EdPsych/Psych/Soc 589
C.J. Anderson

Homework 4: Answer Key

1. For this problem use the data in Table 2.7 on alcohol consumption and infant malformation on page 42 of Agresti (2007).

   (a) The three models:

   Linear Model :

   $$\mathrm{E}\left(\frac{\text{no. } \widehat{\text{present}}}{\text{no. cases}}\right) = .0025 + .0011(\text{alcohol consumption})$$

   As alcohol consumption increases, the expected proportion of infants with malformations increases. For a 1 unit increase in alcohol consumption, the estimated expected proportion increases by .0011.

   See figure.

   The fitted values from the linear model under-predict/estimate the (sample) actual proportion with malformation for alcohol consumption greater than 2 drinks per day (3–5, and ≥ 6).

   Probit Model:

   $$\text{probit}\left(\frac{\text{no. } \widehat{\text{present}}}{\text{no. cases}}\right) = -2.7996 + .1098(\text{alcohol consumption})$$

   As alcohol consumption increases, the estimated expected proportion with malformation increases..

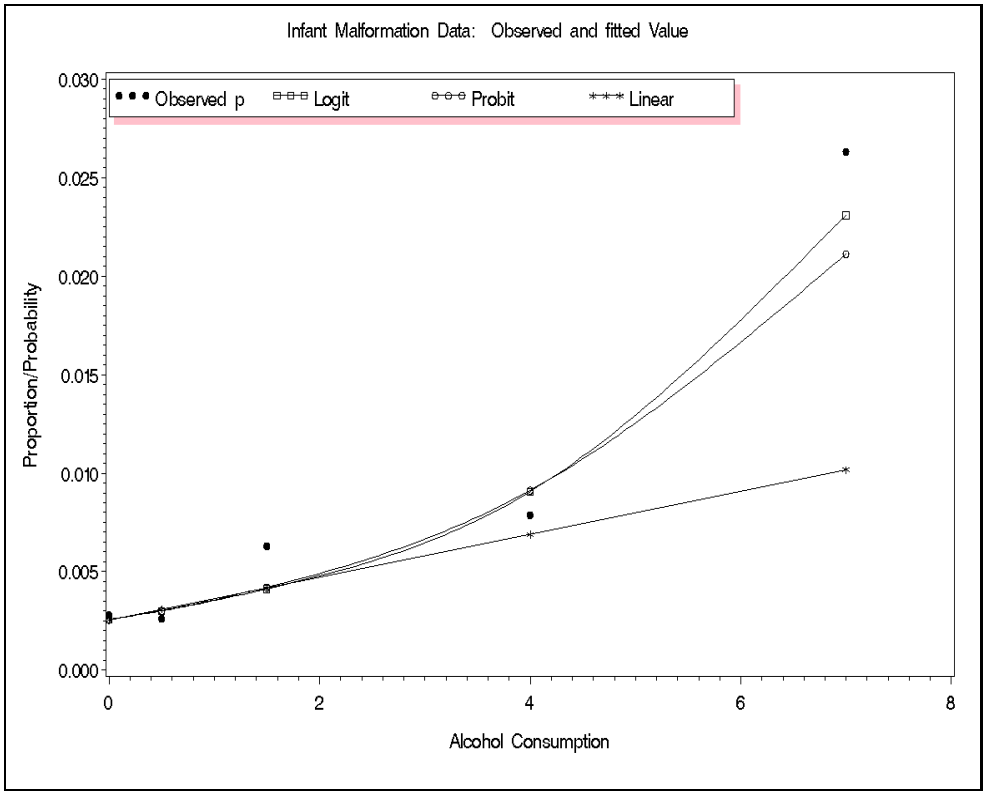
   Logit Model:

   $$\text{logit}\left(\frac{\text{no. } \widehat{\text{present}}}{\text{no. cases}}\right) = \log\left(\frac{\text{no. } \widehat{\text{present}}}{\text{no. absent}}\right) = -5.9605 + .3166(\text{alcohol consumption})$$

   As alcohol consumption increases, the expected proportion with malformation increases. The odds of malformation increase as alcohol consumption increases. For a 1 unit increase in consumption, the odds is $\exp(.3166) = 1.37$ times larger.

