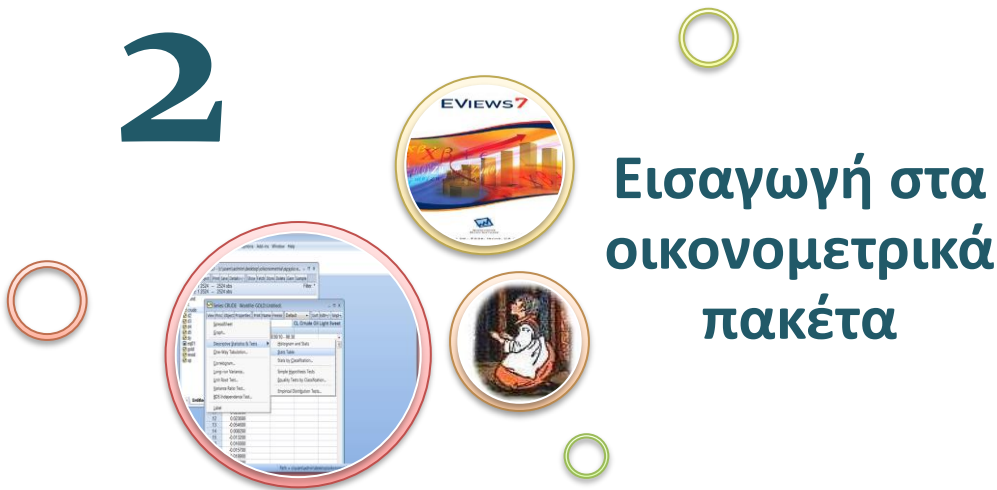


2



Εισαγωγή στα οικονομετρικά πακέτα

Στο βιβλίο αυτό θα δοθεί ιδιαίτερη έμφαση στην ανάλυση, εκτίμηση και παρουσίαση υποδειγμάτων με τη χρήση των οικονομετρικών πακέτων **gretl** και **EViews**. Η επιλογή αυτών των στατιστικών πακέτων έγινε με βάση διάφορα κριτήρια, μεταξύ των οποίων τα εξής: Το **gretl** γιατί είναι διαθέσιμο χωρίς κόστος (και συνεπώς μπορεί να χρησιμοποιηθεί από τους φοιτητές στα πλαίσια των εργαστηρίων και των εργασιών τους) σε πολλές πλατφόρμες Η/Υ και διαθέτει όλα τα χαρακτηριστικά ενός πλήρους εμπορικού οικονομετρικού πακέτου, και το **EViews** γιατί τα τελευταία χρόνια είναι από τα πλέον διαδεδομένα προγράμματα και έχει κατακτήσει μεγάλο μερίδιο αγοράς.

Το **gretl** είναι ένα γρήγορο, πλήρες, αποτελεσματικό και ολοκληρωμένο οικονομετρικό πακέτο λογισμικού, που προσφέρει όλα όσα απαιτούνται για την ανάλυση, εκτίμηση και πρόβλεψη στατιστικών μοντέλων. Το **gretl** δεν πωλείται, αλλά διατίθεται δωρεάν στη διεύθυνση¹ <http://gretl.sourceforge.net> και είναι λογισμικό ανοιχτού κώδικα, που μπορεί να εγκατασταθεί όχι μόνο σε Windows αλλά και σε διάφορες πλατφόρμες του Linux. Το **EViews** είναι ένα εύχρηστο και πολύ διαδεδομένο στατιστικό πακέτο ανάλυσης δεδομένων, το οποίο χρησιμοποιείται ευρύτατα στην ανάλυση οικονομικών και χρηματοοικονομικών χρονοσειρών. Τα τελευταία χρόνια το **EViews** έχει κατακτήσει και άλλους επιστημονικούς τομείς, καθώς οι τελευταίες εκδόσεις του έχουν ενσωματώσει ρουτίνες που βοηθούν στην ανάλυση και επεξεργασία και άλλων τύπων δεδομένων. Σε γενικές γραμμές, το **gretl** και το **EViews** μπορούν να εκτελέσουν τις παρακάτω εργασίες:

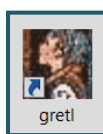
- Στατιστική ανάλυση και αξιολόγηση δεδομένων
- Εκτίμηση γραμμικών και μη γραμμικών υποδειγμάτων
- Εκτίμηση υποδειγμάτων συστημάτων εξισώσεων
- Πρόβλεψη
- Προσομοίωση

2.1. Εισαγωγή στο **gretl**

Η εκκίνηση του **gretl** γίνεται με διπλό κλικ στο εικονίδιό του το οποίο τοποθετείται αυτόματα στην επιφάνεια εργασίας με την εγκατάσταση του προγράμματος:


¹ Για να εγκατασταθεί το **gretl** σε έναν υπολογιστή με Microsoft Windows θα πρέπει να φορτωθεί το κατάλληλο εκτελέσιμο αρχείο από το διαδικτυακό τόπο:

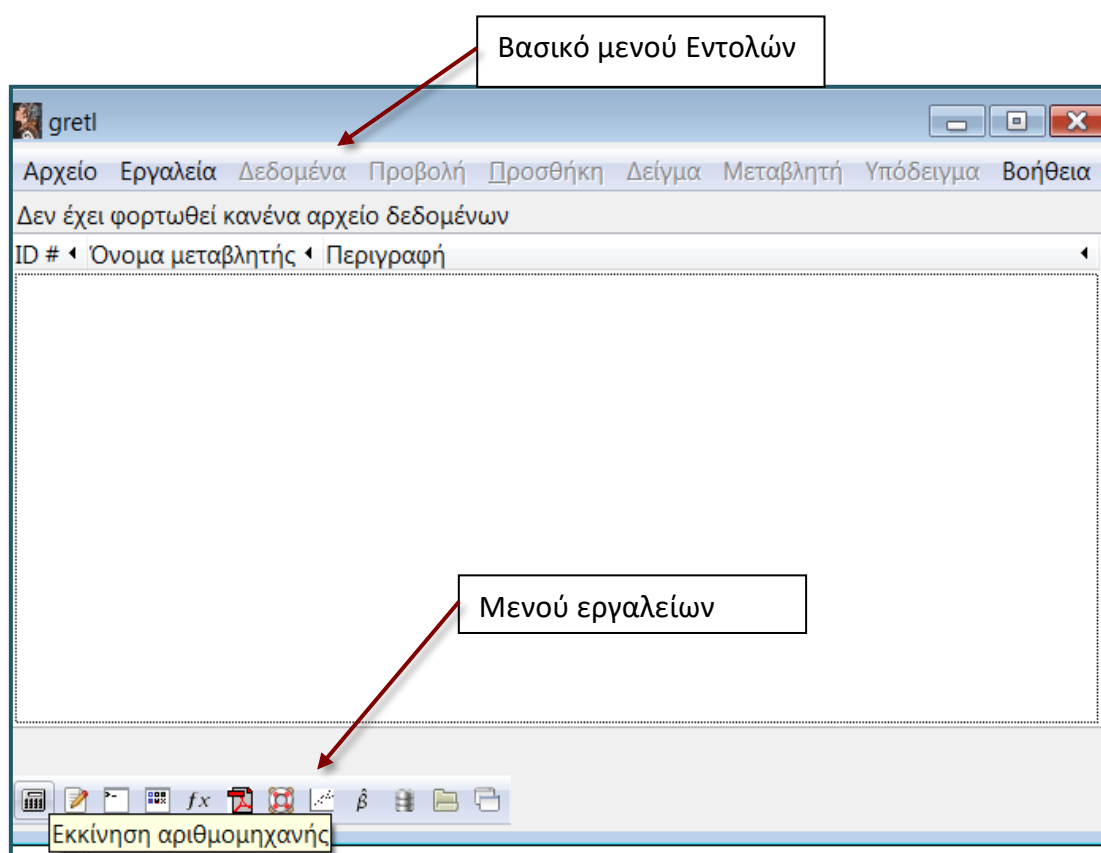
- <http://gretl.sourceforge.net/win32/>



Εικόνα 2.1: Το εικονίδιο του gretl

Το πρόγραμμα περιήγησης του gretl προσφέρεται στα ελληνικά και εύκολα διακρίνουμε το:

- **Βασικό μενού εντολών** διαμέσου του οποίου γίνεται η εισαγωγή, διαχείριση και ανάλυση των δεδομένων, καθώς και η εξαγωγή των τελικών εκτιμήσεων και αποτελεσμάτων.
- **Μενού εργαλείων** το οποίο υποστηρίζει το χρήστη με πλείστα βοηθήματα², όπως είναι για παράδειγμα το πρώτο αριστερά εικονίδιο , η **Αριθμομηχανή**.



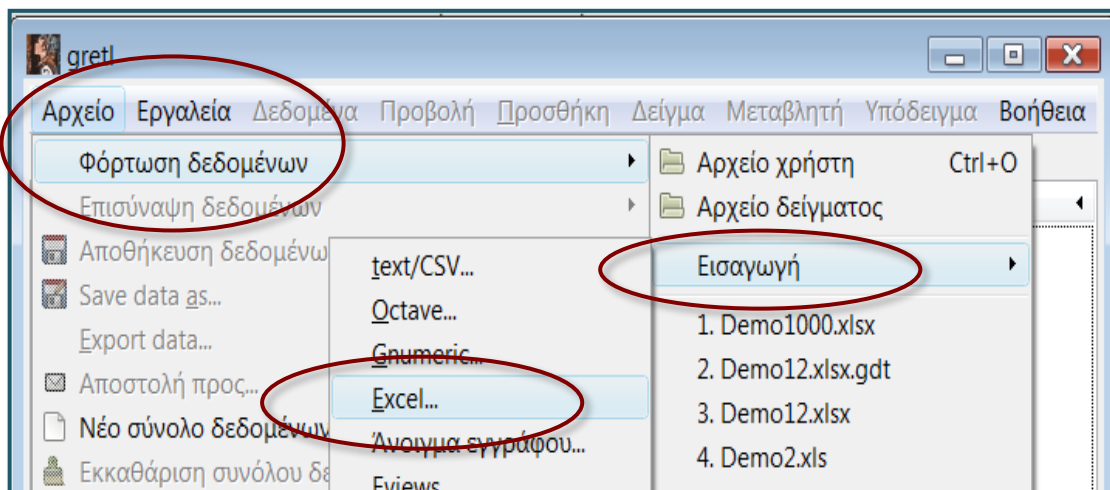
Εικόνα 2.2: Βασική οθόνη περιήγησης του gretl

2.1.1. Εισαγωγή δεδομένων στο gretl

Το gretl υποστηρίζει την εισαγωγή δεδομένων από ποικίλα υπολογιστικά προγράμματα όπως είναι το Excel, το EViews, το SPSS, το Stata, το SAS κλπ. Η διαδικασία εισαγωγής των δεδομένων από το Excel περιλαμβάνει τα εξής βήματα:

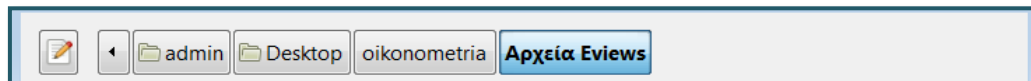
² Όταν το βέλος του ποντικιού πλησιάσει το βοήθημα, τότε εμφανίζεται μια ένδειξη με την ταυτότητα του βοηθήματος, π.χ. **Εκκίνηση αριθμομηχανής**

Επιλέγουμε *Αρχείο/Φόρτωση δεδομένων/Εισαγωγή/Excel...*

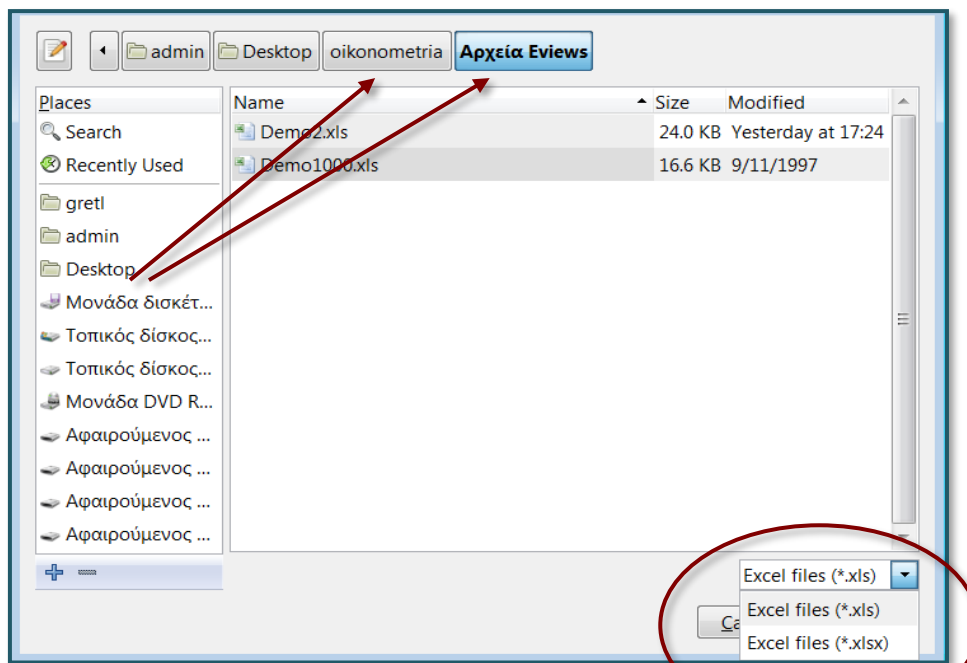


Εικόνα 2.3: Επιλογή τύπου αρχείου για την εισαγωγή δεδομένων

Στο αναδυόμενο παράθυρο διαλόγου αναζητούμε το υπό ανάλυση αρχείο “macro”. Η αναζήτηση του αρχείου δύναται να ξεκινήσει με την επιλογή *Desktop (επιφάνεια εργασίας)*, που βρίσκεται αριστερά του παραθύρου, κάτω από την ένδειξη *places (τοποθεσίες προς αναζήτηση)*. Θα πρέπει να γνωρίζουμε ακριβώς τη διαδρομή στην οποία θα αναζητήσουμε το αρχείο που θέλουμε να εισάγουμε στο πρόγραμμα. Η διαδρομή για την εύρεση του υπό ανάλυση αρχείου απεικονίζεται στο οριζόντιο πλαίσιο, στο πάνω μέρος του παραθύρου:

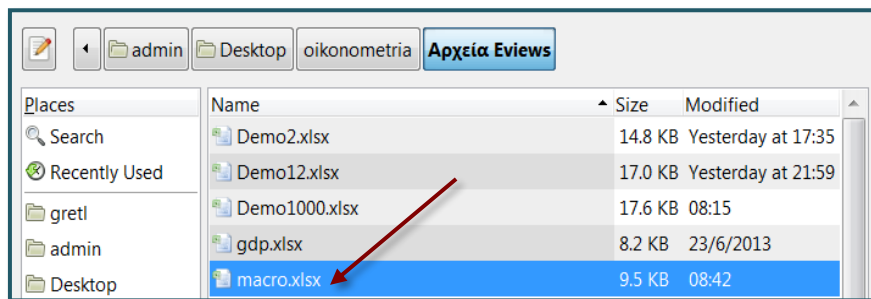


Εικόνα 2.4: Διαδρομή εύρεσης του αρχείου



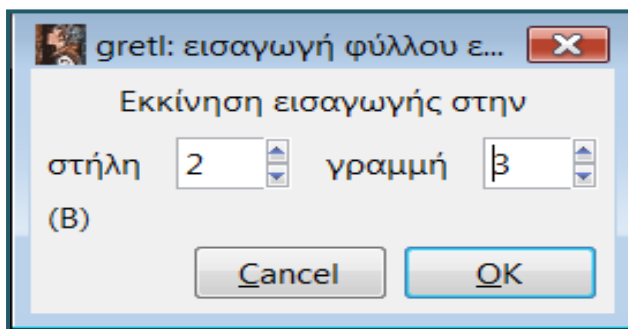
Εικόνα 2.5: Επιλογή διαδρομής αρχείου για την εισαγωγή δεδομένων

Το αρχείο **macro** βρίσκεται στον υποφάκελο **Αρχεία EViews**, αλλά επειδή έχει κατάληξη³ **.xlsx** δεν είναι ορατό (Εικόνα 2.5). Πατώντας το βελάκι ανοίγουμε τον κατάλογο **Excel Files (*.xls)** και επιλέγουμε **Excel files (*.xlsx)**. Έχουν πλέον εμφανιστεί όλα τα αρχεία με κατάληξη **.xlsx** και επιλέγουμε το αρχείο **macro** (Εικόνα 2.6):



Εικόνα 2.6: Επιλογή αρχείου στο gretl

Στο αναδυόμενο παράθυρο ορίζουμε τη **στήλη 2** και **γραμμή 3** του Excel από τις οποίες ξεκινούν οι μεταβλητές και τα δεδομένα.



