Stat 421: Applied Categorical Data Analysis, Spring 2020
Lec. 002
Derek Bean
Homework 8
Due: Friday 4/17/20 by 11:59pm CST through Canvas

Each question worth 10 points. TOTAL: 50 points.

Suggested Readings in Agresti (2nd Ed.)

1. Chapter 3, Section 3.4.
2. Chapter 4, Section 4.1, §4.1.1–§4.1.4.
3. Chapter 4, Section 4.2, §4.2.1–§4.2.4

Problems

1. An experiment analyzes imperfection rates for two processes used to fabricate silicon wafers for computer chips. For treatment A applied to 10 wafers, the number of imperfections are 8, 7, 6, 6, 3, 4, 7, 2, 3, 4. Treatment B applied to the 10 other wafers has 9, 9, 8, 14, 8, 13, 11, 5, 7, 6 imperfections. The wafers are also classified by thickness of silicon coating, Low or High. The first five imperfection counts reported for each treatment refer to chips with Low thickness and the last five refer to chips with High thickness. Consider the model $\log \mu = \alpha + \beta_1 x + \beta_2 z$, where $x = 1$ for treatment B while $x = 0$ for treatment A, and $z = 1$ for Low thickness which $z = 0$ for High. Analyze these data, making inferences about the effects of treatment type and thickness of coating.

2. For the 23 space shuttle flights before the Challenger mission disaster in 1986, the following table shows the temperature (in degrees Fahrenheit) at the time of the flight and whether at least one primary O-ring suffered thermal distress (TD).

Ft	Temp.	TD	Ft	Temp.	TD
1	66	0	13	67	0
2	70	1	14	53	1
3	69	0	15	67	0
4	68	0	16	75	0
5	67	0	17	70	0
6	72	0	18	81	0
7	73	0	19	76	0
8	70	0	20	79	0
9	57	1	21	75	1
10	63	1	22	76	0
11	70	1	23	76	0
12	78	0			

(a) Use logistic regression to model the effect of temperature on the probability of thermal distress. Interpret the effect.

(b) Estimate the probability of thermal distress at 31° F, the temperature at the time of the Challenger flight.

(c) At what temperature does the estimated probability equal 0.50? At that temperature, give a linear approximation for the change in the estimated probability per degree increase in temperature.

(d) Interpret the effect of temperature on the odds of thermal distress.

(e) Test the hypothesis that temperature has no effect, using (i) the Wald test, and (ii) the likelihood-ratio test.

3. The following table taken from Agresti is reportedly from a random sample of subjects selected for a study in Italy investigating the relationship between annual income and whether one possesses a credit card. At each level of annual income, given in millions of lira, the table indicates the total number of subjects, followed by the number of them who possess at least one credit card. (Note: the study is likely from decades ago. Agresti tells us that 1 million lira was worth about 500 Euros around 2007. 500 2007 Euros is worth about 600 2020 Euros, which is worth about $600 American dollars.)

Inc.	No. Cases	Credit Cards	Inc.	No. Cases	Credit Cards	Inc.	No. Cases	Credit Cards	Inc.	No. Cases	Credit Cards
24	1	0	34	7	1	48	1	0	70	5	3
27	1	0	35	1	1	49	1	0	79	1	0
28	5	2	38	3	1	50	10	2	80	1	0
29	3	0	39	2	0	52	1	0	84	1	0
30	9	1	40	5	0	59	1	0	94	1	0
31	5	1	41	2	0	60	5	2	120	6	6
32	8	0	42	2	0	65	6	6	130	1	1
33	1	0	45	1	1	68	3	3			

Source: Based on data in *Categorical Data Analysis*, Quaderni del Corso Estivo di Statistica e Calcolo delle Probabilità, no. 4, Istituto di Metodi Quantitativi, Università Luigi Bocconi, by R. Piccarreta.

The following output is from a logistic regression model fit to the data, treating each income level as an independent binomial sample:

Parameter	Estimate	Standard error
Intercept	-3.5561	0.7169
Income	0.0532	0.0131

(a) Interpret the effect of income on the odds of possessing a credit card.

(b) Conduct a significance test about the income effect.

(c) Construct a 95% confidence interval for the width effect, and interpret it.

4. Hastie and Tibshirani (1990, p. 282) described a study to determine risk factors for kyphosis, which is severe forward flexion of the spine following corrective spinal surgery. The age in months at the time of the operation for the 18 subjects for whom kyphosis was present were 12, 15, 42, 52, 59, 73, 82, 91, 96, 105, 114, 120, 121, 128, 130, 139, 139, 157; and for the 22 subjects for whom kyphosis was absent were 1, 1, 2, 8, 11, 18, 22, 31, 37, 61, 72, 81, 97, 112, 118, 127, 131, 140, 151, 159, 177, 206.

(a) Fit a logistic regression model using age as a predictor of whether kyphosis is present. Test whether age has a significant effect.

(b) Plot the data. Note the differences in dispersion of age at the two levels of kyphosis.

(c) Fit the model $\text{logit}[\pi(x)] = \alpha + \beta_1 x + \beta_2 x^2$. Test the significance of the squared age term, plot the fit, and interpret.

5. For the horseshoe crab data, fit the logistic regression model for $\pi =$ probability of a satellite, using weight as the predictor.

   (a) Report the ML prediction equation.

   (b) Find $\hat{\pi}$ at the weight values 1.20, 2.44, 5.20 kg, which are the sample minimum, mean, and maximum, respectively.

   (c) Find the weight at which $\hat{\pi} = 0.50$.

   (d) At the weight value found in part (c), give a linear approximation for the estimated effect of (i) a 1 kg increase in weight. This represents a relatively large increase, so convert this to the effect of (ii) a 0.10 kg increase, and (iii) a standard deviation increase in weight (0.58 kg).

   (e) Construct a 95% confidence interval to describe the effect of weight on the odds of a satellite. Interpret.

   (f) Conduct the Wald or likelihood-ratio test of the hypothesis that weight has no effect. Report the P-value, and interpret.