

Ασκήσεις Εργαστήριο Στατιστικής Διοίκηση Επιχειρήσεων

1. Στον ακόλουθο πίνακα δίνονται τα δεδομένα που αφορούν το ύψος (σε cm) σε ένα δείγμα 40 μαθητών.

170	180	178	165	170	168	175	175	173	162
160	170	167	177	180	170	182	178	165	178
156	175	172	173	167	187	170	180	178	191
176	169	167	166	179	178	180	164	170	173

Να γίνει ομαδοποίηση των δεδομένων σε 6 τάξεις ίσου πλάτους.

Λύση

Max=191, Min=156 οπότε **R=35, $\delta=R/k=5,833=6$**

Τάξεις
156-162
162-168
168-174
174-180
180-186
186-192

Στο SPSS κατά τα γνωστά.

Παρατήρηση: η ίδια άσκηση να χρησιμοποιηθεί και για περιγραφική στατιστική.

2. Στον ακόλουθο πίνακα δίνονται τα δεδομένα που προέκυψαν από την καταγραφή των ημερήσιων εξόδων (σε ευρώ) σε ένα δείγμα 50 φοιτητών του ΤΕΙ.

30	28	44	32	40	35	26	38	36	33
33	21	30	22	37	33	47	40	32	36
33	32	34	32	39	37	31	23	25	35
39	40	35	29	53	28	35	34	34	26
42	31	37	24	38	29	37	46	31	43

Να γίνει ομαδοποίηση των δεδομένων σε 7 τάξεις ίσου πλάτους.

Λύση

Max=53, Min=21 οπότε **R=32, $\delta=R/k=4,57=5$**

Τάξεις
21-26
26-31
31-36
36-41
41-46
46-51
51-56

Στο SPSS κατά τα γνωστά.

Παρατήρηση: η ίδια άσκηση να χρησιμοποιηθεί και για περιγραφική στατιστική.

3. Πίνακας Διπλής Εισόδου για το φύλο και την απασχόληση στο δείγμα των 30 φοιτητών

Απασχόληση	Άντρας	Γυναίκα
Αθλητισμός	4	2
Διάβασμα	2	2
Διασκέδαση	0	2
Μουσική	3	7
Τηλεόραση	1	2
Υπολογιστές	6	0

Λύση

Στο SPSS εργαζόμαστε ως εξής:

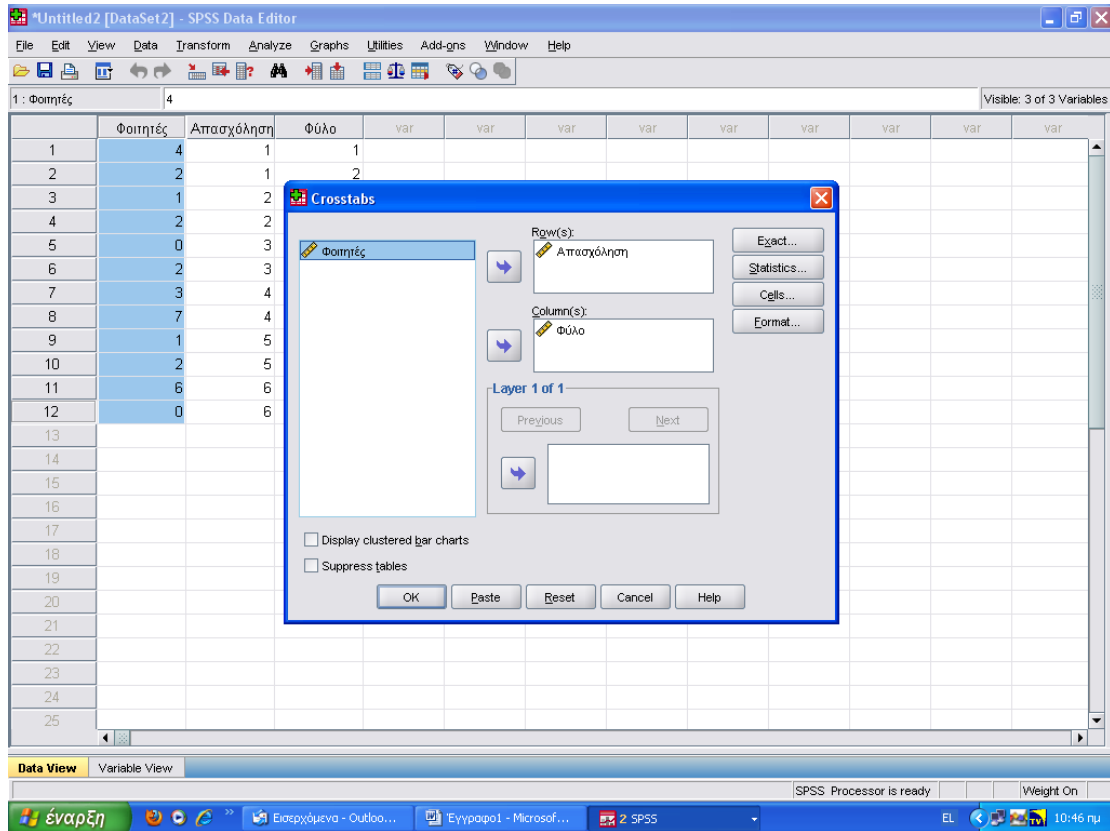
Δημιουργούμε 3 μεταβλητές:

- **Φοιτητές** (Numeric)
- **Απασχόληση** (Numeric, Value, 1-6 για τις διάφορες Απασχολήσεις)
- **Φύλο** (Numeric, Value, 1-2, για τα δυο φύλα)

Μετά περνάμε τα δεδομένα με προσοχή διαβάζοντας τα κατά γραμμή.

Εν συνεχεία, **Data** → **Weight Cases** by *Φοιτητές*

Analyze → **Descriptive Statistics** → **Crosstabs**



Crosstabs

Notes

Output Created		09-Απρ-2014 10:21:38
Comments		
Input	Active Dataset	DataSet2
	Filter	<none>
	Weight	Φοιτητές
	Split File	<none>
	N of Rows in Working Data File	10
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.

Syntax	CROSSTABS /TABLES=Απασχόληση BY Φύλο /FORMAT=AVALUE TABLES /CELLS=COUNT /COUNT ROUND CELL.		
Resources	Processor Time		00:00:00,000
	Elapsed Time		00:00:00,015
	Dimensions Requested		2
	Cells Available		174762

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Απασχόληση * Φύλο	30	100,0%	0	0,0%	30	100,0%

Απασχόληση * Φύλο Crosstabulation

Count		Φύλο		
		Άντρας	Γυναίκα	Total
		Απασχόληση	Αθλητισμός	4
	Διάβασμα	1	2	3
	Διασκέδαση	0	2	2
	Μουσική	3	7	10
	Τηλεόραση	1	2	3
	Υπολογιστές	6	0	6
Total		15	15	30

Παρατήρηση: η ίδια άσκηση να χρησιμοποιηθεί και για περιγραφική στατιστική.

4. Πίνακας Διπλής Εισόδου κατανομής των αγοριών και των κοριτσιών στις διάφορες βαθμίδες εκπαίδευσης

Βαθμίδες Εκπαίδευσης	Αγόρια	Κορίτσια
Προσχολική	55898	52459
Δημοτική	486354	450769
Μέση Γενική	277453	269563
Μέση Τεχνική	113966	17362
ΑΕΙ	59684	35701
Ανώτερη	1740	2668
ΤΕΙ	12094	4992

Λύση

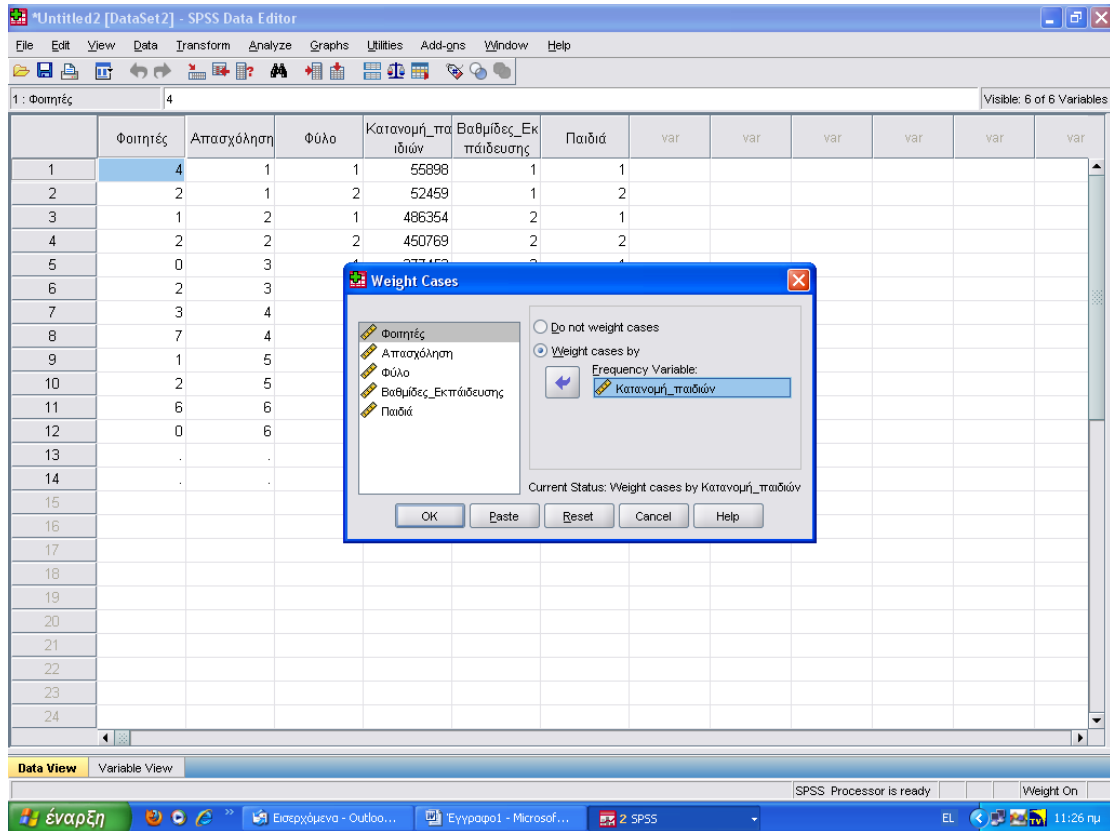
Στο SPSS εργαζόμαστε ως εξής:

Δημιουργούμε 3 μεταβλητές:

