STAT 2132:
Applied Statistical Methods II
HW #3
Due Wednesday, February 9

1. From KNNL: **14.11** (pages 626-627). The analyst believes that logistic regression model, $\text{logit}(\pi_i) = \beta_0 + \beta_1 X_i$, where X_i is the deposit level associated with the i-th bottle, is appropriate for studying the relation between size of deposit and the probability that a bottle will be returned. Do the questions (a), (b), (e), (f) in 14.11 and the following questions:

   (1) What is the estimated odds ratio associated with increasing deposit by 10 cents along with its 95% Wald confidence interval?

   (2) Conduct the deviance and the Pearson χ^2 tests of goodness-of-fit for the appropriateness of the logistic regression. State the hypotheses, decision rules, and conclusions.

2. You are analyzing a survey dataset (cda.txt) of employees of a large national corporation to determine how satisfaction depends on race, gender, age and regional location.

   (a) Build an appropriate logistic regression model starting with a full model with the four covariates and their interactions.

   (b) Conduct a likelihood ratio test to compare your final model with the full model.

   (c) Describe how race and gender are associated with satisfaction (be sure to include confidence intervals).

   (d) What is the most satisfied group of people?

3. For the Bottle return problem from KNNL 14.11:

   (a) Compare the fit from the logistic regression in problem 14.11(a) to the loess fit, and conduct the Hosmer-Lemeshow test of goodness-of-fit for the appropriateness of the logistic regression.

   (b) For the fitted regression function in problem 14.11(a), obtain an approximate 95% confidence interval for the probability of a returner for deposit $x_h = 15$ cents. Interpret your interval.

   (c) Generate the ROC curve based on the logistic model in 14.11(a), and select the cut point c_1 that maximizes the Youden index which is the difference between sensitivity and one minus specificity.

   (d) Obtain the area under the ROC curve to assess the model's predictive power here. What do you conclude?