Assignment

When a student makes a test in RemindoTest, we collect a lot of results that later needs to be processed and analyzed to tell the student how he did and to tell the tutors how the questions performed.

In this assignment, we ask you to show how you would implement these important tasks. Given is an Excel-sheet with the results from 452 students from a real exam. Your job is to calculate the grade for each student and decide if they passed the exam. Besides this, you must show us the quality of each question using standard psychometric values.

Calculate the grade

The grade for an exam is calculated using a so-called 'caesura'. A caesura defines how a score is translated to a grade and when a grade is marked as 'passed' or 'failed'.

In this case, we use a caesura that is based on the total score of the candidate relative to the maximum score that can be obtained. If a candidate scores at least 70% of the maximum score, the candidate passes the exam.

The caesura is defined as follows:

- The grade ranges between 1.0 and 10.0 (with 1 decimal)
- When the student scores 20% (or less) of the available points, he receives a 1.0
- When a student scores 70% of the available points, he receives a 5.5 and passes the exam.
- When a student scores 100% of the available points, he receives a 10.0.

Analytics

After calculating the grade, the tutor wants to know how his questions turned out. Are there any questions that were too easy or too difficult? Two useful statistics are the P'-value and the r_{it}-value. The tutor preferably wants to know both values so he can adjust his exam for next year's batch of students.

P'-value

The p'-value is a very basic statistic, which ranges between 0 and 1 and is the quotient of the average score for all candidates of the given question and the maximum score of the question.

In other words, the P' value for question i is calculated as followed:

$$P'_i = \frac{S_{avg_i}}{S_{max_i}}$$

r_{it}-value

The r_{it}-value of a question is a little bit more complicated. 'r_{it}' is the correlation coefficient (also noted as r) between the scores obtained on the question (notes as i from 'item') and the result on the test (noted as t). A correlation ranges between -1 and +1 and tells the tutor how a high score on a particular question relates to the grade of the student. In other words: When a r_{it}-value is very low, it tells the tutor that the question was answered correctly by students that have a very low grade. When the r_{it}-value is very high, it tells the tutor that the question was answered better by students that received a high grade.

In Excel, a correlation efficient can be easily calculated by using the `CORREL` function, but in PHP no such function exists. For more information how to calculate a correlation coefficient, you can refer to (for example) Wikipedia.

Directions

This assignment is meant to determine your problem solving and programming skills. We are not looking for the perfect solution (or a solution that can be used in our products), but try to assess how you would tackle a assignment like this in a real-life situation.

To accomplish this, we have a couple of directions:

- Keep track of the time you spend on this assignment. We think this assignment could be done in a few hours. Do not spend much more than that, but instead tell us what took too much time and why.
- Presentation of the resulting data is not important, so do not spend to much time on that. Your script can be a very simple script that just outputs to the console or to a very very very simple HTML page. We prefer accuracy.
- You do not have to write all the code yourself. If you know a good external library that solves parts of this assignment, you are free to use it as long as you keep it separate from your own code.
- Feeding the PHP code another Excel document in the same format (but with a different amount of questions/students) should also work.