Εικόνα 2.7: Εκκίνηση διαδικασίας εισαγωγής δεδομένων στο gretl

- Πράγματι, εάν ανοίξουμε το αρχείο **macro.xlsx** θα διαπιστώσουμε ότι⁴

³ Η κατάληξη **.xls** αφορά στα αρχεία Excel μέχρι την έκδοση Office 2003, ενώ η κατάληξη **.xlsx** αφορά στα αρχεία Excel μετά την έκδοση Office 2007.

⁴ Σημειώνεται ότι δύναται να προκύψει μήνυμα σφάλματος στην περίπτωση που:

- τα ονόματα των μεταβλητών είναι με ελληνικούς χαρακτήρες.
- κάποιο από τα πεδία του Excel, εκτός των μεταβλητών και δεδομένων, που βρίσκεται μετά τις ορισθείσες στήλη 2 και γραμμή 3, περιέχει κάποιο αλφαριθμητικό στοιχείο.
- οι μεταβλητές δεν είναι παρατεταγμένες η μία μετά την άλλη (παρεμβολή, για παράδειγμα, μιας άδειας στήλης).
- η εισαγωγή δεν αφορούσε ακριβώς τη γραμμή και στήλη που βρίσκονται τα ονόματα των μεταβλητών και τα δεδομένα. Για παράδειγμα, εάν είχαμε ορίσει παραπάνω **γραμμή 2** αντί **3** θα είχε προκύψει μήνυμα σφάλματος.

Επιπλέον, μπορεί να δημιουργηθούν προβλήματα ορθότητας των δεδομένων εάν στις τοπικές ρυθμίσεις του Window δεν επιλέξουμε ως υποδιαστολή την τελεία **."** Η ρύθμιση αυτή αφορά τον προσδιορισμό του διαχωριστικού χιλιάδων και δεκαδικών. Τα διαχωριστικά αυτά είναι διαφορετικά στις ΗΠΑ και σε άλλες χώρες (χρησιμοποιούν ως διαχωριστικό της χιλιάδας το **,"** και του δεκαδικού την **."**).

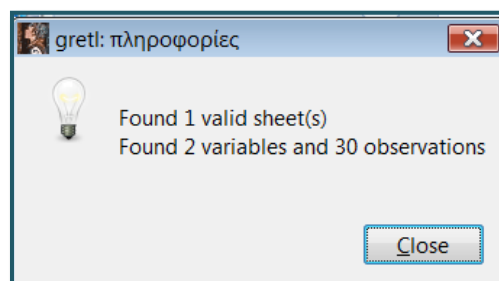
Τέλος, οι ετήσιες χρονολογίες στο Excel θα πρέπει να είναι ορισμένες ως ημερομηνίες διαμέσου της επιλογής **Μορφοποίηση κελιών/Ημερομηνία**, διαφορετικά το gretl μετατρέπει με λανθασμένο τρόπο τη χρονολογία σε ημερομηνία. Εάν επιθυμούμε την εμφάνιση μόνο του έτους στο Excel, τότε αυτό μπορεί να επιτευχθεί με τη συνάρτηση Year.

- στη γραμμή 3 βρίσκονται τα ονόματα των μεταβλητών
- στη στήλη 2 (B) βρίσκονται οι ημερομηνίες.

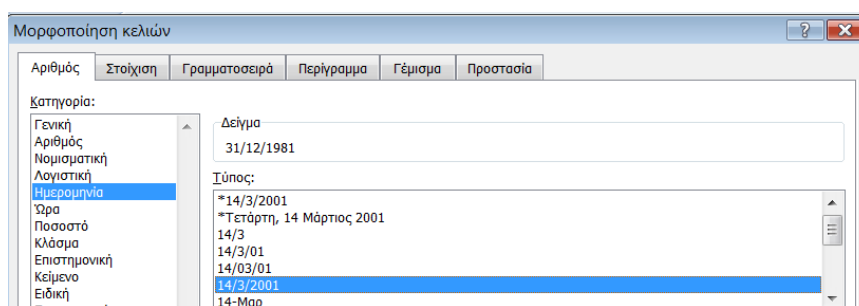
	A	B	C	D	E
1					
2			total (% of total labor force)	(annual %)	
3			Unemployment	GDP growth	
4		31/12/1981	4.0	-1.554	
5		31/12/1982	5.8	-1.133	
6		31/12/1983	7.8	-1.079	
7		31/12/1984	8.1	2.011	
8		31/12/1985	7.8	2.510	
9		31/12/1986	7.4	0.518	
10		31/12/1987	7.4	-2.259	
11		31/12/1988	7.7	4.288	

Εικόνα 2.8: Φύλλο εργασίας με τα δεδομένα προς εισαγωγή στο gretl

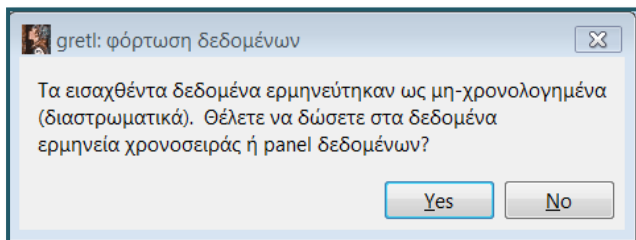
Το gretl με μήνυμα μας πληροφορεί για το πλήθος των δεδομένων και τον αριθμό των μεταβλητών που έχει αποκωδικοποιήσει. Στην προκειμένη περίπτωση, όπως βλέπουμε στην εικόνα 2.9 έχει βρει **2 μεταβλητές των 30 παρατηρήσεων (found 2 variables and 30 observations)**.



Εικόνα 2.9: Παράθυρο πληροφοριών για τα δεδομένα προς εισαγωγή στο gretl

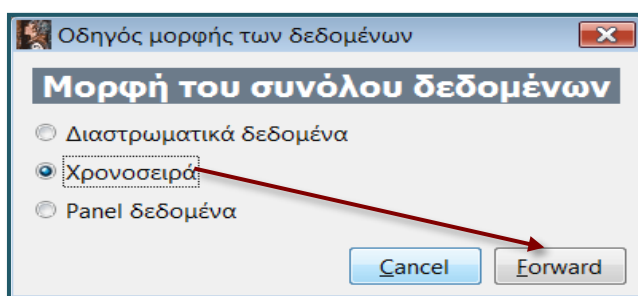


Στη συνέχεια και εφόσον οι μεταβλητές συνοδεύονται από σειρά ημερομηνιών (στήλη B στο Excel), ακολουθεί παράθυρο διαλόγου για τον προσδιορισμό των δεδομένων ως **διαστρωματικά**, **panel**⁵ ή **χρονοσειράς**.



Εικόνα 2.10: Αρχικό παράθυρο προσδιορισμού τύπου δεδομένων στο gretl

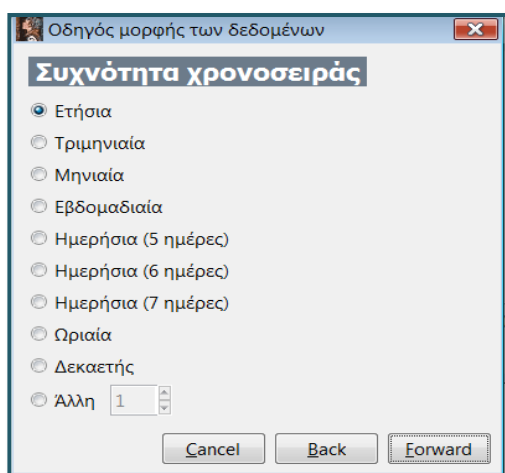
Απαντώντας θετικά (**Yes**) αναδύεται νέο παράθυρο διαλόγου στο οποίο ορίζουμε τη μορφή των παρατηρήσεων. Το gretl δίνει τρεις επιλογές (Εικόνα 2.11).



Εικόνα 2.11: Παράθυρο προσδιορισμού τύπου δεδομένων στο gretl

Στην περίπτωση που θα επιλέξουμε **Χρονοσειρά** και κάνουμε κλικ στην ένδειξη **Forward** (επόμενο), ξεκινά η διαδικασία εισαγωγής δεδομένων χρονοσειράς.

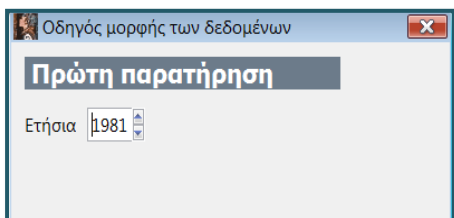
Σε νέο αναδυόμενο παράθυρο ορίζουμε τη συχνότητα της χρονοσειράς. Στην προκειμένη περίπτωση επιλέγουμε **Ετήσια** και πιέζουμε την ένδειξη **Forward** (επόμενο).



Εικόνα 2.12: Ρύθμιση χρονοσειρών στο gretl (1)

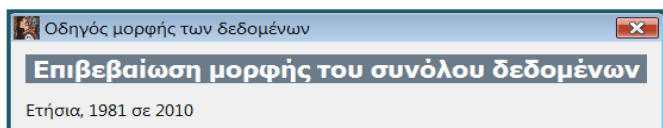
⁵ Ο προσδιορισμός των δεδομένων ως διαστρωματικά στο gretl επηρεάζει τον τρόπο με τον οποίο δομούνται τα δεδομένα. Θα πρέπει να επιλεγεί η μεταβλητή διαστρωμάτωσης. Τα Panel δεδομένα διαφέρουν από τα διαστρωματικά από το γεγονός ότι λαμβάνεται υπόψη και η διάσταση του χρόνου. Για αυτό και θα ζητηθεί να επιλεγεί σε μεταγενέστερο σημείο η μεταβλητή του χρόνου (βλ. παρακάτω διαδικασία επιλογής τύπου δεδομένων).

Σε νέο παράθυρο ορίζουμε την ημερομηνία έναρξης των δεδομένων (**1981**).



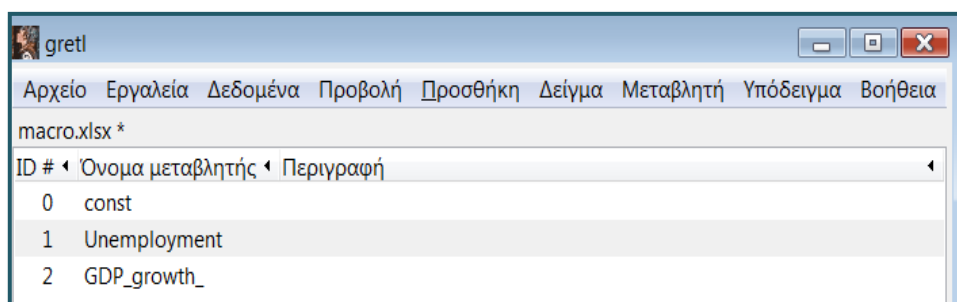
Εικόνα 2.13: Ρύθμιση χρονοσειρών στο gretl (2)

Ακολουθεί μήνυμα επιβεβαίωσης που αφορά στη χρονική διάρκεια των δεδομένων.



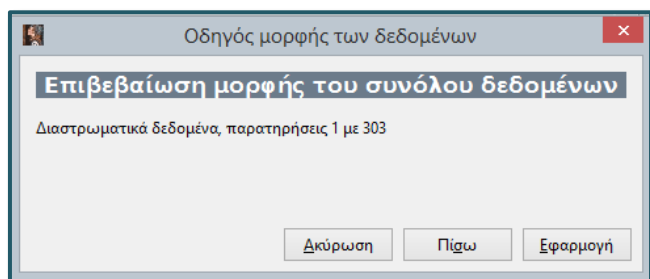
Εικόνα 2.14: Ρύθμιση χρονοσειρών στο gretl (3)

Στο βασικό παράθυρο του gretl έχουν δημιουργεί οι μεταβλητές **Unemployment** και **GDP_growth**.



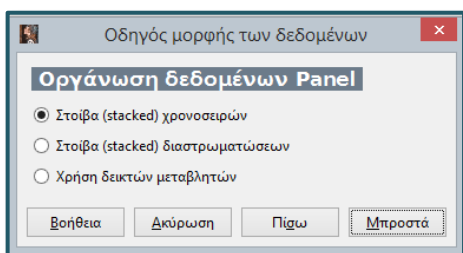
Εικόνα 2.15: Ρύθμιση χρονοσειρών στο gretl (4)

Στην περίπτωση που στο παράθυρο (Εικόνα 2.11) επιλέξουμε **Διαστρωματικά Δεδομένα** και στην συνέχεια κάνουμε κλικ στην ένδειξη **Forward** ξεκινά η διαδικασία εισαγωγής **διαστρωματικών δεδομένων**.



Εικόνα 2.16: Ρύθμιση διαστρωματικών δεδομένων στο gretl

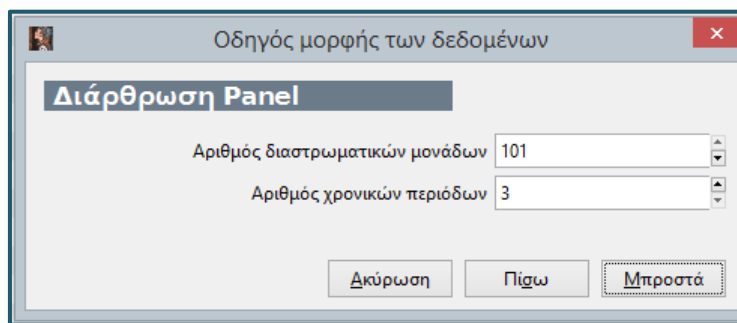
Στην περίπτωση που στο παράθυρο (Εικόνα 2:11) επιλέξουμε **Panel δεδομένα** και στην συνέχεια κάνουμε κλικ στην ένδειξη **Forward** ξεκινά η διαδικασία εισαγωγής **Panel δεδομένων**.



Εικόνα 2.17: Ρύθμιση Panel δεδομένων στο gretl (1)

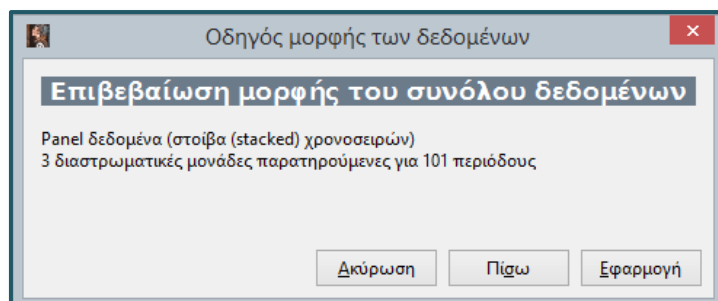
Αυτό το παράθυρο διαλόγου προσφέρει μέχρι τρεις επιλογές όσον αφορά στον ορισμό ενός συνόλου δεδομένων ως panel (διαστρωματικά – διαχρονικά). Οι δύο πρώτες επιλογές προϋποθέτουν ότι τα δεδομένα έχουν ήδη οργανωθεί σε μορφή panel (αν και αυτό δεν μπορεί ακόμα να αναγνωριστεί από το gretl). Η τρίτη επιλογή προϋποθέτει ότι το σύνολο δεδομένων περιέχει τις μεταβλητές που αντιπροσωπεύουν τη δομή του panel.

- **Στοιβα (stacked) χρονοσειρών:** Με αυτήν την επιλογή προσδιορίζεται ο N αριθμός των διαστρωματώσεων στο σύνολο των δεδομένων που αποτελείται από T παρατηρήσεις χρονοσειρών, μία για κάθε μονάδα διαστρωμάτωσης, δηλαδή πρώτα έγινε η κατανομή με βάση τον χρόνο και μετά ανά μεταβλητή διαστρωμάτωσης.



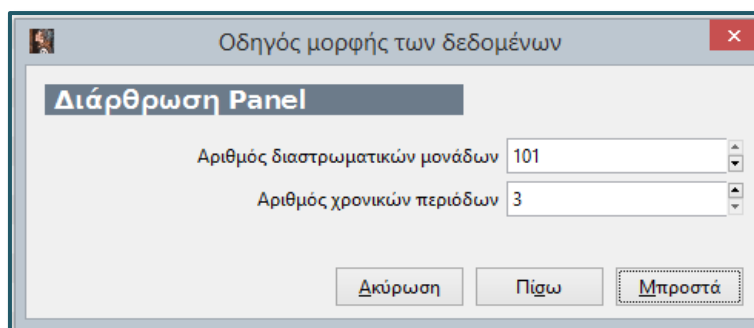
Εικόνα 2.18: Ρύθμιση Panel δεδομένων στο gretl (2)

Στη συνέχεια, γίνεται η επιβεβαίωση της μορφής των δεδομένων.