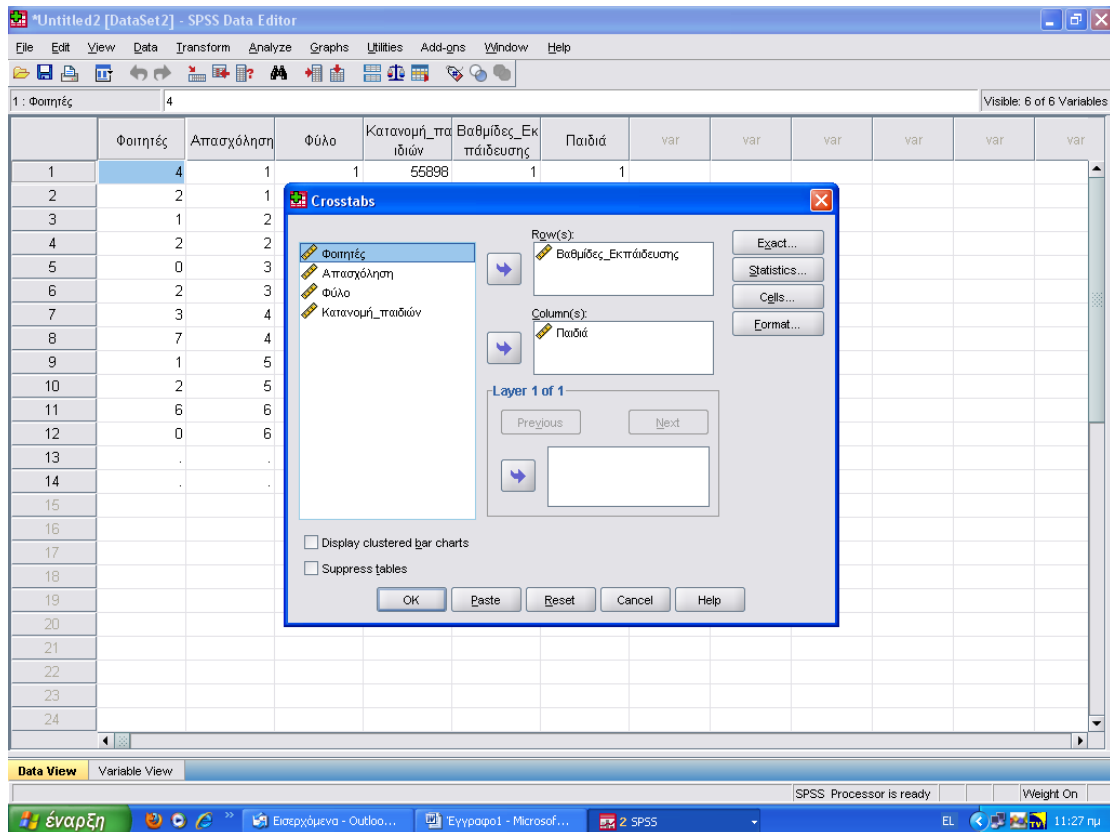
- **Κατανομή παιδιών** (Numeric)
- **Βαθμίδες Εκπαίδευσης** (Numeric, Value, 1-7 για τις διάφορες Βαθμίδες)
- **Φύλο ή Παιδιά**(Numeric, Value, 1-2, για τα δυο φύλα)

Μετά περνάμε τα δεδομένα με προσοχή διαβάζοντας τα κατά γραμμή.

Εν συνεχεία, **Data** → **Weight Cases** by *Κατανομή παιδιών*



Analyze→Descriptive Statistics→Crosstabs



Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Βαθμίδες_Εκπαίδευσης * Παιδιά	1840703	100,0%	0	,0%	1840703	100,0%

Βαθμίδες_Εκπαίδευσης * Παιδιά Crosstabulation

Count		Παιδιά		
		αγόρια	κορίτσια	Total
		Βαθμίδες_Εκπαίδευσης	Προσχολική	55898
	Δημοτική	486354	450769	937123
	Μέση Γενική	277453	269563	547016
	Μέση Τεχνική	113966	17362	131328
	ΑΕΙ	59684	35701	95385
	Ανώτερη	1740	2668	4408
	ΤΕΙ	12094	4992	17086
Total		1007189	833514	1840703

5. Ο ακόλουθος πίνακας καταγράφει τον αριθμό παιδιών σε ένα δείγμα 40 οικογενειών.

0	1	1	0	2	3	2	2	3	1
2	1	2	2	2	4	1	3	2	4
2	2	0	4	2	2	0	1	2	3
2	2	1	2	2	1	4	2	2	2

Να κατασκευαστεί ο πίνακας συχνοτήτων.

Πόσες οικογένειες έχουν το πολύ 2 παιδιά;

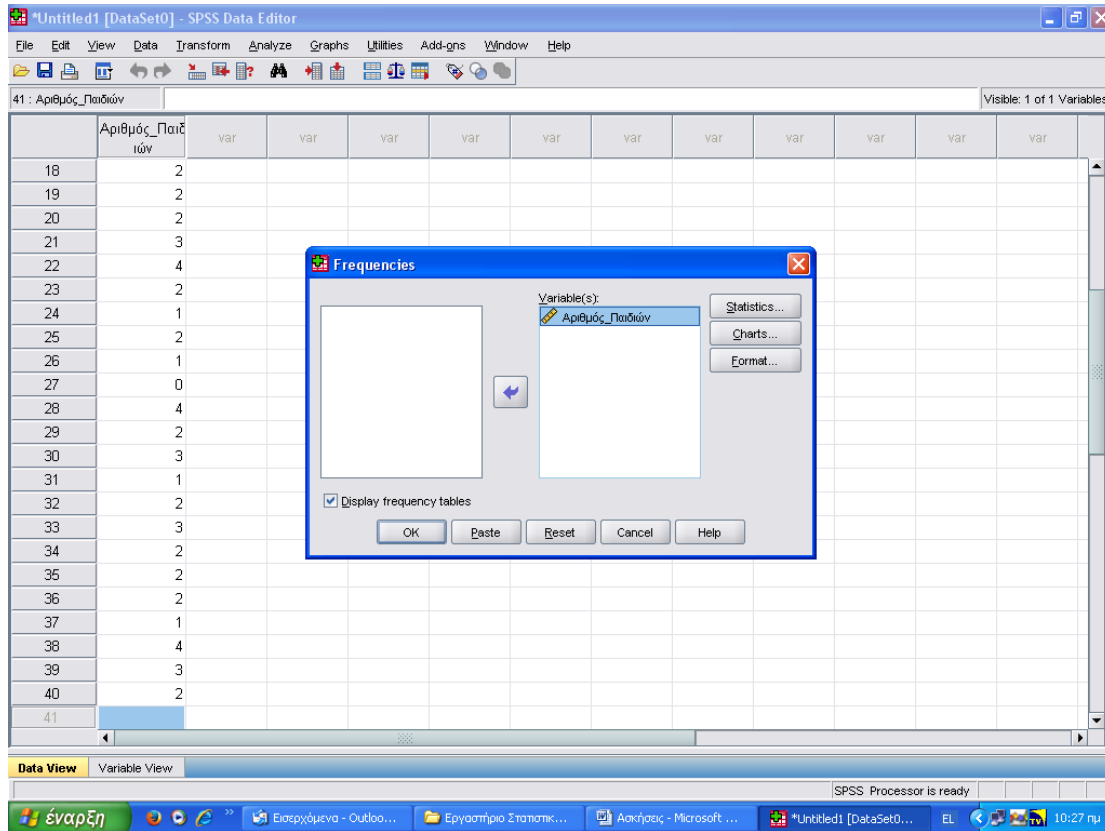
Ποιο είναι το ποσοστό των οικογενειών που έχουν το πολύ 1 παιδί;

Λύση:

Δημιουργούμε τη μεταβλητή «Αριθμός Παιδιών» κατά τα γνωστά. (Numeric, 0 δεκαδικά ψηφία) και εν συνεχεία εισάγουμε τα δεδομένα.

Για τον πίνακα συχνοτήτων εργαζόμαστε ως εξής:

Analyze→Descriptive Statistics→Frequencies



Σε αυτήν την επιλογή υπάρχει και η δυνατότητα υπολογισμού κάποιων στατιστικών μέτρων από την επιλογή Statistics και κατασκευή γραφικών παραστάσεων από την επιλογή Charts. (θα δούμε αυτές τις επιλογές παρακάτω αναλυτικότερα).

Frequencies

Statistics

Αριθμός_Παιδιών

N	Valid	40
	Missing	0

Αριθμός_Παιδιών

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	10,0	10,0	10,0
	1	8	20,0	20,0	30,0
	2	20	50,0	50,0	80,0
	3	4	10,0	10,0	90,0
	4	4	10,0	10,0	100,0
	Total	40	100,0	100,0	

Ο παραπάνω πίνακας παρέχει τις ακόλουθες πληροφορίες:

Frequency→Απλές Συχνότητες

Percent→Σχετικές Συχνότητες

Valid Percent→παρέχει τις ίδιες πληροφορίες με την προηγούμενη στήλη

Cumulative Percent→Σχετικές Αθροιστικές Συχνότητες.

Από τον παραπάνω πίνακα μπορούν να απαντηθούν τα δυο επόμενα ερωτήματα.

Υπάρχουν 32 οικογένειες που έχουν το πολύ 2 παιδιά. (4+8+20).

Για το επόμενο ερώτημα κοιτάμε την τελευταία στήλη, την στήλη των σχετικών αθροιστικών συχνοτήτων (Cumulative Percent). Άρα το ποσοστό των οικογενειών που έχουν το πολύ 1 παιδί είναι 30%.

Παρατήρηση: η ίδια άσκηση να χρησιμοποιηθεί και για τον υπολογισμό των στατιστικών μέτρων.

6. Ο παρακάτω πίνακας περιέχει το ύψος 40 μαθητών.

170	180	178	165	170	168	175	175	173	162
160	170	167	177	180	170	182	178	165	178
156	175	172	173	167	187	170	180	178	191
176	169	167	166	179	178	180	164	170	173

Να γίνει ομαδοποίηση των υψών σε 6 τάξεις και να κατασκευαστεί ο πίνακας συχνοτήτων. (ίδια με παραπάνω, άσκηση 1).

Λύση:

Ακολουθώντας τη γνωστή μεθοδολογία για τη κατασκευή των τάξεων, έχουμε τις εξής τάξεις:

[156,162)

[162, 168)

[168,174)

[174, 180)

[180, 186)

[186,192)

Εν συνεχεία στο SPSS ακολουθούμε ως γνωστό τη διαδικασία: Transform→Recode into different variable, για να ομαδοποιήσουμε τα δεδομένα μας.

