

Τα ψεύδη της Στατιστικής

Μιχαήλ Μανωλόπουλος, Δρ. Παν. Αθηνών. Σχ. Σύμβουλος Μαθηματικών, Παπαδιαμάντη 11, 14343 Ν. Χαλκηδόνα, e-mail: mmanol@sch.gr
Βασίλειος Μανωλόπουλος, Εφαρμοσμένος Φυσικός του Ε.Μ.Π. –M.sc. στην Εφαρμοσμένη Μηχανική, Πολυτεχνειούπολη Ζωγράφου, Ηρώων Πολυτεχνείου 9, 15780 Ζωγράφου, e-mail: synanastrofes@hotmail.com

ΠΕΡΙΛΗΨΗ

Η επιστήμη της στατιστικής ασχολείται με τη συγκέντρωση επεξεργασία και αξιολόγηση της πληροφορίας. Η πολυπλοκότητα της σημερινής πραγματικότητας δίνει την ευκαιρία για εξαπάτηση των μη μυημένων, αφού μια στατιστική μελέτη μπορεί να δώσει σε κάποια στοιχεία όψη που απέχει από την πραγματικότητα και επομένως να παραπληροφορεί. Η σωστή και επαρκής παιδεία προστατεύει τον πολίτη και τον βοηθά να διακρίνει την απάτη από την αλήθεια. Στην παρούσα εργασία θα παρουσιαστούν κίνδυνοι που παραμονεύουν για να εξαπατήσουν τον πολίτη, σε μια στατιστική έρευνα. Θα επικεντρώσουμε την προσπάθειά μας σε παραδείγματα που αφορούν στην παρουσίαση διαγραμμάτων, στην επιλογή του δείγματος, την παρουσίαση των μέτρων θέσεως και διασποράς, τους αριθμοδείκτες στους στατιστικούς ελέγχους, καθώς και στην παρανόηση σκόπιμη ή μη των στατιστικών δεδομένων.

Abstract

The science of statistics is dealing with the collection, processing and evaluation of information. The complexity of today's reality provides an opportunity for using data and statistics to obfuscate reality rather than to illuminate. Proper and adequate education protects citizens and helps them to distinguish fraud from the truth.

In this work, lurking risks to deceive the public in a statistical survey will be presented. We will concentrate our effort on the

presentation of charts, in the selection of the sample, in measures of centrality and variability, on number indexes and on misuse of statistical data, accidental or purposeful.

1. ΕΙΣΑΓΩΓΗ

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write—H.G. Wells¹

Οι κυνικοί λένε σαρκαστικά πως μπορείς να αποδείξεις οτιδήποτε με τη Στατιστική. Άλλοι πάλι υποστηρίζουν πως δεν μπορείς να κάνεις τίποτα με τη Στατιστική. Κάποιοι θυμίζουν ότι η Στατιστική είναι ένας τρόπος για να λείει κανείς ψέματα. «Υπάρχουν τριών ειδών ψέματα: τα ψέματα, τα καταραμένα ψέματα και η στατιστική» (Disraeli²).

Ένα παράδειγμα: Η συντριπτική πλειονότητα των Ελλήνων έχει αριθμό κάτω άκρων μεγαλύτερο από τον μέσο όρο. Πράγματι:

Υπάρχουν 500 συμπατριώτες μας που δεν έχουν κάτω άκρα, ενώ δύομισι χιλιάδες έχουν ένα πόδι. (ακρωτηριασμοί, αναπηρίες κατά την γέννηση, εργατικά ατυχήματα, τροχαία, κλπ-τα νούμερα δεν ανταποκρίνονται στην πραγματικότητα). Οι υπόλοιποι (10 997 000) έχουν δύο πόδια. Μέσος όρος κάτω άκρων των Ελλήνων:

$$\frac{500 \times 0 + 2500 \times 1 + 10997000 \times 2}{11000000} = 1,999682.$$

Η συντριπτική πλειονότητα των Ελλήνων έχει 2 πόδια, $2 > 1,999682$.

