

ΣΕΜΙΝΑΡΙΑ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ – RESEARCH METHODS

Διδάσκων: Αναπλ. καθ. Λογιστικής Ορέστης Βλησμάς

Εφαρμογή 4^η

Διαμόρφωση μεταβλητών στο Stata και περιγραφική στατιστική

Ακολουθήστε τα παρακάτω βήματα:

Βήμα 1^ο:

1. Μεταβείτε στο e-class του σεμιναρίου. Στο folder με ονομασία Laboratory του e-class αναζητήστε 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 αναζητήστε 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 αναζητήστε 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 υπολογιστή σας. Το αρχείο αυτό θα το έχετε ανοικτό στην διάρκεια της εκτέλεσης της εφαρμογής προκειμένου να το συμβουλευέστε αν χρειαστεί.

Βήμα 2^ο: 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 ονομασία **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 εργασία 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 ONOMA_METABΛΗΤΗΣ = TIMH METABΛΗΤΗΣ δίδει εντολή στο STATA να αντικαταστήσει την υφιστάμενη τιμή μίας μεταβλητής με μία νέα τιμή (TIMH METABΛΗΤΗΣ).

2. Η εντολή if inrange (ONOMA_METABΛΗΤΗΣ, TIMH METABΛΗΤΗΣ 1, TIMH METABΛΗΤΗΣ 2) δίδει εντολή στο STATA να εκτελέσει μία εντολή (στην προκειμένη περίπτωση της εντολής replace) αν η τιμή μίας μεταβλητής ευρίσκεται μεταξύ δύο τιμών (TIMH METABΛΗΤΗΣ 1, TIMH METABΛΗΤΗΣ 2).

Βήμα 3^ο: 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:

4.2. Estimation models

Following Dechow et al. (1998), hereafter DKW, I express normal cash flow from operations as a linear function of sales and change in sales in the current period (Eq. (3) in Appendix B). To estimate the model, I run the following cross-sectional regression for every industry and year:

$$\text{CFO}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(S_t/A_{t-1}) + \beta_2(\Delta S_t/A_{t-1}) + \varepsilon_t, \quad (1)$$

where A_t is the total assets at the end of period t , S_t the sales during period t and $\Delta S_t = S_t - S_{t-1}$.

For every firm-year, abnormal cash flow from operations is the actual CFO minus the “normal” CFO calculated using estimated coefficients from the corresponding industry-year model and the firm-year’s sales and lagged assets.¹⁸

Expenses in DKW are expressed as a linear function of contemporaneous sales. Following DKW and allowing for intercepts, the model for normal COGS is estimated as

$$\text{COGS}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta(S_t/A_{t-1}) + \varepsilon_t. \quad (2)$$

Similarly, following DKW, I estimate the model for ‘normal’ inventory growth using the following regression:

$$\Delta \text{INV}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(\Delta S_t/A_{t-1}) + \beta_2(\Delta S_{t-1}/A_{t-1}) + \varepsilon_t, \quad (3)$$

where ΔINV_t is the change in inventory in period t .

I define production costs as $\text{PROD}_t = \text{COGS}_t + \Delta \text{INV}_t$. Using (2) and (3), I estimate normal production costs from the following industry-year regression.¹⁹

$$\text{PROD}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(S_t/A_{t-1}) + \beta_2(\Delta S_t/A_{t-1}) + \beta_3(\Delta S_{t-1}/A_{t-1}) + \varepsilon_t. \quad (4)$$

Under the simplifying assumptions in DKW, discretionary expenses should be also expressed as a linear function of contemporaneous sales, similar to COGS. The relevant regression would then be:

$$\text{DISEXP}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta(S_t/A_{t-1}) + \varepsilon_t,$$

where DISEXP_t is discretionary expenses in period t .

This creates the following problem: if firms manage sales upward to increase reported earnings in any year, they can exhibit unusually low residuals from the above regression in that year, even when they do not reduce discretionary expenses. To avoid this problem, discretionary expenses are expressed as a function of lagged sales. Therefore, to estimate normal discretionary expenses, I run the following regression for every industry and year:

$$\text{DISEXP}_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta(S_{t-1}/A_{t-1}) + \varepsilon_t. \quad (5)$$

4.3. Selection of suspect firm-years

Fig. 1 groups firm-years into intervals based on net income scaled by total assets at the beginning of the year. The histogram of scaled earnings is constructed with widths of 0.005 for the range -0.075 to $+0.075$.²⁰ The histogram in Fig. 1 is similar to that documented by prior literature, with the prominent upward shift in the frequency of firm-years going from the left of zero to the right. Researchers have argued that it is likely that firm-years in the interval just right of zero manage their earnings to report income marginally above zero. Since earnings are scaled by total assets, the discontinuity at zero cannot be explained by Durlsch and Easton (2005), who argue that scaling by market capitalization generates the discontinuity.

To increase the power of my tests to detect real activities manipulation, I concentrate on firm-years in the interval to the immediate right of zero, the *suspect firm-years*. Suspect firm-years have net income scaled by total assets that is greater than or equal to zero but less than 0.005 (interval 16 in the figure). There are 503 suspect firm-years, including 450 unique firms.

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.

Ο 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 πραγματοποιείτε 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 IS=(L.sales-L2.sales)
by id: generate IDrevr=IS/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 ΤΟ δηλώσετε στο STATA. Αυτό γίνεται με τη σύνταξη της εντολής: `xtset ΔΗΛΩΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΠΕΡΙΠΤΩΣΗΣ ΔΗΛΩΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΧΡΟΝΟΥ`. Στην περίπτωση 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 abnormalproducr:
generate abnormalproducr=.
capture program drop abnormalproducr
program abnormalproducr
forvalues i = 1(1)13 {
capture reg producrr last revr Drevr lDrevr if industry==`i'
capture predict abnormalproducrr`i' if industry==`i', resid
replace abnormalproducrr= abnormalproducrr`i' if industry==`i'
capture drop abnormalproducrr`i'
}
end
abnormalproducrr

```

Βήμα 4^ο: Εξαγωγή περιγραφικών στατιστικών

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

- describe (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)

<i>statname</i>	Definition	<i>statname</i>	Definition
<u>mean</u>	mean	p1	1st percentile
<u>count</u>	count of nonmissing observations	p5	5th percentile
<u>n</u>	same as count	p10	10th percentile
<u>sum</u>	sum	p25	25th percentile
<u>max</u>	maximum	<u>median</u>	median (same as p50)
<u>min</u>	minimum	p50	50th percentile (same as median)
<u>range</u>	range = max – min	p75	75th percentile
<u>sd</u>	standard deviation	p90	90th percentile
<u>variance</u>	variance	p95	95th percentile
<u>cv</u>	coefficient of variation (sd/mean)	p99	99th percentile
<u>semean</u>	standard error of mean (sd/ $\sqrt{n}$)	iqr	interquartile range = p75 – p25
<u>skewness</u>	skewness	q	equivalent to specifying p25 p50 p75
<u>kurtosis</u>	kurtosis		

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 υπάρχει 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 5^ο: 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 ONOMA_APXEIOY.doc, replace sum(log)
```

Όπου στη θέση ONOMA_APXEIOY 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 ONOMA_APXEIOY.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 ONOMA_APXEIOY.doc, replace sum(log)
```