(b) Figure comparing observed proportions and fitted values from the linear, probit and logit models, as well as the observed:

The fitted values for the logit and probit models are similar for the first four levels of alcohol consumption (both fit equally well); however, the logit models fits better than the probit for the largest levels of alcohol consumption. Overall, the logit fits better.



2. Problem 3.11 on page 93–94 of Agresti (2007), except for part (c) find the 99% confidence interval for μ_A/μ_B.

(a) & (b) Estimated model is $\log(\hat{\mu}_i) = 1.6094 + 0.5878x_i$. So for each treatment, we have For treatment A:

$$\log(\hat{\mu}_A) = 1.6094 + 0.5878 = 1.6094 \quad \text{so} \quad \hat{\mu}_A = 1.6909$$

For treatment B:

$$\log(\hat{\mu}_B) = 2.197225 \quad \text{so} \quad \hat{\mu}_B = 2.197225$$

From this we can see that

$$\hat{\beta} = \log(\hat{\mu}_B) - \log(\hat{\mu}_A) = 2.197225 - 1.6094- = .5878.$$

and for interpretation

$$\exp(\hat{\beta}) = \exp(.5878) = 1.8$$

On average, the number of imperfections in a wafer given treatment A are 1.8 the number of imperfections given treatment B.

(c) $H_o : \beta = 0$ versus $H_o : \beta \neq 0$,

$$\text{Wald statistic } = \left(\frac{.5878}{.1764}\right)^2 = 11.1, \quad df = 1, \quad \text{and} \quad p - \text{value} < .001.$$

$$\text{Likelihood ratio statistic} \quad = 11.59, \quad df = 1, \quad \text{and} \quad p - \text{value} < .001.$$

With both test statistics, reject H_o; there appears to be a difference in the average number of defects.

(d) To get a 99% confidence interval for μ_B/μ_A, we use the fact $\beta = \log(\mu_B) - \log(\mu_A)$. The 99% CI for β is

$$.5878 \pm 2.576(.1764) \Longrightarrow (.1334, 1.0422)$$

and take the exp of the end points gives us the 99% confidence interval for μ_B/μ_A:

$$(\exp(.1334), \exp(1.0422)) \Longrightarrow (1.143, 2.836)$$

Note if you want the 99% confidence interval for μ_A/μ_B take the inverses of the endpoints and interchange order: $(1/1.14, 1/2.83) \Longrightarrow (.353, .875)$. . . . or you could have used the SAS/GENMOD model option `ALPHA=.01` and read the interval off the SAS output. I did computations in R rather than seek package that give them to me.

Note: Below is another way to test $H_o : \mu_A = \mu_B$ as $H_o : \pi_A = .5$ versus $H_1 : \pi \neq .5$.
$n = 50 + 90 = .140$
$Y =$ number of imperfections from A $= 50$
$p =$ proportions of imperfections from A $= 50/140 = .357$.
Using large procedure and testing proportions,

$$z = \frac{.3571 - .5}{\sqrt{.5(.5)/140}} = \frac{-.1428}{.04225} = -3.3806$$

p-value$< .001$. Reject H_o; the data support the conclusion that the probability of an imperfection from treatment A is not equal to .5.

You could have also computed a 95% (for $\alpha = .05$) confidence interval for $\pi_A = .5$, which equals (.283,.439).

3. 3.12 on page 94 Agresti (2007)
Add in the factor that the wafers also differed in terms of thickness (2 levels), as well as treatment (A and B).

The model with marginal ("main effects") for thickness and treatment fits the data well: df, $G^2 = 16.27$, p-value$= .57$ (or $X^2 = 16.04$, and p-value$= .59$. Furthermore, there are no "large" standardized residuals.

Therefore, go with the model (and interpretation) from problem 3.11.