Τα μέτρα θέσεως τα διαγράμματα οι τάσεις, γενικά τα στατιστικά συμπεράσματα μπορεί να κρύβουν περισσότερα από αυτά που αντικρίζουν τα μάτια μας μπορεί όμως να κρύβονται και πολύ λιγότερα. Ο ενημερωμένος πολίτης, είναι αυτός που κινδυνεύει λιγότερο από τις παρανοήσεις των στατιστικών δεδομένων. Στην ενημέρωση και εγρήγορση του αναγνώστη, και ειδικότερα του καθηγητή της

¹ Ο στατιστικός τρόπος σκέψης, μια μέρα θα είναι τόσο αναγκαίος για τον αποτελεσματικό πολίτη, όσο η γραφή και η ανάγνωση. —H.G. Wells, Άγγλος συγγραφέας έργων επιστημονικής φαντασίας.

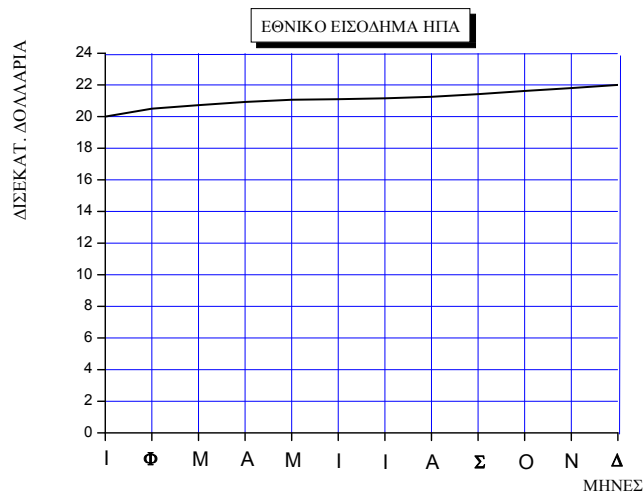
² Benjamin Disraeli (1804-1881). Πρωθυπουργός της Μ. Βρετανίας.

δευτεροβάθμιας εκπαίδευσης, αλλά και του μαθητή του Λυκείου σκοπεύει η παρούσα εργασία. Είναι δε σαφές ότι η σωστή πληροφόρηση βοηθά στην ενδυνάμωση της ίδιας της δημοκρατίας και στο βάθεμά της.

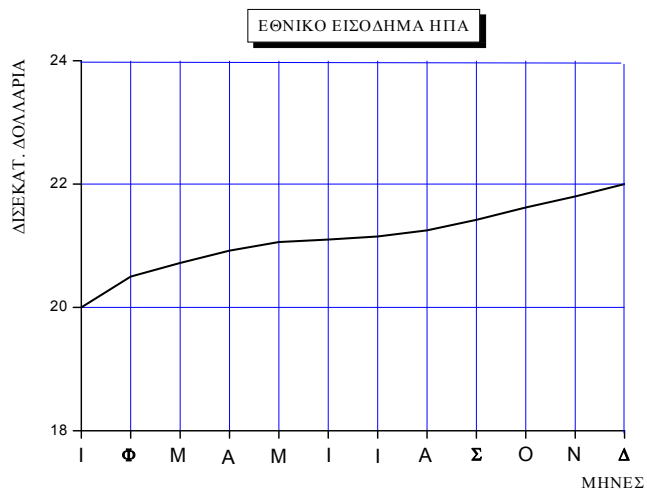
2. ΤΟ ΠΑΡΑΠΛΑΝΗΤΙΚΟ ΔΙΑΓΡΑΜΜΑ

Οι πολλοί αριθμοί μπερδεύουν και ενίοτε τρομοκρατούν. Όταν οι λέξεις δεν ταιριάζουν για μια συγκεκριμένη δουλειά ή δεν μας συμφέρουν μπορούμε να χρησιμοποιήσουμε το διάγραμμα!

