

Many from one:

creating individualized datasets for student assessment from a single one

Abstract

- It is common for the final exam of a statistics course to include questions requiring the selection and application of statistical techniques to a given dataset.
- Due to COVID 19, this past semester final exams were held online.
- To reduce the likelihood of sharing of answers, it would be better if each student received a similar (same variables and structure) but unique (different values) dataset.
- Students of STAT3210: Statistical Techniques in Life Sciences received their final exam questions and datasets (in XLSX format) simultaneously by email.
- Rather than generate 53 (the class size) unique spreadsheets and send them to students individually, we created a single master spreadsheet which is sent to all and produces individualized data upon the entry of a unique identifying number, in this case a student ID number.
- All that is required are simple Excel functions.
- Following the successful application of this "master spreadsheet" last semester, we will use it for take home assignments in 2020-21.

Background

- STAT3210 is offered by the Department of Statistics and traditionally taken by Food and Nutritional Sciences, Biochemistry, Environmental Science majors, among others.
- They learn how to analyse scientific data using appropriate statistical methods, like hypothesis tests and linear regression.
- The format of the 2019/20 STAT3210 final exam was non-invigilated, open-book take-home, with solutions submitted to Veriguide.

The Problem

- Given the nature of the course, it was impossible to avoid a question involving students analysing a dataset using the methods they have learned.
- If students received the same dataset, they may be tempted to share their numerical answers with each other.
- Copying of numerical solutions is often hard to detect.

The Solution

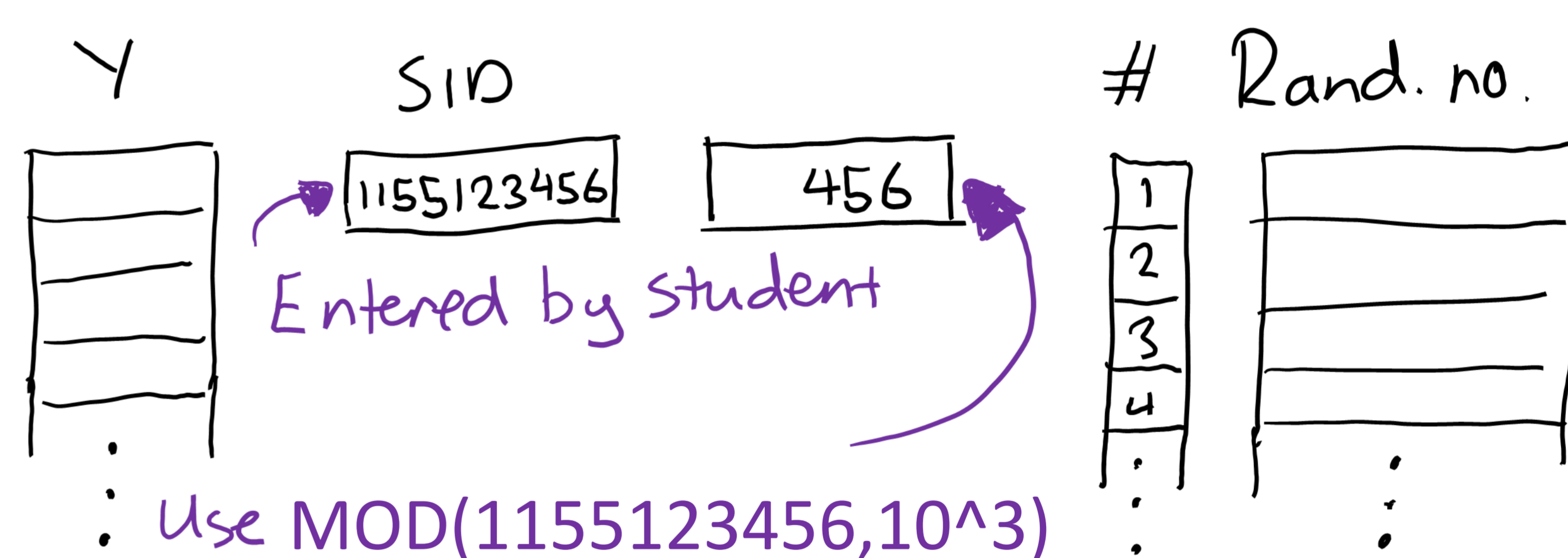
- The dataset was to be distributed in XLSX format.
- So use Excel to create a master spreadsheet, which upon entry of the student's SID, makes it unique.

