INSTRUCTIONS
This worksheet is not assessed! It has 7 questions.

In this exercise you will be performing an contingency analysis on eye and hair colour data using python.

Download 'HairEyeColor.csv' from Blackboard.

HairEyeColor.csv contains two labelled columns. 'Hair Colour' and 'Eye colour.' Each row represents an individual subject.

The 'Hair Colour' column is coded as follows:

1. Dark Brown
2. Light Brown
3. Green
4. Blue

The 'Eye Colour' columns is coded as:

1. Black
2. Brown
3. Red
4. Blond

Write python code to perform the following tasks.

1. Load the HairEyeColor.csv into python using pandas.
2. Create a 2D contingency table containing totalling the number of subjects for each combination of hair colour and eye colour.
3. Calculated the expected number of subjects in each cell in the contingency table if Hair Colour is independent of Eye Colour
4. Calculate chi squared (as shown in lectures).
5. Redo the experiment the easy way using stats.chi2_contingency and check your answers match.
6. Write an appropriate hypothesis and null hypothesis.
7. Calculate p (probability of a Type I Error) using stats.chi2_contingency and determine if Eye Colour is significantly related to Hair Colour.