Part-1

Aim: Revisiting the numpy

Details: One feature of NumPy that is powerful but tricky is the ability to perform broadcasting, which really just refers to repeatedly performing an operation over one or more dimensions.

Tasks:
1. The most basic kind of broadcast is with a scalar, in which you can perform a binary operation (e.g., add, multiply, ...) on an array and a scalar, the effect is to perform that operation with the scalar for every element of the array. To try this out, create a vector 1, 2, . . . , 10 by adding 1 to the result of the arange function.
2. Now, create a 10 × 10 matrix A in which A[i][j] = i + j. You'll be able to do this using the vector you just created, and adding it to a reshaped version of itself.
3. A very common use of broadcasting is to standardize data, i.e., to make it have zero mean and unit variance.
   a. First, create a fake "data set" with 50 examples, each with five dimensions.
   b. import numpy.random as npr
      data = np.exp (npr.randn(50, 5))
   c. Don't worry too much about what this code is doing at this stage of the course, but for completeness: it imports the NumPy random number generation library, then generates a 50 × 5 matrix of standard normal random variates and exponentiates them. The effect of this is to have a pretend data set of 50 independent and identically-distributed vectors from a log-normal distribution.
4. Now, compute the mean and standard deviation of each column. This should result in two vectors of length 5. You'll need to think a little bit about how to use the axis argument to mean and std. Store these vectors into variables and print both of them.
5. Now standardize the data matrix by 1) subtracting the mean off of each column, and 2) dividing each column by its standard deviation. Do this via broadcasting, and store the result in a matrix called normalized. To verify that you successfully did it, compute the mean and standard deviation of the columns of normalized and print them out.

Part-2

Aim: Performing the Hypothesis testing using the Chi-squared test

The table below is a survey response to 4 categorical variables: people in categories from 18–29, 30–44, 45–64 and >65 years, and their movie genre inclination, which is "Action/Adventure", "Romance" and "Biography". Is there any evidence of a relationship between the age group and their movie genre inclination, at 5% significance level?

	Action/Adventure	Romance	Biography	Total
18-29	141	68	4	213
30-44	179	159	7	345
45-64	220	216	4	440
65&older	86	101	4	191
Total	626	544	19	1189