ΙΕΚ ΞΑΝΘΗΣ

Μάθημα : Στατιστική Ι

Υποενότητα : Περιγραφική Στατιστική

Επαμεινώνδας Διαμαντόπουλος
Ιστοσελίδα : http://users.sch.gr/epdiaman/
Email : epdiamantopoulos@yahoo.gr

Στόχοι της υποενότητας

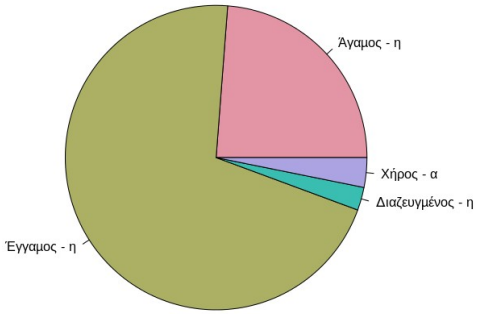
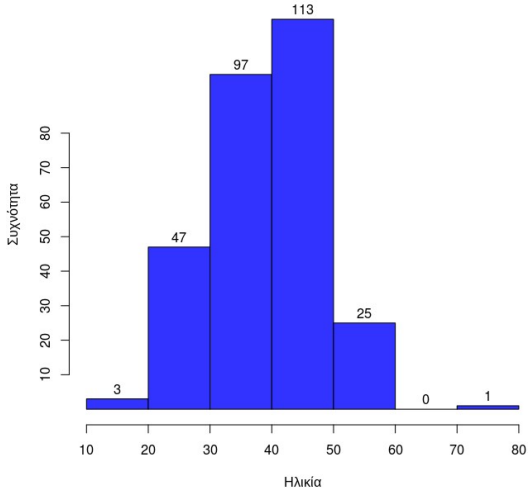
Μετά τη μελέτη αυτής της υποενότητας θα...

1. Γνωρίζεις τι είναι τα περιγραφικά στατιστικά και τους λόγους για τους οποίους χρησιμοποιούνται.
2. Δημιουργείς και ερμηνεύεις στατιστικούς πίνακες.
3. Χρησιμοποιείς διμεταβλητούς πίνακες για να ανιχνεύεις σχέσεις μεταξύ δύο ποιοτικών μεταβλητών.
4. Εφαρμόζεις βασικούς μετασχηματισμούς δεδομένων.
5. Γνωρίζεις και να είσαι σε θέση να ανακαλέσεις βασικά προγράμματα σχεδιασμένα για περιγραφική στατιστική ανάλυση.

Περιγραφική στατιστική ανάλυση

- Περιγραφική Ανάλυση
  - Στοιχειώδεις μετασχηματισμοί των αρχικών δεδομένων με στόχο τον υπολογισμό βασικών μέτρων θέσης και διασποράς, όπως και τη περιγραφή συνολικά της κατανομής των τιμών.
- Ιστόγραμμα
  - Γραφικός τρόπος περιγραφής της κατανομής όπου το ύψος κάθε μίας ράβδου ισούται με τη συχνότητα της κατηγορίας.

Πίνακας 1 **Ενδεικτικά Περιγραφικά Στατιστικά για τα διαφορετικά είδη μεταβλητών.**

Μεταβλητή	Ενδεικτική ανάλυση	Ενδεικτικό διάγραμμα
Ποιοτική (π.χ. φύλο)	Πίνακας Συχνοτήτων και Ποσοστών Επικρατούσα τιμή	
Διατακτική (π.χ. πλήθος παιδιών)		
Διαστημική (π.χ. pH)	Μέση τιμή τυπική απόκλιση τυπικό σφάλμα	
Αναλογική (π.χ. ύψος, βάρος)		

