

Homework #6: Due Friday, December 6.

1. Problem 1, Chapter 15, page 200
2. The **eggprod** dataset in the **faraway** library concerns an experiment where six pullets were placed into each of 12 pens. Four blocks were formed from groups of three pens based on location. Three treatments were applied. The number of eggs produced was recorded.
 - (a) Fit a model for the number of eggs produced with the treatments as fixed effects and the blocks as random effects. Describe the estimated differences between the treatments.
 - (b) Test for the significance of the treatment. Compute the p-value using both the χ^2 -distribution (with a likelihood ratio test) and resampling methods (bootstrapping).
3. The dataset **wbcd** in the **faraway** library comes from a study of breast cancer in Wisconsin. There are 681 cases of potentially cancerous tumors of which 238 are actually malignant. Determining whether a tumor is really malignant is traditionally determined by an invasive surgical procedure. The purpose of this study was to determine whether a new procedure called fine needle aspiration, which draws only a small sample of tissue, could be effective in determining tumor status.
 - (a) Fit a binomial (logistic) regression model with **Class** as the response and the other nine variables as predictors. Report the residual deviance and associated degrees of freedom. Can this information be used to determine if this model fits the data? Explain.
 - (b) Use AIC as the criterion to determine the best subset of variables. You can use the **step** function in **R**.
 - (c) Use the reduced model to predict the outcome for a new patient with predictor variables 1, 1, 3, 2, 1, 1, 4, 1, 1 (same order as above). Give a confidence interval for your prediction.
 - (d) Suppose that a cancer is classified as benign if $p > 0.5$ and malignant if $p < 0.5$. Compute the number of errors of both types that will be made if this method is applied to the current data with the reduced model.
 - (e) Suppose we change the cutoff to 0.9 so that $p < 0.9$ is classified as malignant and $p > 0.9$ as benign. Compute the number of errors in this case. Discuss the issues in determining the cutoff.
 - (f) It is usually misleading to use the same data to fit a model and test its predictive ability. To investigate this, split the data into two parts - assign every third observation to a test set and the remaining two-thirds of the data to a training set. Use the training set to determine the model and the test set to assess its predictive performance. Compare the outcome to the previously obtained results.