Chapter 4 Class Notes – 4 Classes

Class One (Chapter Notes, pp. 1 – 11)

- "GdLMs" (not GLMs): distribution and link function; for logistic: binomial distribution and logit link gives equation (4.3)
- E.g. 4.1 – Polish girls: odds ratio, LD_{50}, scale, residual plot
- E.g. 4.2 – ECMO and respiratory illness in children – odds ratio, LD_{50} makes no sense here, model fits data exactly (i.e., there is no LOF test – just like when fitting a line to two points or to two means)
- E.g. 4.3 – Tobacco budworms – first we need to get the "X scale" right, then compare two groups (Ms versus Fs): like ANOCOV, test of one common line for both genders – not a Full-and-Reduced-F test, but a $-2\Delta LL$ test. Why can we *not* do an F test here?

Class Two (Chapter Notes, pp. 11 – 18)

- E.g. 4.4 – proteins – when we use only protein1, residuals tell us others may be important; stepwise regression tells us proteins protein1, protein2, and protein4 (but not age and p3) are significant; just for fun, dropping p2 and p4 is tested using a $-2\Delta LL$ test on p.12
- E.g. 4.5 lung disease: demonstrates how to interpret ORs
- E.g. 4.6 – pups: wish to see if treatment differs from control, wrong analysis (p = 0.004) collapses over (i.e., ignores) litter-to-litter variability and therefore

overstates the case; two choices for correct analysis: d-scale approach (p = 0.0855) and mixed approach (p = 0.0928) find no significant difference.

Class Three (Chapter Notes, pp. 18-26)

- **Ex. 4.7 – old approach – see the Introduction (Chap. 1, p.10): Breslow-Day declares odds ratios (OR's) are the same for the two faces ($p = 0.5286$), but CMH test declares this common value is not one ($\chi_1^2 = 76.4645$); point estimate is $\widehat{\theta}_{MH} = 18.6537$ and the 95% CI is 7.46 to 46.6. Note the interpretation on p.10 of Chap. 1**
- **In the above, also understand the difference between the above adjusted OR estimate and the marginal estimate given on p.10 of the Introduction (see Chap.1)**
- **Ex. 4.7 – new approach: HA log-linear model (4.5), we focus on the 2-way interactions: accepting that an interaction term is zero is equivalent to accepting that the OR is 1. This model fits the data since the deviance is not large compared to the df. That the 3-way interaction is declared NS is equivalent accepting that the ORs for the 2 faces are equal (this is the HA model counterpart of the Breslow-Day test). To illustrate: here $\widehat{\theta}_{AC} = e^{2.0545} = 7.8$, so after controlling for marijuana usage, the odds a cigarette smokers has drunk alcohol equals 7.8 times the odds that a non-cigarette smoker has drunk alcohol**
- **Ex. 4.8 – a Poisson counterpart of the two-sample t-test**

- **Ex. 4.9 – a Poisson counterpart of the paired t-test; uses a conditional argument: conditional on n = $y_1 + y_2$, y_1 has a binomial distribution with $\pi = \mu_1/(\mu_1 + \mu_2)$. So, testing $\pi = ½$ is the null hypothesis here**
- **Ex. 4.10 – a Poisson counterpart of ANOCOV – the 'offset' – which is log(t) here – is the "covariate" analog.**

Class Four (Chapter Notes, pp. 26-36)

- **Nominal outcomes (hair color, four nucleotides) versus ordinal outcomes (poor, fair, good, excellent) – still denote outcomes by "Y"**
- **Three models are proposed in Section 4.4 and one in the Appendix 4.5.2: BCL in Equation (4.7), PO (extended to UPO in Appendix) in Equation (4.8), AC (top of p.27), and CRA and CRB (on p.33); these models transform the π's on the LHS in different ways and the RHS is still a linear model (in the parms)**
- **Our focus here is on the PO model!**
- **The above models look similar but some fit a given data set better than others, and the interpretations differ in predicted values and odds-ratio interpretations**
- **Ex. 4.11 – 4 ordered outcomes (chronic respiratory disease) and PO model is fit, 3 categorical explanatory variables using 4 dummy variables; Output 4.11b tests whether all variables can be dropped using LR, Score and Wald tests; Output 4.11c "Class" analysis: hard to understand but proportionality is accepted (p = 0.1479)**

- **Output 4.11d is used for <u>odds-ratio interpretations:</u>**
  - **Those with no job exposure to pollution have odds of being in the less serious (as opposed to more serious) respiratory direction 2.37 times the odds for those exposed to pollution on the job**
  - **<u>Preferred (but equivalent)</u>: the odds of being in the more serious (as opposed to less serious) respiratory direction for those with job exposure to pollution is 2.37 times the odds of being in the more serious respiratory direction for those not exposed to pollution on the job**
- **Ex. 4.12 – sometimes PO and BCL models don't fit – GOF tests are rejected in both cases – CRB model fits these data "best" and fit is shown on p.35**
- **In § 4.5.1 (p.31), a "hyper-parameter" is introduced to help us pick the scale; this parameter is θ_6 in Equations 4.9 - 4.10. When $\theta_6 = 1$ we use the original scale. When $\theta_6 = 0$ we use the log scale (any base is okay … easiest to use the natural log) – for budworms, output 4.13 indicates CI (-1.74,0.54) for F and (-0.70,1.15) for M – if they both included 1, we'd use dose scale – both CI's include 0 so that's why we used the log scale on pp.7-11. Note this is not unique, since ½ is in both intervals, so we could have used the square root scale.**