Παράδειγμα Πίνακα Συχνοτήτων

Predominant religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Animist	4	3,7	3,7	3,7
	Buddhist	7	6,4	6,5	10,2
	Catholic	41	37,6	38,0	48,1
	Hindu	1	,9	,9	49,1
	Jewish	1	,9	,9	50,0
	Muslim	27	24,8	25,0	75,0
	Orthodox	8	7,3	7,4	82,4
	Protstnt	16	14,7	14,8	97,2
	Taoist	2	1,8	1,9	99,1
	Tribal	1	,9	,9	100,0
	Total	108	99,1	100,0	
Missing	missing	1	,9		
Total		109	100,0		

Η αθροιστική σχετική συχνότητα ερμηνεύεται ευκολότερα στις διατακτικές (ordinal) μεταβλητές

Number of Cylinders

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3 Cylinders	4	1,0	1,0	1,0
	4 Cylinders	207	51,0	51,1	52,1
	5 Cylinders	3	,7	,7	52,8
	6 Cylinders	84	20,7	20,7	73,6
	8 Cylinders	107	26,4	26,4	100,0
	Total	405	99,8	100,0	
Missing	System	1	,2		
Total		406	100,0		

Παράδειγμα Πίνακα Συχνοτήτων

Πίνακας : Ποιοι είναι οι σημαντικότεροι πόλοι έλξης της περιοχής σας τους οποίους αναγνωρίζουν περισσότερο οι τουρίστες / επισκέπτες (πχ Χιονοδρομικό κέντρο, Δάσος, Φιλοξενία ...):

Απόκριση	Συχνότητα	Ποσοστό
Το χιονοδρομικό κέντρο Περτουλίου	83	87,4%
Το δάσος του Περτουλίου	44	46,3%
Η φιλοξενία και η καλή εξυπηρέτηση	40	42,1%
Το φυσικό περιβάλλον της περιοχής	36	37,9%
Το καλό φαγητό, γαστρονομία	12	12,6%
Η εύκολη πρόσβαση	4	4,2%
δραστηριότητες βουνού	3	3,2%
Η ψυχαγωγία, αναψυχή που μπορεί να βρει ο επισκέπτης στην περιοχή	3	3,2%
Το Περτούλι	2	2,1%
Ξεκούραση	2	2,1%
Η Ελάτη	2	2,1%
Κοντινοί προορισμοί	1	1,1%
Το πανελλήνιο αντάμωμα των Σαρακατσαναίων που γίνεται στα Περτουλιότικα λιβάδια	1	1,1%
Το καλό φαγητό, γαστρονομία και μπαρ	1	1,1%

Σχεδιασμός Τουριστικού Μάρκετινγκ και διαχείριση τουριστικής ανάπτυξης : Η περίπτωση της περιοχής Ελάτης – Περτουλίου και του ορεινού τμήματος του Νομού Τρικάλων, 2011

Διμεταβλητός Πίνακας Συχνοτήτων (Cross-Tabulation)

- Διμεταβλητός Πίνακας Συχνοτήτων (Cross-Tabulation)
  - Χρησιμοποιείται για την ανίχνευση εξάρτησης μεταξύ δύο ποιοτικών ή δύο διακριτών μεταβλητών με σχετικά λίγες τιμές
  - Οι τιμές της μίας μεταβλητής ορίζει τις γραμμές ενώ οι τιμές της άλλης ορίζει τις στήλες.
- Πίνακας συμπτώσεων (contingency table)
  - Πίνακας που περιέχει τις συχνότητες εμφάνισης συνδυασμών τιμών δύο ή περισσοτέρων μεταβλητών.
- Περιθώριες συχνότητες (Marginals)
  - Τα αθροίσματα των γραμμών και των στηλών ενός διμεταβλητού πίνακα που εμφανίζονται στο περιθώριο του πίνακα.

