A white rectangular sign with black text

Description automatically generated

**ΣΕΜΙΝΑΡΙΑ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ – RESEARCH METHODS**

**Διδάσκων:** Επ. καθ. Λογιστικής Ορέστης Βλησμάς

Εφαρμογές στο STATA

Εφαρμογή 4η   
Διαμόρφωση μεταβλητών στο Stata και περιγραφική στατιστική

Ακολουθήστε τα παρακάτω βήματα:

**Βήμα 1ο**:

1. Μεταβείτε στο e-class του σεμιναρίου. Στο folder με ονομασία Laboratory του e-class αναζη-τήστε το αρχείο με ονομασία Application\_4.dta. Αποθηκεύεται το τοπικά στον υπολογιστή σας.
2. Μεταβείτε στο e-class του σεμιναρίου. Στο folder με ονομασία Laboratory του e-class αναζη-τήστε το αρχείο με ονομασία Code\_Application\_4.txt. Αποθηκεύεται το τοπικά στον υπολογιστή σας και ανοίξτε το στο πρόγραμμα λογισμικού **STATA**.
3. Μεταβείτε στο e-class του σεμιναρίου. Στο folder με ονομασία Laboratory του e-class αναζη-τήστε το αρχείο με ονομασία Roychowdhury\_Earnings management through real activities manipulation.pdf. Αποθηκεύεται το τοπικά στον υπολογιστή σας.
4. Μεταβείτε στο e-class του σεμιναρίου. Στο folder με ονομασία Library\_Stata αναζητήστε και ανοίξτε το αρχείο με ονομασία Xristikos\_Odigos\_Stata. Αποθηκεύεται το τοπικά στον υπολο-γιστή σας. Το αρχείο αυτό θα το έχετε ανοικτό στην διάρκεια της εκτέλεσης της εφαρμογής προκειμένου να το συμβουλεύεστε αν χρειαστεί.

**Βήμα 2ο**: Διαμόρφωση μεταβλητής δηλωτικής του κλάδου δραστηριοποίησης

Στο αρχείο Application\_4.dta περιλαμβάνονται δεδομένα που αφορούν την ερευνητική εργασία Roychowdhury (2006). Το δείγμα εταιρειών έχει λάβει δηλωτικό κλάδου δραστηριότητας κατά Standard Industrial Classification (SIC). Η σχετική μεταβλητή έχει την ομομασία sic1. Το SIC είναι ένα σύστημα ταξινόμησης βιομηχανιών με τετραψήφιο κωδικό ως μέθοδος τυποποίησης της ταξινόμησης του κλάδου για στατιστικούς σκοπούς σε όλους τους οργανισμούς. Ιδρύθηκε στις Ηνωμένες Πολιτείες το 1937 και χρησιμοποιείται από κυβερνητικούς φορείς για την ταξινόμηση των κλάδων δραστηριότητας.

ΓΕΝΙΚΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ:

**1.** Το σύστημα κατηγοριοποίησης SIC αποδίδει ένα τετραψήφιο κωδικό κλάδου δραστηριότητας σε κάθε μία εταιρεία. Όταν εκπονείτε τη διπλωματική εργασίας συνίσταται σφόδρα μαζί με όλες τις άλλες μεταβλητές που θα κατεβάσετε από τις συνδρομητικές βάσεις δεδομένων να βεβαιωθείτε ότι έχετε κατεβάσει και τη μεταβλητή που είναι δηλωτική του κλάδου δραστηριότητας (κωδικό για SIC ή παρεμφερή κωδικό).

**2.** Πρόσφατα το σύστημα κατηγοριοποίησης SIC αντικαταστάθηκε από το σύστημα κατηγοριοποίησης North American Industry Classification System (NAICS code). Ωστόσο, αρκετοί κυβερνητικοί φορείς και ρυθμιστικές αρχές εξακολουθούν να κάνουν χρήση του συστήματος κατηγοριοποίησης SIC.

**3.** Συνίσταται η ομαδοποίηση των κλάδων δραστηριότητας σε γενικότερες ομάδες δεδομένου ότι τα συστήματα κατηγοριοποίησης SIC και NAICS διαμορφώθηκαν για λόγους εθνικοστατιστικής καταγραφής και επεξεργασίας δεδομένων. Η ομαδοποίηση **ΔΕΝ** γίνεται με **ΤΥΧΑΙΟ** και αυθαίρετο τρόπο. Συνήθως ακολουθείται ένας αποδεκτός τρόπος ομαδοποίησης. Έχει επικρατήσει να αξιοποιείται η ομαδοποίηση κατά Fama και French.