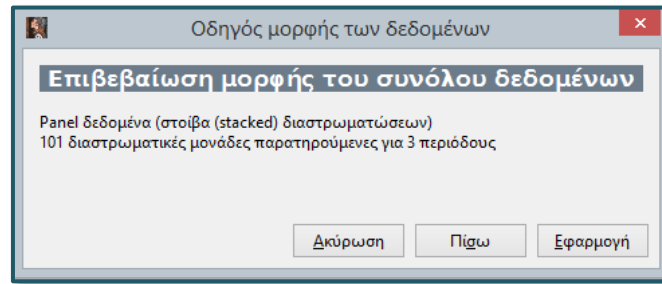
Εικόνα 2.19: Ρύθμιση Panel δεδομένων στο gretl (3)

- **Στοιβα (stacked) διαστρωματώσεων:** Με αυτή την επιλογή προσδιορίζεται ο T αριθμός των διαστρωματώσεων στο σύνολο των δεδομένων που αποτελείται από N παρατηρήσεις χρονοσειρών, δηλαδή πρώτα έγινε η διαστρωμάτωση με βάση μια μεταβλητή και μετά η κατανομή στο χρόνο.



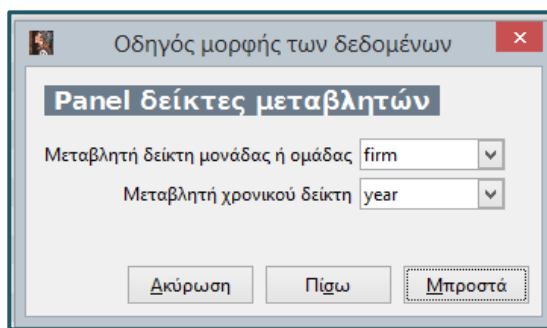
Εικόνα 2.20: Ρύθμιση Panel δεδομένων στο gretl (4)

Στη συνέχεια, γίνεται η επιβεβαίωση της μορφής των δεδομένων.



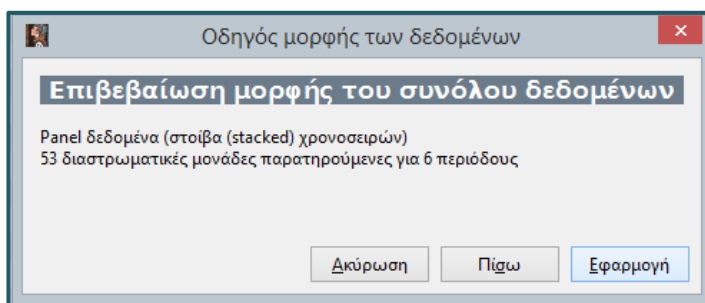
Εικόνα 2.21: Ρύθμιση Panel δεδομένων στο gretl (5)

- **Χρήση δεικτών μεταβλητών:** Με βάση αυτήν την επιλογή τα δεδομένα είναι οργανωμένα σύμφωνα με τον κλασικό τρόπο, δηλαδή όπως οι χρονοσειρές, αλλά περιέχονται δύο μεταβλητές. Η μία μεταβλητή αφορά τον χρόνο και η άλλη τη διαστρωμάτωση. Το πρόγραμμα τροποποιεί τις σειρές δεδομένων και τις φέρνει σε μορφή panel. Το επόμενο βήμα θα είναι να επιλέξουμε αυτές τις δύο μεταβλητές.



Εικόνα 2.22: Ρύθμιση Panel δεδομένων στο gretl (6)

Οι μεταβλητές διαστρωμάτωσης (panel) δεν πρέπει να έχουν αρνητικές ακέραιες τιμές ή ελλείπουσες τιμές. Αν υπάρχουν τέτοιες μεταβλητές στα δεδομένα, αυτή η επιλογή δεν είναι διαθέσιμη.



Εικόνα 2.23: Ρύθμιση Panel δεδομένων στο gretl (7)

Σημειώνεται ότι με ανάλογο τρόπο γίνεται η εισαγωγή των δεδομένων και από τα άλλα προαναφερόμενα υπολογιστικά προγράμματα.

2.1.2. Δημιουργία νέων μεταβλητών στο gretl

Έστω ότι ζητείται η δημιουργία δυο μεταβλητών, αύξηση του ΑΕΠ (GDP growth) και ανεργία

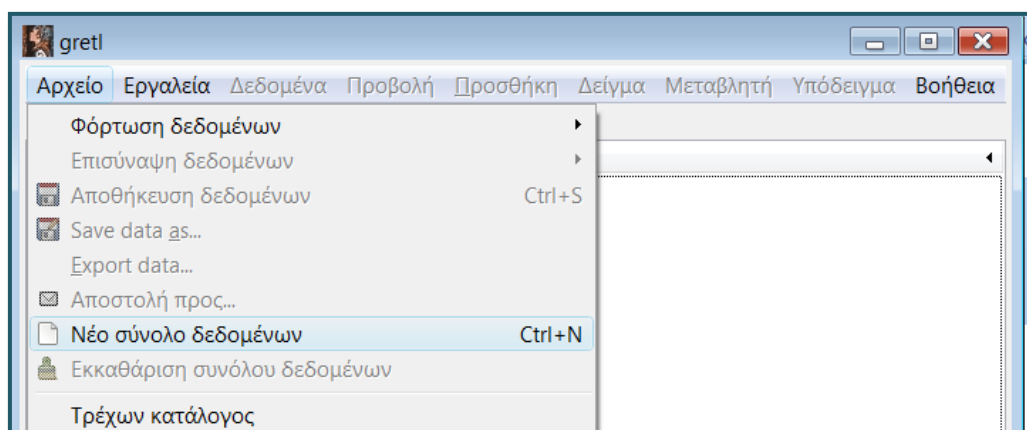
(Unemployment) με τα εξής δεδομένα:

Πίνακας 2.1: Ανεργία και ΑΕΠ στην Ελλάδα (πηγή: Eurostat)

Έτος	Ανεργία – Unemployment of total labor force (%)	ΑΕΠ – GDP growth (annual %)
2001	10,2	4,197
2002	10,3	3,439
2003	9,7	5,943
2004	10,5	4,368
2005	9,9	2,280
2006	8,9	5,511
2007	8,3	3,536
2008	7,7	-0,214
2009	9,5	-3,136
2010	12,5	-4,943

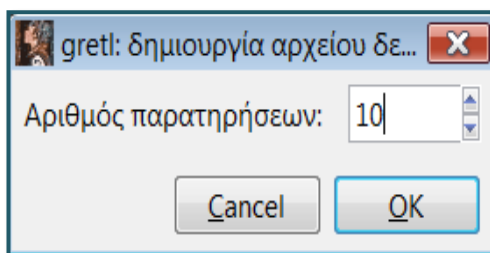
Για να δημιουργήσουμε τις εν λόγω μεταβλητές στο gretl εργαζόμαστε ως εξής:

1. Επιλεγούμε από το μενού εντολών **Αρχείο/Νέο Σύνολο Δεδομένων**⁶



Εικόνα 2.24: Δημιουργία μεταβλητών στο gretl (1)

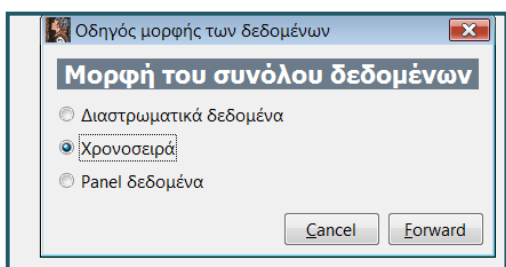
2. Στο αναδυόμενο παράθυρο ορίζουμε τον αριθμό των δεδομένων της νέας μεταβλητής (**10**).



Εικόνα 2.25: Δημιουργία μεταβλητών στο gretl (2)

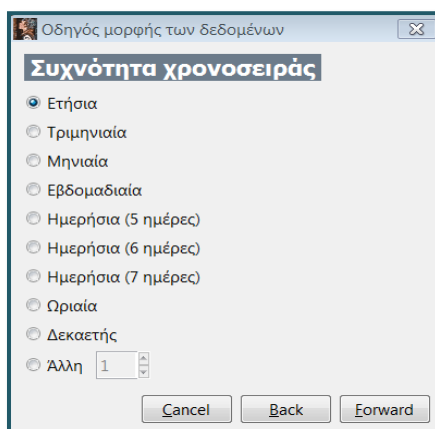
⁶ Το ίδιο αποτέλεσμα επιτυγχάνουμε και με το συνδυασμό των πλήκτρων **Ctrl** και **N**.

3. Στο νέο αναδυόμενο παράθυρο ορίζουμε τη μορφή των παρατηρήσεων. Στην προκειμένη περίπτωση επιλέγουμε **Χρονοσειρά**.



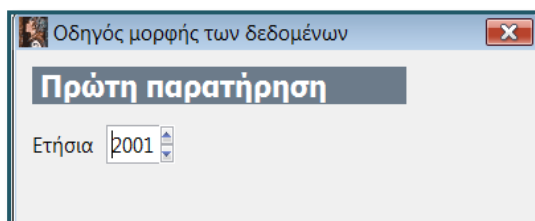
Εικόνα 2.26: Δημιουργία μεταβλητών στο gretl (3)

4. Έπεται η επιλογή της κατάλληλης συχνότητας στην ορισθείσα χρονοσειρά. Επιλέγουμε **Ετήσια**.



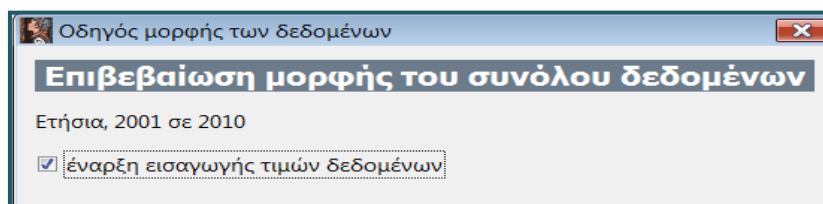
Εικόνα 2.27: Δημιουργία μεταβλητών στο gretl (4)

5. Στο επόμενο παράθυρο ορίζουμε την ημερομηνία της πρώτης παρατήρησης. Στην προκειμένη περίπτωση καταχωρούμε το **2001**.



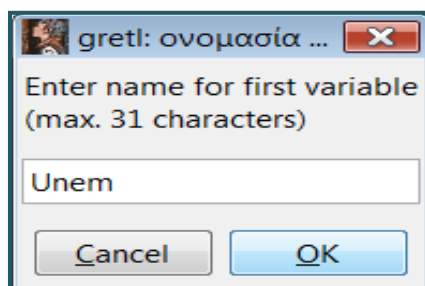
Εικόνα 2.28: Δημιουργία μεταβλητών στο gretl (5)

6. Ακολουθεί μήνυμα επιβεβαίωσης του επιλεχθέντος χρονικού διαστήματος (**Ετήσια, 2001 σε 2010**), καθώς και επιλογή που αφορά στην άμεση ή μη εισαγωγή των δεδομένων. Στην προκειμένη περίπτωση τσεκάρουμε την **έναρξη εισαγωγής τιμών δεδομένων**.





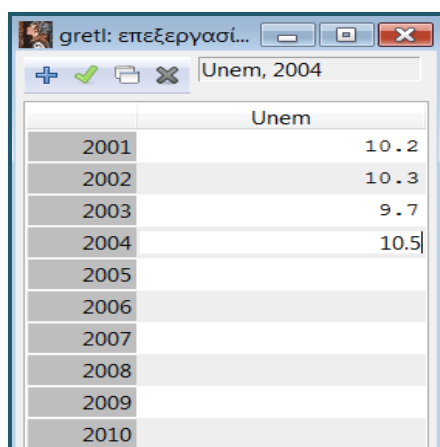
Εικόνα 2.29: Δημιουργία μεταβλητών στο gretl (6)

7. Στο επόμενο παράθυρο ορίζουμε το όνομα της νέας μεταβλητής **Unem**.



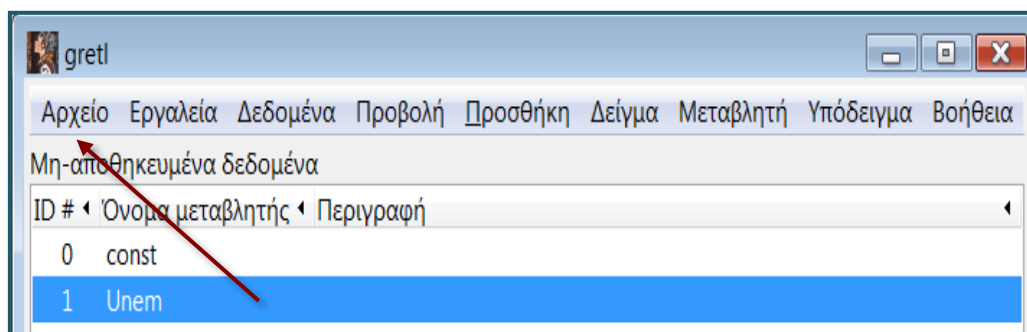
Εικόνα 2.30: Δημιουργία μεταβλητών στο gretl (7)

8. Στη συνέχεια εμφανίζεται παράθυρο στο οποίο εισάγουμε τις τιμές της επιλεγθείσας μεταβλητής **Unem** (Ανεργία) του Πίνακα 2.1. Για να αποθηκευτούν οι καταχωρημένες τιμές στην ορισθείσα μεταβλητή πιέζουμε το σύμβολο , διαφορετικά θα βγει ανάλογο μήνυμα όταν επιλέξουμε έξοδο από το παράθυρο μέσω του συμβόλου .



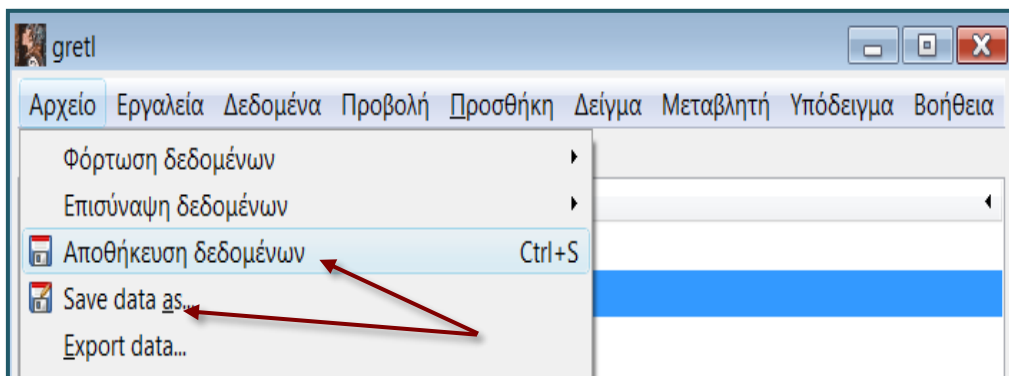
Εικόνα 2.31: Δημιουργία μεταβλητών στο gretl (8)

9. Η μεταβλητή εμφανίζεται πλέον στο βασικό μενού του gretl. Σημειώνεται ότι στο προηγούμενο βήμα καταχωρήθηκαν – αποθηκεύτηκαν οι τιμές στο όνομα της μεταβλητής, παρόλα αυτά απαιτείται και η αποθήκευση της μεταβλητής σε αρχείο, διαφορετικά η έξοδος από το βασικό μενού θα έχει ως αποτέλεσμα την απώλεια των δεδομένων.



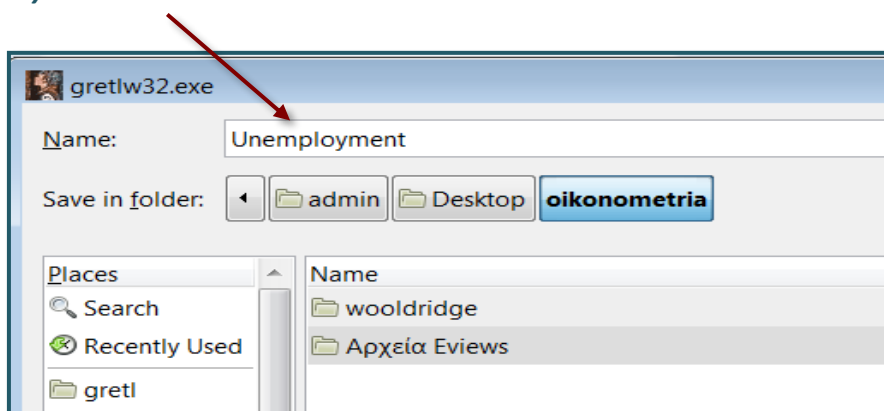
Εικόνα 2.32: Δημιουργία μεταβλητών στο gretl (9)

10. Επιλέγουμε είτε **Αποθήκευση δεδομένων**, είτε **Save data as...**



Εικόνα 2.33: Δημιουργία μεταβλητών στο gretl (10)

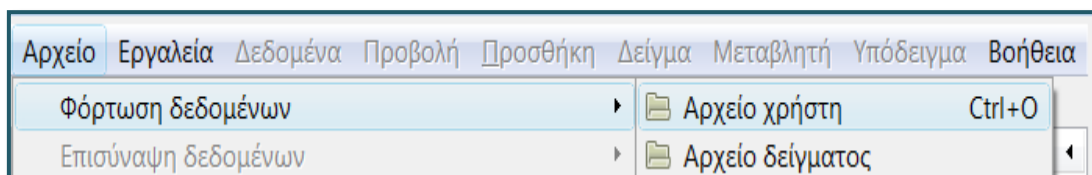
11. Στο αναδυόμενο παράθυρο, αφού επιλέξουμε τη διαδρομή του φακέλου που θα αποθηκεύσουμε το αρχείο⁷, ορίζουμε το όνομα του αρχείου. Στην προκειμένη περίπτωση γράφουμε **Unemployment**.