Παράδειγμα διμεταβλητού πίνακα (1)

Πίνακας : Πίνακας συμπτώσεων στις απόκρίσεις μεταξύ του στρεσογόνου γεγονότος στην παιδική ηλικία και του φόβου για την ασθένεια

			Φοβάστε την ασθένεια;		Σύνολο
			Ναι	Όχι	
Στρεσογόνο γεγονός παιδικής ηλικίας	Ναι	Συχνότητα	30	21	51
		Αναμενόμενη Συχνότητα	26,2	24,8	51,0
	Όχι	Συχνότητα	6	13	19
		Αναμενόμενη Συχνότητα	9,8	9,2	19,0
Σύνολο		Συχνότητα	36	34	70

(Αντιμετώπιση του πόνου, 2010)

Παράδειγμα διμεταβλητού πίνακα (2)

Πίνακας : Πίνακας συμπτώσεων της ικανοποίησης σχετικά με τον καταμερισμό εργασίας στην ομάδα και στα έτη εργασίας

		Έτη εργασίας					
	Απόκριση	Λιγότερα από 3	3 έως 5	6 έως 10	11 έως 15	Περισσότερα από 15	Σύνολο
Υπάρχει ικανοποιητικός καταμερισμός της εργασίας μέσα στην ομάδα	Διαφωνώ Πλήρως		1	1			2
	Διαφωνώ		1				1
	Ούτε συμφωνώ ούτε συμφωνώ	1	1	5		1	8
	Συμφωνώ	1	2	11	5	2	21
	Συμφωνώ πλήρως			2	1	1	4
	Σύνολο	2	5	19	6	4	36

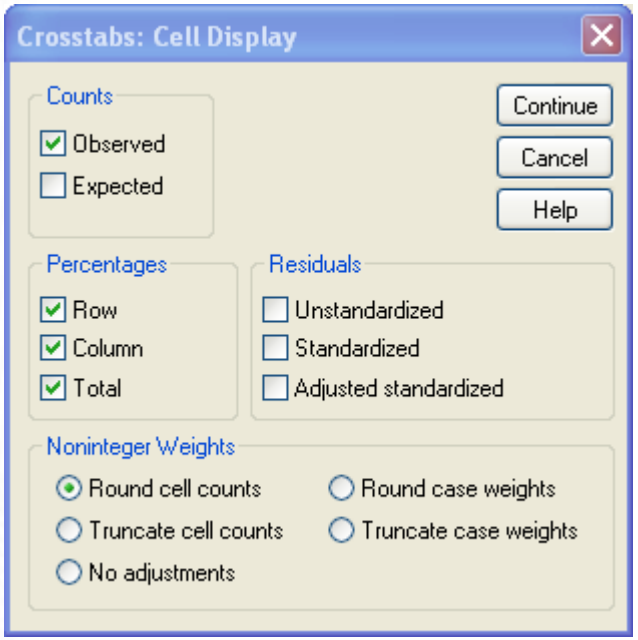
(Διερεύνηση αντιλήψεων για το ρόλο και τη λειτουργία της επικοινωνίας στους επαγγελματίες του ΚΕΘΕΑ, 2010)

Πίνακας 2 **Σχέση του φύλου του εργαζόμενου και της κατηγορίας προσωπικού στην οποία ανήκει (όπως παράγονται από το SPSS)**

Φύλο * Κατηγορία προσωπικού Crosstabulation

Count

		Κατηγορία προσωπικού			Total
		Υπάλληλος	Επιστάτης	Διευθυντής	
Φύλο	Γυναίκα	206	0	10	216
	Άνδρας	157	27	74	258
Total		363	27	84	474