**4.** Συνήθως στις διπλωματικές εργασίες με θεματολογία λογιστικής θα πρέπει να εξαιρεθούν οι λογιστικές οντότητες που δραστηριοποιούνται στο τραπεζικό κλάδο και εν γένει στον χρηματοοικονομικό τομέα. Ο λόγος είναι ότι σημαντικά λογιστικά μεγέθη του έχουν ιδιάζουσα συμπεριφορά και κατά συνέπεια η συμπερίληψη τους μας οδηγεί σε παραπλανητικά συμπεράσματα για λογιστικά θέματα ευρύτερης σημασίας.

Για να περιορισθεί ο αριθμός των κλάδων δραστηριοποίησης θα υιοθετηθεί η κατηγοριοποίηση 10 κλάδων δραστηριοποίησης κατά Fama και French. Για το λόγο αυτό πραγματοποιείστε τις ακόλουθες ενέργειες:

**Ενέργεια 1 –** Απόρριψη εταιρειών που δραστηριοποιούνται στο τραπεζικό κλάδο και εν γένει στον χρηματοοικονομικό τομέα, δεν έχουν κωδικό τομέα δραστηριοποίησης και γενική εκκαθάριση:

drop if sic1==.

drop if sic1>=4400 & sic1<5000

drop if sic1>5999 & sic1<6501

drop if sales<0

drop if tas<0

Με την εντολή drop if απορρίπτονται οι εταιρείες που πληρούν συγκριμένες συνθήκες:

Δραστηριοποιούνται στο τραπεζικό κλάδο και εν γένει στον χρηματοοικονομικό τομέα (drop if sic1>=4400 & sic1<5000, και drop if sic1>5999 & sic1<6501)

Δεν έχουν κωδικό τομέα δραστηριοποίησης (drop if sic1==.)

Γενική εκκαθάριση αν οι πωλήσεις τους (sales) ή το σύνολο του ενεργητικού (tas) είναι μηδέν (drop if sales<0, και drop if tas<0).

**Ενέργεια 2 –** Δημιουργία βοηθητικής και κύριας μεταβλητής:

generate twodigit= int(sic1/100)

generate industry=.

Άρα δημιουργείτε η βοηθητική μεταβλητή twodigit η οποία περιορίζει την κωδικοποίηση της sic1 σε δύο ψηφία. Επίσης δημιουργείτε μία μεταβλητή χωρίς τιμές η οποία θα χρησιμοποιηθεί ως δηλωτική του κλάδου δραστηριοποίησης κάθε εταιρείας του δείγματος σύμφωνα με την κατηγοριοποίηση 10 κλάδων δραστηριοποίησης κατά Fama και French. Και οι δύο μεταβλητές θα αξιοποιηθούν (twodigit, industry) θα αξιοποιηθούν στην επόμενη ενέργεια.

**Ενέργεια 3 –** Μετατροπή της μεταβλητής industry ως δηλωτική του κλάδου δραστηριότητας σύμφωνα με την κατηγοριοποίηση 10 κλάδων δραστηριοποίησης κατά Fama και French:

replace industry=1 if inrange(twodigit, 1, 14)

replace industry=2 if inrange(twodigit, 15, 17)

replace industry=3 if inrange(twodigit, 20, 21)

replace industry=4 if inrange(twodigit, 22, 23)

replace industry=5 if inrange(twodigit, 24, 27)

replace industry=6 if inrange(twodigit, 28, 30)

replace industry=7 if inrange(twodigit, 31, 34)

replace industry=8 if inrange(twodigit, 35, 39)

replace industry=9 if inrange(twodigit, 40, 47)

replace industry=10 if inrange(twodigit, 50, 52)

replace industry=11 if inrange(twodigit, 53, 59)

replace industry=12 if inrange(twodigit, 70, 79)

replace industry=13 if inrange(twodigit, 80, 99)

ΠΑΡΑΤΗΡΗΣΕΙΣ:

**1.** Η εντολή replace ΟΝΟΜΑ\_ΜΕΤΑΒΛΗΤΗΣ = ΤΙΜΗ ΜΕΤΑΒΛΗΤΉΣ δίδει εντολή στο STATA να αντικαταστήσει την υφιστάμενη τιμή μίας μεταβλητής με μία νέα τιμή (ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ).

