Informatics 1 Data & Analysis
Tutorial 8

Week 10, Semester 2, 2011

- Please attempt all questions on this worksheet in advance of the tutorial, and bring with you all work, including printouts of code and other results. Tutorials cannot function properly unless you do the work in advance.
- Data & Analysis tutorial exercises are not assessed, but they are a compulsory and important part of the course. If you do not do the exercises then you are unlikely to pass the exam.
- Attendance at tutorials is obligatory: if you are ill or otherwise unable to attend one week then email your tutor, and if possible attend another tutorial group in the same week.

Introduction

In this tutorial we will perform statistical analysis over data on students' nationality, physical exercise and sleep. The data for this tutorial was collected in the first DA lecture on 11th January using an anonymous questionnaire asking students for their: hours of physical exercise per week; hours of sleep the previous night; whether of UK, EU (and not UK) or Other nationality.

Reading

For this tutorial, you will need to carry out two specific statistical tests. Look at the following sets of lecture slides for information about how to do these.

- Pearson's correlation coefficient: Unstructured Data 4 — Note III:51–67
- χ^2 (chi-squared) tests: Unstructured Data 5 — Note III:68–89

These are available from the course web page.

You will also need suitable tables of critical values and significance levels. For the χ^2 test, see the lecture slides. For Pearson's correlation coefficient, see the link on the course web page.

Question 1: Statistical analysis of numerical data

Download the file data.pdf from the course homepage. This contains the results of the anonymous questionnaire.

(a) Extract a sample of 7–12 students from this data.

(b) Based on your sample, estimate the mean, and standard deviations for both daily sleep and weekly exercise hours.

(c) Draw a scatter plot showing the sleep and weekly exercise hours for each student in your sample. Visually, does there appear to be any correlation between sleep and exercise hours? If so, is it positive or negative?

(d) Based on your sample, estimate the correlation coefficient between daily sleep and weekly exercise hours for Informatics 1 students. Is there a significant correlation? Is it positive or negative?

Question 2: Statistical analysis of categorical data

The following are some statistics from the the file data.pdf

EU students who exercise at least 2.5 hours per week	32
EU students who exercise less than 2.5 hours per week	18
Overseas students who exercise at least 2.5 hours per week	13
Overseas students who exercise less than 2.5 hours per week	1
EU students who exercise at least 5 hours per week	22
EU students who exercise less than 5 hours per week	28
Overseas students who exercise at least 5 hours per week	6
Overseas students who exercise less than 5 hours per week	8

Here "EU students" here combines the UK and EU-not-UK categories; while "Overseas" contains all other nationalities.

(a) Compile the relevant contingency tables to investigate correlation between: (a) nationality and exercising at least 2.5 hours per week; (b) nationality and exercising at least 5 hours per week.

(b) Calculate the corresponding tables of expected frequencies.

(c) Calculate the corresponding χ^2 values.

(d) Are the two χ^2 tests reliable? If yes, are there correlations? At what significance levels?

(e) Estimate the mean weekly exercise of EU students and the mean weekly exercise of other students.

(f) Which information do you find more informative: the answer to question (d) or the answer to question (e)?

(g) *Optional:* Revisit the data file and look for a correlation between EU/Overseas and reporting 7 hours sleep or less, or more than 7 hours sleep, in the 24 hours before the survey.