

STAT 645 – Homework 5

Due: March 31

1. Scientists agree that preserving certain habitats in their natural state is necessary to slow the accelerating rate of species extinctions. But, they are undecided on how to construct such reserves. Specifically, is it better to have many small reserves or a few large ones? Observe the following study.

Island	Area (km^2)	Species at risk	Extinctions
Ulkokrunni	185.80	75	5
Maakrunni	105.80	67	3
Ristikari	30.70	66	10
Isonkivenletto	8.50	51	6
Heitakraasukka	4.80	28	3
Kraasukka	4.50	20	4
Lansiletto	4.30	43	8
Pihlajakari	3.60	31	3
Tyni	2.60	28	5
Tasasenletto	1.70	32	6
Raiska	1.20	30	8
Pohjanletto	0.70	20	2
Toro	0.70	31	9
Luusiletto	0.60	16	5
Vatunginletto	0.40	15	7
Vatunginnokka	0.30	33	8
Tiirakari	0.20	40	13
Ristikarenletto	0.07	6	3

Table 1: Island area, number of bird species present in 1949, and number of those not present in 1959; Krunnit Islands study.

Our goal is to fit a binomial regression with logit link for the probability of extinction as a function of Area. Use any software you wish.

- a. Plot the $\text{logit}(\text{proportion of extinction})$ as a function of Area. Why does a linear regression model not seem appropriate?
- b. Plot the $\text{logit}(\text{proportion of extinction})$ as a function of $\log(\text{Area})$. Does a linear regression model seem appropriate?
- c. After making a log transformation to Area, fit a logistic regression curve to the data. What are $\hat{\beta}_0$ (intercept), $\hat{\beta}_1$ (slope), the estimated standard errors of both, and the confidence intervals for both?
- d. Perform a ‘Drop-in-Deviance’ test for this model fit versus the intercept only model. What are the null and alternate hypotheses this statistic is testing? What is the general form for this test statistic in this case (this should involve writing out the deviances and their difference)? What is the associate p-value? Is this a valid test to use in the situation and why?

- e. Interpret the estimated coefficient on $\log(\text{Area})$. Get the fitted probability for Pihlajakari and for Maakrunni. Which island would we predict to have a greater rate of extinction?
- f. It seems possible that there is a bit of a quadratic slope in the data as well. Get the model fit with the additional quadratic term. Give a formal test for the inclusion of a quadratic term.
- g. What conclusions might you draw from this study? What is the scope of this inference (that is, can your conclusions be extrapolated to other situations with different islands or habitats)?
- h. Look at the overall Deviance/Pearson Chi Squared goodness of fit test. What are the null and alternate hypotheses this statistic is testing? What is the general form for this test statistic in this case (this should involve writing out the deviance, for example)? What does this test indicate as far as model fit? If warranted, go through the steps related to overdispersion¹.

¹Important: The steps in the question are out of order for doing an actual analysis. My motivation is I didn't want some people to do this problem with overdispersion, others without it. Remember, always look for overdispersion first.

2. For each of the following, discuss when the topic applies, for what it can be used, and how to use it. Consider both the explanatory and response variables and include any relevant reference distributions under the null hypothesis.
 - a. Deviance goodness-of-fit test
 - b. Hosmer-Lemeshow test
 - c. Drop-in-deviance test
 - d. AIC