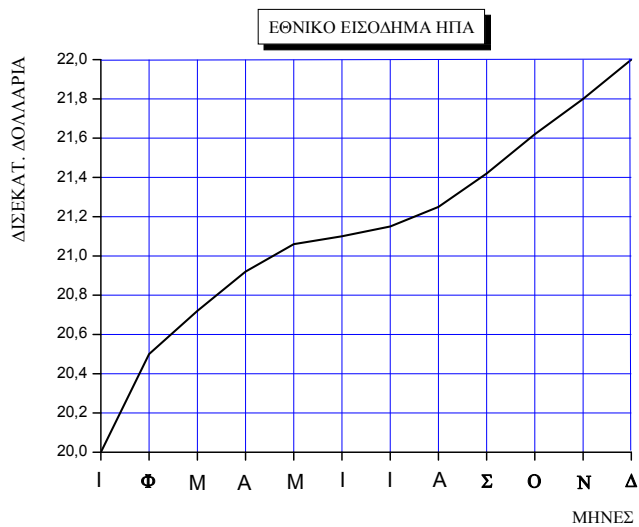
Στο διάγραμμα του σχ. 2.1 παρουσιάζεται η αύξηση του εθνικού εισοδήματος των ΗΠΑ σε ένα χρόνο (Huff 1983, Κεφ. 5). Το διάγραμμα αρχίζει από το 0 και η αύξηση (10%) έχει την πραγματική της διάσταση. Το σημαντικό είναι πως στα σχ. 2.2 και 2.3 οι συντεταγμένες των σημείων είναι οι ίδιες, αλλά ο κάθετος άξονας αρχίζει στο 1^ο από το 18 και στο 2^ο από το 20. Η εντύπωση που δημιουργείται είναι εντελώς διαφορετική. Στο σχ. 2.2 η αύξηση δείχνει σημαντική, ενώ στο 2.3 είναι εντυπωσιακή και όλα αυτά χωρίς λόγια.



Σχήμα 2.1. Το εθνικό εισόδημα των ΗΠΑ μέσα σε ένα χρόνο. (Αύξηση 10%)



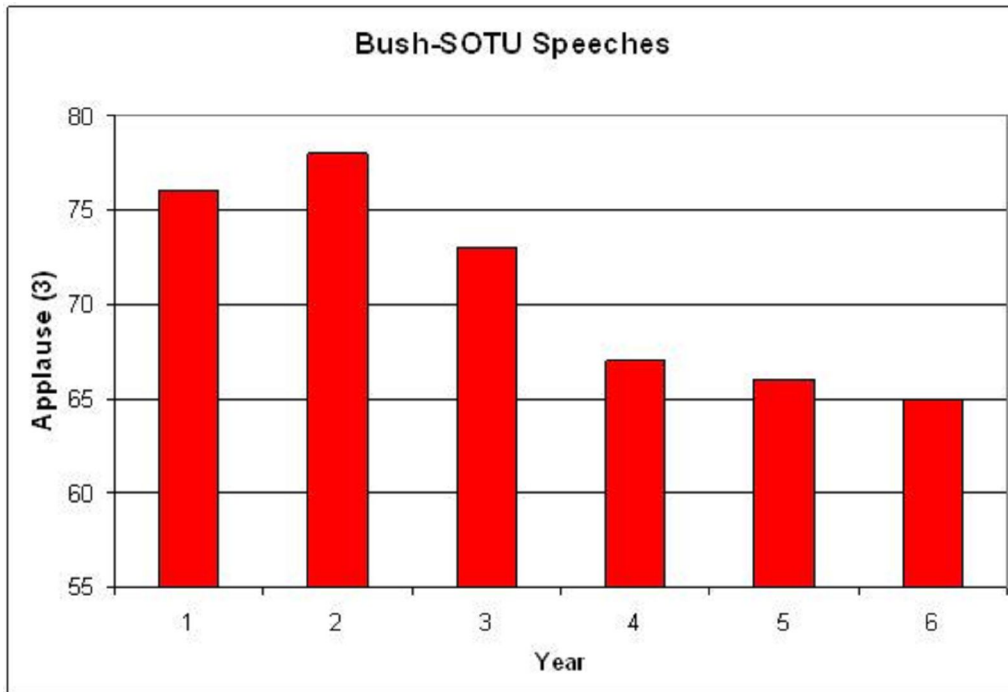
Σχήμα 2.2. Η αύξηση φαίνεται μεγαλύτερη.