Method

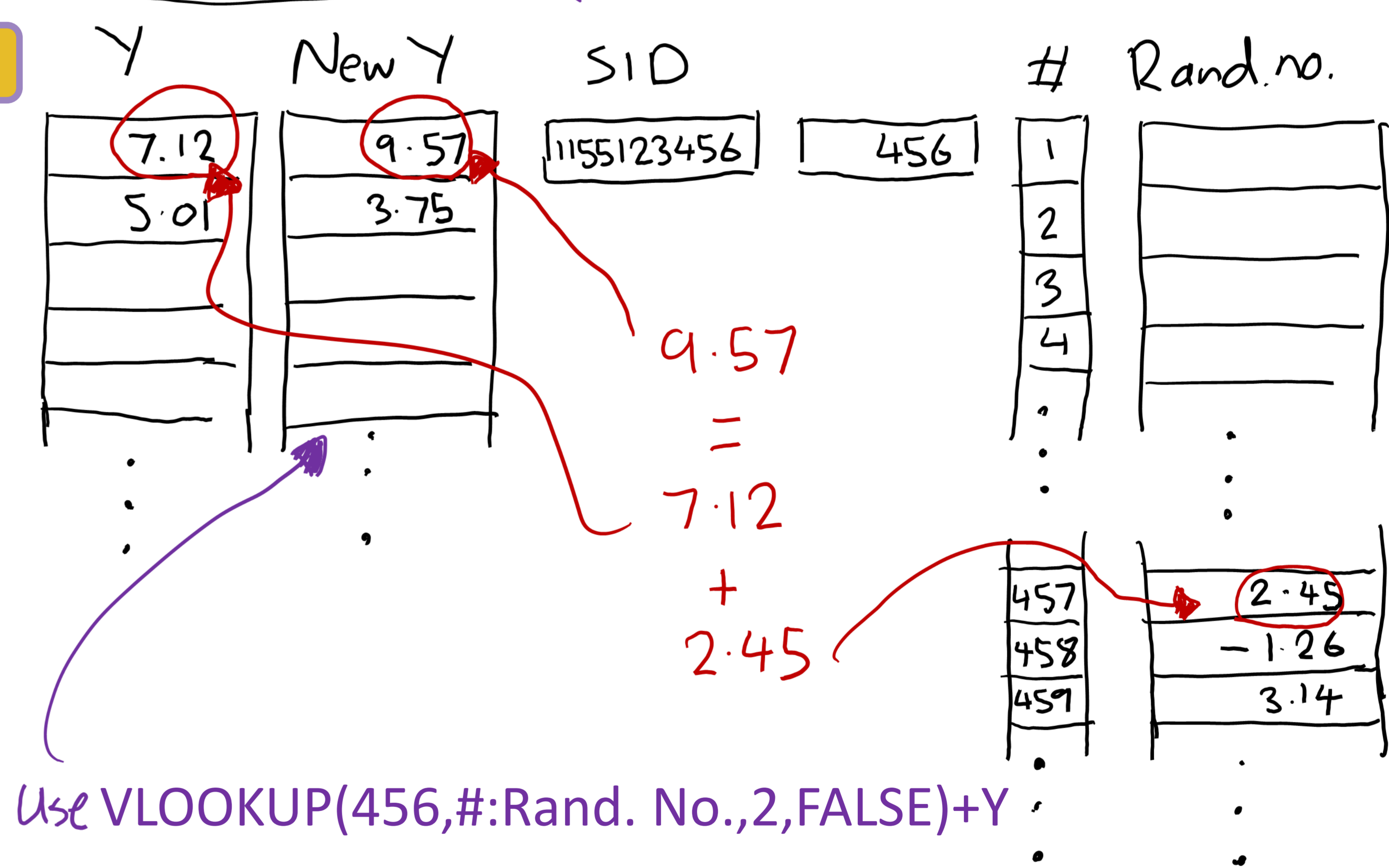
- Start with original data
- Generate random numbers



- Enter SID, find last 3 digits



- Combine 457th to (n+456)th random numbers with original data Voila!



Other Advantages

- Changing data not only changes numerical results, but (possibly) conclusions of hypothesis test results end even which tests are appropriate.
- Easy to mark: set up all possible tests in master spreadsheet, marker enters SID to reveal correct answers.

Future plans

- Incorporate into assignments, where applicable.
- Adapt for different types of data: ordinal, nominal.

