Bivariate Analyses

Héctor Sanz[1,2*], **Isaac Subirana**[3,1,4], **Joan Vila**[1,3]

1. Institut Municipal d'Investigació Mèdica (IMIM), Barcelona, Spain
2. Institut d'Investigació Biomèdica Girona (IDIBGI), Girona, Spain
3. CIBER Epidemilogía y Salud Pública
4. Statistics Department, University of Barcelona, Barcelona, Spain

*Contact author: hsanz@imim.es

Keywords: Software Design, Bivariate Table, LaTeX, Descriptive Analysis

In many studies, especially epidemiological studies, it is important to compare characteristics between independent groups of individuals. Usually these comparisons are presented in the form of tables of descriptive statistics where rows are characteristics, and each column is a group. Tables of this form are usually called bivariate. For example, the comparison between treated and untreated patients (column variable), in terms of age, history of hypertension, triglyceride levels, etc (row variables). Usually the number of characteristics is large, and thus construction of the table is laborious. And if, as often happens, the results must be presented stratified by sex, the process is even more laborious.

For these reasons, we have implemented a function which quickly and efficiently generates bivariate tables in plain text, or as LaTeX instructions. Depending on the nature of the row variable, different statistics can be calculated:

- for numeric, normally distributed variables: mean and standard deviation are calculated, and a Students t-test/ANOVA test is performed.

- for non-normally distributed numeric variables: median, and first and third quartiles are calculated, and a Mann-Whitney U or Kruskal-Wallis test is performed.

- for categorical variables, frequencies are presented, and a Chi-square or Fishers Exact test is performed, as appropriate.

In order to facilitate construction of the bivariate table, we have implemented a graphical interface which allows the user to easily modify a variety of options: for example, specify how many decimals, select individuals to include and variables to analyse, choose whether to display absolute or relative frequencies, etc. All this functionality is also available using R syntax.

To illustrate its way of working, our function uses a data set resulting from a cross-sectional study[1].

Bivariate Table

2010-02-24
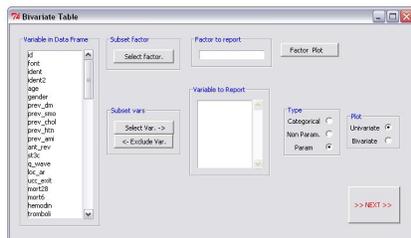
Variable	bbloc_dis		p.value
	No	Yes	
age	68.9 (12.1)	62.3 (12.5)	<0.001
prev_htn: Yes	2045 (49.6%)	2795 (48.6%)	0.323
kill2c: III-IV	1163 (28.4%)	361 (6.32%)	<0.001
delay_r: <3h	577 (15.7%)	1077 (22.4%)	<0.001
delay_r: >=3 & <6h	454 (12.4%)	646 (13.4%)	<0.001
delay_r: >=6h	165 (4.50%)	188 (3.91%)	<0.001
delay_r: Not performed or delayed >12h	2473 (67.4%)	2901 (60.3%)	<0.001
r_ini_mo	140 (62.0-340)	120 (60.0-245)	<0.001

Table 1: My first table

References

1. Tomas M, Vazquez E, Fernandez-Fernandez JM, Subirana I, Plata C, Heras M, Vila J, Marrugat J, Valverde MA, Senti M. (2008). Genetic variation in the KCNMA1 potassium channel alpha subunit as risk factor for severe essential hypertension and myocardial infarction. *J Hypertens*, 26(11):2147-53.