Για τη κατασκευή του πίνακα συχνοτήτων ακολουθούμε τη γνωστή διαδικασία **Analyze→Descriptive**

Statistics→Frequencies, εισάγοντας τη μεταβλητή «τάξεις του ύψους»

		ύψους			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	3	7,5	7,5	7,5
	2,00	8	20,0	20,0	27,5
	3,00	11	27,5	27,5	55,0
	4,00	15	37,5	37,5	92,5
	5,00	1	2,5	2,5	95,0
	6,00	2	5,0	5,0	100,0
	Total	40	100,0	100,0	

Παρατήρηση: η ίδια άσκηση να χρησιμοποιηθεί και για τον υπολογισμό των στατιστικών μέτρων και δημιουργία γραφικών παραστάσεων.

7. Ο αριθμός παιδιών σε 50 οικογένειες δίνεται στον παρακάτω πίνακα:

0	2	1	2	0	5	0	2	1	1
1	2	0	3	1	2	3	0	1	2
2	1	0	2	4	2	1	2	3	2
1	0	2	4	2	2	1	0	5	1
2	1	2	2	2	3	3	1	4	2

Να κατασκευαστεί ο πίνακας συχνοτήτων.

8. Ο ακόλουθος πίνακας περιλαμβάνει τους χρόνους σε sec, που απαιτήθηκαν για την εκτέλεση ενός προγράμματος σε 40 διαφορετικούς υπολογιστές.

24	36	36	36	42	43	51	53	64	64
72	76	76	79	87	87	90	90	94	100
102	105	108	112	115	120	120	120	120	130
136	138	141	144	144	148	148	153	159	169

Να γίνει ομαδοποίηση των δεδομένων σε 5 τάξεις ίσου πλάτους και να κατασκευαστεί ο πίνακας συχνοτήτων.

9. ακόλουθος πίνακας καταγράφει τον αριθμό παιδιών σε ένα δείγμα 40 οικογενειών.

0	1	1	0	2	3	2	2	3	1
2	1	2	2	2	4	1	3	2	4
2	2	0	4	2	2	0	1	2	3
2	2	1	2	2	1	4	2	2	2

Να υπολογιστούν τα βασικότερα αριθμητικά μέτρα και η κατάλληλη γραφική παράσταση. (ίδια με 5 άσκηση)

Λύση:

Analyze→Descriptive Statistics→Frequencies

Στην επιλογή **Statistics** έχουμε τη δυνατότητα να υπολογίσουμε τεταρτημόρια, ποσοστημόρια, κάποια μέτρα κεντρικής τάσεως, κάποια μέτρα μεταβλητότητας και κάποια μέτρα που αφορούν την κατανομή των δεδομένων.

Quartiles→Τεταρτημόρια

Percentiles→Ποσοστημόρια

Mean→Μέση Τιμή

Median→Διάμεσος

Mode→Κορυφή

Sum→Άθροισμα των παρατηρήσεων

Std. Deviation→Τυπική Απόκλιση

Variance→Διασπορά

Range→Εύρος (Εκταση)

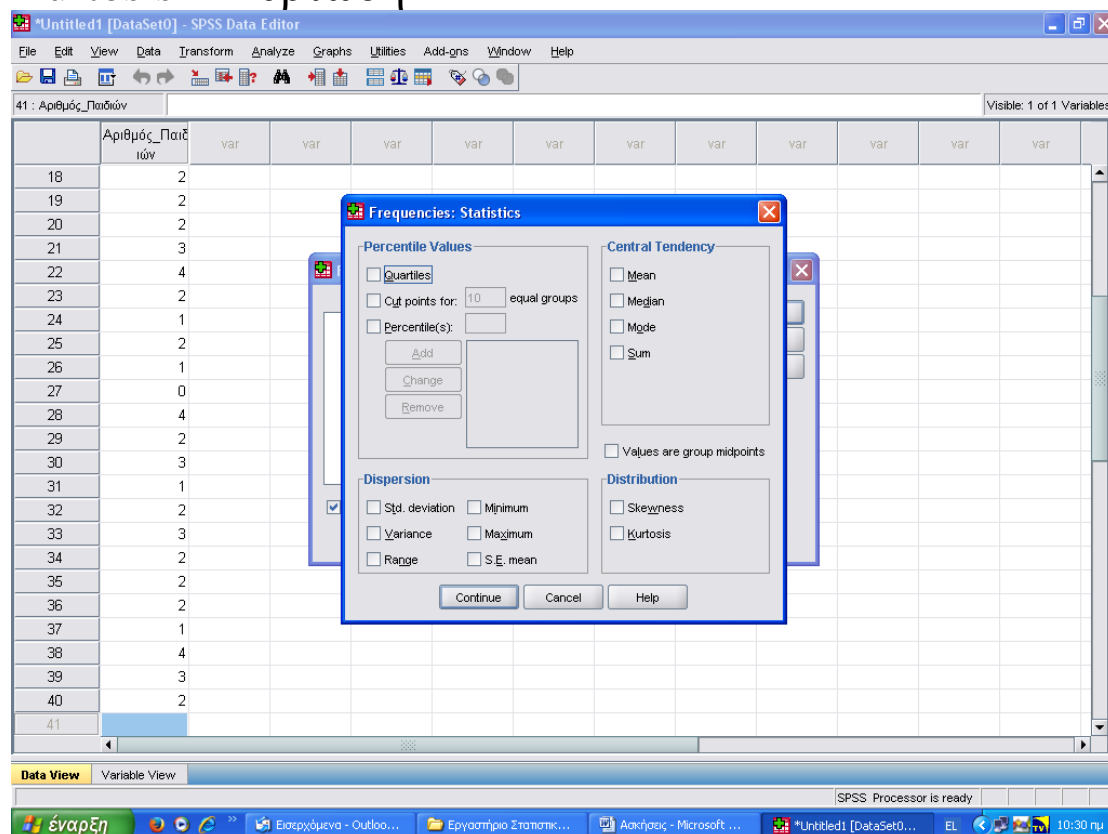
Maximum→Μέγιστη Τιμή

Minimum→Ελάχιστη τιμή

S.E. of Mean→Τυπικό Σφάλμα

Skewness→Ασυμμετρία

Kurtosis→Κύρτωση

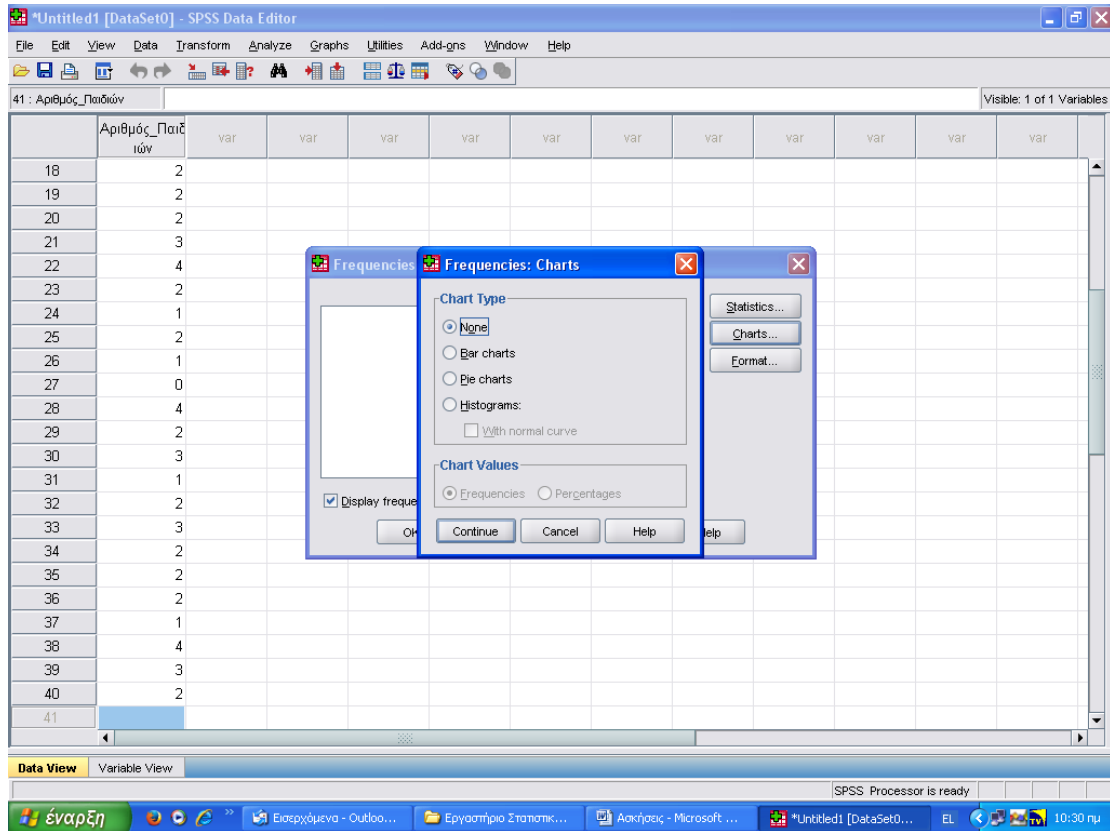


Στην επιλογή **Gcharts** μπορούμε να κατασκευάσουμε γραφικές παραστάσεις απλών και σχετικών συχνοτήτων.