Εικόνα 2.34: Δημιουργία μεταβλητών στο gretl (11)

2.1.3. Επεξεργασία μεταβλητών στο gretl

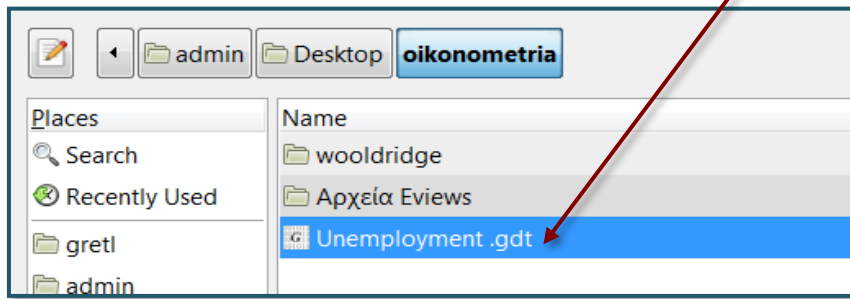
Η πρώτη εργασία που καλούμαστε να εκτελέσουμε, για να επεξεργαστούμε τις μεταβλητές που έχουμε ήδη αποθηκεύσει σε ένα αρχείο, είναι να φορτώσουμε το εν λόγω αρχείο. Ανοίγουμε το αρχείο επιλέγοντας **Αρχείο/Φόρτωση δεδομένων/Αρχείο χρήστη**.



Εικόνα 2.35: Φόρτωση δεδομένων στο gretl (1)

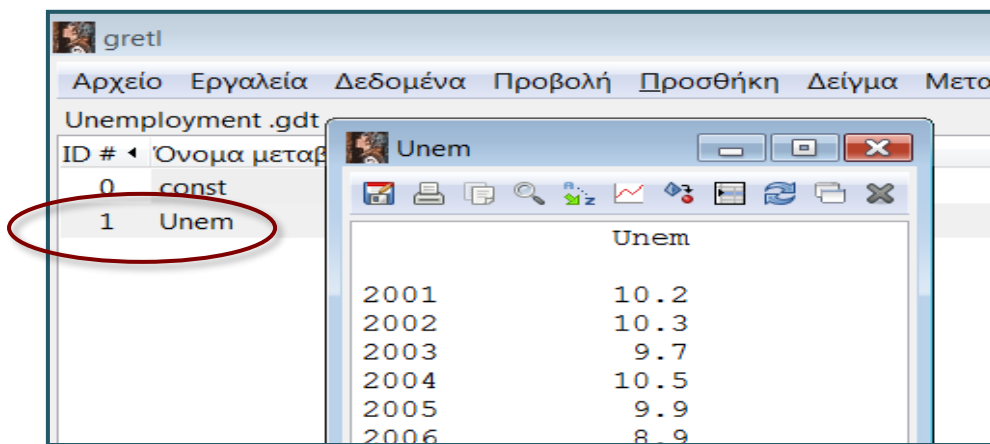
⁷ Τα αρχεία του gretl έχουν κατάληξη **.gdt**.

Έστω ότι ενδιαφερόμαστε να αναλύσουμε τη μεταβλητή **Unem** του αρχείου **Unemployment**. Επιλέγουμε στο αναδυόμενο παράθυρο **Unemployment.gdt**.




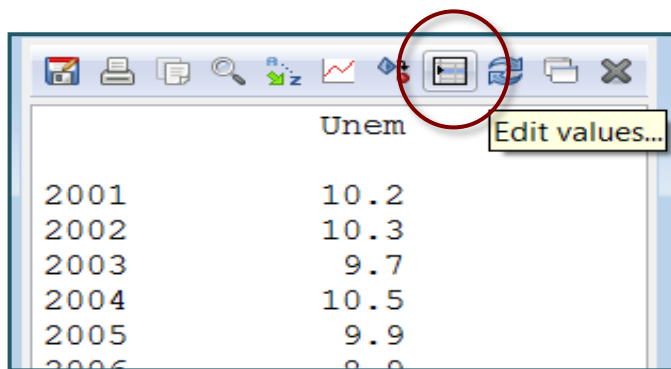
Εικόνα 2.36: Φόρτωση δεδομένων στο gretl (2)

Με διπλό κλικ στο όνομα της μεταβλητής **Unem** αναδύεται παράθυρο που εμφανίζει τις τιμές της και παρέχει μεταξύ άλλων και τις παρακάτω δυνατότητες επεξεργασίας και παρουσίασης:



Εικόνα 2.37: Φόρτωση δεδομένων στο gretl (3)

- **Edit values** (επεξεργασία τιμών) με αναφορά το εικονίδιο  (Εικόνα 2.38). Η επιλογή αυτή έχει ως αποτέλεσμα ένα νέο παράθυρο επεξεργασίας τιμών.




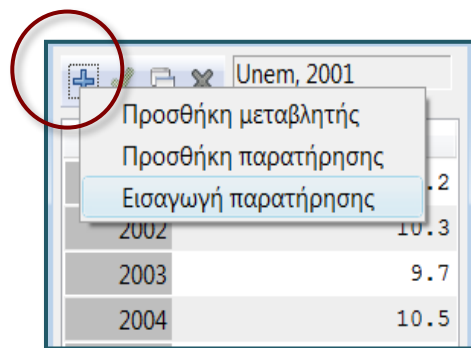
Εικόνα 2.38: Επεξεργασία δεδομένων στο gretl (1)

Στο νέο παράθυρο μπορούμε να αλλάξουμε τις τιμές της υπό ανάλυση μεταβλητής, με απλή εισαγωγή νέων τιμών στα σχετικά κελιά.


	Unem
2001	10.2
2002	10.3
2003	5
2004	10.5
2005	9.9
2006	8.9
2007	8.3
2008	7.7

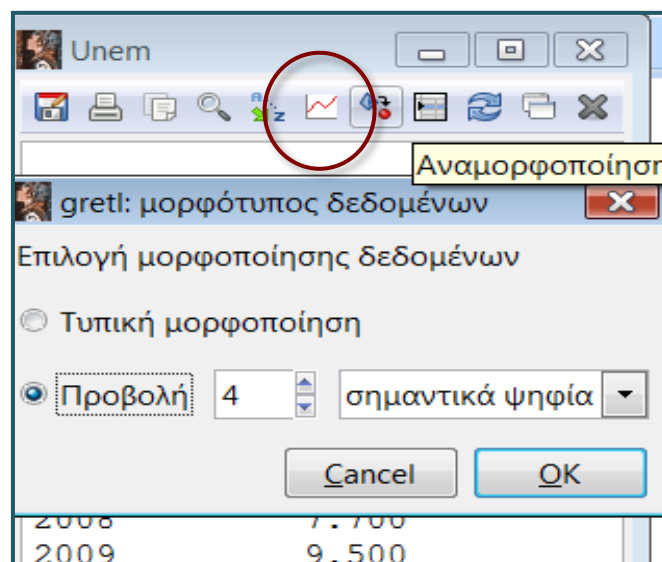
Εικόνα 2.39: Επεξεργασία δεδομένων στο gretl (2)

Επιπλέον, με την επιλογή του εικονιδίου  δίνεται η δυνατότητα είτε της προσθήκης νέας μεταβλητής, είτε της προσθήκης ή εισαγωγής νέας παρατήρησης.




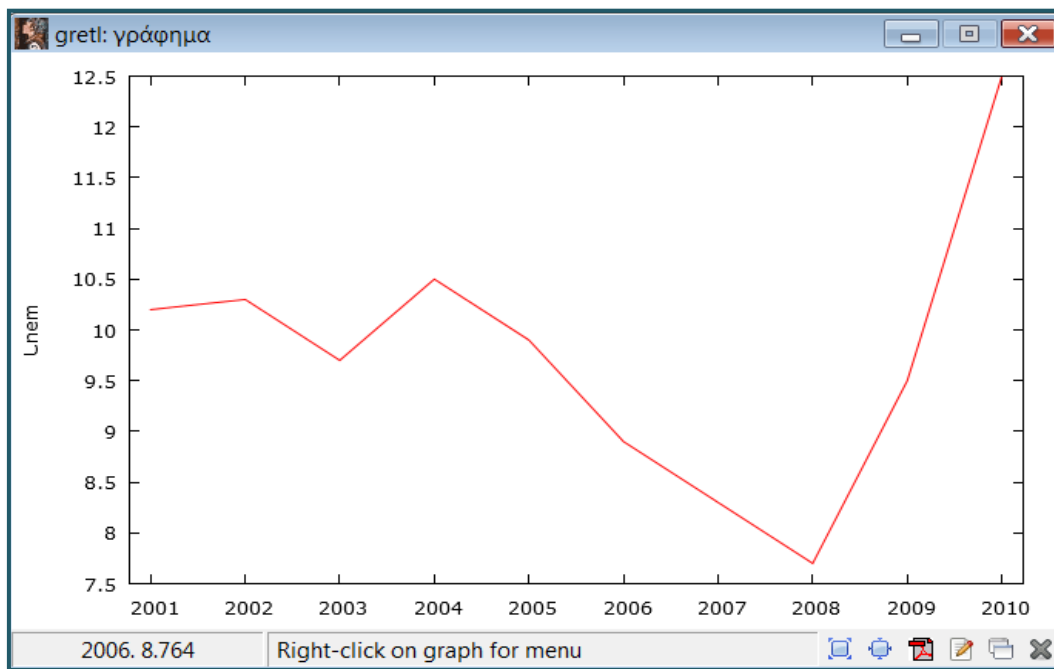
Εικόνα 2.40: Εισαγωγή παρατήρησης στο gretl

- **Αναμορφοποίηση** - εικονίδιο  (Εικόνα 2.41). Η επιλογή αυτή δίνει τη δυνατότητα της παρουσίασης των δεδομένων με περισσότερα ή λιγότερα ψηφία. Στην προκειμένη περίπτωση στη θέση **Προβολή** επιλέγουμε **4 σημαντικά ψηφία** και ο αριθμός 12,5 εμφανίζεται πλέον ως 12,50.



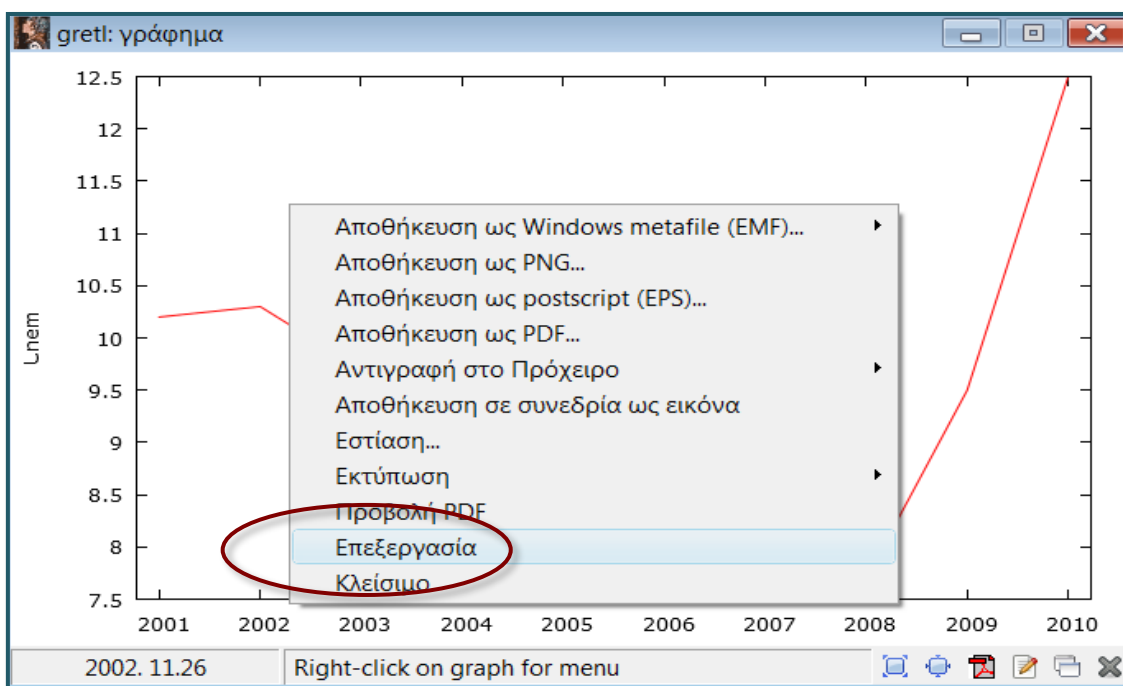
Εικόνα 2.41: Αναμορφοποίηση δεδομένων στο gretl

- **Γράφημα** με αναφορά στο εικονίδιο . Η επιλογή αυτή παράγει το γράφημα των σχετικών δεδομένων.



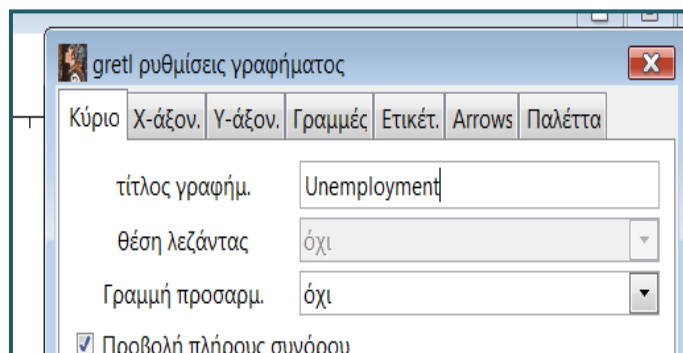
Εικόνα 2.42: Αναμορφοποίηση δεδομένων στο gretl

Για να επεξεργαστούμε το γράφημα κάνουμε δεξί κλικ πάνω στο γράφημα και στον αναδυόμενο κατάλογο επιλέγουμε **Επεξεργασία**.



Εικόνα 2.43: Δημιουργία γραφήματος στο gretl

Εάν, για παράδειγμα, θελήσουμε να προσθέσουμε τίτλο στο γράφημα, τότε στο νέο αναδυόμενο μενού επιλέγουμε την καρτέλα **Κύριο** και δίπλα στην ένδειξη **τίτλος γραφήμ.** πληκτρολογούμε τη λέξη **Unemployment**⁸.

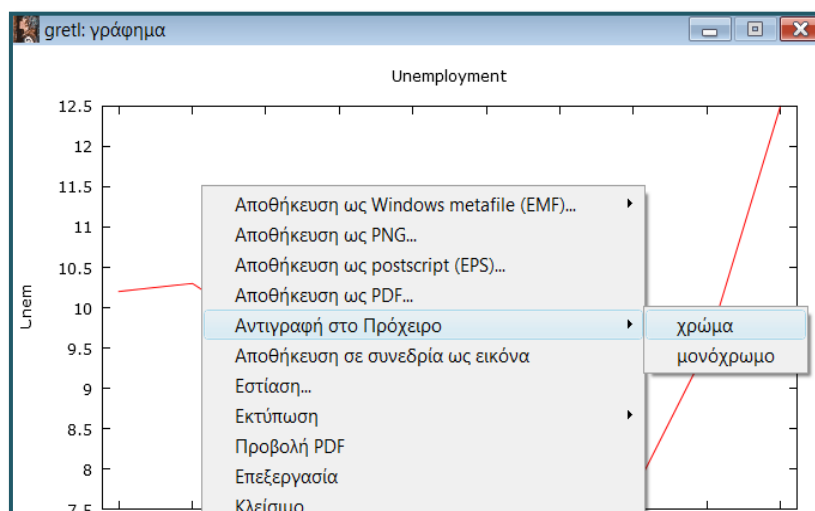


Εικόνα 2.44: Επεξεργασία γραφήματος στο gretl

Για να αντιγράψουμε έγχρωμα το γράφημα σε έγγραφο Word επιλέγουμε **Αντιγραφή στο Πρόχειρο/χρώμα**. Με δεξί κλικ και επικόλληση αντιγράφουμε το γράφημα σε Word. Με αριστερό κλικ στην περιοχή του κάθε τίτλου μπορούμε να αλλάξουμε τον κάθε τίτλο. Στην προκειμένη περίπτωση αντικαθιστούμε το **“Unemployment”** με το **“Ανεργία”**.