Σχήμα 2.3. Η αύξηση τώρα είναι «εντυπωσιακή».

Τα δύο επόμενα διαγράμματα (σχ. 2.4, 2.5) είναι από το διαδίκτυο (<http://jdasovic.com/2008/01/lies-damned-lies-and-excel-charts/>), κάτω από τον τίτλο «ψέματα, καταραμένα ψέματα, διαγράμματα excel». Η πηγή του διαγράμματος είναι ο δικτυακός τόπος του δημοκρατικού κόμματος των ΗΠΑ: <http://www.democrats.org>. Το SOTU είναι το αρκτικόλεξο του State of the Union και πρόκειται για τις ομιλίες που απευθύνει ο πρόεδρος των ΗΠΑ κατά καιρούς για την κατάσταση στην χώρα του. Ο οριζόντιος άξονας αφορά έτη ο κάθετος την αποδοχή που είχε στην χώρα του ο λόγος του προέδρου. Το συμπέρασμα που προκύπτει από το διάγραμμα του σχήματος 2.4 είναι πως η αποδοχή του Bush από τον κόσμο με εξαίρεση τον Ιανουάριο του 2003 (λίγο πριν την εισβολή στο Ιράκ), πέφτει σημαντικά. Το ερώτημα είναι γιατί ο κάθετος άξονας αρχίζει από το 55, μια αυθαίρετη επιλογή του προγράμματος Excel; Τι θα συμβεί αν κάποιος επιλέξει ως αρχή του κάθετου άξονα το 0; Η απάντηση φαίνεται στο σχήμα 2.5. Η εντύπωση που δημιουργείται με το νέο διάγραμμα είναι διαφορετική. Η πτώση της αποδοχής του προέδρου δεν φαίνεται σημαντική και δείχνει να βρίσκεται στα όρια του τυχαίου σφάλματος³. Ποιο διάγραμμα είναι το σωστό; Ίσως η ορθότερη απάντηση να ήταν μετά τη σύγκριση με αντίστοιχα διαγράμματα άλλων προέδρων.

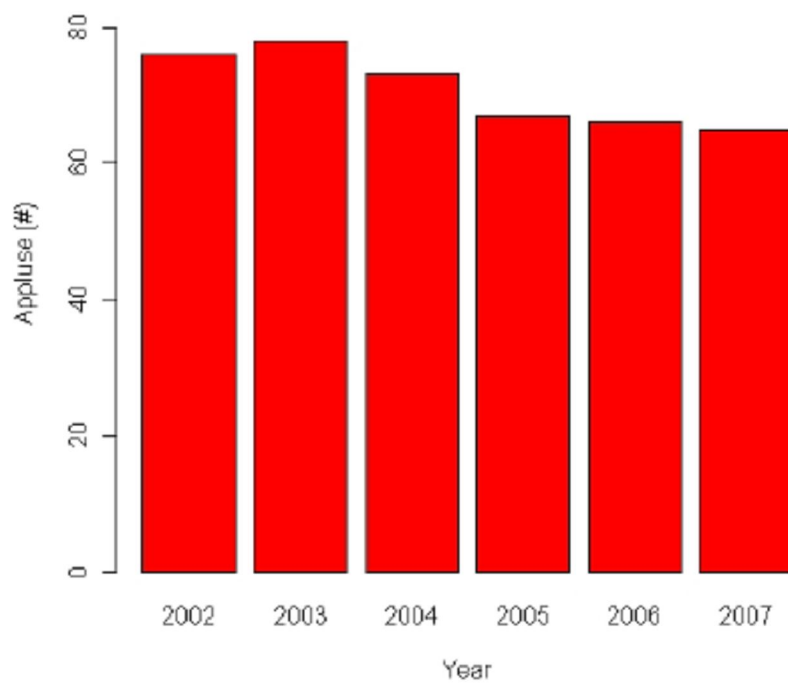
³ Τυχαία σφάλματα (σε αντίθεση με τα συστηματικά): πρόκειται για σφάλματα που εξαρτώνται από διάφορους περιστασιακούς κυρίως παράγοντες, όπως είναι η κούραση του ατόμου, η έλλειψη συγκέντρωσης λόγω περιστασιακής συναισθηματικής φόρτισης και άλλοι. Οι αποκλίσεις που παράγει το τυχαίο σφάλμα δεν συσσωρεύονται προς μία κατεύθυνση αλλά κυμαίνονται προς διάφορες κατευθύνσεις, οπότε δεν παράγουν μεροληψία υπέρ ή κατά μίας συγκεκριμένης κατηγορίας. Σύμφωνα με τη στατιστική θεωρία εάν ένα φαινόμενο είναι τυχαίο τότε η οριακή κατανομή που θα προκύψει (μετά από άπειρες επαναλήψεις των προσπαθειών) θα είναι μια κανονική κατανομή.



