STATISTICS 4C03 / 6C03

Assignment #4

The data sets and documentation are found in R in the MASS library. You can do the analysis in R or export the data to any package you prefer.

---

1. Suppose that $Y_1 \sim \text{Bin}(m_1,\pi_1)$ and $Y_2 \sim \text{Bin}(m_2,\pi_2)$ are independent. Show that the probability mass function for $Y_+ = Y_1 + Y_2$ can be written as a function of π_1 and π_2 times a polynomial in the odds ratio. Use this result to derive the non-central hypergeometric distribution, that is, the distribution of Y_1 given Y_+. How does this simplify when the odds ratio is 1? Hint: You can simplify the formula by letting $M(\beta)$ denote the moment generating function for the hypergeometric distribution, evaluated at the log odds ratio.

2. Analyze the Copenhagen Housing Conditions Survey (data frame `housing` in the MASS library).

3. Analyze the Age of Menarche data (data frame `menarche` in the MASS library). Compare logit, probit and arcsine-root transformation analyses. Discuss the adequacy of the fit. Give an interpretation of the coefficients you fitted in the probit analysis.

4. Analyze the data on Diabetes in Pima Indian Women (data frames `Pima.tr`, `Pima.te` in the MASS library). Find a model that fits as well as possible without overfitting. Test the goodness of fit of the model with a Hosmer-Lemeshow test.