**Issues in analysing the data from the Maths education questionnaire:**

When designing a questionnaire, you should really run a pilot version, to see if there are any problems with it before administering it to a large sample. In the course of analysing the data from our questionnaire, it became apparent that there were various issues that we either hadn't thought about, or hadn't thought about enough.

**Problems with the sample:**

In theory, our sample should consist of 250 completed questionnaires, since that's the total number of students enrolled on the course.

In practice, we ended up with a total of 163 completed questionnaires. This is a good response rate (though perhaps not so impressive when one appreciates that this was a "captive" audience). *What distinguishes the people who submitted their questionnaire from those that did not?*

**Problems with the questions:**

The data from students who were not born in the UK have been omitted from the spreadsheet, for two reasons.

(a) Question 3, "Place of birth (town, country)" was intended to distinguish home students from overseas students. There were 116 students whose place of birth identified them as "home" (UK) students. (For this purpose, the Channel Islands students were considered to be UK).

There were 47 students who were not born in the UK. *However, does this mean that they are necessarily overseas students, who were born and educated in another country? How could we fix this problem?*

(b) While the "qualifications" questions were suitable for the UK students, they proved problematic in the case of overseas students because different countries have rather different educational systems that produce qualifications that are hard to equate. *How could we fix this problem?*

Question 8. "What subjects did you study for your A-levels? (Please give grades)". To simplify analysis, and to stand a chance of seeing the overall picture, these data have been coded as follows. Each person has been classified according to (a) whether or not they have A-level Psychology and (b) whether or not they have at least one Science A-level. For the purposes of this coding, Maths, Biology, Chemistry and Physics were classed as Sciences, and anything else was classed as not Sciences. (Psychology was not counted as a Science in the "At least one Science A-level" variable because it was already taken into account in the "Do you have A-level Psychology" variable). This coding means that we lose information about what specific A-levels people have (and their grades) but we can get a general idea of whether "some kind of Science experience at A-level" affects the scores on the other variables.

Questions 10 and 11 produced some ambiguous answers. For one thing, "Public school" was assumed to be "state school"! When people responded "state school", it wasn't clear whether they meant "grammar" or "comprehensive". It was assumed that they meant the latter unless they wrote "grammar", but this may have introduced a small amount of error into the responses for these questions.

For questions 12-20, note that the original questions varied in their presentation: for some, the scale increased from left to right while for others it decreased from left to right (this was to stop people from responding in an "automatic" way to the questions - but did it also possibly increase the likelihood of error s due to respondents misreading the scale on one or more questions? The fact that a low number on the scale means different things in different contexts needs to be kept in mind when interpreting the results.

It would have been better to have numbered the boxes. Firstly this would have reduced respondents' uncertainty about whether or not the end boxes were actually part of the scale itself. (Were they seven-point scales or five-point scales? It wasn't clear from the questionnaire itself). Secondly, when entering the data into SPSS, it would have the questionnaire easier to read!

**The nature of the variables:**

**Which of the variables are continuous data, and which are categorical?**

**Age -** entered as a date. This needs to be converted into a number before descriptive statistics can be performed.

**Gender, birthplace** and **entrant type:** self-explanatory.

**MathsGCSEgrade -** entered as a letter grade. How could you change this so that you could use it as an ordinal measure of maths ability?

**A\_level\_Maths -** entered as a letter grade. How could you change this so that you could use it as an ordinal measure of maths ability?

**A\_level\_Psychology -** simply whether or not the student possesses it.

**Science\_A\_levels -**simply whether or not the A-levels included at least one Science A-level. (In practice, this was usually Biology or Maths, and occasionally Chemistry).

**School\_type\_primary -** state (including Church schools) or private.

**School\_type\_secondary -** state (including Church schools) or private.

**School\_type\_A-levels -** state (including Church schools) or private.

**Primary\_teacher\_patience -**  1 = very patient, 7 = very impatient

**Primary\_teacher\_kindness -**  1 = very kind, 7 = very unkind

**Primary\_teacher\_explanatory -**  this refers to the teacher's perceived ability to explain maths concepts. 1 = very poor, 7 = very good

**Primary\_teacher\_knowledge -** this refers to the teacher's perceived knowledge of maths. 1 = poor, 7 = very good.

**Secondary\_teacher\_patience -** 1 = very patient, 7 = very impatient

**Secondary\_teacher\_kindness -** 1 = very kind, 7 = very unkind

**Secondary\_teacher\_explanatory:** this refers to the teacher's perceived ability to explain maths concepts. 1 = very poor, 7 = very good

**Secondary\_teacher\_knowledge:** this refers to the teacher's perceived knowledge of maths. 1 = poor, 7 = very good.

**Lifetime\_rating:** respondent's ratings of their lifetime experiences of learning maths. 1 = very good, 7 = very poor.

**Maths\_confidence:** 1 = very low confidence, 7 = very high confidence.

**Arithmetic\_ability:** 1 = very poor , 7 = very hig**h.**

**Course\_confidence:** confidence about coping with the course . 1 = very high, and 7 = very low.

**Statistics\_anxiety:** anxiety about learning statistics: 1 = not all anxious, 7 = very highly anxious.

**Essay\_ability:** rating of ability to write essays. 1 = very poor, and 7 = very high.

**Essay\_anxiety:** anxiety about writing Psychology essays. 1 = not at all anxious, and 7 = very anxious.

**StatsAnxietycategorica**l: the data in *statistics anxiety* were turned into a categorical variable by dichotomising them. and a "high anxiety" group. People who gave ratings of 1, 2 or 3 were classified as belonging to the "low anxiety" group, and people who gave ratings of 5, 6 or 7 were classified as belonging to the "high anxiety" group. (People who gave a rating of 4 were discarded).

**Possible analyses:**

It's up to you what you do with these data. Use the descriptive statistics, crosstabs and correlation commands to ask theoretically-motivated questions. Here's one to get you started: is statistics anxiety (high or low) associated with Maths GCSE grade (A\*, A or B)? From the table, it looks like a sizeable majority of A\* grade students have low statistics anxiety. People with A grades are more evenly divided between the two categories. People with B grades tend to occur in the "high anxiety" group more often than the "low anxiety" group. What does the Chi-Square test result tell you?

|  |
| --- |
| **MathsGCSEgrade \* Statistics anxiety dichotomised Crosstabulation** |
|  | Statistics anxiety dichotomised | Total |
| low (rating of 1 2 or 3) | high (rating of 5 6 or 7) |
| MathsGCSEgrade | A\* | **Count** | **16** | **3** | **19** |
| Expected Count | 10.5 | 8.5 | 19.0 |
| Residual | 5.5 | -5.5 |  |
| A | **Count** | **18** | **17** | **35** |
| Expected Count | 19.3 | 15.7 | 35.0 |
| Residual | -1.3 | 1.3 |  |
| B | **Count** | **9** | **15** | **24** |
| Expected Count | 13.2 | 10.8 | 24.0 |
| Residual | -4.2 | 4.2 |  |
| Total | **Count** | **43** | **35** | **78** |
| Expected Count | 43.0 | 35.0 | 78.0 |

|  |
| --- |
| **Chi-Square Tests** |
|  | Value | df | Asymp. Sig. (2-sided) |
| Pearson Chi-Square | 9.705a | 2 | .008 |
| Likelihood Ratio | 10.488 | 2 | .005 |
| Linear-by-Linear Association | 8.900 | 1 | .003 |
| N of Valid Cases | 78 |  |  |
| a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.53. |