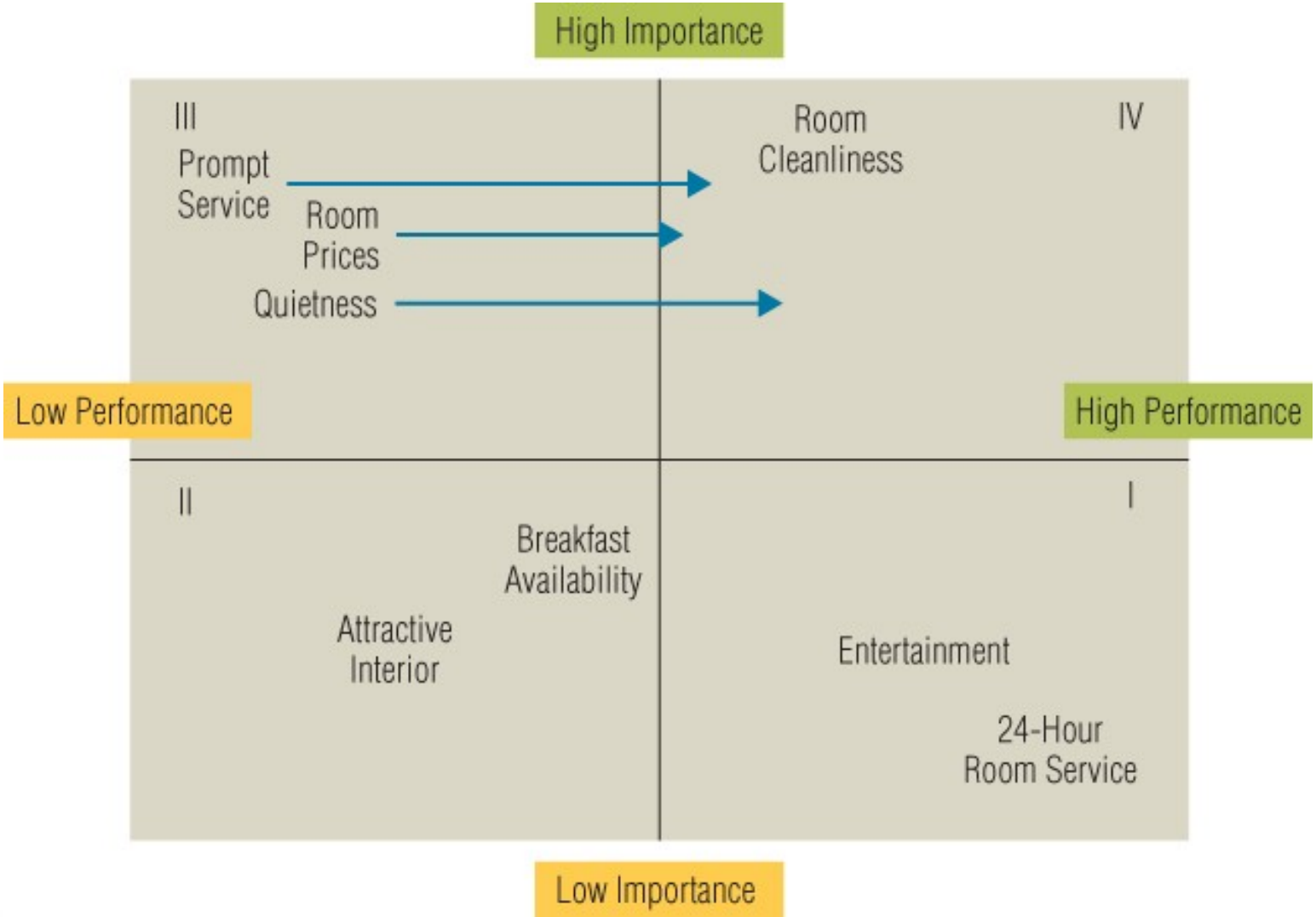
Φύλο * Κατηγορία προσωπικού Crosstabulation

			Κατηγορία προσωπικού			Total
			Υπάλληλος	Επιστάτης	Διευθυντής	
Φύλο	Γυναίκα	Count	206	0	10	216
		% within Φύλο	95,4%	,0%	4,6%	100,0%
		% within Κατηγορία προσωπικού	56,7%	,0%	11,9%	45,6%
		% of Total	43,5%	,0%	2,1%	45,6%
	Άνδρας	Count	157	27	74	258
		% within Φύλο	60,9%	10,5%	28,7%	100,0%
		% within Κατηγορία προσωπικού	43,3%	100,0%	88,1%	54,4%
		% of Total	33,1%	5,7%	15,6%	54,4%
Total		Count	363	27	84	474
		% within Φύλο	76,6%	5,7%	17,7%	100,0%
		% within Κατηγορία προσωπικού	100,0%	100,0%	100,0%	100,0%
		% of Total	76,6%	5,7%	17,7%	100,0%