Bar Charts→Ραβδογράμματα ή Ακιδωτά διαγράμματα

Histogramms→Ιστόγραμμα

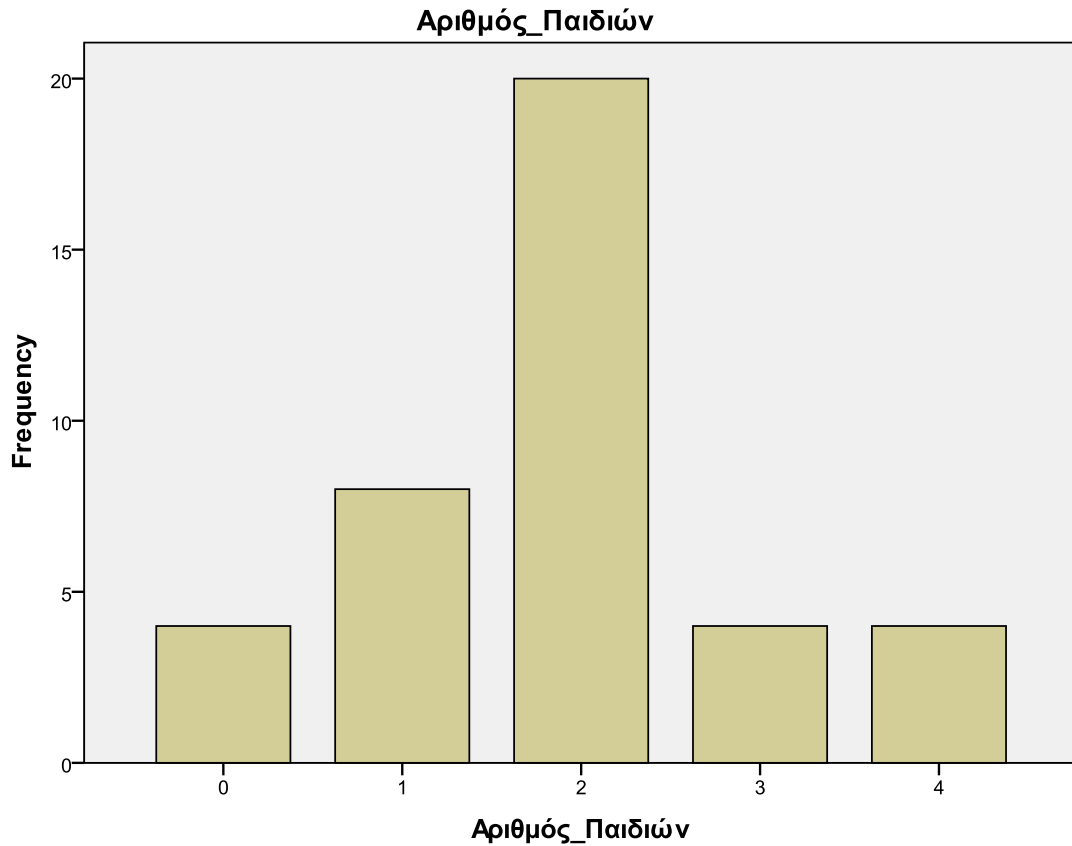
Pie Charts→Κυκλικό διάγραμμα



Statistics

Αριθμός_Παιδιών

N	Valid	40
	Missing	0
Mean		1,90
Median		2,00
Mode		2
Variance		1,118
Range		4
Minimum		0
Maximum		4



9. Με τα δεδομένα της άσκησης 5 να υπολογιστούν ο αριθμητικός μέσος, η διάμεσος, η κορυφή, η τυπική απόκλιση, η διασπορά , μέγιστη, ελάχιστη τιμή, εύρος και τα τεταρτημόρια.

Λύση:

Analyze→Descriptive Statistics→Frequencies

Statistics

Αριθμός_Παιδιών

N	Valid	40
	Missing	0
Mean		1,90
Median		2,00
Mode		2
Std. Deviation		1,057

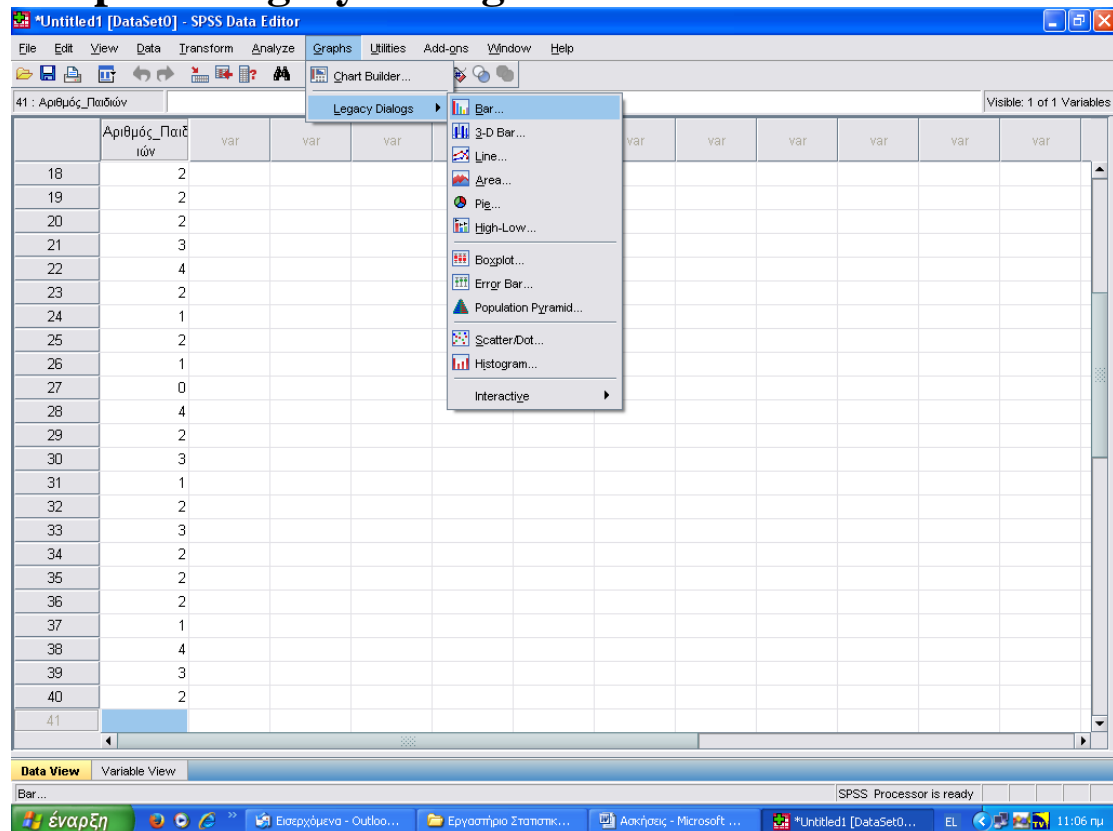
Variance		1,118
Range		4
Minimum		0
Maximum		4
Percentiles	25	1,00
	50	2,00
	75	2,00

10. Στην παραπάνω άσκηση να κατασκευαστεί το ραβδόγραμμα.

Λύση:

Η κατασκευή του ραβδογράμματος μπορεί να γίνει και από το Frequencies αλλά και από το Graphs.

Graphs→Legacy Dialogs



*Untitled1 [DataSet0] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

41 : Αριθμός_Παιδιών Visible: 1 of 1 Variables

	Αριθμός_Παιδιών	var	var	var	var	var	var	var	var	var	var	var
18	2											
19	2											
20	2											
21	3											
22	4											
23	2											
24	1											
25	2											
26	1											
27	0											
28	4											
29	2											
30	3											
31	1											
32	2											
33	3											
34	2											
35	2											
36	2											
37	1											
38	4											
39	3											
40	2											
41												

Bar Charts

Simple
Clustered
Stacked

Data in Chart Are

Summaries for groups of cases
 Summaries of separate variables
 Values of individual cases

Define Cancel Help

Data View Variable View

SPSS Processor is ready

έναρξη Εισαγόμενα - Outloo... Εργαστήριο Στατιστικ... Ασκησης - Microsoft ... *Untitled1 [DataSet0... EL 11:07 πμ

11. Μια εταιρεία κατασκευάζει μηχανικές αντλίες και τις προμηθεύει σε πελάτες. Μια απαίτηση του πελάτη είναι η απόκλιση να είναι το πολύ 0,002 ίντσες από τη μέση διάμετρο. Επιλέχτηκε ένα τ.δ 20 αντλιών και καταγράφηκε η διάμετρος σε ίντσες

4,01	4,00	4,02	4,02	4,03	4,00	3,98	3,99	3,99	4,01
3,99	3,98	3,97	4,00	4,02	4,01	4,02	4,00	4,01	3,99

Να υπολογιστεί μέση διάμετρος, η διάμεσος, η κορυφή και μέσω κατάλληλου διαγράμματος να σχολιαστεί το είδος της κατανομής και να χαρακτηριστεί η μεταβλητότητα των δεδομένων.

12. Παρακάτω δίνονται οι ηλικίες 50 μελών ενός κοινωνικού προγράμματος

83	51	66	61	85	65	54	56	92	60
65	87	68	64	51	70	75	66	74	68
44	55	78	69	98	67	82	77	79	62
38	88	76	99	84	47	60	42	66	74
91	71	83	80	68	65	51	56	73	55

Να υπολογιστεί η μέση ηλικία, ο ξακρισμένος μέσος, τα τεταρτημόρια και ο συντελεστής μεταβλητότητας.

Ασκήσεις Πιθανότητας

Στο SPSS οι πιθανότητες υπολογίζονται από τους πίνακες συνάφειας, και αν πρόκειται για πιθανότητες κατανομών από την επιλογή:

Transform → Compute

Στην επιλογή **Function** υπάρχει **PDF & Noncentral PDF** που υπολογίζει τη **συνάρτηση πιθανότητας** διάφορων κατανομών (εμείς θα ασχοληθούμε με BINOMIAL, POISSON, NORMAL) και υπάρχει **CDF & Noncentral CDF** που υπολογίζει τη **συνάρτηση κατανομής** διάφορων κατανομών (εμείς θα ασχοληθούμε με BINOMIAL, POISSON, NORMAL).

1. Αν για μια διωνυμική τ.μ. είναι $n = 10, p = 0.3$ να υπολογιστούν οι πιθανότητες $P(X = 3), P(X = 5), P(X = 8)$

Λύση

The screenshot shows the SPSS Data Editor interface. A 'Compute Variable' dialog box is open, allowing the user to define a new variable. The 'Target Variable' is set to 'p1'. The 'Numeric Expression' is 'PDF.BINOM(3,10,0.3)'. The 'Function group' is 'PDF & Noncentral PDF'. The 'Functions and Special Variables' list includes 'Pdf Binom'. The data grid shows variables 'Μετοχές', 'αύξηση_τιμής', 'Μερίσματα', and 'pdfn'. The status bar at the bottom indicates 'SPSS Processor is ready' and 'Weight On'.

PDF(ΒΑΖΟΥΜΕ ΤΟΝ ΑΡΙΘΜΟ ΠΟΥ ΘΕΛΟΥΜΕ ΝΑ ΥΠΟΛΟΓΙΣΟΥΜΕ ΤΗΝ ΠΙΘΑΝΟΤΗΤΑ, ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΤΗΣ ΚΑΤΑΝΟΜΗΣ) ‘

Εδώ συγκεκριμένα PDF (3, 10, 0.3) και μετά πατάμε OK και το αποτέλεσμα βγαίνει στο Data Editor και ισούται με 0.2668

Όμοια και τα υπόλοιπα

$$\text{PDF}(5, 10, 0.3)=0.1029$$

$$\text{PDF}(8, 10, 0.3)=0.0014$$

2. Αν για μια διωνυμική τ.μ. είναι $n = 25, p = 0,7$ να υπολογιστούν οι πιθανότητες $P(X = 18), P(X = 15), P(X \leq 20), P(X \geq 16)$.