**2.** Η εντολή if inrange (ΟΝΟΜΑ\_ΜΕΤΑΒΛΗΤΗΣ, ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ 1, ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ 2) δίδει εντολή στο STATA να εκτελέσει μία εντολή (στην προκειμένη περίπτωση της εντολή replace) αν η τιμή μίας μεταβλητής ευρίσκεται μεταξύ δύο τιμών (ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ 1, ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ 2).

**Βήμα 3ο**: Δημιουργία πρόσθετων μεταβλητών Roychowdhury (2006)

Με βάση αναφοράς την ερευνητική εργασία του Roychowdhury (2006) θα πρέπει να δημιουργηθούν κάποιες πρόσθετες μεταβλητές. Για το λόγο αυτό αρχικά πρέπει να μελετήσουμε το σχετικό κείμενο της ερευνητικής εργασίας:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Από την ανάγνωση του κειμένου διαπιστώνουμε ότι πρέπει να δημιουργηθούν τρεις μεταβλητές για την χειραγώγηση λογιστικών κερδών μέσω λειτουργικών αποφάσεων (real earnings management) ως:

1. Ασυνήθιστα χαμηλές καθαρές ταμειακές ροές από λειτουργικές δραστηριότητες εξαιτίας της μόχλευσης των πωλήσεων.
2. Υπερβάλλουσα μείωση των δαπανών που υπόκεινται στη διακριτική ευχέρεια των διοικητικών στελεχών.
3. Υπερβάλλουσα μείωση του ανά μονάδα κόστους εξαιτίας της αδικαιολόγητης αύξησης της παραγωγής.

Για διευκόλυνση θα χρησιμοποιήσουμε το γενικό όρο μεταβλητές REM όταν αναφερόμαστε γενικά στις μεταβλητές αυτές και τα ακόλουθα ακρωνύμια για κάθε μία επιμέρους μεταβλητές ως εξής:

1. abnormalcfor: ασυνήθιστα χαμηλές καθαρές ταμειακές ροές από λειτουργικές δραστηριότητες εξαιτίας της μόχλευσης των πωλήσεων.
2. abnormalDiscrr: υπερβάλλουσα μείωση των δαπανών που υπόκεινται στη διακριτική ευχέρεια των διοικητικών στελεχών.
3. abnormalproducr: υπερβάλλουσα μείωση του ανά μονάδα κόστους εξαιτίας της αδικαιολόγητης αύξησης της παραγωγής.

O υπολογισμός των μεταβλητών αυτών γίνεται με εκτίμηση μοντέλων παλινδρόμησης των οποίων οι εκτιμηθείσες τιμές αφορούν το κανονικό μέγεθος και τα κατάλοιπα τις μεταβλητές abnormalcfor, abnormalDiscrr και abnormalproducr. Επίσης θα πρέπει να δημιουργηθεί μία επιπλέον μεταβλητή που θα σηματοδοτεί τα «ύποπτα» έτη (με την ονομασία (suspect). Επιπλέον θα πρέπει να δημιουργηθούν μία σειρά από μεταβλητές που είναι αναγκαίες για την παλινδρόμηση.

Για το λόγο αυτό πραγματοποιείστε τις ακόλουθες ενέργειες:

**Ενέργεια 1 –** Δημιουργία αναγκαίων μεταβλητών για την εκτίμηση των παλινδρομήσεων που θα αποτελέσουν βάση για το υπολογισμό των μεταβλητών abnormalcfor, abnormalDiscrr και abnormalcfor. Δημιουργία μεταβλητής suspect:

xtset id year

by id: generate cfor=cfo/L.tas

by id: generate last=1/L.tas

by id: generate revr=sales/L.tas

by id: generate Drev=(sales-L.sales)

by id: generate Drevr=(Drev/L.tas)

by id: generate lS=(L.sales-L2.sales)

by id: generate lDrevr=lS/L.tas

by id: generate Dinv=(inv-L.inv)

egen produc= rowtotal(cogs Dinv) if cogs!=. | Dinv!=.

by id: generate producr=produc/L.tas

egen Discr= rowtotal (rd sga) if rd!=. | sga!=.

