
Εισαγωγή στο Stata

Βιοστατιστική στην επιδημιολογική έρευνα II

A. Αναλυτής, Βιοστατιστικός, ΕΤΕΠ
Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής,
Ιατρική Σχολή ΕΚΠΑ
aanalit@med.uoa.gr

Εισαγωγή

- <http://www.stata.com>
 - Πλήρης γλώσσα προγραμματισμού – «στατιστικό πακέτο»
 - Διαθέσιμο για Windows, Unix και Mac
 - μία άδεια για όλες τις πλατφόρμες
 - τα αρχεία δεδομένων (datasets) λειτουργούν και στα τρία λειτουργικά
-

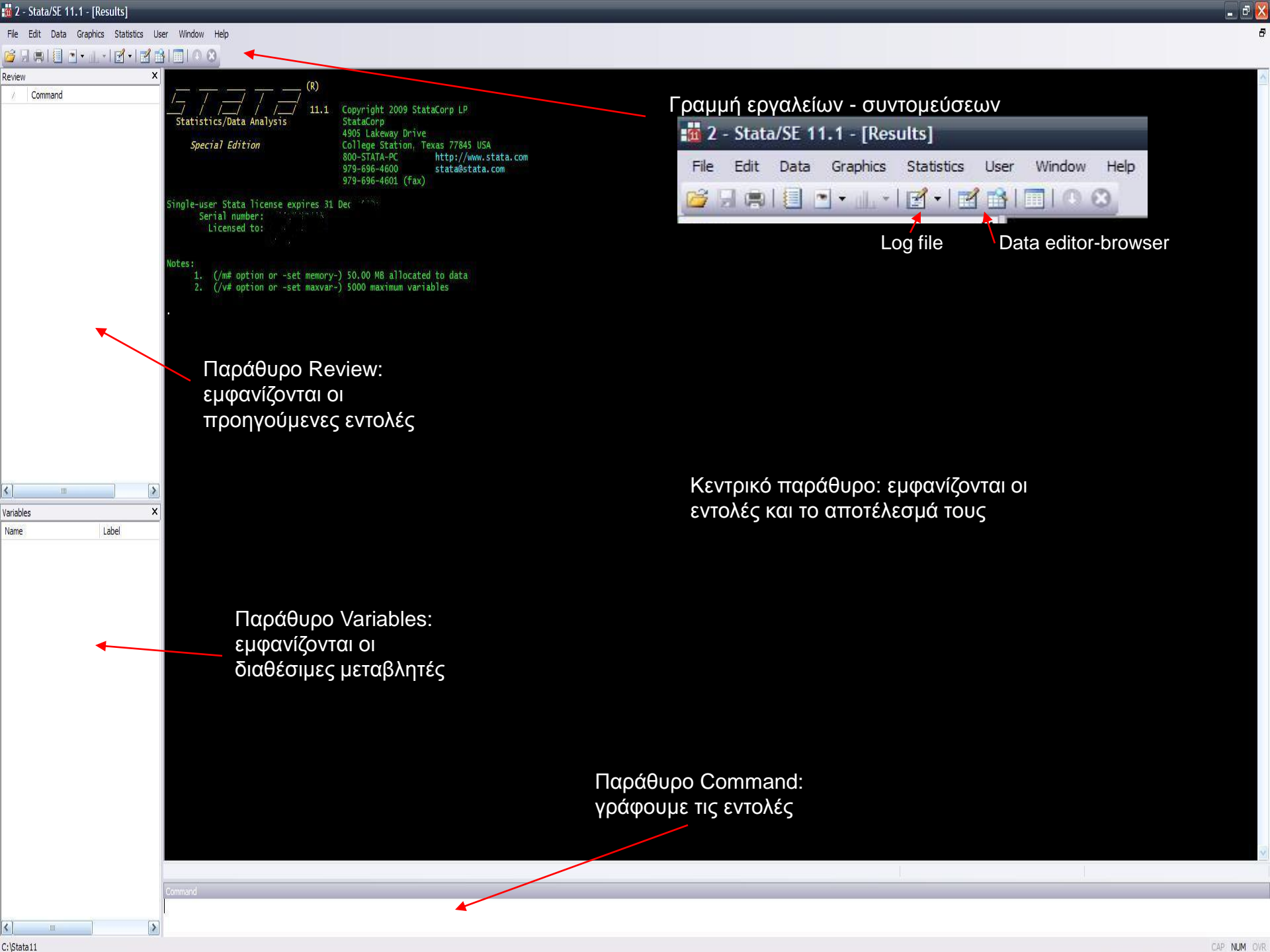
- 4 βασικές εκδόσεις: small, IC, SE, MP

- ίδιες αναλυτικές δυνατότητες - ρουτίνες
- διαφορετικές δυνατότητες σχετικά με το μέγεθος των δεδομένων