Διμεταβλητός Πίνακας Συχνοτήτων

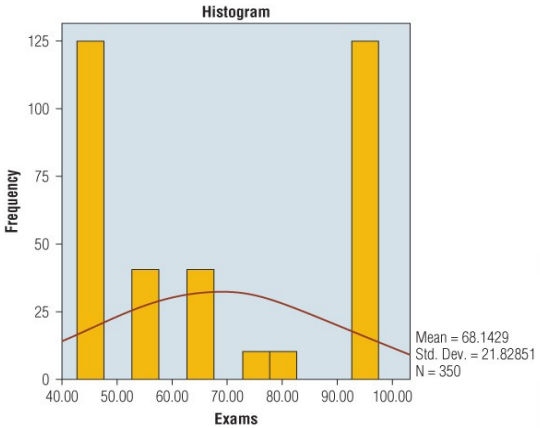
- Πόσοι διμεταβλητοί πίνακες;
  - Κάθε μία από τις ερωτήσεις μπορεί να γίνει και ένας παράγοντας διερεύνησης και ερμηνείας.
  - Αν η ερευνητική υπόθεση αφορά σχέση μεταξύ δύο ποιοτικών ή κατηγορικών μεταβλητών τότε ο διμεταβλητός πίνακας είναι η κατάλληλη επιλογή.
- Τετραγωνική Ανάλυση (quadrant analysis)
  - Επέκταση του διμεταβλητού πίνακα στην οποία οι αποκρίσεις δύο μεταβλητών εμφανίζονται στα τέσσερα μέρη ενός διμεταβλητού πίνακα.
    - Συσχέτιση της σημαντικότητας των παραγόντων όπως αξιολογούνται από τους πελάτες έναντι της πραγματικής επίδοσης του οργανισμού

Πίνακας 3 Τετραγωνική ανάλυση σημαντικότητας – επίδοσης ενός ξενοδοχείου.



