

HW2 for STAT 818R Fall 2026, due Monday 10/19 on ELMS

Eric Slud

10/6/2026

(1). Show how to compute the `glmgee` Wald test (chi-squared) values for logistic-regression estimating-equation Models 1 vs 2, 2 vs 3, and 4 vs 5 using the `crabs` data as was done in the script `GLM-Model-Fitting.RLog` in the `RScripts` Directory on the course web-page for Model 3 vs 4. `vspace3mm`

(2). Find the likelihood ratio test and chi-squared test test-statistics in a multinomial 3×3 table (with cell probabilities $p_{j,k}$) for the hypothesis $H_0 : p_{jk} \propto \exp(aj + bk + cjk)$ versus the general alternative. Find and interpret the p-values for these statistics in the 3×3 table with first row (3, 7, 18), second row (5, 18, 17), and 3rd row (9, 35, 42). *Note:* you should find the likelihood-maximizing values $\hat{a}, \hat{b}, \hat{c}$ under H_0 either using a numerical-maximizing function like `nlm` or `optim` in R, or else set up a multiparameter Newton-Raphson iteration to find an iterative numerical solution of the likelihood equations. Then substitute these MLEs to create the likelihood-ratio or chi-square statistics.

(3). Formulate a hypothesis test that the coefficient $c = 0$ assuming only that the score ('quasilikelihood') estimating equation has the form it took in Problem 2, and tell (in formulas and with the numbers given) how you would do the estimating-equation Wald test.

(3). Regard the $n \times p$ design matrix $\mathbf{Z}$ in a binomial regression problem for data $Y_1, \dots, Y_n$ as fixed. Use the formulas for logistic and probit regression to show that the observed (whole-dataset, not per-observation) information matrix with the probit link depends on the data $\mathbf{Y}$, while that from the logit link does not.

(4). Simulate a dataset from an Overdispersed Poisson model with $Y_i \sim \text{Poisson}(\lambda_i \exp(\beta Z_i))$ where $\lambda_i \sim \text{Gamma}(\gamma, \gamma)$ are iid, $i = 1, \dots, 300$, with $Z_i \sim \mathcal{N}(3, 1)$, $\beta = 1, \gamma = 3$. Use GLM software in R to demonstrate lack of fit for a Poisson regression estimating equation.

(5). Read the UCLA dataset on school absences into an R data frame using the `haven` package (which you must install) and the statement `ucla_absences <- read_sas("https://stats.idre.ucla.edu/wp-content/uploads/2016/02/poisson.sas7bdat")`.

Do a variable selection and data analysis with counts of days of school absences as response variable, based on Poisson or (if needed because of overdispersion) Negative Binomial Regression for these data and report your conclusions about a suitable model.