

---

# Web-Created Real Data Worksheets

---

**Neville Davies and  
Bradley Payne**  
Nottingham Trent  
University

neville.davies@ntu.ac.uk

bradley.payne@ntu.ac.uk

The Royal Statistical Society (RSS) Centre for Statistical Education runs an on-going international *CensusAtSchool* project that involves, inter-alia, collecting and analysing school pupils' responses to about 18 simple questions. These fall into three areas of interest, namely details about (1) the pupils, (2) the pupil's household and (3) the pupil's school. The data collected comprise measured, counted and categorical types. See, for example, Connor et al (2000). After suitable anonymisation the data is being added to a large database and random samples from the responses can be taken by visiting the results section of the web site at [www.censusatschool.ntu.ac.uk](http://www.censusatschool.ntu.ac.uk).

Other material, for example the original questionnaires and classroom-based activity worksheets, suitable for teaching statistics and data handling at school level, can be downloaded free of charge in both Adobe .pdf and Word .doc format. In the latter format, the worksheets can be altered to suit local teaching needs and some, after due alteration, may be suitable for use in non-specialist statistics courses at Higher Education (HE) level. There is no copyright on these worksheets, but the RSS Centre for Statistical Education would like to be informed of any innovative alterations to them so that it can make them available for general educational use through dissemination from its web site. Countries currently involved in the project are the UK, Australia, South Africa and Norway, with others, such as New Zealand, set to participate in the near future. The main web site allows access to the sister web sites in these countries.

The *CensusAtSchool* database contains a rich source of material that can be used for teaching and learning at *all* levels. In this article we describe the results of the initial stage of a project that uses data from the database to create worksheets suitable for use in HE by:

- students who wish to carry out activities that will help them learn a range of statistical topics;
- teachers of statistics who may wish to utilise unique worksheets, with solutions, that are generated from randomly selected real data, in their teaching.

The database contains approximately 60,000 entries from the UK, 16,000 from Queensland, Australia and 45,000 from a pilot exercise that has been run in South Africa. More data is being added regularly.

## ***A Dynamic Worksheet Creator***

The five regional statistics and OR workshops, run in late 2000 to ask the UK statistical community their needs and wants from the LTSN MSOR network, yielded several popular requests. The full results of this exercise were reported by Davies (2000). One key need was that teachers of statistics felt they wanted easier access to real data and that, if possible, corresponding access to routine exercises derived from the data that would help them to get students to better learn a range of statistical topics. The notion of taking real data and creating meaningful teaching material from it is sometimes referred to as *adding educational value to raw data*. There are many web sites that, taken together, allow thousands of data sets to be downloaded to PCs and workstations, but there are very few that provide help with teaching and learning activities from the downloaded data.



This is a joint project between  
RSS Centre for Statistical  
Education and LTSN Maths,  
Stats & OR Network



We have developed a browser-enabled worksheet creator that adds value to the *CensusAtSchool* data and allows students and teachers to create and download worksheets, with solutions, based on random samples of values of some of the variables in the database. The web page that is available at <http://worksheet.edev.ntu.ac.uk> gives access to a prototype dynamic worksheet creator that links to the *CensusAtSchool* database. We are in the process of designing and writing several worksheets that will cover a broad range of statistical topics, but the one that we describe here initially covers learning and teaching scatter plotting, simple linear regression and related concepts. On proceeding to the next web page access is given to different worksheet topics.

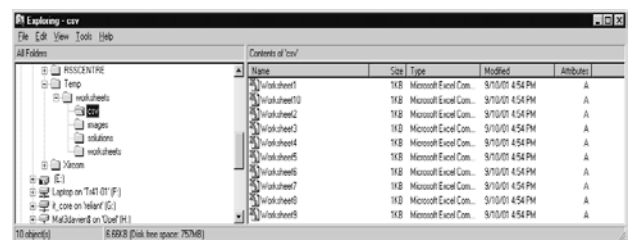


The screen snap above displays a version of the topic selection window in which we have highlighted a worksheet title, the content of which will be generated from the database containing the responses from the school pupils about their heights and length of right foot measured in cm.

After proceeding to the next web page the user can select one of three levels of difficulty, the sample size (up to 200) to be taken from the database for each unique worksheet and, if a teacher has logged in, the number of worksheets required. In the case of a student login to the site, a worksheet control panel allows the worksheet, suggested solution and table of the random sample of responses to be viewed within the web browser. These may be printed as hard copy or saved to the hard disk of the user's local machine. The data can be saved in an Excel file or, alternatively, as a comma separated variable .csv file. In the case of a teacher-login, there is an extra level of security requiring an email address, to which the multiple worksheets will be sent as an attachment, a postal address, phone details and a password. Once all details are verified, the system creates the appropriate number of unique worksheets, their solutions, the data sets as .csv files and any images needed within the worksheets. It then zips them into one file and returns it as an attachment to the email address of the teacher.

