Due by 4pm on Wednesday, April 30. Please leave your homework on the table before class begins on Friday or leave it in the dropbox outside my office. Do not forget to attach the honor code.

(1) The following table refers to a prospective study of maternal drinking and congenital malformations. After the first 3 months of pregnancy, the women in the sample completed a questionnaire about alcohol consumption. Following childbirth, observations were recorded on the presence or absence of congenital sex organ malformations. Alcohol consumption, measured as average number of drinks per day, is an explanatory variable with ordered categories. Malformation, the response variable, is nominal.

Table 2.7. Infant Malformation and Mother's Alcohol Consumption

Alcohol Consumption	Malformation: Absent	Malformation: Present	Total	Percentage Present	Standardized Residual
0	17,066	48	17,114	0.28	−0.18
<1	14,464	38	14,502	0.26	−0.71
1–2	788	5	793	0.63	1.84
3–5	126	1	127	0.79	1.06
≥ 6	37	1	38	2.63	2.71

Source: B. I. Graubard and E. L. Korn, *Biometrics*, **43**: 471–476, 1987. Reprinted with permission from the Biometric Society.

(a) (5 points) Fit a linear probability model by taking x to be the mother's alcohol consumption and Y to be whether a baby has sex organ malformation, with scores $(0, 0.5, 1.5, 4.0, 7.0)$ for alcohol consumption
(b) (5 points) State the prediction equation, and interpret the intercept and slope.
(c) (5 points) Use the model fit to estimate the (i) probabilities of malformation for alcohol levels 0 and 7.0, (ii) relative risk comparing those levels.
(d) (5 points) Fit a logistic regression model. Report the prediction equation. Interpret the sign of the estimated effect.
(e) (5 points) Fit a probit model. Report the prediction equation. Interpret the sign of the estimated effect.

(2) 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) (5 points) Fit a logistic regression model using age as a predictor of whether kyphosis is present. Test whether age has a significant effect.
(b) (5 points) Plot the data. Note the difference in dispersion of age at the two levels of kyphosis.
(c) (5 points) 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.

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

Table 4.10. Data for Problem 4.5 on Space Shuttle

Ft	Temperature	TD	Ft	Temperature	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	58	1
12	78	0			

Note: Ft = flight no., TD = thermal distress (1 = yes, 0 = no).

Source: Data based on Table 1 in S. R. Dalal, E. B. Fowlkes and B. Hoadley, *J. Am. Statist. Assoc.*, **84**: 945–957, 1989. Reprinted with the permission of the American Statistical Association.

(a) (10 points) Use logistic regression to model the effect of temperature on the probability of thermal distress. Interpret the effect.
(b) (5 points) Estimate the probability of thermal distress at 31° F, the temperature at the time of the Challenger flight.
(c) (5 points) 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) (5 points) Interpret the effect of temperature on the odds of thermal distress.
(e) (5 points) Test the hypothesis that temperature has no effect, using (i) the Wald test, (ii) the likelihood-ratio test.

(4) For the horseshoe crab data Table 3.2, available at `https://users.stat.ufl.edu/~aa/intro-cda/data/`, fit the logistic regression model for $\pi =$ probability of a satellite, using weight as the predictor.

(a) (10 points) Using $x =$ weight and $Y =$ number of satellites, fit a Poisson loglinear model. Report the prediction equation.
(b) (5 points) Estimate the mean of Y for female crabs of average weight 2.44 kg.
(c) (5 points) Use $\hat{\beta}$ to describe the weight effect. Construct a 95% confidence interval for β and for the multiplicative effect of a 1 kg increase.
(d) (5 points) Conduct a Wald test of the hypothesis that the mean of Y is independent of weight. Interpret.
(e) (5 points) Conduct a likelihood-ratio test about the weight effect. Interpret.