HW3 for STAT 818R Fall 2026, due Saturday 11/14 on ELMS

Eric Slud 10/28/2026

(1). In the `GLM-Poisson-Multinomial.RLog` script in the `Rscripts` directory of the course web-page, we had the following code to generate a 4×6 multinomial contingency table with table-total $n = 650$ and a set of cell probabilities p_{ij} that definitely do **not** follow a row-column-independence model.

```
eta2 = outer(1:4, 1:6, function(x,y) c(.37, .22, 0.29, -.88)[x] +
                       c(.6, -.26, .35, .79, -.85, -.63)[y] +
                       1.5*(x>2 & y>3) - 1.5*(x <= 2 & y <= 3))
  a2 = -log(sum(exp(c(etamat))))
  pmat2 = exp(a2 + eta2)
  rv2 = apply(pmat2,1,sum)
  cv2 = apply(pmat2,2,sum)
  summary(c(pmat2-outer(rv2,cv2)))    ## definitely not zero
  n=650
  set.seed(27281)
  ndata = array(table(sample(24, n, replace=T, prob=c(pmat2))), c(4,6))
```

This exercise is to follow the steps of that script, using the `R` model-fitting function `glm` and appropriate use of its output components,

(a) Find the best (parsimonious, well-fitting) log-linear model you can using the 4x6 dataset `ndata` generated above. Check the fit using analysis-of-deviance type goodness of fit tests, and explain why you think you have found the most parsimonious loglinear model that passes goodness-of-fit tests.

paar **(b)** Find the standard errors (and display them in a 4×6 table) for each of the cell probabilities fitted by the model you found in (a).

(2). Check the theoretical formula (9) in the handout (h), 'Handout22Ktables-MH.pdf', summed over k as the theoretical exression for $E(R - \theta S)^2$ with and without this last term (10). You should do this check via simulation of a large set of replicates (maybe 10^5) of one kind of $2 \times 2 \times K$ table with $K = 3$ and large n_{ijk} (say, all > 50) and another kind of table with large K (say, > 100) and small entries n_{ijk}

(nonzero, but with table totals ≤ 20). Use these simulations also to show that the two answers (with and without this term) can be meaningfully different, and to check (via confidence intervals based on the simulations) the consistency of the estimator $\hat{V}$, given 3 lines after the formula (10) in the handout, for the variance of $\hat{\theta}^{MH} = R/S$.

(3). Load the `R` package `metadat` and from it find the dataset `dat.nissen2007` consisting of the results of 42 separate clinical studies on the effect of an antidiabetic drug ' rosiglitazone' on myocardial infarction (MI) and cardiovascular death. Use our techniques involving Poisson regression and/or Cochran-Mantel-Haenszel statistics in $2 \times 2 \times K$ tables to assess the effects of this drug on MI and (separately) on cardiovascular death. This is an exercise (#6.29) drawn from Agresti's (2013) book, asking for a re-analysis of a categorical-data Meta-analysis done in a 2010 paper by S. Nissen and K. Wolski in *Archives of Internal Medicine* **14**, 1191-1201, also available through *Pubmed*. Explain your model assumptions and the hypothesis you are testing, and present the `R` code and results of the analysis.