Humboldt-Universität zu Berlin

Ladislaus von Bortkiewicz Chair for Statistics

Data analysis I

Sigbert Klinke

May 24, 2016

- Quick-R: `http://www.statmethods.net`
- UCLA Institute for digital research and education

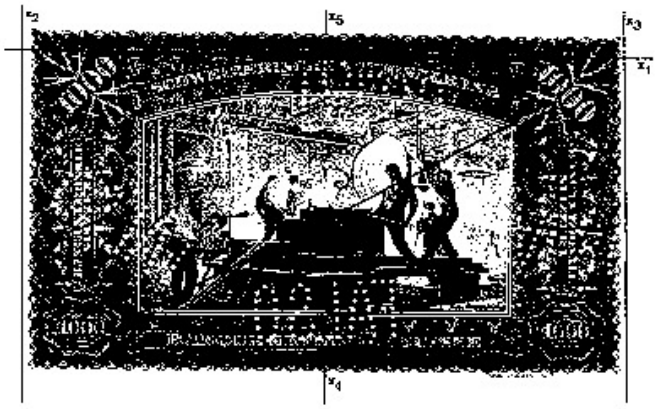
Choose R, SPSS or any other statistical package to fulfill the following tasks. If you do not know the software good enough then use the help of your software or search for help in the internet.

1. Read in the `ALLBUS2014` data set (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften).

`V7`	ERHEBUNGSGEBIET <WOHNGEBIET>: WEST - OST
`V10`	WIRTSCHAFTSLAGE, BEFR. HEUTE
`V11`	WIRTSCHAFTSLAGE, BEFR. IN 1 JAHR
`V417`	BEFR.: NETTOEINKOMMEN, OFFENE ABFRAGE
`V418`	BEFR.: NETTOEINKOMMEN, LISTENABFRAGE
`V419`	BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>
`V868`	BUNDESLAND, IN DEM BEFRAGTE<R> WOHNT

2. (a) Run a χ^2 independence test between the variables BUNDESLAND and BEFRAGTE NETTOEINKOMMEN. What do you observe? Could you improve your result?
   (b) Compare the variables WIRTSCHAFTSLAGE, BEFR. HEUTE (`V10`) and WIRTSCHAFTSLAGE, BEFR . IN 1 JAHR (`V11`). Are they independent?
   (c) Redo the last exercise with a random sample of 50% and 10% of the observations. What do you notice?
   (d) For each of the last two exercises compute Cohen's w.
3. Read in the `BANK2` data set. The data are taken from the book of Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A practical approach.* London: Chapman & Hall. They are about the measurement of 200 genuine and forged old Swiss bank notes.

WIDTH (X1)	Width of the bank note
LEFT (X2)	Height of the bank note, measured on the left
RIGHT (X3)	Height of the bank note, measured on the right
LOWER (X4)	Distance of inner frame to the lower border
UPPER (X5)	Distance of inner frame to the upper border
DIAGONAL (X6)	Length of the diagonal



4. The first half of the observations consists of genuine bank notes, the second half of forged bank notes.

   (a) Which level of measurement do the six variabels have?

   (b) By numerical inspection answer: would you agree that all variables are continuous and metric? (see `unique`)

   (c) Compute some (appropriate) descriptive statistics (e.g. central tendency, dispersion). What do you notice?

   (d) For each group (genuine, forged) compute the same statistics as before. Is there any single variable which might help to distingish between forged and genuine bank notes?

   (e) Compute an effect size for the means in each group.

   (f) Compute a test on the mean and median for the forged bank notes. Assume that the hypothetical value is the mean/median of the genuine bank notes.