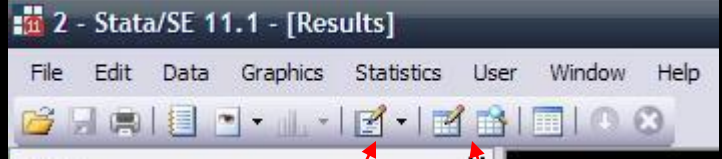
Έκδοση	Μεταβλητές	Παρατηρήσεις	Parallel processing	Λειτουργικό
Stata MP	32,767	20 billion*	Ναι	Win Mac, Unix
Stata SE	32,767	2.14 billion*	Όχι	Win, Mac, Unix
Stata IC	2,047	2.14 billion*	Όχι	Win, Mac, Unix
Small Stata (φοιτητές)	99	1,200	Όχι	Win, Mac, Unix

* Καθορίζεται από τη διαθέσιμη μνήμη

-
- Portable – δεν χρειάζεται εγκατάσταση
 - Γραμμή εντολών και (από την έκδοση 8) παραθυρικό περιβάλλον
 - Πλήρες εγχειρίδιο χρήσης (pdf)
 - Ενεργή κοινότητα χρηστών με ρουτίνες ανάλυσης και από χρήστες
 - Ανταγωνιστικό σε σχέση με άλλα στατιστικά πακέτα (R, SPSS, SAS) (δομή εντολών, ευκολία εκμάθησης, υποστήριξη, κόστος)
-



Γραμμή εργαλείων - συντομεύσεων



Log file Data editor-browser

Παράθυρο Review:
εμφανίζονται οι
προηγούμενες εντολές

Κεντρικό παράθυρο: εμφανίζονται οι
εντολές και το αποτέλεσμα τους

Παράθυρο Variables:
εμφανίζονται οι
διαθέσιμες μεταβλητές

Παράθυρο Command:
γράφουμε τις εντολές

Βασικές αρχές

- Χρήση κατάλληλων ονομάτων στις μεταβλητές (case sensitive), τις ετικέτες αλλά και στα αρχεία δεδομένων
- Προσοχή κατά τον χειρισμό των δεδομένων για να μην χαθούν τα αρχικά δεδομένα
- Χρήση συντομεύσεων στις εντολές
- Διατήρηση αρχείου των εντολών (.do file) /αποτελεσμάτων (.log file) με επαρκή σχολιασμό

Εισαγωγή στο Stata (εφαρμογή)

Διαχείριση δεδομένων- Γραφική απεικόνιση

- Το αρχείο δεδομένων bwt.xls περιλαμβάνει τις παρακάτω μεταβλητές από ένα δείγμα 100 νεογνών στις ΗΠΑ:

Μεταβλητή	Περιγραφή
id	A/A
age	Age of mother (yrs)
heightm	Height of mother (m)
weightkg	Weight of mother (kg)
race	Race of mother (1=White, 2=Black, 3=Other)
smoke	Mother smokes during pregnancy 0=no και 1 =yes
birthweight	Birth weight (g)
prog1	Progesterone during 16th week
prog 2	Progesterone during 27th week

- Σκοπός της μελέτης είναι να διερευνηθούν οι παράγοντες που σχετίζονται με το βάρος του νεογνού κατά τη γέννηση

Εισαγωγή στο Stata (εφαρμογή)

Διαχείριση δεδομένων- Γραφική απεικόνιση

- Δείτε την κατανομή όλων των τιμών των μεταβλητών της έρευνας

Format εντολής Stata:

[prefix:] command [varlist][if expression][in range][weight][, options]

```
cd "C:\Users\...\Desktop\Stata"
```

```
** read data from excel
```

```
describe
```

```
summarize age heightm weightkg birthweight prog1 prog2
```

```
tabulate race
```

```
tab smoke
```

- Δώστε **label** σε κάθε μεταβλητής όπως δίνεται στον παραπάνω Πίνακα

```
label variable age "Age of mother (Kg)"
```

```
lab var heightm "Height of mother (m)"
```

```
...
```