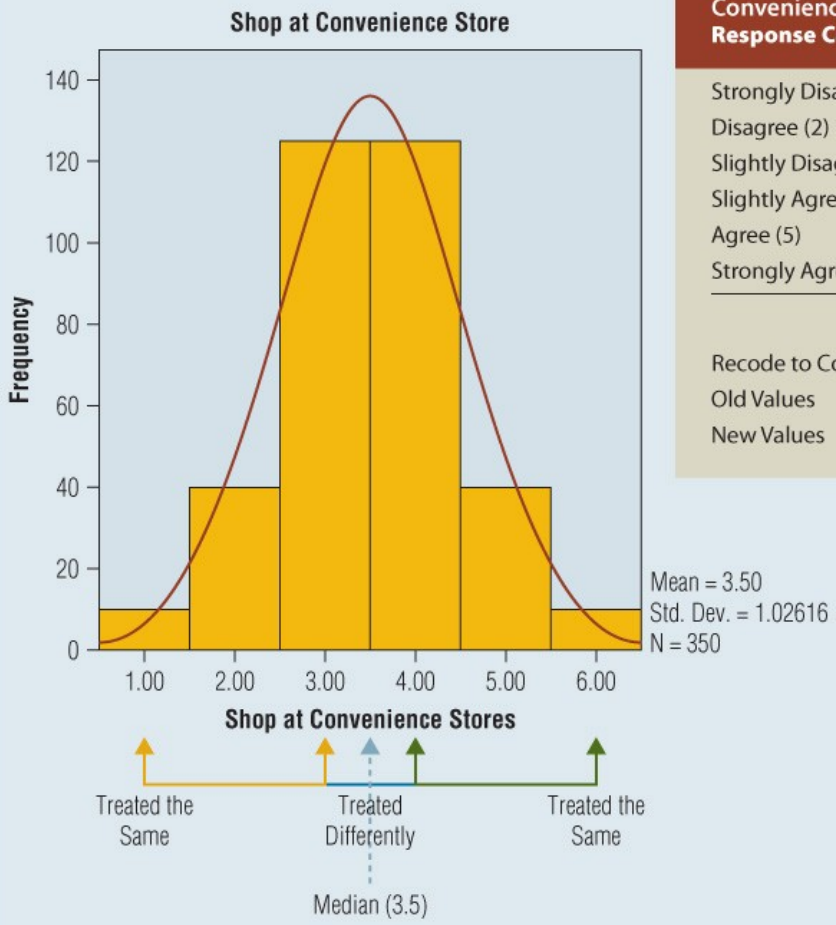
Μετασχηματισμοί των δεδομένων

- Διαχωρισμός στη διάμεσο (median split).
  - Διαχωρισμός των παρατηρήσεων σε δύο κατηγορίες ανάλογα με τη θέση τους σχετικά με τη διάμεσο.
  - Διαδικασία προτεινόμενη αν η κατανομή χαρακτηρίζεται από την εμφάνιση δύο επικρατουσών τιμών.



  - Χρήση του διαχωρισμού σε συνεχείς μεταβλητές αφαιρεί πληροφορίες από τα δεδομένα με αποτέλεσμα πιθανό αναληθές συμπέρασμα.

Πίνακας 4 **Προβληματικός διαχωρισμός στη διάμεσο σε κατανομή με μία επικρατούσα τιμή.**



Frequency Distribution: X1 = I Do most of my Shopping at Convenience stores.

Response Category (Code)	Counts	Cumulative Percentage
Strongly Disagree (1)	10	2.86%
Disagree (2)	40	14.29%
Slightly Disagree (3)	125	50.00%
Slightly Agree (4)	125	85.71%
Agree (5)	40	97.14%
Strongly Agree (6)	10	100.00%

Median = 3.5

Recode to Complete Data Transformation:

Old Values	1	2	3	4	5	6
New Values	1	1	1	2	2	2

Αριθμοδείκτες (Index Numbers)

- Αριθμοδείκτες (Index Numbers)
  - Τιμές ή μεταβλητές μετασχηματισμένες έτσι ώστε να φαίνεται η σχέση τους με μία βασική ποσότητα.
- Αριθμοδείκτες τιμών/αξιών.
  - Απλοί μετασχηματισμοί με τους οποίες είναι δυνατή η διαχρονική παρακολούθηση των τιμών της μεταβλητής ή η σύγκριση των τιμών δύο ή περισσοτέρων μεταβλητών.
  - Τιμάριθμος

Πίνακας 4 Τιμάριθμος

Πίνακας 1
Μηνιαίες μεταβολές Δείκτη Τιμών Καταναλωτή
(Έτος βάσης: 2009=100,0)