Λύση:

Για τις δυο πρώτες όπως παραπάνω.

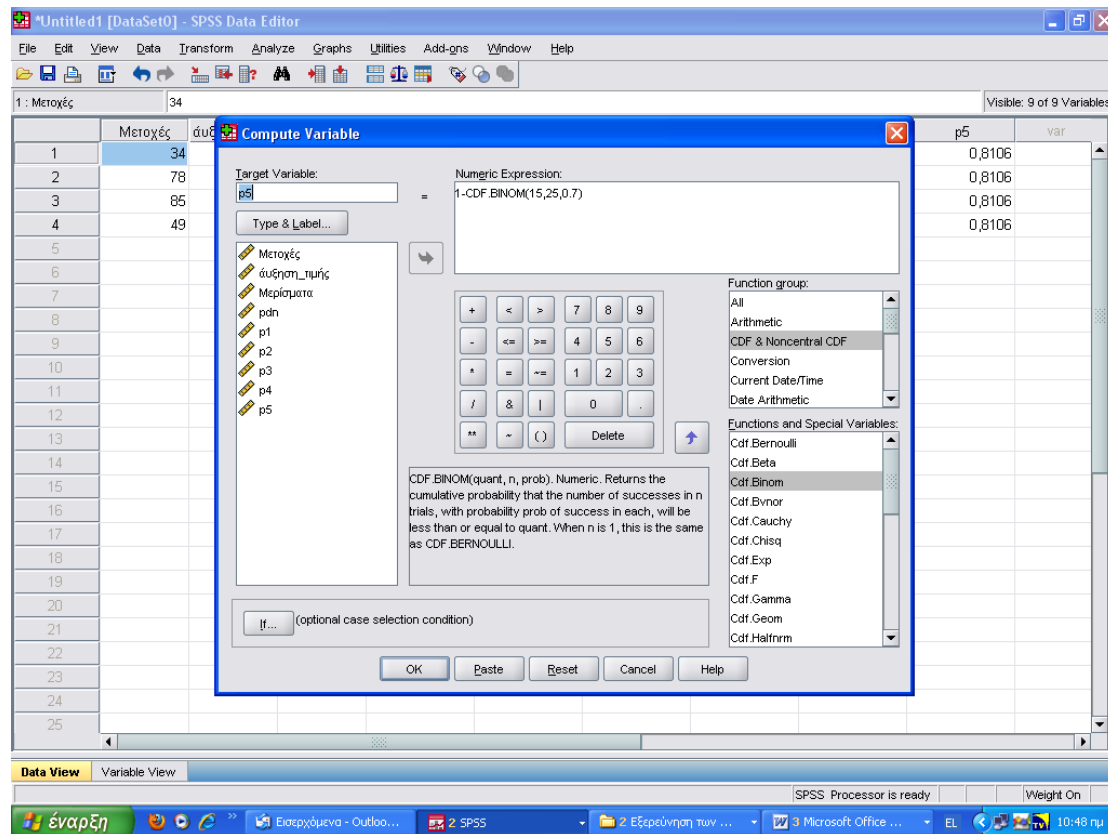
Για την επόμενη:

The screenshot shows the SPSS Data Editor interface. A 'Compute Variable' dialog box is open, allowing the user to define a new variable. The 'Target Variable' is set to 'p3' and the 'Numeric Expression' is 'CDF.BINOM(20,25,0.7)'. The 'Function group' is 'CDF & Noncentral CDF'. The dialog box also shows a list of functions and special variables on the right, including Cdf.Bernoulli, Cdf.Beta, Cdf.Binom, Cdf.Bvnr, Cdf.Cauchy, Cdf.Chisq, Cdf.Exp, Cdf.F, Cdf.Gamma, Cdf.Geom, and Cdf.Halfnm. The data grid shows variables 'Μετοχές' and 'αύξηση_τιμής' with values 34, 78, 85, 49 for the first four rows.

$$P(X \leq 20)=0.9095$$

Για την επόμενη: $P(X \geq 16) = 1 - P(X < 16) = 1 - P(X \leq 15)$. Άρα θα υπολογίσουμε την $P(X \leq 15)$.

Από το SPSS θα υπολογίσουμε τη $P(X \leq 15)$



Και το αποτέλεσμα είναι ίσο με 0,8106

Παρατήρηση

Παρόμοια διαδικασία ακολουθείται και για τις άλλες κατανομές, απλά επιλέγουμε τις σωστές παραμέτρους ανάλογα με το είδος της κατανομής κάθε φορά.

3. Ένας φοιτητής που σύντομα ολοκληρώνει τις σπουδές του στην ειδικότητα της λογιστικής, υπολογίζει με βάση τους βαθμούς του και την πρακτική άσκηση που έχει κάνει, ότι έχει πιθανότητα 70% να δεχθεί μια προσφορά εργασίας από κάθε εταιρεία που θα στείλει το βιογραφικό του. Αν στείλει μόνο σε 4 εταιρείες το βιογραφικό του, ποια η πιθανότητα να μην δεχθεί καμία προσφορά εργασίας.

Λύση

X = να δεχθεί προσφορά εργασίας, $X \sim B(4, 0.70)$ και ζητείται η πιθανότητα $P(X = 0)$

Ακολουθούμε τη διαδικασία που περιγράφηκε παραπάνω και έχουμε: 0,0081

Οι μεγάλες εταιρείες λογισμικού διαθέτουν τηλεφωνική γραμμή υποστήριξης, που οι πελάτες μπορούν να καλούν όταν χρειάζονται βοήθεια ή έχουν προβλήματα με το λογισμικό. Λόγω του μεγάλου αριθμού των τηλεφωνημάτων, συχνά οι πελάτες που καλούν χρειάζεται να μπου σε αναμονή μέχρι να είναι διαθέσιμος ένας υπάλληλος για να τους απαντήσει. Αν για μια εταιρεία το ποσοστό των κλήσεων που μπαίνουν σε αναμονή είναι 20% και αν υποθέσουμε ότι 100 πελάτες καλούν την τηλεφωνική υποστήριξη της εταιρείας αυτής, ποια η πιθανότητα να μπου σε αναμονή περισσότεροι από 25 πελάτες;

Λύση

X = ο αριθμός των πελατών που μπαίνουν σε αναμονή, $X \sim B(100, 0.20)$ και ζητείται η πιθανότητα $P(X \geq 25)$. Δηλ.
 $P(X \geq 25) = 1 - P(X < 25) = 1 - P(X \leq 24)$

Ακολουθούμε τη διαδικασία που περιγράφηκε παραπάνω και έχουμε τα ακόλουθα αποτελέσματα:

$$P(X \leq 24) = 0,868647$$

Άρα η ζητούμενη πιθανότητα είναι $1 - 0,868647 = 0,131353$

Κάποιος παρατήρησε ότι το πλήθος των τυπογραφικών λαθών σε νέα βιβλία διαφέρει σημαντικά από βιβλίο σε βιβλίο. Από μια έρευνα διαπιστώθηκε ότι το πλήθος των τυπογραφικών λαθών είναι μια κατανομή Poisson με αναμενόμενη τιμή 1,5 λάθος ανά 100 σελίδες. Αν επιλεγούν τυχαία 100 σελίδες ενός νέου βιβλίου, ποια η πιθανότητα να μην βρεθεί κανένα τυπογραφικό λάθος; Ποια η πιθανότητα να υπάρχουν πέντε ή λιγότερα τυπογραφικά λάθη;

Λύση

X = ο αριθμός των τυπογραφικών λαθών σε 100 σελίδες τότε $X \sim P(\lambda)$ με $\lambda = 1,5$ και ζητείται να υπολογιστούν οι πιθανότητες $P(X = 0) = 0,2231, P(X \leq 5) = 0,9955$.

4. Ο αριθμός των χρηστών μιας αυτόματης ταμειακής μηχανής σε μια τράπεζα είναι μια κατανομή Poisson με 1,5 χρήστες ανά 5 λεπτά. Να υπολογιστούν οι ακόλουθες πιθανότητες:

1. Κανένας χρήστης στα επόμενα 5 λεπτά
2. Πέντε ή λιγότεροι χρήστες στα επόμενα 15 λεπτά
3. Τρεις ή περισσότεροι χρήστες στα επόμενα 10 λεπτά

Λύση

X αριθμός χρηστών μιας αυτόματης ταμειακής μηχανής

$X \sim P(\lambda)$ με $\lambda = 1,5$ ανά 5 λεπτά

1. Ζητάμε $P(X = 0)$

Ακολουθούμε τη γνωστή διαδικασία επιλέγοντας για μέση τιμή 1,5 και για τιμή $x=0$.
Οπότε έχουμε:

$$P(X = 0) = 0,223130$$

2. Ζητάμε $P(X \leq 5)$ αλλά η μέση τιμή θα έχει αλλάξει γιατί ζητάμε στα επόμενα 15 λεπτά άρα η παράμετρος της κατανομής θα είναι τώρα 4,5

Cumulative Distribution Function

Poisson with mean = 4,5

$$P(X \leq 5) = 0,702930$$

3. Ζητάμε $P(X \geq 3)$ αλλά η μέση τιμή θα έχει αλλάξει γιατί ζητάμε στα επόμενα 10 λεπτά άρα η παράμετρος της κατανομής θα είναι τώρα 3.
 $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2)$

Οπότε:

$$P(X \leq 2) = 0,423190$$

$$\text{Άρα } 1 - 0,423190 = 0,57681$$

5. Έστω μια επένδυση που οι αποδόσεις έχουν κανονική κατανομή με μέσο 10% τυπική απόκλιση 5%.

1. Να υπολογιστεί η πιθανότητα αρνητικής απόδοσης (ζημίας)
2. Να υπολογιστεί η πιθανότητα αρνητικής απόδοσης (ζημίας) αν η τυπική απόκλιση είναι 10%

Λύση

Μια επένδυση παρουσιάζει ζημία αν η απόδοση είναι αρνητική και η πιθανότητα αρνητικής τιμής είναι $P(X < 0)$

$$P(X \leq 0) = 0,0227501$$

Αν αλλάξει η τιμή της τυπικής απόκλισης έχουμε:

$$P(X \leq 0) = 0,158655$$

Παρατήρηση

Παρατηρούμε ότι μια αύξηση της τυπικής απόκλισης μας οδηγεί σε αύξηση της πιθανότητας ζημίας.