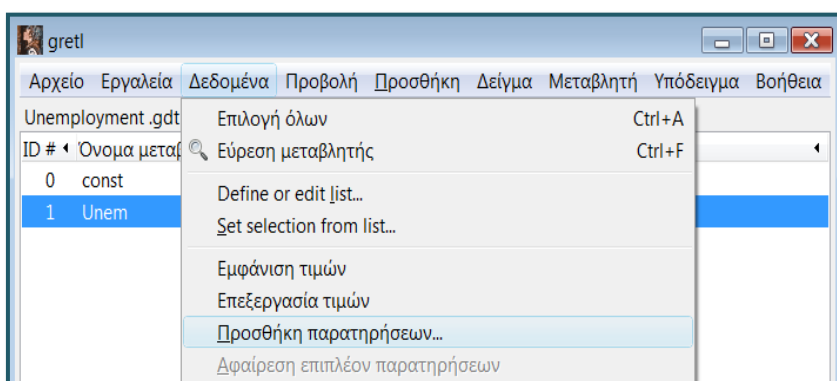
Σημειώνεται ότι πολλές από τις παραπάνω εργασίες δύναται εναλλακτικά να επιτευχθούν και από το βασικό μενού του gretl. Για παράδειγμα, η εντολή **Δεδομένα/Εμφάνιση τιμών ή Επεξεργασία τιμών ή Προσθήκη παρατηρήσεων...** δίνουν ακριβώς τις ίδιες δυνατότητες παρουσίασης και επεξεργασίας των δεδομένων με τις επιλογές που εξετάστηκαν στην αρχή της ενότητας.

Θα πρέπει επίσης αν σημειώσουμε ότι εκτεταμένη παρουσίαση της δημιουργίας και επεξεργασίας γραφημάτων διαφόρων τύπων θα γίνει στο επόμενο κεφάλαιο. Στην παρούσα θέση απλώς παρουσιάζουμε τις βασικές δυνατότητες του προγράμματος ως προς την εισαγωγή και το χειρισμό των μεταβλητών.



Εικόνα 2.45: Αντιγραφή γραφήματος στο gretl

⁸ Το gretl μπορεί να δεχτεί ελληνικούς χαρακτήρες, αλλά κατά τη διαδικασία της αντιγραφής και επικόλλησης παρουσιάζει σφάλμα.



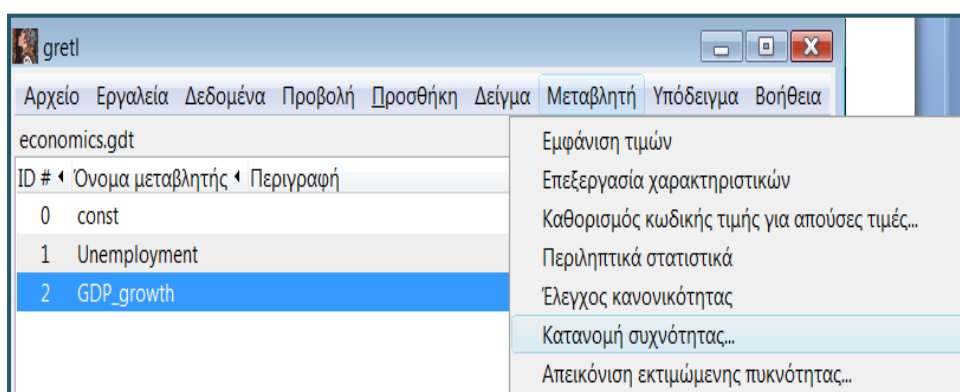
Εικόνα 2.46: Μενού επιλογών για τα δεδομένα στο gretl

Το gretl διαθέτει πληθώρα και άλλων εντολών και διεργασιών, που εύκολα ο ερευνητής, όταν αποκτήσει οικειότητα με το πρόγραμμα, μπορεί να ανασύρει για να βελτιώσει την παρουσίαση και επεξεργασία των δεδομένων.

2.1.4. Δημιουργία ιστογράμματος με το gretl

Το **ιστόγραμμα** αποτελεί μια γραφική μέθοδο αναπαράστασης της κατανομής μιας συνεχούς μεταβλητής, η οποία έχει ομαδοποιηθεί σε τάξεις (κλάσεις) δεδομένων⁹. Η επιφάνεια της κάθε ράβδου του ιστογράμματος (σημειώνεται ότι η ράβδος καλείται **ιστός**) αντανακλά τη συχνότητα εμφάνισης των δεδομένων της κάθε τάξης, ενώ το ύψος του ισούται με το λόγο της συχνότητας προς το εύρος των τιμών που αντιπροσωπεύει η κάθε ράβδος. Όσο υψηλότερο το ύψος της κάθε ράβδου, τόσο μεγαλύτερη είναι και η συχνότητα των δεδομένων και αντίθετα, όσο χαμηλότερο το ύψος, τόσο μικρότερη είναι και η συχνότητα των δεδομένων. Τα ιστογράμματα μας επιτρέπουν εκτός των άλλων να βγάλουμε συμπεράσματα και σχετικά με την κατανομή του γεννήτορα πληθυσμού.

Η κατασκευή ιστογράμματος με το gretl πραγματοποιείται ως εξής: Αφού κάνουμε κλικ πάνω στην υπό ανάλυση μεταβλητή, επιλέγουμε την εντολή **Μεταβλητή/Κατανομή συχνότητας...** από το βασικό μενού εντολών.

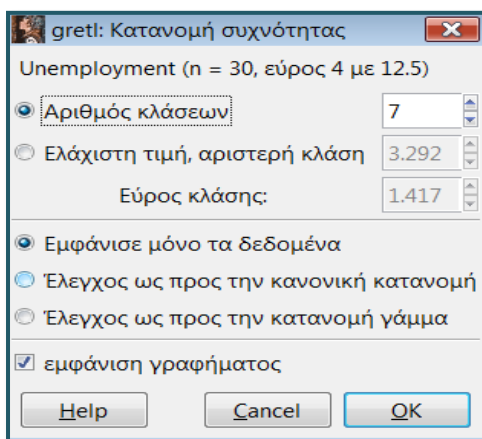


Εικόνα 2.47: Κατανομή συχνότητας στο gretl (1)

Στο αναδυόμενο παράθυρο μπορούμε να προσδιορίζουμε τα χαρακτηριστικά του ιστογράμματος όπως Αριθμό κλάσεων (**7**), Ελάχιστη τιμή (της μικρότερης) κλάσης (**3,292**) και Εύρος κλάσης (**1,417**). Στην

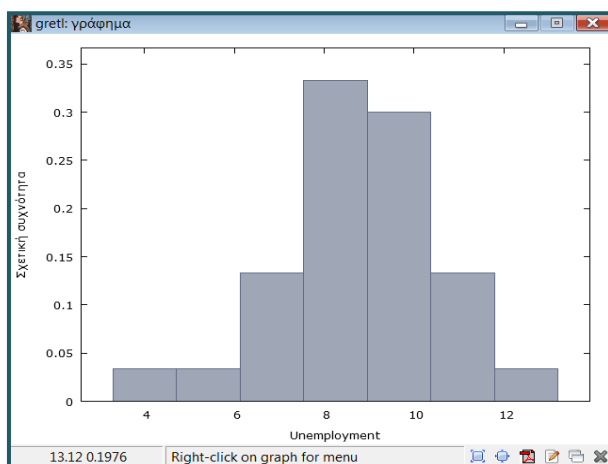
⁹ Συγκεκριμένα, τα δεδομένα διατάσσονται ομαδοποιημένα σε αύξουσα σειρά στον οριζόντιο άξονα.

προκειμένη περίπτωση ακολουθούμε τις προεπιλογές του gretl.



Εικόνα 2.48: Κατανομή συχνότητας στο gretl (2)

Εφόσον έχουμε επιλέξει (τσεκάρει) την **εμφάνιση γραφήματος**, στο πρώτο αναδυόμενο παράθυρο προβάλλεται το γράφημα του ιστογράμματος.



Εικόνα 2.49: Δημιουργία ιστογράμματος στο gretl

Στο δεύτερο αναδυόμενο παράθυρο παρουσιάζεται ο πίνακας συχνοτήτων του ιστογράμματος.

διάστημα	κεντρ. τιμή	συχνότητα	σχετ.	αθροιστ.
< 4.7083	4.0000	1	3.33%	3.33% *
4.7083 - 6.1250	5.4167	1	3.33%	6.67% *
6.1250 - 7.5417	6.8333	4	13.33%	20.00% ****
7.5417 - 8.9583	8.2500	10	33.33%	53.33% *****
8.9583 - 10.375	9.6667	9	30.00%	83.33% *****
10.375 - 11.792	11.083	4	13.33%	96.67% ****
>= 11.792	12.500	1	3.33%	100.00% *

Εικόνα 2.50: Πίνακας κατανομής συχνοτήτων στο gretl

Πολύ αναλυτικότερα για την κατασκευή ιστογραμμάτων με το gretl (αλλά και με το Eviews) θα μιλήσουμε στο επόμενο κεφάλαιο.

2.1.5. Στατιστικά Περιγραφικά Μέτρα

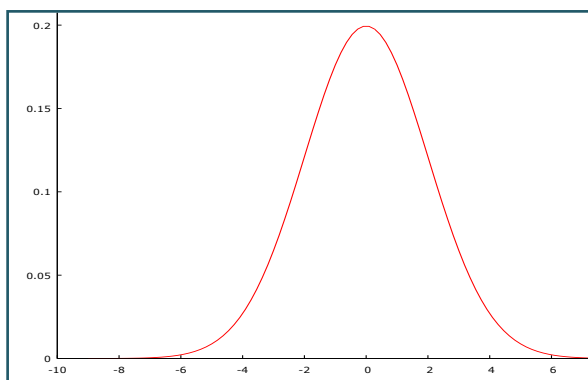
Για να περιγραφεί ικανοποιητικά η κατανομή μιας **τυχαίας μεταβλητής** θα πρέπει να υπολογισθούν τα **στατιστικά μέτρα**:

- **Κεντρικής τάσης και θέσης.** Τα μέτρα κεντρικής τάσης¹⁰, ιδιαίτερα στην περίπτωση της **κανονικής κατανομής**¹¹, δείχνουν την τάση συγκέντρωσης των τιμών προς μια αντιπροσωπευτική τιμή. Το πιο διαδεδομένο μέτρο κεντρικής τάσης είναι η **μέση τιμή (mean value ή average)**. Τα μέτρα θέσης¹² προσδιορίζουν χαρακτηριστικές θέσεις στο εύρος κατανομής των δεδομένων. Το πιο διαδεδομένο μέτρο θέσης είναι η **διάμεσος**¹³ (**median**).
- **Διασποράς.** Τα μέτρα διασποράς¹⁴ μας πληροφορούν γενικώς περί της συγκέντρωσης ή απομάκρυνσης των δεδομένων, συνήθως με σημείο αναφοράς τη μέση τιμή. Το πιο διαδεδομένο μέτρο διασποράς είναι η διακύμανση και κατ' επέκταση η τυπική απόκλιση, που είναι η τετραγωνική ρίζα της διακύμανσης.
- **Ασυμμετρίας.** Τα μέτρα ασυμμετρίας ποσοτικοποιούν το μέγεθος της ασυμμετρίας των δεδομένων μιας κατανομής γύρω από τη μέση τιμή.
- **Κύρτωσης.** Η κύρτωση μετρά τη συγκέντρωση των παρατηρήσεων γύρω από το μέσο.

¹⁰ Τα σημαντικότερα μέτρα κεντρικής τάσης είναι:

- Ο αριθμητικός μέσος
- Ο γεωμετρικός μέσος
- Ο αρμονικός μέσος

¹¹ Η κανονική κατανομή παρουσιάζει τη μεγαλύτερη συγκέντρωση των δεδομένων στο αριθμητικό μέσο (μέσο όρο) και όσο απομακρυνόμαστε από το μέσο εμφανίζεται συμμετρικά όλο και μικρότερη συχνότητα, με αποτέλεσμα οι ουρές της κατανομής να πλησιάζουν ασυμπτωτικά τον οριζόντιο άξονα. Η μορφή της είναι καμπανοειδής (κωδωνοειδής). Στο Παράρτημα του κεφαλαίου μπορείτε να δείτε συνοπτικά τα χαρακτηριστικά της κανονικής κατανομής.



¹² Τα σημαντικότερα μέτρα θέσης είναι:

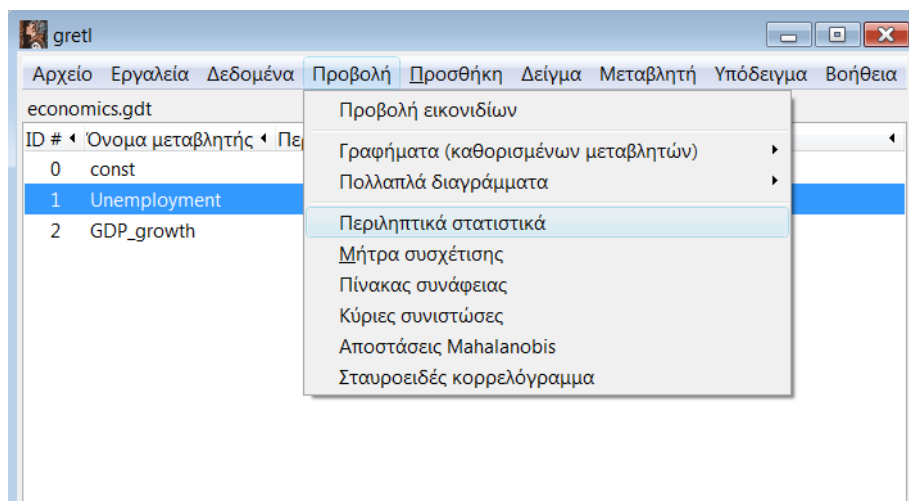
- Η διάμεσος
- Τα τεταρτημόρια
- Τα δεκατημόρια
- Τα εκατοστημόρια (εκατοστηαία σημεία ή εκατοστηαίες θέσεις)

¹³ Η διάμεσος είναι η κεντρική τιμή των δεδομένων όταν αυτά διατάσσονται κατά αύξουσα σειρά.


¹⁴ Τα σημαντικότερα μέτρα διασποράς είναι

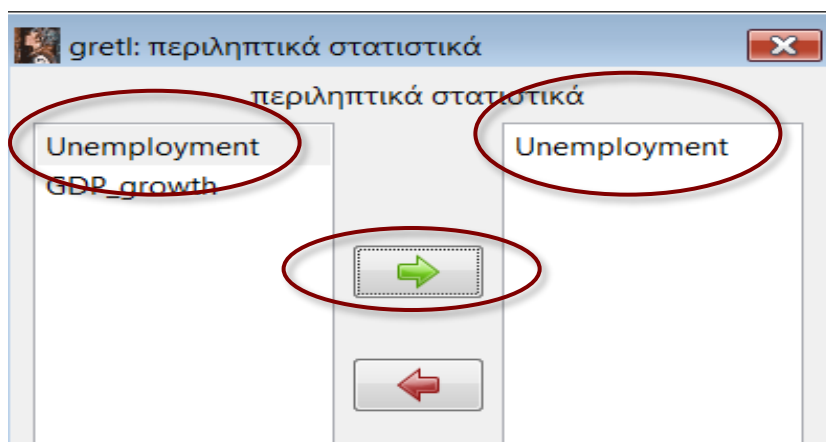
- Η διακύμανση (variance)
- Η τυπική απόκλιση (standard deviation)
- Το εύρος μεταβολής (range)
- Το ενδοτεταρτημοριακό εύρος (interquartile range)
- Ο συντελεστής μεταβλητότητας (coefficient of variation)

Το gretl παρέχει τη δυνατότητα υπολογισμού των βασικών στατιστικών **περιληπτικών μέτρων (descriptive statistics)** της κατανομής μιας τυχαίας μεταβλητής. Έστω ότι ζητείται η εκτίμηση των περιληπτικών στατιστικών μέτρων της μεταβλητής **Unemployment** του αρχείου **economics.gdt**. Τα βήματα που θα ακολουθήσουμε καθώς και η ερμηνεία των σχετικών αποτελεσμάτων έχουν ως εξής: Επιλέγουμε την εντολή **Προβολή/Περιληπτικά στατιστικά**¹⁵.



Εικόνα 2.51: Περιληπτικά στατιστικά στο gretl (1)

Στο αναδυόμενο παράθυρο με τίτλο **περιληπτικά στατιστικά** επιλέγουμε τη μεταβλητή **Unemployment** και πατώντας το πράσινο βελάκι  τοποθετούμε την υπό ανάλυση μεταβλητή στο δεξιό πίνακα.