	Ομάδες αγαθών & υπηρεσιών	Σταθμίσεις ΕΟΠ 08/09 (ο/οο)	2011	2011	Δ (%)	2010	2010	Δ(%)
			Νοέμβριος	Οκτώβριος		Νοέμβριος	Οκτώβριος	
1	Διατροφή και μη αλκοολούχα ποτά	171,22	104,20	103,40	0,8	100,43	100,25	0,2
2	Αλκοολούχα ποτά και καπνός	34,80	126,28	126,26	0,0	120,81	120,00	0,7
3	Ένδυση και υπόδηση	86,40	110,36	110,13	0,2	107,76	107,71	0,0
4	Στέγαση	115,88	118,53	117,79	0,6	108,42	107,89	0,5
5	Διαρκή αγαθά - Είδη νοικοκυριού και υπηρεσίες	73,63	103,92	103,87	0,0	103,52	103,22	0,3
6	Υγεία	70,73	100,40	100,63	-0,2	101,40	101,33	0,1
7	Μεταφορές	132,39	122,78	123,43	-0,5	119,55	118,93	0,5
8	Επικοινωνίες	45,56	104,71	104,71	0,0	104,05	104,05	0,0
9	Αναψυχή - Πολιτιστικές – δραστηριότητες	47,77	100,73	100,80	-0,1	101,54	101,44	0,1
10	Εκπαίδευση	30,41	101,16	101,22	-0,1	102,12	102,12	0,0
11	Ξενοδοχεία-Καφέ - Εστιατόρια	114,09	106,64	106,46	0,2	103,32	103,53	-0,2
12	Άλλα αγαθά και υπηρεσίες	77,11	108,83	108,70	0,1	106,67	106,54	0,1
Γενικός Δείκτης Τιμών Καταναλωτή		1000,00	109,73	109,56	0,2	106,61	106,37	0,2

Σημείωση : 1. Οι δείκτες υπολογίζονται με άπειρα δεκαδικά ψηφία και δημοσιεύονται με στρογγυλοποίηση δύο δεκαδικών ψηφίων.
2. Οι ποσοστιαίες μεταβολές υπολογίζονται βάσει δεικτών με άπειρα δεκαδικά ψηφία και δημοσιεύονται με στρογγυλοποίηση ενός δεκαδικού