Εισαγωγή στο Stata (εφαρμογή)

Διαχείριση δεδομένων- Γραφική απεικόνιση

- Δώστε **label** στις τιμές της μεταβλητή *race* 1=White, 2=Black, 3=Other και της *smoke* 0=no και 1 =yes

```
label define rl 1 "White" 2 "Black" 3 "Other"  
label values race rl  
lab def sl 0 "no" 1 "yes"  
lab val smoke sl
```

Ορίστε για τις μεταβλητές *age* και *weightkg* ως **missing value** τη τιμή 999

```
replace age=. if age==999  
replace weightkg=. if weightkg==999  
summarize age weightkg  
histogram age  
hist weightkg, normal bin(7)  
graph pie, over(race) plabel(_all percent)
```

Εισαγωγή στο Stata (εφαρμογή)

Διαχείριση δεδομένων- Γραφική απεικόνιση

- Δείτε την κατανομή των φυλών ξεχωριστά για καπνίστριες και μη

```
tabulate race if smoke==0
```

```
tabulate race if smoke==1
```

εναλλακτικά

```
sort smoke
```

```
by smoke: tab race
```

- Δημιουργείστε μία νέα μεταβλητή με όνομα *lowbwt* (**label** *low birth weight*) και τιμές 0= « \geq 2500g» και 1=«<2500g»

```
generate lowbwt=1 if birthweight<2500
```

```
replace lowbwt=0 if birthweight>=2500
```

εναλλακτικά

```
recode birthweight (0/2499.99=1) (2500/10000=0), gen(low)
```

Εισαγωγή στο Stata (εφαρμογή)

Διαχείριση δεδομένων- Γραφική απεικόνιση

- Δημιουργήστε μία καινούργια μεταβλητή, το δείκτη μάζας σώματος της μητέρας, με όνομα `bmi = weight/(height*height)`

```
gen bmi=weightkg/(heightm*heightm)
```

- Δημιουργείστε μία νέα μεταβλητή με όνομα `bmi4` (**label** `bmi WHO`) και τιμές 1= «<18.5 kg/m²», 2= «18.5 – <25 kg/m²», 3= «25 – <30 kg/m²» και 4= «≥30 kg/m²»

```
recode bmi (0/18.499=1 "underweight") (18.5/24.999=2 "normal") (25/29.999=3 "overweight") (30/500=4 "obese"), gen(bmi4)
```

- Εξετάστε με γράφημα τη σχέση μεταξύ βάρους γέννησης και δείκτη μάζας σώματος της μητέρας

```
scatter birth bmi
```

Εισαγωγή στο Stata (εφαρμογή)

Έλεγχοι υποθέσεων – t-test

- Εξετάστε πιθανή συσχέτιση του βάρους γέννησης με τις καπνιστικές συνήθειες της μητέρας

```
log using lab.log
sort smoke
by smoke: sum birthweigh
hist birthwe, by(smoke) normal
by smoke: swilk birthwe
graph box birthwe, over(smoke)
ttest birthwei, by(smoke)
```

- Εξετάστε πιθανή διαφορά στις τιμές της προγεστερόνης στις 2 χρονικές στιγμές μέτρησης

```
gen dprog= prog1- prog2
hist dprog, normal
ttest prog1= prog2
```

Εισαγωγή στο Stata (εφαρμογή)

Έλεγχοι υποθέσεων – χ^2 , regression

- Εξετάστε πιθανή συσχέτιση της φυλής με τις καπνιστικές συνήθειες

```
graph pie, over(race) by(smoke) plabel(_all percent)  
tab race smoke, column expected chi2 exact
```

- Ποιοι παράγοντες επηρεάζουν το βάρος γέννησης;

```
scatter birthweight age  
graph box birthw, over(race)
```

```
regress birthweight age  
regress birthweight height  
reg birth smoke  
reg birth i.race  
reg birth age heigh i.race
```

Εισαγωγή στο Stata (εφαρμογή)

Έλεγχοι υποθέσεων – logistic regression

- Σχετίζεται κάποιος παράγοντας με χαμηλό βάρος γέννησης;

```
generate lowbwt=1 if birthweight<2500
```

```
replace lowbwt=0 if birthweight>=2500 & birthweight!=.
```

```
logit lowbwt age, or
```

```
logit lowbwt i.race, or
```

```
logit lowbwt age height i.race, or
```

```
log close
```