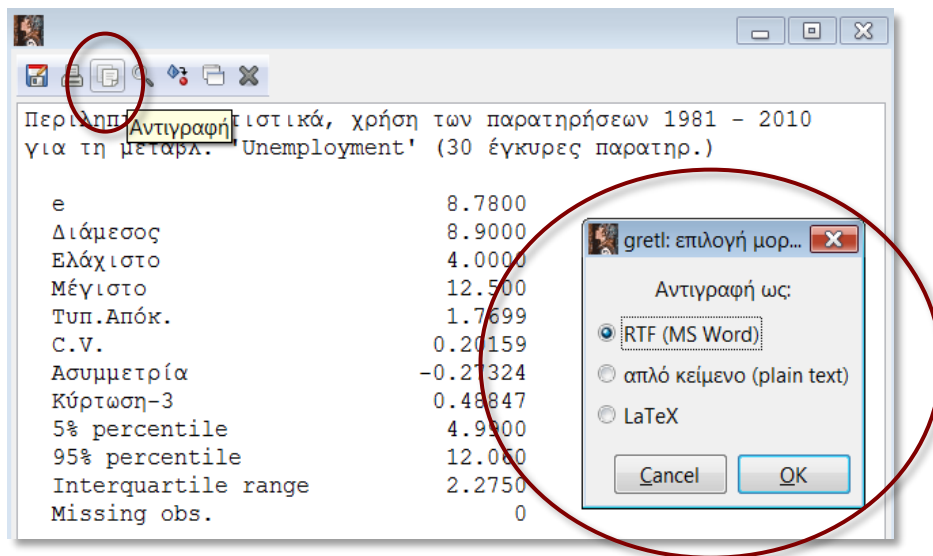
Εικόνα 2.52: Περιληπτικά στατιστικά στο gretl (2)

Στη συνέχεια, εμφανίζεται ο πίνακας των περιληπτικών στατιστικών της μεταβλητής **Unemployment**. Για να αντιγράψουμε τον πίνακα σε έγγραφο Word πιέζουμε το σχετικό εικονίδιο που αφορά στην αντιγραφή¹⁶



¹⁵ Το ίδιο αποτέλεσμα παίρνουμε και με την εντολή **Μεταβλητή/Περιληπτικά στατιστικά**

¹⁶ Η ίδια δυνατότητα δίνεται και με δεξί κλικ του ποντικιού πάνω στην περιοχή του πίνακα και επιλογή **αντιγραφή** στο αναδυόμενο παράθυρο.



Εικόνα 2.53: Περιληπτικά στατιστικά στο gretl (3)

Τα αποτελέσματα σε έγγραφο Word (αφού τα μορφοποιήσουμε με την επιλογή Όλα τα περιγράμματα του Word) εμφανίζονται ως εξής¹⁷:

Summary Statistics, using the observations 1981 - 2010
για τη μεταβλ. Unemployment (30 έγκυρες παρατηρ.)

Μέσος	Διάμεσος	Ελάχιστο	Μέγιστο
8.78000	8.90000	4.00000	12.5000
Τυπ.Απόκ.	C.V.	Ασυμμετρία	Κύρτωση-3
1.76994	0.201588	-0.273237	0.488471
5% Perc.	95% Perc.	IQ range	Missing obs.
4.99000	12.0600	2.27500	0

Η συνοπτική ερμηνεία των εν λόγω περιγραφικών στατιστικών μέτρων έχει ως εξής:

- Ο **μέσος όρος** ενός στατιστικού δείγματος¹⁸ ή πληθυσμού είναι το πηλίκο του αθροίσματος των τιμών δια του σχετικού πλήθους. Σε ένα δείγμα μεγέθους n ο δειγματικός μέσος υπολογίζεται με τον τύπο:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Μέσος} = \bar{x} = 8.78$$

- Η **διάμεσος**¹⁹ είναι η τιμή που χωρίζει την κατανομή των τιμών μιας μεταβλητής σε δύο ίσα μέρη, με την προϋπόθεση ότι έχουμε κατατάξει τις τιμές σε **αύξουσα** ή **φθίνουσα σειρά**.

¹⁷ Παρατηρούμε ότι η πρώτη επιλογή του αντιγραμμένου πίνακα σε Word αφορά στο **μέσο** (8,78), τον οποίο το gretl στον προηγούμενο πίνακα εμφανίζει με την ονομασία e.

¹⁸ Δείγμα καλείται ένα οποιοδήποτε αντιπροσωπευτικό τμήμα ενός στατιστικού πληθυσμού

$$\text{Διάμεσος} = M_d = 8,9$$

- **Ελάχιστο** είναι η **ελάχιστη τιμή** της κατανομής.
- **Μέγιστο** είναι η **μέγιστη τιμή** της κατανομής.
- Η **τυπική απόκλιση** εκφράζει τη διασπορά (τη μέση απόσταση) των τιμών από τη μέση τιμή²⁰ και ισούται με τη ρίζα της διακύμανσης²¹. Η **τυπική απόκλιση δείγματος**²² ισούται με:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Η **πληθυσμιακή τυπική απόκλιση** διαφέρει από την **δειγματική τυπική απόκλιση** στον παρονομαστή, καθώς οι τετραγωνικές αποκλίσεις από το μέσο διαιρούνται $n - 1$ **βαθμούς ελευθερίας**²³ (βλέπε παράρτημα του κεφαλαίου).

$$\text{Τυπική απόκλιση} = s = 1,76$$

- **C.V. (Coefficient of Variation)** είναι ο **συντελεστής μεταβλητότητας**, ο οποίος μετρά τη σχετική μεταβλητότητα μιας κατανομής παρατηρήσεων, καθώς υπολογίζεται ως το πηλίκο της τυπικής απόκλισης προς το μέσο όρο: $CV = \frac{s}{\bar{x}}$

Με άλλα λόγια, ο συντελεστής μεταβλητότητας εξομαλύνει τη διασπορά (τυπική απόκλιση) με βάση το μέγεθος των παρατηρήσεων (μέσο όρο), για να είναι δυνατή η σύγκριση της διασποράς

¹⁹ Η διάμεσος, ως στατιστικό μέτρο θέσης, είναι λιγότερο ευαίσθητη σε ακραίες τιμές. Για παράδειγμα εάν έχουμε τις τιμές 1, 2, 5, 10, 1000 τότε ο μέσος όρος ($\bar{x} = 203,6$) θα επηρεαστεί πάρα πολύ από την τιμή 1000, ενώ η διάμεσος ($M_d = 5$) θα μείνει ανεπηρέαστη.

²⁰ Έστω μια μεταβλητή X που έχει τιμές 21 και 1, τότε η μέση τιμή είναι:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{21 + 1}{2} = 11$$

Η τυπική απόκλιση είναι:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\frac{(1 - 11)^2 + (21 - 11)^2}{2}} = 10$$

και όπως εύκολα παρατηρούμε δείχνει την απόσταση της μέσης τιμής 11 από τις τιμές 1 και 21, γεγονός που εύληπτα αναπαριστά την έννοια της τυπικής απόκλισης “ως μέσης απόστασης των τιμών από τη μέση τιμή”. Φυσικά όσο μεγαλύτερη η “απόσταση” των τιμών από τη μέση τιμή, τόσο μεγαλύτερη και η τυπική απόκλιση ή αλλιώς η διασπορά.

²¹ Η τυπική απόκλιση πληθυσμού δίνεται από τον τύπο:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

²² Η τυπική απόκλιση στο gretl υπολογίζεται με τον εν λόγω τύπο της δειγματικής τυπικής απόκλισης.

²³ Η μείωση του παρονομαστή σημαίνει ταυτόχρονα και αύξηση του μεγέθους της δειγματικής τυπικής απόκλισης για να διορθωθεί το σχετικό πιθανό σφάλμα (το δείγμα είναι πιθανό να μην έχει τις ακραίες τιμές του πληθυσμού και επομένως να υποεκτιμά την πραγματική τυπική απόκλιση).

κατανομών που διαφέρουν όχι μόνο ως προς την κλίμακα των παρατηρήσεων, αλλά και ως προς τις μονάδες μέτρησης (π.χ. μια κατανομή μπορεί να αφορά μέτρα, ενώ μια άλλη κιλά)²⁴.

$$C.V. = 0,20$$

- Η **ασυμμετρία** μετρά το βαθμό της συμμετρικότητας των δεδομένων γύρω από τη μέση τιμή. Η θετική ασυμμετρία ($G > 0$) υποδηλώνει την επέκταση (ουρά) των τιμών προς τα δεξιά, αντίθετα η αρνητική ασυμμετρία ($G < 0$) υποδηλώνει την επέκταση των τιμών προς τα αριστερά, ενώ τέλος, η συμμετρία ($G = 0$) υποδηλώνει τη συμμετρική (κανονική) κατανομή των δεδομένων γύρω από το μέσο. Ο τύπος εύρεσης της ασυμμετρίας στο gretl είναι: $G = \frac{1}{N} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\hat{s}} \right)^3$

όπου \hat{s} είναι εκτιμητής της τυπικής απόκλισης, που ισούται με: $\hat{s} = s \sqrt{\frac{n-1}{n}}$

$$\text{Asυμμετρία} = G = -0,27 \text{ αρνητική}$$

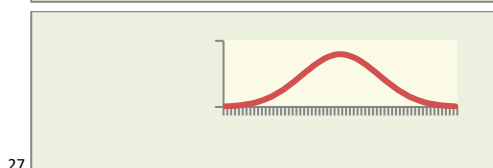
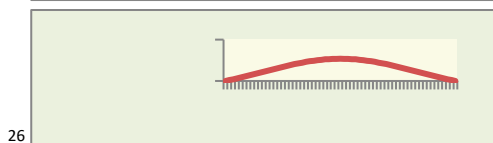
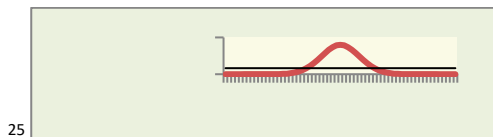
- Η **κύρτωση** δείχνει κατά πόσο τα δεδομένα της κατανομής σχηματίζουν έντονη κορυφή στο μέσο τους, δηλαδή, με την κύρτωση μετράμε τη συγκέντρωση των παρατηρήσεων γύρω από το μέσο. Εάν τα δεδομένα σχηματίζουν έντονη κορυφή, τότε η κατανομή λέγεται λεπτόκυρτη²⁵ ($K > 3$), ενώ εάν τα δεδομένα έχουν πεπλατυσμένη κορυφή η κατανομή λέγεται πλατύκυρτη²⁶ ($K < 3$). Η μέση κατάσταση (κανονική κατανομή) λέγεται μεσόκυρτη²⁷ ($K = 3$). Στον τύπο εύρεσης της κύρτωσης του gretl ο αριθμός 3 αφαιρείται και συνεπώς το όριο για τον χαρακτηρισμό της κύρτωσης είναι το μηδέν, δηλαδή αν $K > 0$ τότε η κατανομή είναι λεπτόκυρτη, αν $K < 0$ τότε η κατανομή είναι πλατύκυρτη και αν $K = 0$ τότε η κατανομή είναι μεσόκυρτη:

$$K = \frac{1}{N} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\hat{s}} \right)^4 - 3$$

²⁴ Έστω ότι η τυπική απόκλιση εκφράζει κίνδυνο και οι δυο σειρές Y και X έχουν τυπικές αποκλίσεις 20 και 10, τότε θα επιλέγαμε τη σειρά X. Εάν όμως η μέση τιμή της Y είναι 20, ενώ της X είναι 8, τότε με βάση το συντελεστή μεταβλητότητας θα επιλέγαμε την Y καθώς:

$$CV = \frac{s}{y} = \frac{20}{20} = 1$$

$$CV = \frac{s}{\bar{x}} = \frac{10}{8} = 1,25$$



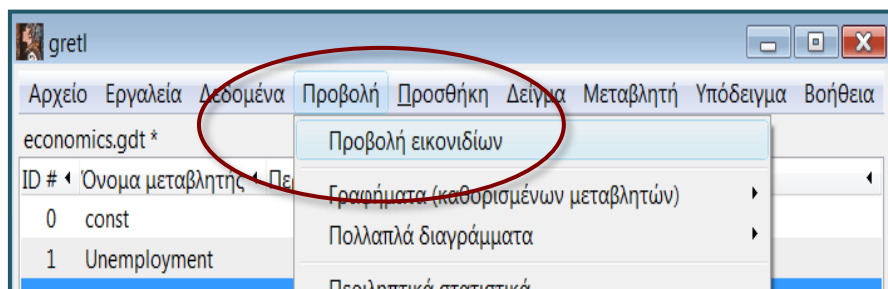
όπου \hat{s} είναι εκτιμητής της τυπικής απόκλισης, που όπως και στην περίπτωση της ασυμμετρίας ισούται με: $\hat{s} = s \sqrt{\frac{n-1}{n}}$

Κύρτωση -3 = K = 0,49 λεπτόκυρτη

- Τα **εκατοστιαία σημεία** χωρίζουν την κατανομή σε εκατό ίσα τμήματα²⁸. Το k εκατοστιαίο σημείο χωρίζει τα δεδομένα σε δύο τμήματα: στο πρώτο τμήμα έχουμε $k\%$ των δεδομένων και στο δεύτερο τμήμα έχουμε $(100 - k)\%$ των δεδομένων. **5% Perc. = 4,99** κάτω από την τιμή 4,99 βρίσκεται το 5% των δεδομένων και πάνω από την εν λόγω τιμή το 95% των δεδομένων. **95% Perc. = 12,06** αντιστοίχα κάτω από την τιμή 12,06 βρίσκεται το 95% των δεδομένων και πάνω από την εν λόγω τιμή το 5% των δεδομένων.
- Το **ενδοτεταρτημοριακό εύρος (IQR – InterQuartile Range)** είναι ένα μέτρο της στατιστικής διασποράς και είναι ίσο με τη διαφορά μεταξύ των άνω και κάτω τεταρτημορίων, $IQR = Q_3 - Q_1$. Το IQR υπολογίζει που βρίσκεται το 50% των κεντρικών δεδομένων και είναι ένας δείκτης που συχνά τον προτιμούν οι ερευνητές γιατί μπορεί να δείξει τη διασπορά των τιμών. Το IQR μπορεί να φανεί καθαρά σε ένα **θηκόγραμμα (box-plot)** στο οποίο φαίνονται επίσης σε συνδυασμό με το IQR οι **ακραίες τιμές (outliers)**²⁹. Στο επόμενο κεφάλαιο θα γίνει αναλυτική παρουσίαση κατασκευής θηκογραμμάτων με το gretl, καθώς επίσης και με το Eviews.

2.1.6. Αρχεία συνεδρίας (Sessions Files)

Η **συνεδρία (session)** στο gretl είναι μια διαδικασία που εξασφαλίζει στο χρήστη την παρακολούθηση, καταγραφή και αναθεώρηση βασικών εργασιών. Η ειδοποιός διαφορά με την **αποθήκευση δεδομένων (Save)** είναι ότι η αποθήκευση μιας συνεδρίας αφορά όχι μόνο δεδομένα, αλλά και εκτιμήσεις μοντέλων, μεταβολές στατιστικών κλπ. Για να δούμε τα βασικά εικονίδια (εργασίες) που παρακολουθεί μια συνεδρία επιλέγουμε **Προβολή/Προβολή εικονιδίων**.



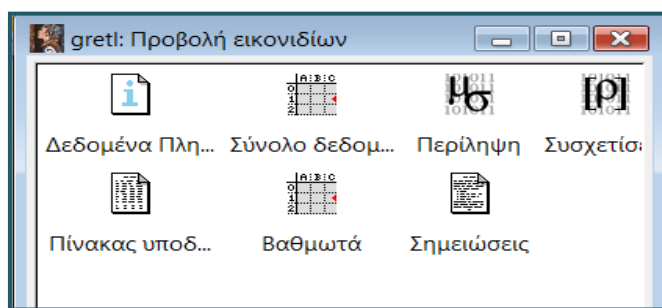
Εικόνα 2.54: Προβολή εικονιδίων στο gretl (1)

²⁸ Για να βρούμε ένα εκατοστιαίο σημείο k υπάρχει μια εμπειρική μέθοδος σύμφωνα με την οποία κάνουμε τα εξής βήματα:

- Διατάσσουμε τα δεδομένα κατά αύξουσα σειρά.
- Πολλαπλασιάσουμε το k ως δεκαδικό με το πλήθος των δεδομένων n .
 - Εάν το αποτέλεσμα είναι ακέραιος αριθμός, μετράμε τα δεδομένα από το μικρότερο προς το μεγαλύτερο μέχρι να βρούμε την τιμή που αντιστοιχεί σε αυτό που μας έδωσε το βήμα 2. Το k εκατοστιαίο σημείο είναι το ημίθροισμα της τιμής που μόλις βρήκαμε και της επόμενης.
 - Εάν όμως το αποτέλεσμα του πολλαπλασιασμού είναι δεκαδικός, τότε στρογγυλεύουμε στον πλησιέστερο ακέραιο και μετράμε τα δεδομένα από το μικρότερο προς το μεγαλύτερο μέχρι να βρούμε την τιμή που αντιστοιχεί σε αυτόν τον ακέραιο. Αυτή η τιμή είναι το k εκατοστιαίο σημείο.

