit will require calculating the parameter(s) for the distribution given the information provided. (5 pts)

i) What assumption of Pearson's product moment correlation coefficient is violated here? Name one alternative the authors might have used. (10 pts)

10. (55 pts)
This question will focus on the analysis completed for the following paper:

> RESEARCH ARTICLE
>
> **Same law, diverging practice: Comparative analysis of Endangered Species Act consultations by two federal agencies**
>
> **Megan Evansen[1,2]*, Ya-Wei Li[2,3], Jacob Malcom[2]**
>
> **1** Sustainable Development and Conservation Biology Graduate Program, College Park, MD, United States of America, **2** Center for Conservation Innovation, Defenders of Wildlife, Washington, DC, United States of America, **3** Environmental Policy Innovation Center, Washington, DC, United States of America

In this paper, the authors use a Generalized Linear Model to model the completeness of an Endangered Species Act (ESA) listing, as described below:

> For formal consultations, we selected the four core sections from the Handbook to score the completeness of each biological opinion: "Status of the Species," "Environmental Baseline," "Effects of the Action," and "Cumulative Effects." Although not an exhaustive list of biological opinion sections, these four sections contain the bulk of the information and analysis of the species and proposed action. The Status of the Species and Environmental Baseline sections received a score from 0–5 and the Effects of the Action and Cumulative Effects sections were given a score from 0–2 based on how well they met the specific requirements for that section by the Handbook. Rating the completeness of these core sections of the biological opinion was straightforward because the criteria described by the Handbook allowed for a simple present/absent scoring system. For some analyses, these present/absent scores were summed for each of the four core sections.

In short, completeness across all sections was recorded as a discrete number from 0 to 14. The model is described more fully below (irrelevant sections have been redacted for simplicity).

Our goal was to understand patterns and associations of variation in consultation completeness. We used summary statistics (mean and standard deviation) and Pearson's correlations to describe patterns. To examine relationships between completeness and associated factors, we used two modeling approaches: a binomial generalized linear model [GLM; 11] to identify predictors of the proportions of total possible points,

We considered six variables that were most likely to affect consultation completeness: the Service performing the consultation, whether the consultation was formal or informal, the year the consultation took place, the species of sea turtle assessed, the type of action assessed, and whether the consultation was part of a programmatic consultation (see Glossary). We incorporated these variables into a global model (Model 1) of all variables and eight additional subset candidate models for the analysis of overall completeness using the GLM (Table 1). We also considered that the particular office within the Service might be an important predictor of consultation completeness. However, given that our focus is on the potential differences between the Services and that the offices are nested within the Services, the office variable was not included in our candidate model set.

we calculated the response variable as the proportion of possible points for each consultation.

The models considered are listed in the table below. All effects are considered purely additive, no interactions.

Model Type	Model Num.	Predictors
GLM Binom*	1	Service + Formal + Year + Action_type + Programmatic + total_duration
	2	Service + Formal + Year + Programmatic + total_duration
	3	Service + Formal + Year + Action_type + total_duration
	4	Service + Formal + Year + total_duration
	5	Service + Formal
	6	Service
	7	Formal
	8	total_duration
	9	Service + Formal + Programmatic + total_duration

More information on the 6 model covariates: **Service**: Binary (FWS or NMFS), **Formal**: Binary (Yes or No), **Year**: Continuous, **Action_type**: Discrete (17 action types, examples include "Install fishing pier", "Beach nourishment", Maintenance dredging", etc.), **Programmatic**: Binary (Yes or No), **total_duration**: Continuous.

a) Write out the full model equation for Model 4. (10 pts)

b) How many parameters does this model have? (5 pts)

c) Name one pair of models in the table that could be tested by a Likelihood Ratio Test and one pair of models that could not be tested using a Likelihood Ratio Test. (10 pts)

d) Why might someone be skeptical about their use of a Binomial Generalized Linear Model to model the response variable of completeness? What are they assuming? (5 pts)

e) Given that the response variable is a measure of completeness, do you think Action_type should be modelled as a random effect or as a fixed effect (and why)? (5 pts)

f) The authors do not use "Office" as a Covariate because it is nested within "Service". Why would this be a problem for modelling? (10 pts)

g) The authors describe their model selection procedure as follows:

> We carried out model selection [14] based on Akaike's Information Criterion adjusted for small sample sizes (AIC_C) using the AICcmodavg package [15]. We considered models with $\Delta AIC_c >$ 2.0 as having strong support [14]. All analyses were done in R 3.3 [16] and are available as a package vignette in the project's OSF repository (https://doi.org/10.17605/OSF.IO/KAJUQ).