Η αύξηση της τυπικής απόκλισης αύξησε το πλάτος της κατανομής συμμετρικά και προς τις δυο κατευθύνσεις, που σημαίνει ότι αυξήθηκε και η πιθανότητα να δώσει η επένδυση μεγαλύτερα κέρδη.

Επειδή όμως οι επενδυτές τείνουν να ανησυχούν κυρίως για τον επενδυτικό κίνδυνο και αν μη δίνουν σημασία ιδιαίτερη στις πιθανότητες ψηλότερων κερδών, η τυπική απόκλιση χρησιμοποιείται ως δείκτης επενδυτικού κινδύνου.

Γενικές ασκήσεις στις Κατανομές

Άσκηση

Σε μια δημοσκόπηση που πραγματοποιήθηκε το 52% θεωρεί ότι το πιο σημαντικό θέμα είναι «η προστασία του περιβάλλοντος», το 36% «η εξασφάλιση των ενεργειακών αποθεμάτων», το 6% ότι και τα δυο θέματα είναι εξίσου σημαντικά, και το υπόλοιπο 6% δεν έχει γνώμη. Αν επιλέξουμε τυχαία 100 άτομα να υπολογιστούν οι πιθανότητες:

1. 50 τουλάχιστον να πιστεύουν ότι το σημαντικότερο θέμα είναι η προστασία του περιβάλλοντος
2. 30 ή λιγότεροι να πιστεύουν ότι το σημαντικότερο θέμα είναι η εξασφάλιση των ενεργειακών αποθεμάτων
3. 5 ή λιγότεροι να μην έχουν γνώμη

Λύση: Διωνυμική

Άσκηση

Ο αριθμός των χρηστών μιας αυτόματης ταμειακής μηχανής σε μια τράπεζα είναι κατά μέσο όρο 1,5 χρήστες ανά 5 λεπτά. Να υπολογιστούν οι πιθανότητες:

1. Κανένας χρήστης στα επόμενα 5 λεπτά
2. Πέντε ή λιγότεροι χρήστες στα επόμενα 15 λεπτά
3. Τρεις ή περισσότεροι χρήστες στα επόμενα 10 λεπτά

Λύση: Poisson

Άσκηση

Σύμφωνα με μια έρευνα το 45% των γυναικών που εργάζονται έχουν πλήρες ωράριο. Αν επιλεγούν τυχαία 50 εργαζόμενες γυναίκες. Να βρεθεί η πιθανότητα τουλάχιστον 19 από αυτές να έχουν πλήρες ωράριο.

Λύση: Διωνυμική

Άσκηση

Σύμφωνα με μια έρευνα το 45% των γυναικών που εργάζονται έχουν πλήρες ωράριο. Αν επιλεγούν τυχαία 50 εργαζόμενες γυναίκες. Να βρεθεί η πιθανότητα τουλάχιστον 19 από αυτές να **μην** έχουν πλήρες ωράριο.

Λύση: Διωνυμική

Άσκηση

Οι επισκέψεις σε μια ιστοσελίδα συμβαίνουν τυχαία και ανεξάρτητα με μέση συχνότητα 4 ανά εβδομάδα.

1. Να βρεθεί η πιθανότητα να δεχθεί η ιστοσελίδα στη διάρκεια μιας εβδομάδας 10 ή περισσότερες επισκέψεις
2. Να βρεθεί η πιθανότητα να δεχθεί η ιστοσελίδα στη διάρκεια δυο εβδομάδων 20 ή περισσότερες επισκέψεις.

Λύση: Poisson

Άσκηση

Η διάρκεια των υπεραστικών τηλεφωνημάτων μιας επιχείρησης είναι μια τ.μ. που ακολουθεί την κανονική κατανομή με μέσο 6,3 λεπτά και τυπική απόκλιση 2,2 λεπτά. Ποια η πιθανότητα ένα υπεραστικό τηλεφώνημα να έχει διάρκεια:

1. Από 5 έως 10 λεπτά
2. Περισσότερο από 7 λεπτά
3. Λιγότερο από 4 λεπτά.

Συσχέτιση-Παλινδρόμηση

1. Στον ακόλουθο πίνακα δίνονται οι βαθμοί 20 σπουδαστών σε μια πρόοδο για το μάθημα της Στατιστικής και ο βαθμός των τελικών εξετάσεων. Να κατασκευαστεί το διάγραμμα διασποράς. Τι διαπιστώνεται από το διάγραμμα για την σχέση των δυο μεταβλητών; Να υπολογιστεί και να ερμηνευτεί ο συντελεστής συσχέτισης. Να προσαρμοστεί η ευθεία παλινδρόμησης.

Βαθμός Προόδου	Βαθμός εξετάσεων
65	85
50	77
75	90
70	84
68	61
58	70
49	76
92	78
68	80
93	92
38	56
92	95
59	60
66	87
61	83
71	77
68	79
82	74
85	94
67	80

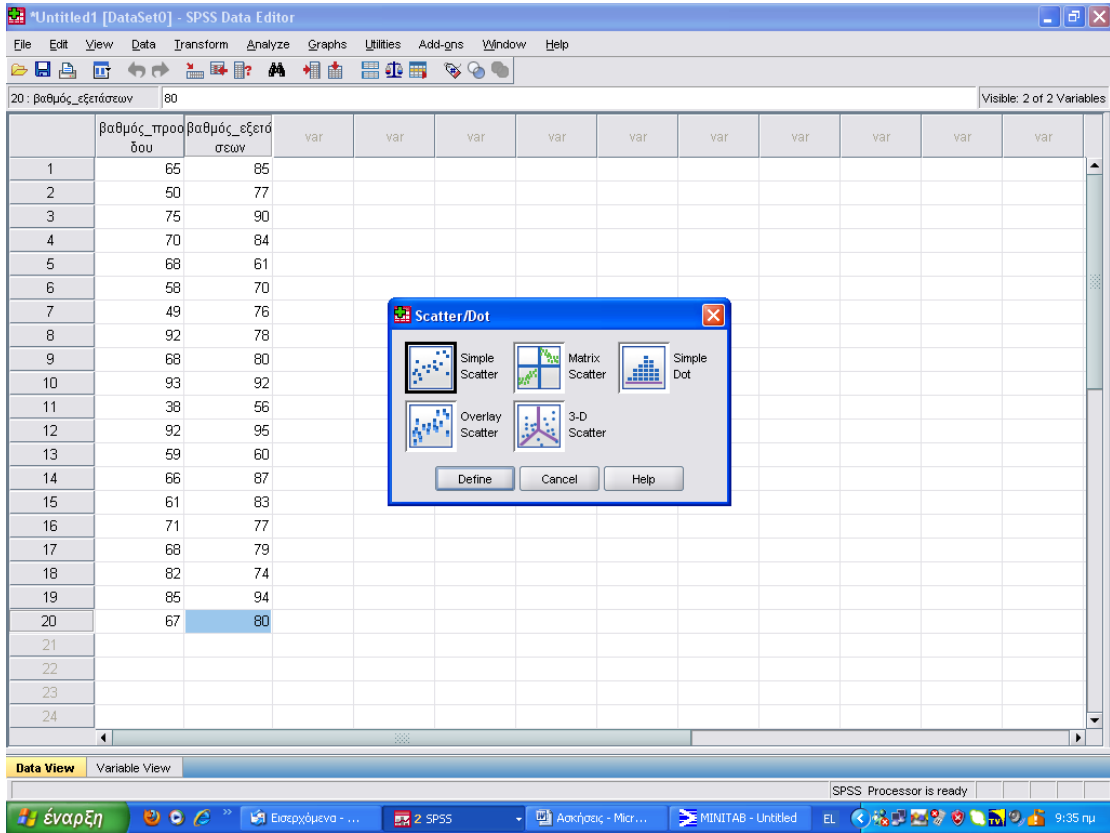
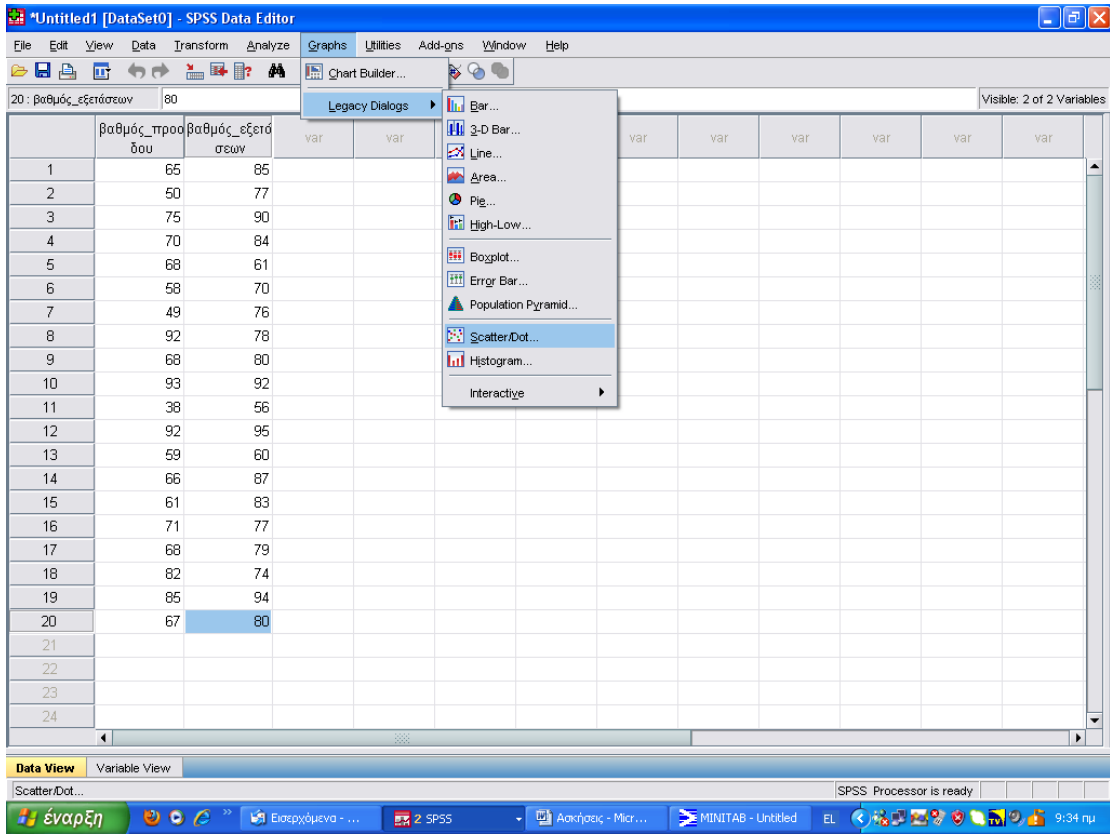
Λύση