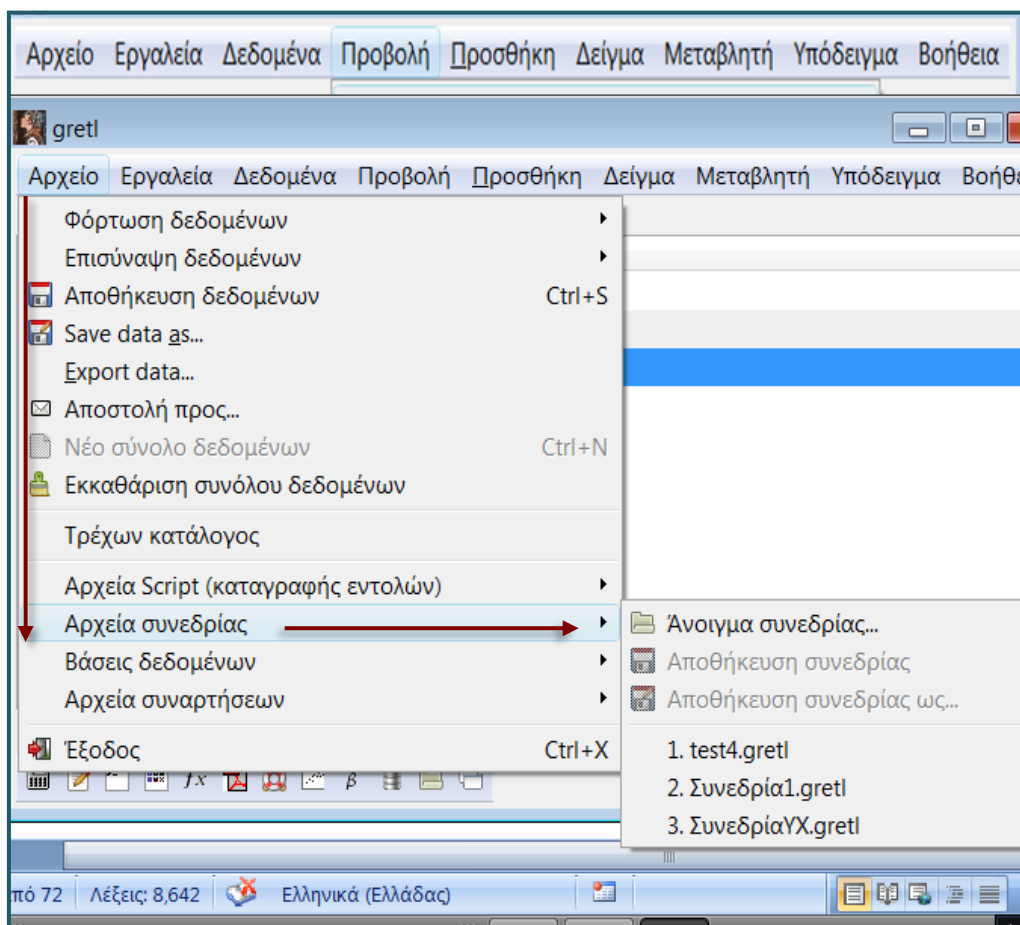
²⁹ Οι ακραίες τιμές είναι οι παρατηρήσεις κάτω από $Q_1 - 1,5 (IQR)$ ή πάνω από $Q_3 + 1,5 (IQR)$

Το αναδυόμενο πλαίσιο εικονιδίων περιέχει εικονίδια³⁰ τα οποία αφορούν **Δεδομένα**, **Περίληπτικά Στατιστικά**, **Συσχετίσεις** κλπ.



Εικόνα 2.55: Προβολή εικονιδίων στο gretl (2)



Εάν επιλέξουμε την εντολή **Αρχείο/ Αρχεία συνεδρίας** θα διαπιστώσουμε ότι η εντολή **Αποθήκευση συνεδρίας** δεν είναι διαθέσιμη, καθώς δεν υπέστησαν καμία μεταβολή ως προς το περιεχόμενό τους τα προαναφερόμενα εικονίδια.

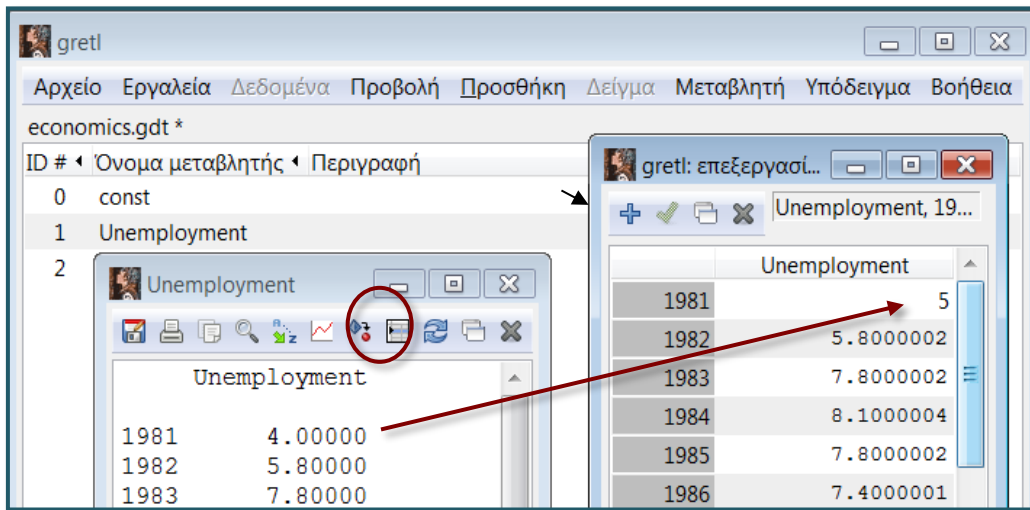


Εικόνα 2.56: Άνοιγμα συνεδρίας στο gretl (2)

³⁰ Το πληροφοριακό περιεχόμενο των οποίων ενεργοποιείται με διπλό κλικ στο σχετικό εικονίδιο.

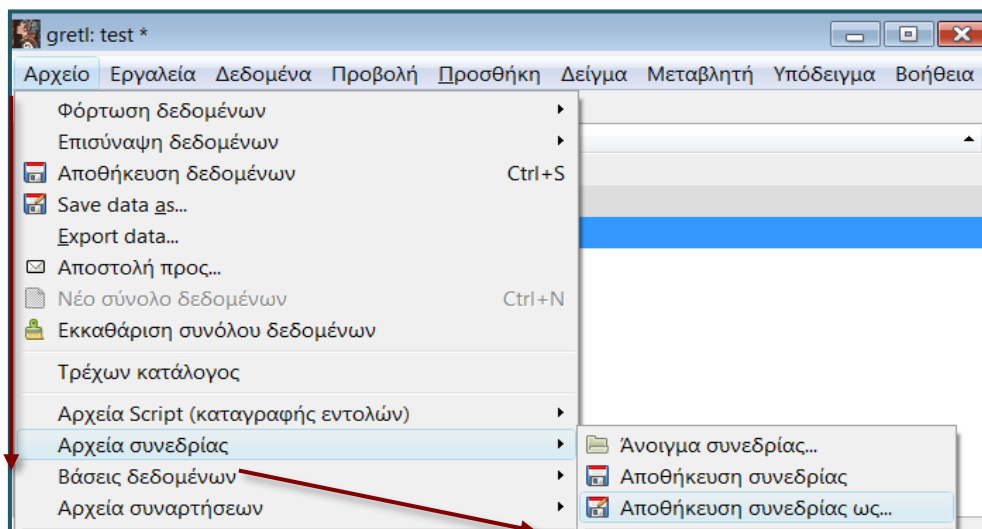
Έστω ότι με τις ακόλουθες ενέργειες μεταβάλλουμε το πρώτο στοιχείο της μεταβλητής **Unemployment** από 4 σε 5. Αναλυτικά:

- ανοίγουμε με διπλό κλικ τη μεταβλητή **Unemployment**
- επιλέγουμε το εικονίδιο  που αφορά στην επεξεργασία τιμών³¹ (Edit values).
- μεταβάλλουμε το πρώτο στοιχείο από 4 σε 5
- οριστικοποιούμε την καταχώρηση πιέζοντας το σχετικό σύμβολο 



Εικόνα 2.57: Άνοιγμα συνεδρίας στο gretl (3)

Θα παρατηρήσουμε ότι πλέον η εντολή **Αρχείο/Αρχεία συνεδρίας/Αποθήκευση συνεδρίας ως...** είναι διαθέσιμη, καθώς τα δεδομένα υπέστησαν μεταβολή και συνεπώς το περιεχόμενο των εν λόγω εικονιδίων επίσης μεταβλήθηκε και επομένως έχει νόημα να αποθηκευτεί.

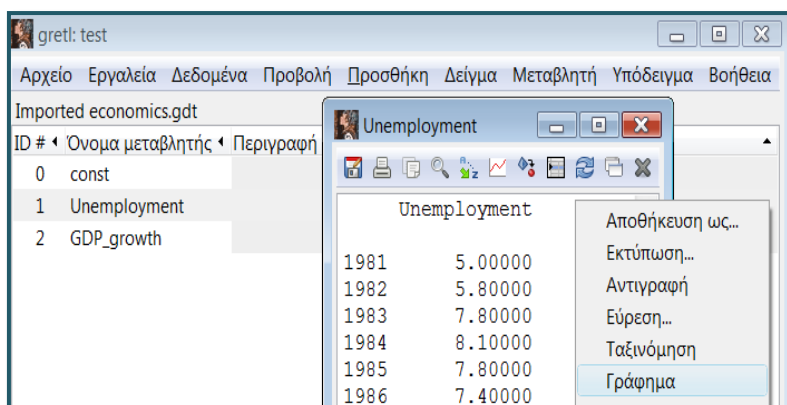


Εικόνα 2.58: Αποθήκευση συνεδρίας ως... στο gretl

³¹ Εναλλακτικά μπορεί να επιτευχθεί από τον αναδυόμενο κατάλογο όταν επιλέξουμε δεξί κλικ πάνω στο παράθυρο των τιμών της μεταβλητής.

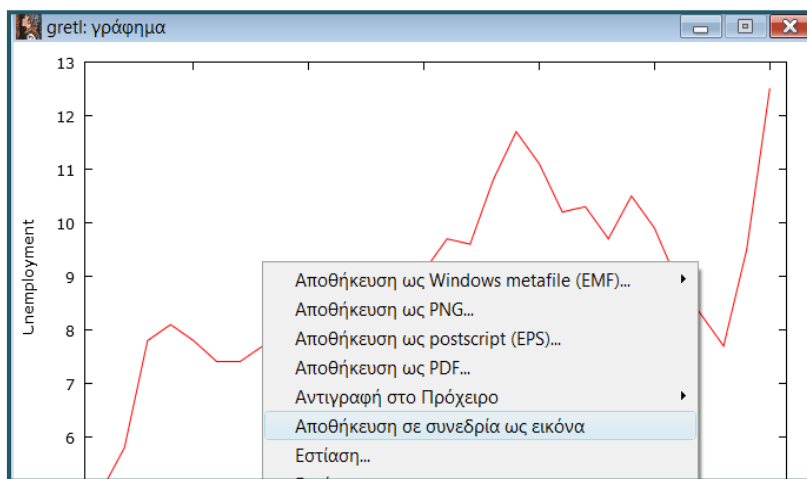
Σημειώνεται ότι μπορούμε να εισαγάγουμε και νέα εικονίδια σε μια συνεδρία που αφορούν στην παραγωγή γραφημάτων και οικονομετρικών εκτιμήσεων. Για παράδειγμα, έστω ότι ενδιαφερόμαστε να δημιουργήσουμε γράφημα της μεταβλητής **Unemployment** που να μπορούμε να το ανασύρουμε στο μέλλον μέσα από τη διαδικασία της σχετικής συνεδρίας. Τα βήματα που ακολουθούμε είναι τα εξής:

- ανοίγουμε με διπλό κλικ τη μεταβλητή **Unemployment**
- με δεξί κλικ πάνω στο παράθυρο των τιμών της μεταβλητής αναδύεται κατάλογος από τον οποίο επιλέγουμε **Γράφημα**³².



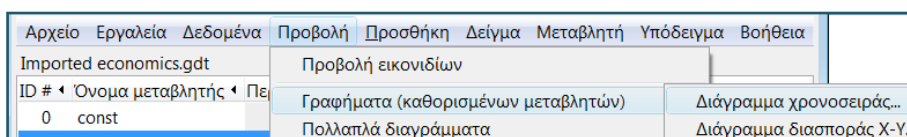
Εικόνα 2.59: Διαμόρφωση εικονιδίων συνεδρίας στο gretl

- Με δεξί κλικ στο παραγόμενο γράφημα αναδύεται κατάλογος από τον οποίο επιλέγουμε **Αποθήκευση συνεδρίας ως εικόνα**.

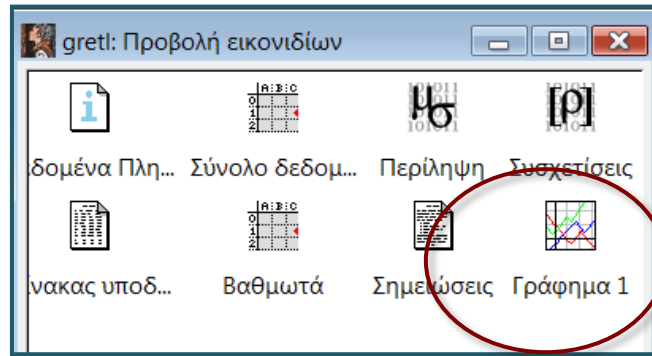


Εικόνα 2.60: Αποθήκευση σε συνεδρία ως εικόνα στο gretl (1)

³² Επιπλέον, το γράφημα μπορεί να παραχθεί με την εντολή **Προβολή/Γραφήματα (καθορισμένων μεταβλητών)/Διάγραμμα Χρονοσειράς..**

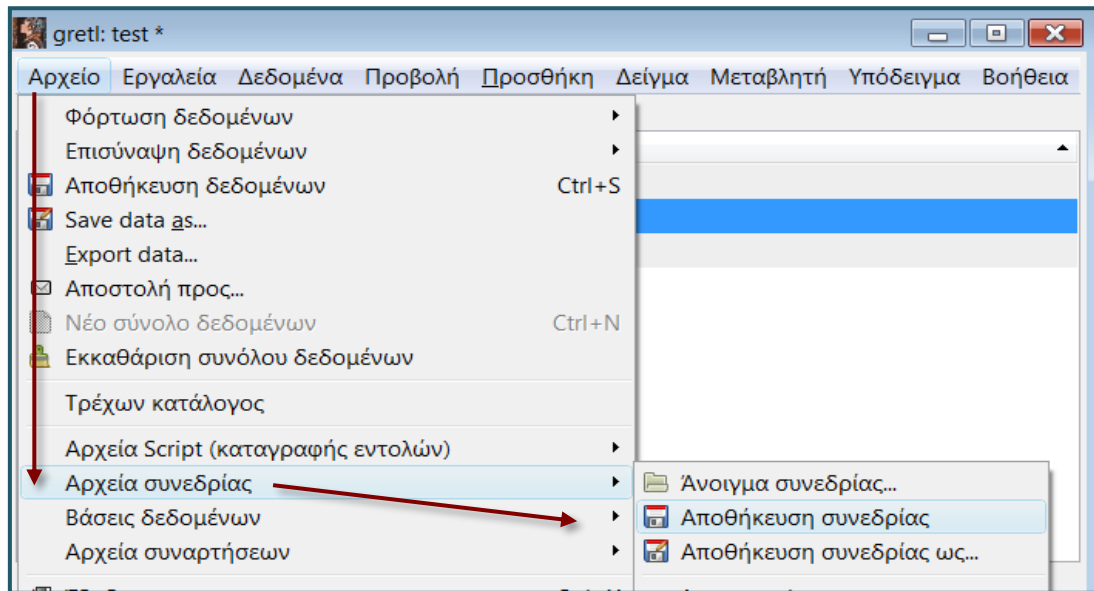


- Αναδύεται το πλαίσιο εικονιδίων που συμπεριλαμβάνει και το εικονίδιο του παραπάνω γραφήματος ως **Γράφημα 1**.



Εικόνα 2.61: Αποθήκευση σε συνεδρία ως εικόνα στο gretl (2)

- Με την εντολή **Αρχείο/Αρχεία συνεδρίας/Αποθήκευση συνεδρίας** αποθηκεύουμε το γράφημα στην τρέχουσα συνεδρία.



Εικόνα 2.62: Αποθήκευση συνεδρίας στο gretl

A.1. Βαθμοί ελευθερίας (degrees of freedom)

Εάν τα διαθέσιμα δεδομένα αφορούν n παρατηρήσεις, τότε μπορούμε να υποστηρίξουμε πως κάθε παρατήρηση αποτελεί και μία πληροφορία. Ο αριθμός των **βαθμών ελευθερίας (β.ε.)** είναι ουσιαστικά ο αριθμός των ανεξάρτητων πληροφοριών – δεδομένων που χρησιμοποιούνται για να γίνει ένας υπολογισμός. Ο αριθμός των βαθμών ελευθερίας είναι επίσης και ένα μέτρο για το κατά πόσο σίγουροι είμαστε ότι το δείγμα είναι αντιπροσωπευτικό του συνόλου του πληθυσμού, συνήθως όσο περισσότεροι βαθμοί ελευθερίας τόσο πιο σίγουροι είμαστε ότι το δείγμα είναι αντιπροσωπευτικό. Στη στατιστική, λοιπόν, οι βαθμοί ελευθερίας είναι ο αριθμός των τιμών – πληροφοριών που διαμορφώνονται ελεύθερα, δίχως περιορισμό, σε έναν υπολογισμό. Η εκτίμηση μιας παραμέτρου σε έναν υπολογισμό (τύπο) αφαιρεί και έναν βαθμό ελευθερίας, καθώς περιορίζει την ελεύθερη διαμόρφωση μιας τιμής.

