**Data mining**

**Lab 3**

[**Contingency Table Analysis**](http://www.dataminingbook.info/pmwiki.php/Main/ContingencyTableAnalysis)

Download the datafile **adult.data.txt**. The description of the data, and its attributes, is available at **http://archive.ics.uci.edu/ml/datasets/Adult**. This is a selection of the Census data from 1994, and it has 48842 instances over 14 categorial, real and integer attributes.

Compute the contingency matrix for variables education and race, and compute the χ2 statistic using your own function, i.e., write a function that takes as input two categorical column-vectors, and returns the χ2 value and its p-value. At the 99% confidence level, are education and race dependent?