**Receiving and Using the Exercise Worksheets**

At the teacher's machine, after unzipping the attached file in the usual way, a TEMPorary folder is used, or created, if necessary. Within the TEMP folder on the local machine, sub folders containing the .csv data files, worksheets, solutions and images are then created. The snapped screen below shows the typical content of the csv sub-folder that, in this example, contains 10 unique data files that a teacher could make available for student use.



The teacher can, of course, choose to use the unique worksheets in any way that is optimal for the course or module being delivered. For example, each worksheet could be emailed to individual students for use at a PC or workstation within or outside the university. Alternatively they could be printed and distributed in hard copy form. In both cases the teacher will have received the data and suggested solution for each of the exercise worksheets. For those colleagues who use university-wide computer-based e-delivery systems such as WebCT or Blackboard, it is possible to incorporate the electronic versions of the worksheets and solutions

into such commercial software, although we have not done this ourselves.

The .csv files of the uniquely generated data can themselves be read into standard commercial statistical software, such as Minitab, SPSS, SAS and so forth. If these, or other, are the favourite analysis packages for a particular course or module, further exercises and analyses can be generated for learning and teaching by using them to perform more statistical analyses.

### Method and Production of Worksheets

An example of part of a prototype worksheet that we have developed, that covers exercises in simple linear regression, is displayed below. The variables are height and length of right foot, and the data comes from a random sample of 100 school pupils' responses to the questionnaire administered in primary and secondary schools in the East Midlands area. The exercise worksheet contains some fairly routine questions about estimating the simple linear regression line of foot size on height, interpretation of the scatter plot and prediction of foot size from a randomly generated height (none of these questions is displayed in the web page below). A version of R (see, for example, the review by Ripley, 2000) running on our server and a customised version of RWeb<sup>1</sup> have been programmed to carry out the routine calculation of the regression estimates and predictions for each worksheet. As the data is generated these are added to the solutions sheet. The exercise worksheets will also contain questions about the nature of any outliers (either or both of high-residual or influential types) *only if* they are present in the random sample taken. The outlier detection will be achieved automatically by programming R to do this in the usual way, but questions about the outliers are only added to the exercise worksheet if the detection is positive for one or both types. The created worksheets are thus highly dependent upon the data that is randomly selected.



### Summary and Future Activity

We are in the early stages of a project in which we are developing a way of crafting unique worksheets from real data. We believe that such a facility will be very useful to both teachers and students of statistics, especially if it is important to guarantee that each student, in carrying out statistical analyses, works with a different set of data. If that data is real and interesting then students could get extra motivation from working with it.

Students are able to access the facility from a web browser and practice their data analytic skills as often as they like, either at times when they have access to the Internet, or when they can work through paper-based versions in the traditional way. Teachers can generate as many unique worksheets and solutions as they wish and these can be disseminated to students electronically or using paper-based methods. For both students and teachers solutions to the unique worksheets are available.

We intend to extend the coverage and usage of the worksheets, both by creating them for key parts of many syllabuses in statistics at different higher education levels and by using other sources of data. For example in the latter case, over the next 3 years the RSS Centre for Statistical Education will be working closely with the Office for National Statistics and we will develop the web-based creator of worksheets and solutions described here to add value to data that can be downloaded from the National Statistics web site.

We invite colleagues who are interested in trying out the individual student worksheets to visit <http://worksheet.edev.ntu.ac.uk> and use the on-line reply facility to send back comments. For those who are interested in helping us to evaluate the web-based multiple worksheet generator, please contact us by email from the above web page, for a user name and corresponding password.

### References

- [1] Connor, D, Davies, N and Holmes, P (2000). CensusAtSchool 2000. *Teaching Statistics*, 22, 66-70.
- [2] Davies, N (2000). Feedback from The Stats/OR community. *MSOR Connections*, 1, 3-4.
- [3] Ripley, B D (2000). The R Project in Statistical Computing. *MSOR Connections*, 1, 23-25.

<sup>1</sup> RWeb is a web-based interface to R developed by Jeff Banfield at Montana University. See for example [www.math.montana.edu/RWeb](http://www.math.montana.edu/RWeb)