Παράδειγμα 1: Εάν έχουμε μια ομάδα 5 παικτών μπάσκετ και 5 αντίστοιχες θέσεις, τότε ο προπονητής είναι ελεύθερος να τοποθετήσει τον καθένα σε όποια θέση επιθυμεί. Αν ένας παίχτης τοποθετηθεί σε μία ελεύθερη θέση, τότε τη δεσμεύει και οι ελεύθερες θέσεις είναι πλέον 4.

Παράδειγμα 2: Έστω ότι γνωρίζουμε τη μέση τιμή τριών παρατηρήσεων: $\mu = 7$ και την τιμή των δύο από τις τρεις παρατηρήσεις: 10 και 8. Για να βρούμε τη μέση τιμή γνωρίζουμε ότι θα πρέπει να αθροίσουμε τις τρεις τιμές και να διαιρέσουμε με το πλήθος τρία, δηλαδή

$$\frac{\overbrace{10 + 8 + x}^{\text{όπου } x \text{ η } \text{άγνωστη τιμή}}}{3} = 7 \Rightarrow x = 3$$

Εάν αντί για δύο γνωστές τιμές είχαμε μία (10) και οι υπόλοιπες δύο ήταν άγνωστες: x και y τότε:

$$\frac{10 + x + y}{3} = 7 \Rightarrow x + y + 10 = 21 \Rightarrow x + y = 11 \Rightarrow x = y - 11$$

γεγονός που σημαίνει ότι εφόσον οριστεί η τιμή για το y τότε η τιμή του x είναι δεδομένη. Οι βαθμοί ελευθερίας της παραπάνω σχέσης είναι 1, διότι μία τιμή είναι ελεύθερη προς διαμόρφωση (y), από τη στιγμή που αυτή οριστεί, τότε το αποτέλεσμα είναι δεδομένο.

Παράδειγμα 3: Έστω ότι σε δείγμα 50 παρατηρήσεων έχουμε γνωστή και ίση τη μέση τιμή $\bar{x} = 30$, αλλά οι τιμές των παρατηρήσεων είναι άγνωστες. Η γνώση (εκτίμηση) ή αλλιώς ο περιορισμός της μέσης τιμής σημαίνει και την απώλεια ενός βαθμού ελευθερίας, καθώς πλέον μόνο οι 49 παρατηρήσεις μπορούν να πάρουν ελεύθερα διάφορες τιμές, η 50οστή δεσμεύεται από τη μέση τιμή. Συγκεκριμένα,

$$\begin{aligned} \frac{x_1 + x_2 + \dots + x_{49} + x_{50}}{50} = 30 &\Rightarrow x_1 + x_2 + \dots + x_{49} + x_{50} = 30 \cdot 50 \Rightarrow \\ &\Rightarrow x_1 + x_2 + \dots + x_{49} + x_{50} = 1500 \end{aligned}$$

Εφόσον προσδιοριστούν οι 49 τιμές τότε η 50οστή εξάγεται από την παραπάνω σχέση, εάν για παράδειγμα $x_1 + x_2 + \dots + x_{49} = 1450$, τότε

$$1450 + x_{50} = 1500 \Rightarrow x_{50} = 1500 - 1450 = 50$$

Συνεπώς, όπως εύκολα γίνεται αντιληπτό η τελευταία τιμή έχει καθοριστεί ως αποτέλεσμα της σχέσης.

Συνοψίζοντας, οι βαθμοί ελευθερίας εξαρτώνται από το μέγεθος των υπό ανάλυση δειγμάτων και από τον αριθμό των εκτιμώμενων παραμέτρων. Εξετάζονται συνοπτικά οι ακόλουθες περιπτώσεις:

- **Ένα μόνο δείγμα:** Έστω ότι ζητείται σε ένα δείγμα μεγέθους n παρατηρήσεων, η εκτίμηση της δειγματικής διακύμανσης, γεγονός που προϋποθέτει την εκτίμηση μιας παραμέτρου (μέσος όρος).

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$$

Η παρουσία της μέσης τιμής στην εκτίμηση της διακύμανσης δεσμεύει μια τιμή και επομένως η δειγματική διακύμανση υφίσταται απώλεια ενός βαθμού ελευθερίας ($n - 1$).

- **Δύο δείγματα:** Έστω ότι ζητείται το διάστημα εμπιστοσύνης της διαφοράς των μέσων δυο

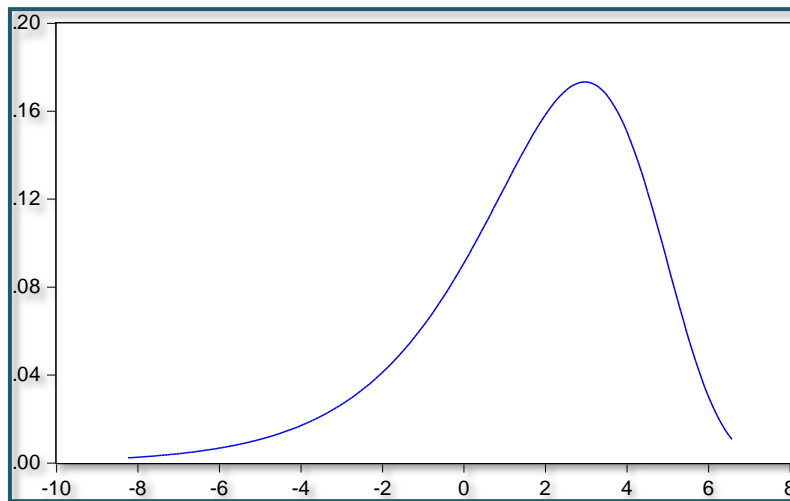
δειγμάτων n_1 και n_2 . Στην περίπτωση αυτή, η εκτίμηση της δειγματικής διακύμανσης δίνεται από τον τύπο:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

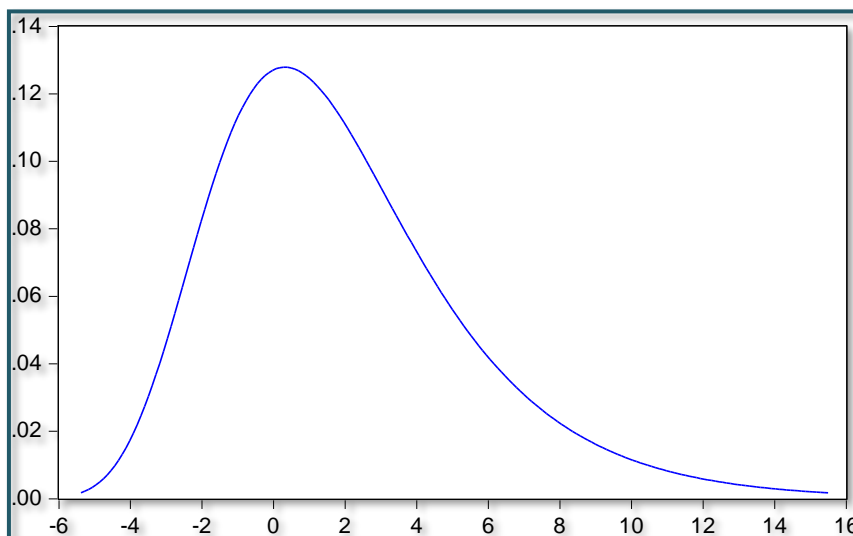
Υπάρχουν $n_1 + n_2$ παρατηρήσεις και υπάρχουν δύο μέσοι που έχουν εκτιμηθεί. Αυτό αφήνει $n_1 + n_2 - 2$ βαθμούς ελευθερίας για την εκτίμηση της διακύμανσης.

A.2. Κανονική κατανομή (normal distribution)

Τα δεδομένα μιας οποιασδήποτε μεταβλητής δύναται να είναι κατανεμημένα με ποικίλους τρόπους. Μπορεί για παράδειγμα να παρουσιάζουν αρνητική ή θετική ασυμμετρία όπως παρακάτω:

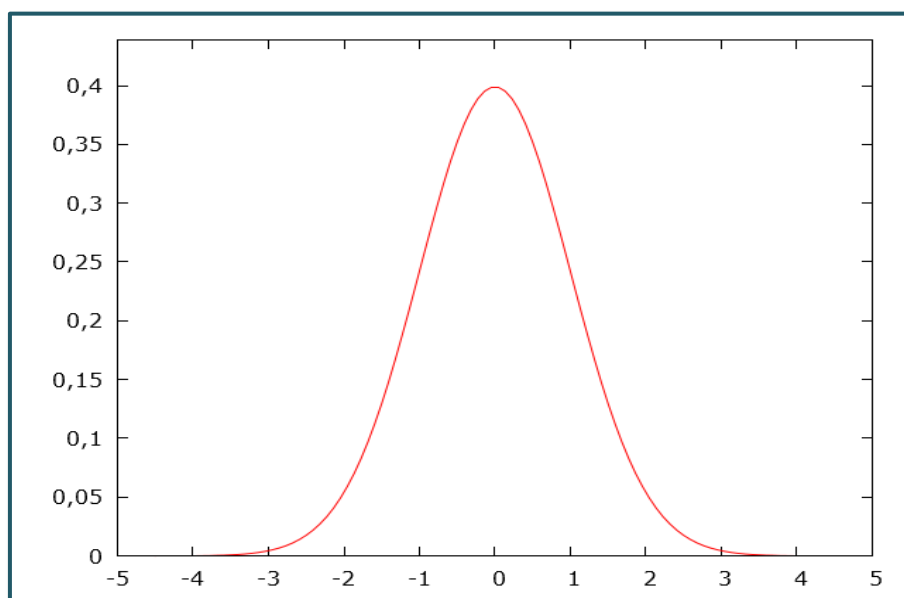


Γράφημα A.1: Αρνητική ασυμμετρία



Γράφημα A.2: Θετική ασυμμετρία

Αλλά υπάρχουν και πολλές περιπτώσεις όπου τα δεδομένα τείνουν να βρίσκονται συμμετρικά γύρω από μια κεντρική τιμή (μέσο όρο) σχηματίζοντας μια καμπάνα:



Γράφημα A.3: Κανονική κατανομή $N(0,1)$

Συνοπτικά τα χαρακτηριστικά της κανονικής κατανομής είναι:

- Ο οριζόντιος άξονας είναι η ευθεία των πραγματικών αριθμών. Το συνολικό εμβαδόν ανάμεσα στην κανονική καμπύλη και τον οριζόντιο άξονα είναι 1.
- Οι τιμές που μπορεί να πάρει η τυχαία μεταβλητή είναι άπειρες και επομένως η πιθανότητα να πάρουμε μία συγκεκριμένη τιμή είναι $\frac{1}{\infty} = 0$. Αυτό που αναζητούμε λοιπόν είναι η πιθανότητα να είμαστε πάνω ή κάτω από μία συγκεκριμένη τιμή a ή η πιθανότητα να είμαστε ανάμεσα σε δύο συγκεκριμένες τιμές a και β . Δηλαδή, υπολογίζουμε τις πιθανότητες $P(X < a)$ ή $P(X > a)$ ή $P(a < X < \beta)$.
- Ο μέσος χωρίζει την κατανομή σε δύο ίσα τμήματα, με άλλα λόγια η κανονική κατανομή είναι συμμετρική ως προς το μέσο της.
- Ο μέσος, η διάμεσος και η επικρατούσα τιμή ταυτίζονται.

Οι πιθανότητες στην κανονική κατανομή ακολουθούν το λεγόμενο εμπειρικό κανόνα που λέει τα εξής:

- Περίπου το 68% των τιμών βρίσκονται σε διάστημα μιας τυπικής απόκλισης πάνω και κάτω από το μέσο. Δηλαδή το διάστημα $(\mu - \sigma, \mu + \sigma)$ περιέχει περίπου το 68% των τιμών της κατανομής
- Το διάστημα $(\mu - 2\sigma, \mu + 2\sigma)$ περιέχει περίπου το 95% των τιμών της κατανομής
- Το διάστημα $(\mu - 3\sigma, \mu + 3\sigma)$ περιέχει περίπου το 99,7% των τιμών της κατανομής

Τα μέτρα που χαρακτηρίζουν την κανονική κατανομή είναι η μέση τιμή μ και η διακύμανση σ^2 , επομένως συμβολίζουμε:

$$X \sim N(\mu, \sigma^2)$$

και λέμε ότι η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση σ^2 . Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε την τυπική απόκλιση σ . Όλες οι κανονικές κατανομές, ανεξάρτητα από την τιμή που έχουν ο μέσος και η διακύμανση, έχουν τις ίδιες ιδιότητες και σχηματίζουν την ίδια βασική μορφή καμπάνας.

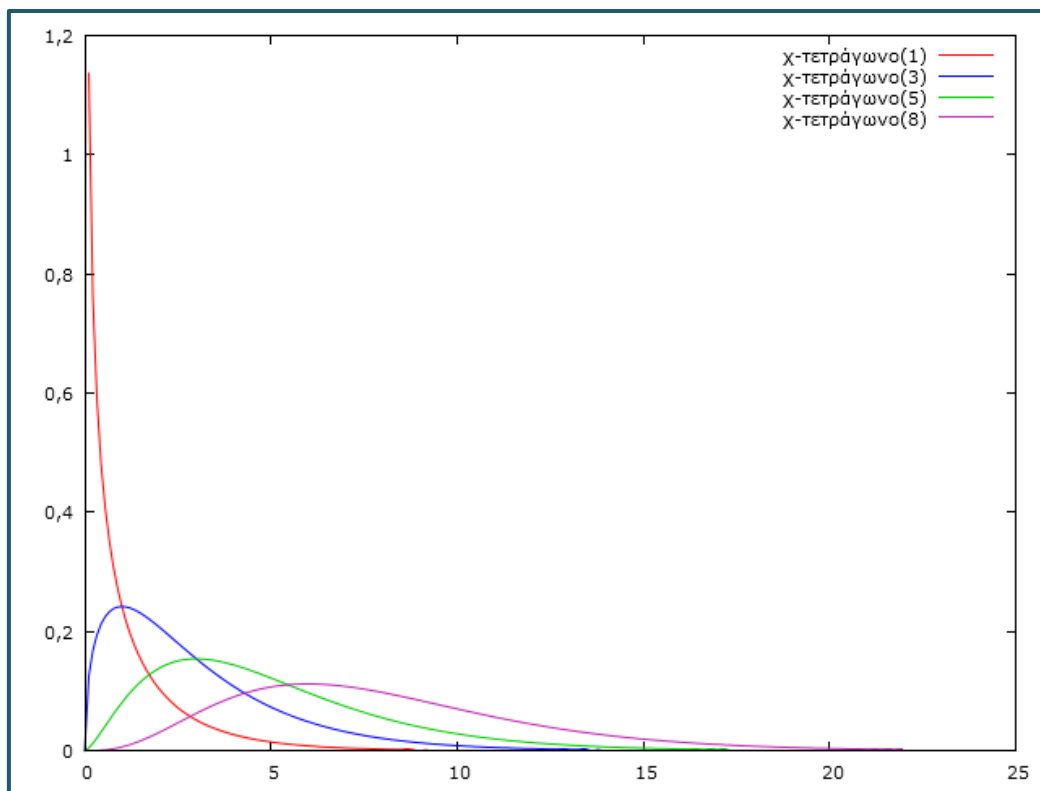
A.3. Η κατανομή χ^2 (chi square distribution)

Εάν έχουμε έναν αριθμό n τυχαίων μεταβλητών που όλες ακολουθούν την τυπική κανονική κατανομή, τότε το άθροισμα τετραγώνων τους ακολουθεί την κατανομή χ_n^2 (διαβάζουμε “χι τετράγωνο με n βαθμούς ελευθερίας”). Δηλαδή, έστω ότι,

$$Y = X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2$$

Με $X_i \sim N(0, 1)$ για $i = 1, 2, 3, \dots, n$. Τότε η Y ακολουθεί την κατανομή χ^2 με n βαθμούς ελευθερίας και γράφουμε: $Y \sim \chi_n^2$

Η κατανομή χ_n^2 έχει διαφορετικό σχήμα ανάλογα με τους βαθμούς ελευθερίας. Στο Γράφημα A.4 έχουμε τη γραφική παράσταση τεσσάρων κατανομών χ_n^2 και συγκεκριμένα για $n = 1$, $n = 3$, $n = 5$ και $n = 8$. Η κατασκευή έγινε με το gretl.



Γράφημα A.4: Η κατανομή χ^2 με 1, 3, 5 και 8 βαθμούς ελευθερίας