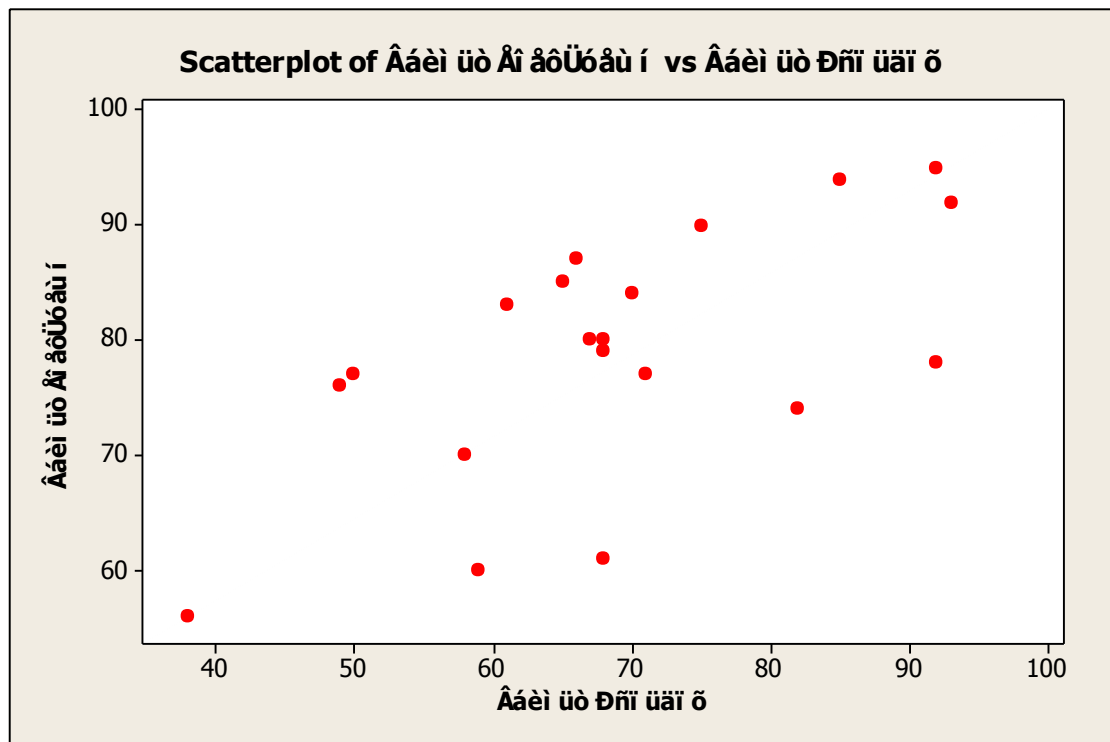
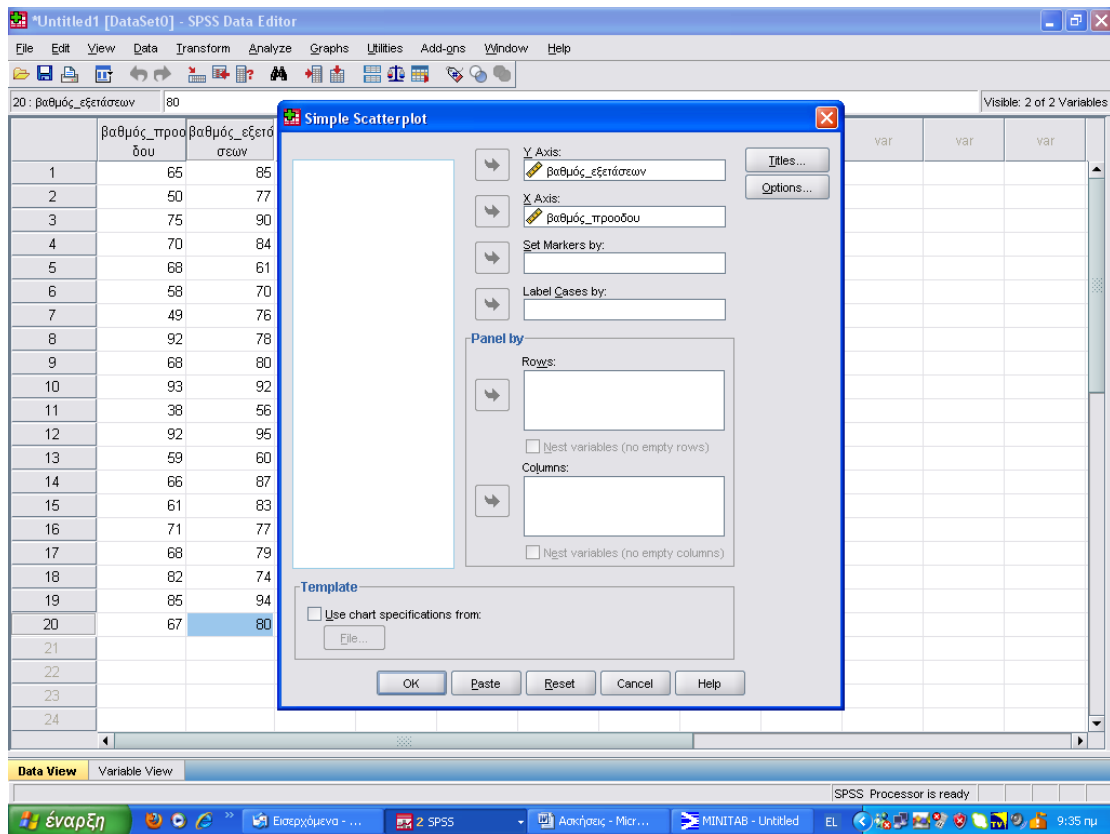
Για το διάγραμμα διασποράς έχουμε:

Graphs→**Legacy Dialogs**→**Scatter**

Επιλέγουμε Simple→**Define**→

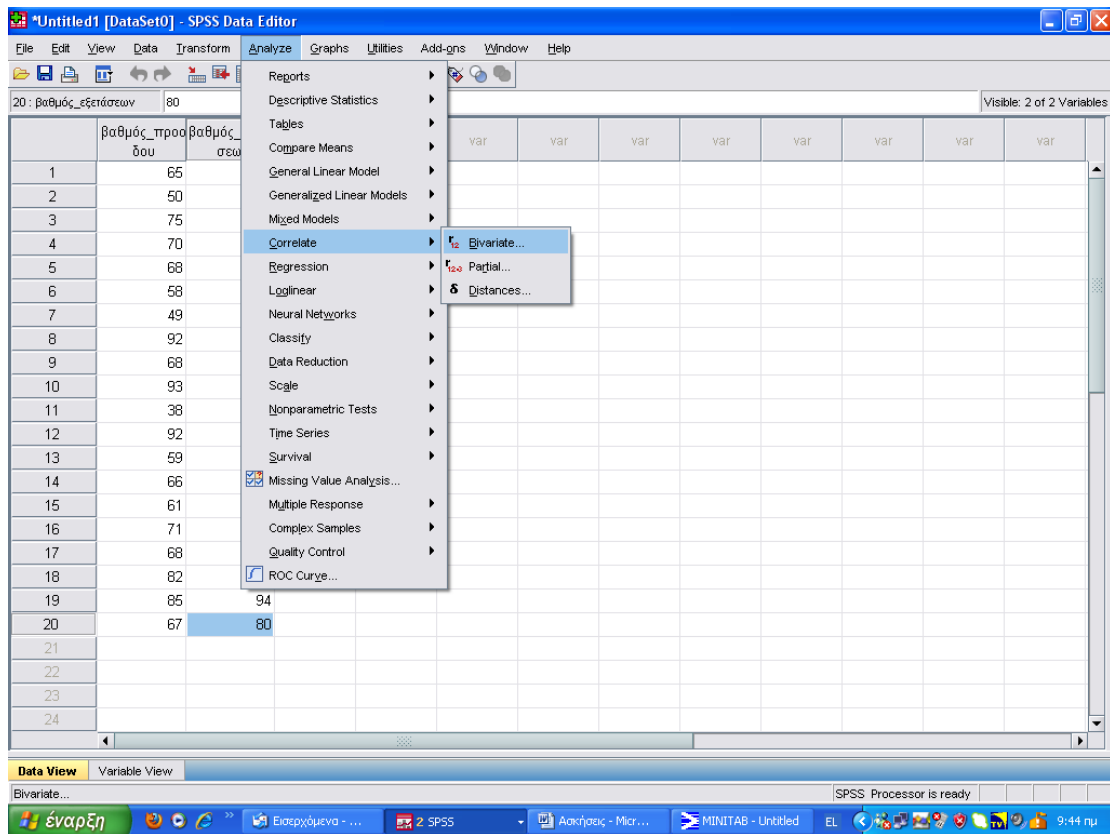
Στον άξονα Y βάζουμε την εξαρτημένη μεταβλητή και στον άξονα X την ανεξάρτητη και OK.



