**HOMEWORK 10**

*Chi-square Tests for Independence and Logistic regression*

Reading: *STAT2: Building Models for a World of Data*, second portion of Chapter 11.4 labeled as Review: Two-sample z-test and Chi-square Test for a Two-way Table, and sections 9.1, 9.2, and 9.4.

Notes:

* Round all numbers to 3 decimal places unless otherwise specified.

11.23 – Pines: thorny cover and browse. The study is described in question 11.22 and 11.23. The pines15.jmp data set has only the pines planted at 15 foot spacing. We consider the association between thorny cover (cover95, 4 categories) and whether the tree was browsed in 1995 (deer95).

Cover95 is an ordinal variable. The first set of questions analyze the data as if it were categorical. Use the **Cover95** variable, not the Cover95 2 variable for these questions.

1. Draw a mosaic plot that shows the conditional probability that a tree is browsed for each of the 4 cover95 categories. Select the correct assignment of variables to draw this plot.

X = cover95, Y = deer95

1. Based on your mosaic plot, answer true or false to each of these statements

The marginal probability of deer browse is within 5% of 70% FALSE

The marginal probability of “not deer browsed” is within 5% of 70% TRUE

The conditional probability of deer browse is higher in category 3 of cover95 than in category 0 FALSE

There are fewer trees in category 3 of cover95 than in category 0 TRUE

1. Use a Chi-square test to test whether deer browse (deer95) is independent of thorny cover (cover95).

Report the degrees of freedom for this test: 3

Report the value of the chi-square statistic: 4.856

Report the p-value for this test: .1826

1. Report the conditional probability of deer browse in category 3 of cover95 (report as a probability, not a percent) .1912
2. Evaluate the conditions for appropriate use of a Chi-square test, answer Yes or No
   1. Did the study use one of the appropriate types of random sample? Yes, data collected for fixed period of time (1955)
   2. Are all the expected counts sufficiently large? yes

The second set of questions treat cover95 as a continuous variable with values 0, 1, 2, 3. Use the **Cover95 2** variable in the JMP dataset, which is numeric / continuous, for these questions.

1. Fit a logistic regression that uses the numerical cover95 value to predict the probability that a tree is browsed. I suggest you use Fit model and make sure to set the target level to **1** (tree is browsed).

Report the intercept of that logistic regression: -.5803

Report the slope of that logistic regression: -.2236

1. Fit a logistic regression that uses the numerical cover95 value to predict the probability that a tree is **not browsed**. I suggest you use Fit model and make sure to set the target level to **0** (tree is not browsed).

Report the slope of this logistic regression: .2236

1. If deer browse (deer95) is independent of thorny cover (cover95), what value do you expect for the logistic regression slope: 0

1. Use the logistic regression to test whether deer browse is independent of thorny cover.

Report the p-value for this test: .0435

1. Use the logistic regression to predict the probability of deer browse for each cover category.

When thorny cover = 0, the predicted probability of deer browse is: 0.358863115352697

When thorny cover = 1, the predicted probability of deer browse is: 0.309182356427203

When thorny cover = 2, the predicted probability of deer browse is: 0.263552122501657

When thorny cover = 3, the predicted probability of deer browse is: 0.22248729627777

1. It is reasonable to ask whether a linear logistic regression is sufficient, i.e., is there lack of fit to a linear logistic regression. Evaluate this by fitting a model with both a linear (cover95) and quadratic term (cover95\*cover95).

Report the p-value for the quadratic coefficient: .3913

1. Based on the quadratic regression, select the appropriate conclusion:

A linear regression fits the data