by id: generate Discrr=Discr/L.tas

by id: generate revr2=L.sales/L.tas

by id: generate accr=ni-cfo

by id: generate accrtas=(accr/l.tas)

by id: generate ppetas=(l.ppe/l.tas)

generate suspect=0

replace suspect=1 if (ni/L.tas)<=0.005 & (ni/L.tas)>=0

ΠΑΡΑΤΗΡΗΣΕΙΣ

**1.** Όταν ανοίγεται ένα αρχείο dta το οποίο εμπεριέχει δεδομένα τύπου πάνελ (και φυσικά η δομή των δεδομένων είναι τύπου Long) ΘΑ ΠΡΕΠΕΙ ΟΠΩΣΔΗΠΟΤΕ να TO δηλώσετε στο STATA. Αυτό γίνεται με τη σύνταξη της εντολής: xtset ΔΗΛΩΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΠΕΡΙΠΤΩΣΗΣ ΔΗΛΩΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΧΡΟΝΟΥ. Στην περίπτωσή μας η μεταβλητή id εκτελεί τον ρόλο της δηλωτικής μεταβλητής περίπτωσης και η μεταβλητή year τη δηλωτική μεταβλητή χρόνου.

**2.** Όταν πριν από μία εντολή δηλώνεται by id: δίδει την οδηγία στο STATA να την εκτελέσεις ξεχωριστά για κάθε περίπτωση της δηλωτικής μεταβλητής id.

**3.** Η εντολή egen είναι πολύ δυναμική. Για περισσότερες λεπτομέρειες αναζητήστε μελετήσε το παράρτημα του αρχείου με ονομασία Xristikos\_Odigos\_Stata.

**Ενέργεια 2 –** Δημιουργία μεταβλητής abnormalcfor:

generate abnormalcfor=.

capture program drop abnormalcfor

program abnormalcfor

forvalues i = 1(1)13 {

capture reg cfor last revr Drevr if industry==`i'

capture predict abnormalcfor`i' if industry==`i', resid

replace abnormalcfor= abnormalcfor`i' if industry==`i'

capture drop abnormalcfor`i'

}

end

abnormalcfor

**Ενέργεια 3 –** Δημιουργία μεταβλητής abnormalDiscrr:

generate abnormalDiscrr=.

capture program drop abnormalDiscrr

program abnormalDiscrr

forvalues i = 1(1)13 {

capture reg Discrr last revr2 if industry==`i'

capture predict abnormalDiscrr`i' if industry==`i', resid

replace abnormalDiscrr= abnormalDiscrr`i' if industry==`i'

capture drop abnormalDiscrr`i'

}

end

abnormalDiscrr

**Ενέργεια 4 –** Δημιουργία μεταβλητής abnormalcfor:

generate abnormalproducr=.

capture program drop abnormalproducr

program abnormalproducr

forvalues i = 1(1)13 {

capture reg producr last revr Drevr lDrevr if industry==`i'

capture predict abnormalproducr`i' if industry==`i', resid

replace abnormalproducr= abnormalproducr`i' if industry==`i'

capture drop abnormalproducr`i'

}

end

abnormalproducr

**Βήμα 4ο**: Εξαγωγή περιγραφικών στατιστικών

Εφόσον έχετε εκτελέσει τα παραπάνω βήματα δοκιμάστε τις ακόλουθες εντολές στο STATA:

* dercibe (με και χωρίς αναφορά σε συγκεκριμένες μεταβλητές)
* summarize (με και χωρίς αναφορά σε συγκεκριμένες μεταβλητές)
* codebook (με και χωρίς αναφορά σε συγκεκριμένες μεταβλητές)
* by vrCAT, sort: summarize
* tab (για μία μεταβλητή). Δοκιμάστε την επιλογή plot sort.
* tab1 (για περισσότερες μεταβλητές). Δοκιμάστε τις επιλογές row nofreq, row, column.
* tabstat vr1 vr2 vr3, s(mean semean median sd var skew k count sum range min max)

A table of information with text

Description automatically generated with medium confidence

Έστω ότι επιθυμείτε να εξετασθεί η μέσου όρου μεταξύ δύο ανεξάρτητων μεταβλητών που αφορούν το συνολικό δείγμα παρατηρήσεων (έστω των μεταβλητών vs1 και vs2), με τη χρήση της εντολής ttest. Η σύνταξη της εντολής στο **STATA** θα ήταν:   
ttest vr1==vr2

Στο παράθυρο Results θα εμφανισθεί το αποτέλεσμα της εντολής.