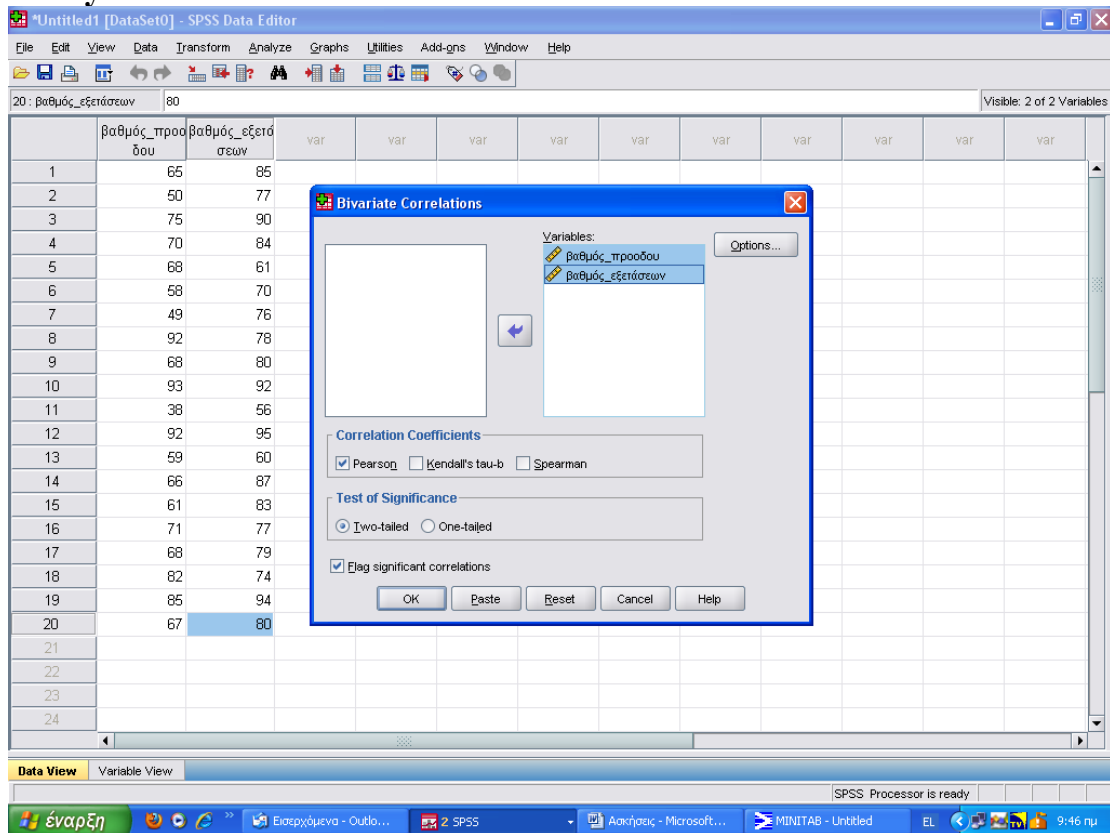


Από το διάγραμμα διαπιστώνουμε ότι υπάρχει μια ανοδική πορεία στο νέφος σημείων. Οι δυο μεταβλητές σχετίζονται θετικά και μια πιθανή σχέση μεταξύ τους είναι η γραμμική, καθώς το νέφος «ακολουθεί» ευθεία γραμμή.

Υπολογισμός του γραμμικού συντελεστή συχέτισης.



Analyze → Correlate → Bivariate



Εισάγουμε και τις δυο μεταβλητές και OK.

Correlations

		βαθμός_προοδο	βαθμός_εξετάσε
		U	ωV
βαθμός_προοδου	Pearson Correlation	1	,633**
	Sig. (2-tailed)		,003
	N	20	20
βαθμός_εξετάσεων	Pearson Correlation	,633**	1
	Sig. (2-tailed)	,003	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

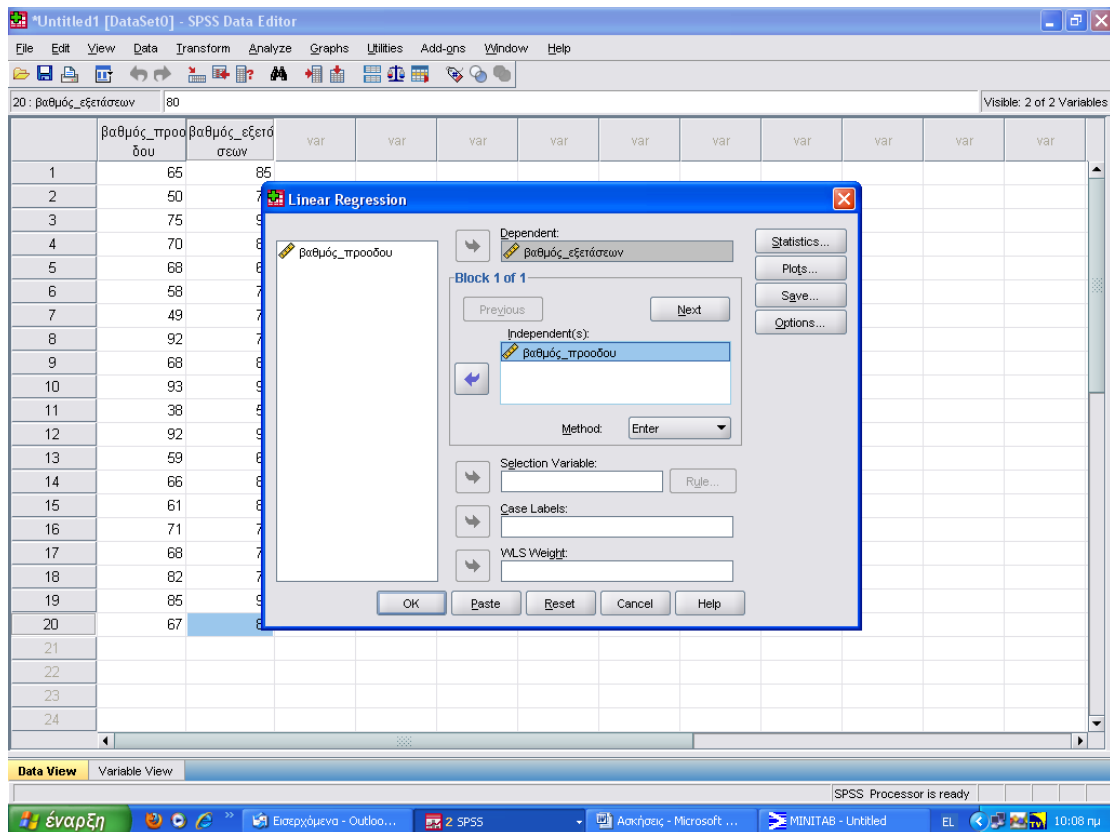
Ο γραμμικός συντελεστής συσχέτισης είναι ίσος με 0,633 που δηλώνει ότι υπάρχει μέτρια θετική σχέση μεταξύ του βαθμού της τελικής εξέτασης και του βαθμού της προόδου.

Τώρα για την προσαρμογή της ευθείας:

Analyze→Regression→Linear

The screenshot shows the SPSS Data Editor interface. The 'Analyze' menu is open, and the path 'Analyze -> Regression -> Linear' is highlighted. The data table below shows the following values:

	βαθμός_προοδου	βαθμός_εξετάσεων
1	65	
2	50	
3	75	
4	70	
5	68	
6	58	
7	49	
8	92	
9	68	
10	93	
11	38	
12	92	
13	59	
14	66	
15	61	
16	71	
17	68	
18	82	
19	85	94
20	67	80
21		
22		
23		
24		



Εισάγουμε την εξαρτημένη μεταβλητή στο **Dependent** και την ανεξάρτητη στο **Independent** και **Ok**.

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	βαθμός_προοδου ^a		. Enter

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,633 ^a	,401	,368	8,655

a. Predictors: (Constant), βαθμός_προοδου

Το **RSquare** είναι ο συντελεστής προσδιορισμού που ερμηνεύει το ποσοστό της **Y** που ερμηνεύεται από την **X**, δηλ. το 40% του βαθμού της τελικής εξέτασης ερμηνεύεται από τον βαθμό της προόδου και το υπόλοιπο 60% από άλλους παράγοντες.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	903,504	1	903,504	12,062	,003 ^a
	Residual	1348,296	18	74,905		
	Total	2251,800	19			

a. Predictors: (Constant), βαθμός_προόδου

b. Dependent Variable: βαθμός_εξετάσεων

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	46,730	9,463		4,938	,000
	βαθμός_προόδου	,467	,135	,633	3,473	,003

a. Dependent Variable: βαθμός_εξετάσεων

Από εδώ παίρνουμε τις εκτιμήσεις των παραμέτρων της ευθείας

$$\hat{\alpha} = 46,730$$

$$\hat{\beta} = 0,467$$

άρα

$$\hat{y} = 46.730 + 0.467x$$

Ερμηνείες

Αν ο βαθμός της προόδου είναι μηδέν τότε ο βαθμός της τελικής εξέτασης θα είναι ίσος με 46,730

Αν ο βαθμός της προόδου αυξηθεί κατά μια μονάδα θα έχουμε αύξηση του βαθμού της τελικής εξέτασης κατά 0,467.

Παρατήρηση

Ο γραμμικός συντελεστής συσχέτισης και το β έχουν το ίδιο πρόσημο. Επίσης στην απλή γραμμική παλινδρόμηση ισχύει η σχέση

$$R^2 = r \Leftrightarrow r = \pm\sqrt{R^2}$$

Επίσης η εκτιμηθείσα ευθεία μπορεί να χρησιμοποιηθεί για πρόβλεψη.

Η παραπάνω διαδικασία να εφαρμοστεί στις ακόλουθες ασκήσεις

2. Έστω τα δεδομένα:

X	38	26	48	22	40	15	30	33
Y	10	11	16	8	12	5	10	11

3. Έστω τα δεδομένα:

X	145	164	154	138	130	154	148	152	162	144
Y	41	50	49	36	34	45	42	43	46	46

Όπου X το ύψος σε cm10 μαθητών και Y το βάρος τους σε kg.

4. Δίνεται ο παρακάτω πίνακας

Βάρος X	Χοληστερίνη Y
84	398
73	268
65	245
70	263
76	340
69	253
63	288
72	296
79	329

75	295
67	273
89	354
65	264
57	254
59	225

Να εκτιμηθεί η ευθεία παλινδρόμησης.

Τι ποσοστό της μεταβλητότητας της χοληστερίνης εξηγείται από τη γραμμική σχέση με το βάρος;

Να εκτιμηθεί το επίπεδο της χοληστερίνης ενός ατόμου που έχει βάρος 62 κιλά.

Επαναληπτικές Ασκήσεις

1. Τα ακόλουθα νούμερα αφορούν αριθμό υπαλλήλων σε τ.δ. 25 εταιρειών. Να κατασκευαστεί και να σχολιαστεί το θηκόγραμμα.

35 6 12 11 23 32 18 92 5 17 11
44 14 9 53 15 29 12 58 17 61 38
45 28 10

2. Έστω τα παρακάτω 3 δείγματα. Να κατασκευαστούν και να σχολιαστούν τα αντίστοιχα θηκογράμματα.

ΔΕΙΓΜΑ Ι	ΔΕΙΓΜΑ ΙΙ	ΔΕΙΓΜΑ ΙΙΙ
4	1	1
7	5	3
10	10	10
13	15	17
16	19	19

3. Στον παρακάτω πίνακα δίνεται η κατανομή σε κατηγορίες του προσωπικού μιας Βιομηχανίας. Να κατασκευαστεί το κατάλληλο διάγραμμα.

Κατηγορία προσωπικού	Άτομα
Επιστημονικό	54
Υπαλληλικό	75
Εργατικό	490