What mistake do they make here? (5 pts)

h) Why do the authors use the phrase "strong support" in this paragraph? (5 pts)

11. (90 pts)
Consider the following R code, which generates two groups of random draws. The x variable is a vector denoting the group (group 1 vs. group 2) and the y variable is a vector of the response data. A linear model is fit to these data, followed by an ANOVA.

```
> x<-c(rep(1,times=10),rep(2,times=10))
> x<-as.factor(x)
> y<-c(rnorm(n=10,mean=0,sd=2),rnorm(n=10,mean=1,sd=2))
> fit<-lm(y~x)
> summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.5762 -1.2758  0.7781  1.4756  3.1000 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6010     0.7454   0.806    0.431
x             0.4954     1.0542   0.470    0.644

Residual standard error: 2.357 on 18 degrees of freedom
Multiple R-squared:  0.01212,	Adjusted R-squared:  -0.04276 
F-statistic: 0.2209 on 1 and 18 DF,  p-value: 0.644

> anova(fit)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
x          1   1.227  1.2272  0.2209  0.644
Residuals 18 100.015  5.5564
```

a) Write the full equation describing this linear model. (5 pts)

b) What kind of contrast is represented in this model? (5 pts)

c) State the two null hypotheses being tested. (5 pts)

d) If the second line of code had not been included, what would the two estimated parameters have been (approximately)? In words, state the interpretation of the two parameters in this scenario (in which line 2 had been deleted). (15 pts)

Given the code provided (including Line 2) for the following questions, what is the R code to calculate the following quantities (Hint: your answer will involve the expression "residuals(fit)"):

e) the mean squared error (5 pts):

f) the root mean squared error (5 pts):

g) the residual sum of squares (also known as the sum of squared error) (5 pts):

h) the residual standard error (5 pts):

i) Which of the above four quantities are indicated by the R output provided? (There may be more than one.) What is the value of those quantities? (10 pts)

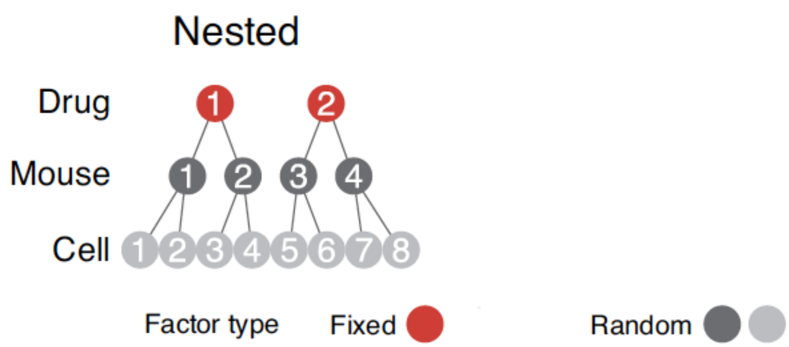
j) Why is the uncertainty of the intercept smaller than the uncertainty of the term for x? You can use words or a combination of words and equations to answer this question. (10 pts)

k) What is R^2 for this model? (I'm looking for a number.) What is the interpretation of R^2 and how is it calculated? (10 pts)

l) Describe a scenario in which R^2 is low but the p-value with the covariate is highly significant. (5 pts)

m) Describe a scenario in which R^2 is high but the p-value with the covariate is not statistically significant. (5 pts)

12. (38 pts)
A student in Biometry wants to make sure she understands nested ANOVA, and she has decided to follow's Prof. Lynch's advice and learn through simulation. Having carefully read the Krzywinski et al. (2014) essay from Points of Significance, this student wants to write some R code to simulate the following nested ANOVA design, involved two drugs (fixed effect), two mice per drug (random), two cells per mouse (random), and three replicate measurements of each cell. She will simulate some data (the response as measured in different cells) assuming the difference between drugs 1 and 2 has an effect of 2.7, and $\sigma^2_{mouse} = 1.1$, $\sigma^2_{cell} = 0.4$, and $\sigma^2_{\varepsilon} = 0.1$.



a) (3 pts) How many data points is involved in this experiment? (This is the same number of total data points the student's code will generate.)

b) (25 pts) Write some code (either actual R code or pseudocode (i.e. an ordered list of steps that could be converted to code), whichever is easier) to generate the data this student needs.

(LEFT BLANK FOR WORK)

c) (5 pts) What R command(s) would you use to fit this ANOVA model. Please include all the inputs needed and make sure your answer matches the naming of variables for part b so this enterprising student is not confused.

d) (5 pts) With only two mice per drug, what is the advantage of treating mouse as a random effect rather than as a fixed effect?