Ενδέχεται να μας ενδιαφέρει να ελέγξουμε διαφορά μέσου όρου της ίδιας μεταβλητής υπολογισμένοι για δύο υποσύνολα δεδομένων από το συνολικό δείγμα παρατηρήσεων. Υποθέστε ότι η μεταβλητή vrBIN είναι δυαδική και αναφέρεται στο αν υπαρχει μία ιδιότητα σε μία παρατήρηση. Τότε μπορούμε να ελέγχουμε αν ο μέσος όρος της μεταβλητής vr1 διαφοροποιείται μεταξύ δύο υποσυνόλων δεδομένων εκ των οποίων το ένα έχει την ιδιότητα και το άλλο δεν την έχει. Η σύνταξη της εντολής στο **STATA** θα ήταν:   
ttest vr1, by vrBIN

Στο παράθυρο Results θα εμφανισθεί το αποτέλεσμα της εντολής.

Ενδέχεται να επιθυμούμε να εξετάσουμε αν ο μέσος όρος μίας μεταβλητής είναι ίσος ή διάφορος από μία υποθετική τιμή (έστω z). Η σύνταξη της εντολής στο **STATA** θα ήταν:   
ttest vr1=z

**Βήμα 5ο**: Εξαγωγή αποτελεσμάτων στο MSWORD

Εφόσον έχετε εκτελέσει τα παραπάνω βήματα θα πρέπει να εξάγεται τα αποτελέσματα σας.

Η εξαγωγή αποτελεσμάτων περιγραφικής στατιστικής από το **STATA** μπορεί να γίνει με τη χρήση της εντολής outreg2. Η εντολή αυτή όμως θα πρέπει να εγκατασταθεί στο **STATA** και για το λόγο αυτό θα πρέπει να εκτελεσθεί ο ακόλουθος κώδικας στο **STATA**:

ssc install outreg2

Δοκιμάστε τις ακόλουθες εντολές στο STATA:

outreg2 using ΟΝΟΜΑ\_ΑΡΧΕΙΟΥ.doc, replace sum(log)

Όπου στη θέση ΟΝΟΜΑ\_ΑΡΧΕΙΟΥ δίνεται μία ονομασία με λατινικούς χαρακτήρες. Προσοχή γιατί η επιλογή replace θα αντικαταστήσει προϋπάρχουν αρχείο με το ίδιο όνομα.

Εάν θέλετε να εξαγάγετε τον πίνακα αποτελεσμάτων στο Excel, χρησιμοποιήστε την επέκταση \*.xls αντί να χρησιμοποιήσετε \*.doc

Αν επιθυμείτε να εκτελεσθεί η εντολή για συγκεκριμένες μεταβλητές (να εξαχθούν τα αποτελέσματα της περιγραφικής στατιστικής για επιλεγμένες μεταβλητές) τότε προσθέτετε στο τέλος της επιλογή keep () και μέσα στην παρένθεση βάζετε τα ονόματα των μεταβλητών (εστω vr1, vr2 και vr3). Η γενική σύνταξη της έχει ως εξής:

outreg2 using ΟΝΟΜΑ\_ΑΡΧΕΙΟΥ.doc, replace sum(log) keep(vr1 vr2 vr3)

Η εντολή outreg2 μπορεί να εκτελεσθεί και για κατηγορίες δεδομένων. Για παράδειγμα, αν στα δεδομένα μας υπάρχει μία μεταβλητή με ονομασία vrCAT η οποία κατηγοριοποιεί τα δεδομένα μας σε πέντε μεγάλες κατηγορίες με βάση κάποιο κριτήριο (όπως για παράδειγμα το μέγεθος της εταιρείας σε πολύ μικρή, μικρή, μεσαία, μεγάλη και πολύ μεγάλη με δηλωτικές αριθμητικές τιμές 1, 2, 3, 4 και 5 αντίστοιχα). Με την επιλογή by δύναται να λάβουμε ξεχωριστούς πίνακες με τα περιγραφικά στατιστικά μεταβλητών για κάθε μία κατηγορία δεδομένων:

by vrCAT, sort: outreg2 using ΟΝΟΜΑ\_ΑΡΧΕΙΟΥ.doc, replace sum(log)