**Understanding Data and Statistical Design (60117)**

**Lab 11: Multiple Logistic Regression**

**This lab is not assessed.**

**QUESTION 1.**

This week we continue the analysis of the history of cardiovascular disease in individuals with a multiple logistic regression model. The variables we consider are summarised in the table below.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
|  | categorical | state of general health: 1 (very good), 2 (good), 3 (average), 4 (poor), 5 (very poor) |
|  | continuous | age of individual (years) |
|  | continuous | systolic blood pressure |
|  | binary dummy | history of cardiovascular disease: 0 (no), 1 (yes) |

The data is from a British Panel Survey consisting of the responses of 3916 individuals (available in lab11.csv on Canvas).

We are going to model against , and , including interaction between and . To represent in our model, we need 4 binary dummy variables that we will code as

The target model to fit on the log-odds scale is

where

or if we wish to make the dependence on the predictors explicit

Now fit the multiple logistic regression model described above.

1. Write down the fitted logistic regression for on the log-odds scale, on odds scale and on the probability scale.

Log odds: -7.641911

Odds scale:

Prob:

1. Provide interpretations of the estimates for and on the log-odds scale and for and on the odds scale.

**:** Predicted log odds of a history of cardiovascular disease for those with h1 and zero age and syst.

**:** Predicted difference in log-odds of a history of CVD with good health compared to those with very good health of the same age with zero blood pressure.

**:** Predicted multiple of the odds of a history of CVD for each additional year of age (with constant sys) for all health categories.

**:** Precited multiple of the odds ratio for each additional unit in syst for those with poor health compared to those with very good health

1. Using significance level , document a test to determine if each additional year in age is associated with less than a 0.0375 increase in the log-odds of a history of cardiovascular disease. Write down the hypotheses, the test statistic and p-value, the result of the test with reason and a conclusion in non-mathematical language.

A black screen with white text

Description automatically generated

Rejct null

There is strong evidence each additional year is associated with less than a 0.0375 in the log-odds of a history of cvd.

1. Using significance level , document a test to determine if the fitted logistic regression model is significant. Write down the hypotheses, the test statistic and p-value, the result of the test with reason and a conclusion in non-mathematical language.

A black screen with white text

Description automatically generated

1. Using significance level , document a test to determine if the interaction term is significant. Write down the hypotheses, the test statistic and p-value, the result of the test with reason and a conclusion in non-mathematical language.
2. Using the rule

determine if the fitted regression model predicts a history of cardiovascular disease for those with

* , and
* , and .

1. Calculate the odds of high blood pressure for those with

* , and
* , and .

**Calculating the prop p/1-p. Using the values from the previous point**