Advanced Statistical Analysis in R

Alistair Walsh

September 12, 2014

To do basic data handling, cleaning and summary statistics in R, we need some data. We can use any of the datasets included with R. From the help file of the R datasets package

Details

Student is the pseudonym of William Sealy Gosset. In his 1908 paper he wrote (on page 13) at the beginning of section VI entitled Practical Test of the forgoing Equations:

Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically. The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by W. R. MacDonell (Biometrika, Vol. I., p. 219). The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book, which thus contains the measurements of 3000 criminals in a random order. Finally, each consecutive set of 4 was taken as a sample750 in alland the mean, standard deviation, and correlation of each sample determined. The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, giving us the z of Section III.

The table is in fact page 216 and not page 219 in MacDonell(1902). In the MacDonell table, the middle finger lengths were given in mm and the heights in feet/inches intervals, they are both converted into cm here. The midpoints of intervals were used, e.g., where MacDonell has 4' 7"9/16 – 8"9/16, we have 142.24 which is 2.54*56 = 2.54*(4' 8").

MacDonell credited the source of data (page 178) as follows: The data on which the memoir is based were obtained, through the kindness of Dr Garson, from the Central Metric Office, New Scotland Yard... He pointed out on page 179 that : The forms were drawn at random from the mass on the office shelves; we are therefore dealing with a random sampling.

Source

http://pbil.univ-lyon1.fr/R/donnees/criminals1902.txt thanks to Jean R. Lobry and Anne-Batrice Dufour.