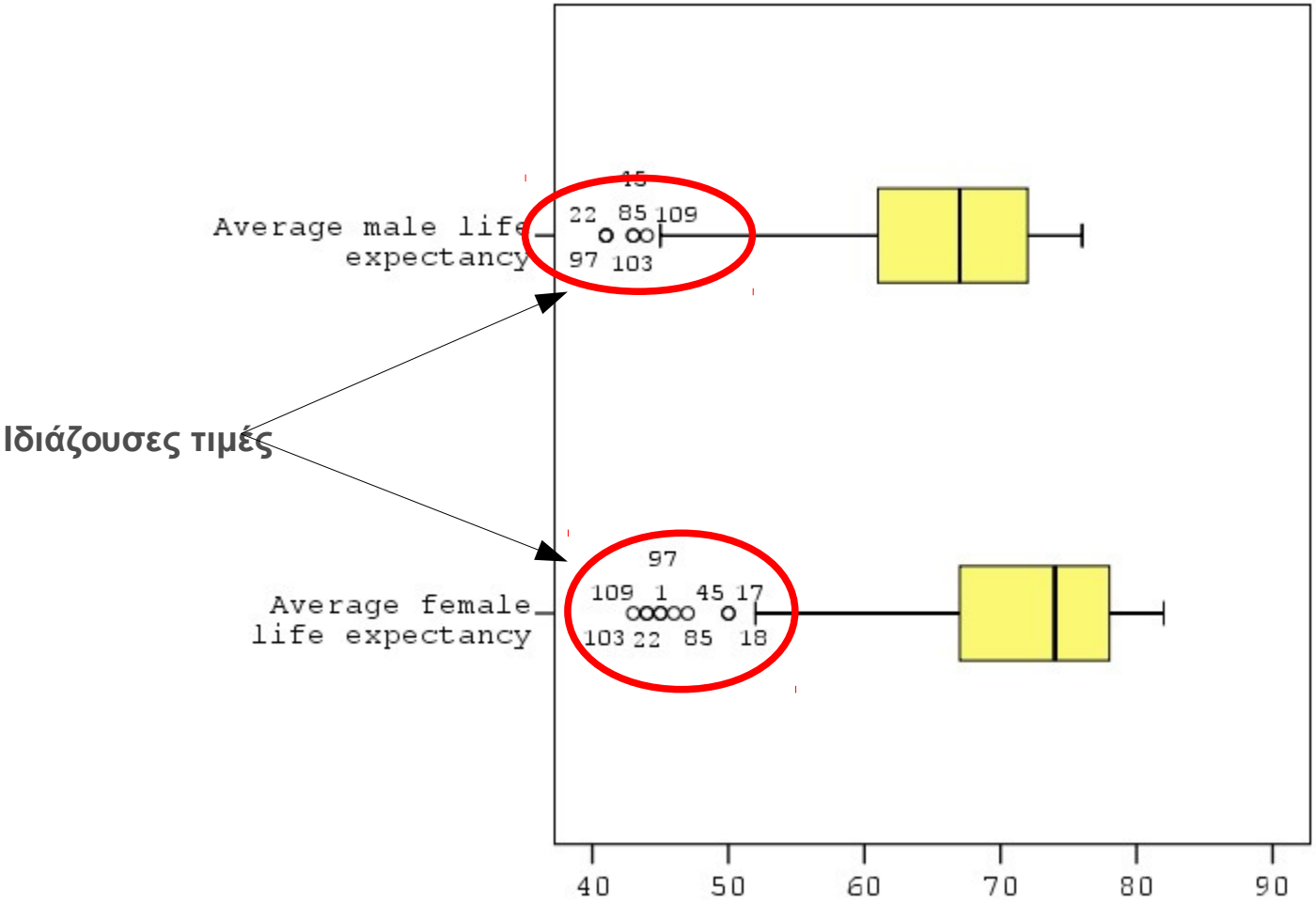
Προγράμματα Στατιστικής Ανάλυσης

- Πακέτα Λογισμικού
  - Λογιστικά φύλλα(Spreadsheets)
    - Excel
  - Στατιστικά προγράμματα :
    - SAS
    - SPSS (*Statistical Package for Social Sciences*)
    - MINITAB
    - R - Project

Διαγράμματα και αναπαράσταση από ειδικά προγράμματα λογισμικού

- Θηκογράμματα
  - Γραφική αναπαράσταση της μορφής της κατανομής με τη βοήθεια της διαμέσου και των τεταρτημορίων.
- Ενδοτεταρτημοριακό εύρος
  - Μέτρο μεταβλητότητας.
- Ιδιάζουσα τιμή (Outlier) – Ακραία τιμή (Extreme Value)
  - Μία τιμή που λαμβάνει ασυνήθιστα μεγάλη ή μικρή τιμή σε σχέση με τα υπόλοιπα δεδομένα.
  - Δεν υπάρχει σωστός και λάθος τρόπος για την διαχείρισή αυτών των τιμών.

Πίνακας 9 Θηκόγραμμα (box plot) δημιουργημένο από υπολογιστή.



(Προσδόκιμο ζωής ανδρών και γυναικών, δείγμα 109 χωρών, 1995)

Παρουσίαση περιγραφικών στατιστικών

Πρέπει να δίνεται πάντα η μέση τιμή (M) και ένα μέτρο διασποράς όπως η τυπική απόκλιση (SD) ή το τυπικό σφάλμα (SE). Τα στατιστικά μπορούν να δίνονται στο κείμενο ή σε πίνακα αλλά όχι και με τους δύο τρόπους.

Ενδεικτικά

- “Η μέση ηλικία των ερωτώμενων ήταν 25.5 έτη (SD = 7.94).”

- “Οι ερωτώμενοι ήταν από 18 έως 70 ετών (M = 25.5 SD = 7.94). Η ηλικία δεν ήταν κανονικά κατανεμημένη, δείκτης ασυμμετρίας 1.87 (SE = 0.05) και δείκτης κυρτότητας 3.93 (SE = 0.10)”

- “Στην έρευνα συμμετείχαν 98 άνδρες και 132 γυναίκες, από 17 έως 25 ετών (άνδρες : M = 19.2, SD = 2.32, γυναίκες : M = 19.6, SD = 2.54).”