
Introduction to R for Social Sciences

Εισαγωγή και αποθήκευση δεδομένων

Αναστάσιος Εμβαλωτής & Αικατερίνη Σαργιώτη

Εισαγωγή δεδομένων από αρχείο Excel

- Αποθήκευση του αρχείου δεδομένων σε αρχείο .xls ή .xlsx
- Εγκατάσταση του πακέτου **readxl**
 - *install.packages("readxl")*
- Εισαγωγή δεδομένων ακολουθώντας τα βήματα:
 - **File** → **Import Dataset** → **From Excel**

Εισαγωγή δεδομένων από αρχείο SPSS

- Αποθήκευση του αρχείου δεδομένων σε αρχείο .spv
- Εγκατάσταση του πακέτου **haven**
 - `install.packages("haven")`
- Εισαγωγή δεδομένων ακολουθώντας τα βήματα:
 - **File** → **Import Dataset** → **From SPSS**
- Η R εκτυπώνει 1000 δεδομένα
 - Αλλαγή ορίου εκτύπωσης δεδομένων με την εντολή `options(max.print=x)`

Επιλογή μεταβλητών από σύνολο δεδομένων

- Νέο αρχείο δεδομένων με συγκεκριμένες μεταβλητές αλλά όλα τα υποκείμενα (περιπτώσεις)
 1. Ορισμός μεταβλητών νέου αρχείου
 - `myvars = c("var_a", "var_b", "var_c",...)`
 2. Ανάγνωση του νέου αρχείου δεδομένων συναρτήσει του παλιού και των μεταβλητών που ορίστηκαν
 - `dataset_new = dataset_old[myvars]`

Επιλογή περιπτώσεων από σύνολο δεδομένων

- Νέο αρχείο δεδομένων με όλες τις μεταβλητές αλλά συγκεκριμένα υποκείμενα (περιπτώσεις)
 - Χρήση της συνάρτησης `subset()`
1. Ανάγνωση του νέου αρχείου συναρτήσει του αρχικού και της περίπτωσης που θέλει να συμπεριλάβει
 - `dataset_new_2 = subset(dataset_new, var_a == value)` → η μεταβλητή που ορίζει τις περιπτώσεις είναι αριθμητική
 - `dataset_new_2 = subset(dataset_new, var_a == "value")` → η μεταβλητή που ορίζει τις περιπτώσεις είναι αλφαριθμητική

Επιλογή περιπτώσεων από σύνολο μεταβλητών

- Νέο αρχείο δεδομένων με συγκεκριμένες μεταβλητές και συγκεκριμένα υποκείμενα (περιπτώσεις)
 - Χρήση της συνάρτησης `subset()`
1. Ανάγνωση του νέου αρχείου συναρτήσει του αρχικού, της περίπτωσης που θέλει να συμπεριλάβει και των μεταβλητών που θα απομονωθούν στο νέο αρχείο
 - `dataset_new_3a = subset(dataset_new_2, var_a == value, select=c(var_c, var_d))`
 - `dataset_new_3b = subset(dataset_new_2, var_a == value & var_b == value, select=c(var_c, var_d))`

Αποθήκευση δεδομένων σε αρχείο

- Αποθήκευση δεδομένων σε αρχείο tab-delimited
 - `write.table()`
- Αποθήκευση δεδομένων σε αρχείο CSV
 - `write.csv()`
- Στις παρενθέσεις ο χρήστης γράφει το όνομα του αρχείου, ο προορισμός αποθήκευσης και η μορφή αποθήκευσης
- Και οι δύο παραπάνω μορφές μπορούν να εισαχθούν σε SPSS, excel, κλπ.