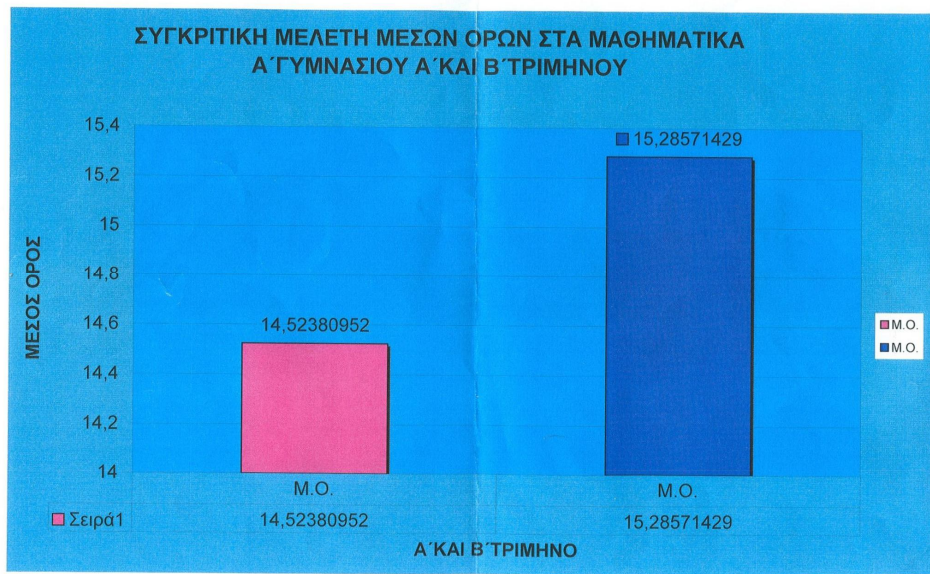
Σχήμα 2.4. Η αποδοχή από τον κόσμο των ομιλιών του προέδρου Bush, κατά τα έτη 2002 έως 2007. Το διάγραμμα στον κάθετο άξονα αρχίζει από το 55.

Τέλος το ραβδόγραμμα του σχήματος 2.6 προέρχεται από μια μελέτη των μαθητών του 1^{ου} Γυμνασίου Θηβών, όπου ο δαίμων του Excel, έχει βάλει το χέρι του «αλλοιώνοντας» την πραγματικότητα. Η πρώτη ράβδος έχει ύψος περίπου 14,5 ενώ η δεύτερη 15,3. Το ραβδόγραμμα εξαπατά ως προς την πραγματική αύξηση του μέσου όρου, αν κάποιος επικεντρώσει την προσοχή του στην εικόνα και όχι στους αριθμούς.

Bush SOTU-Speeches Applause Trend

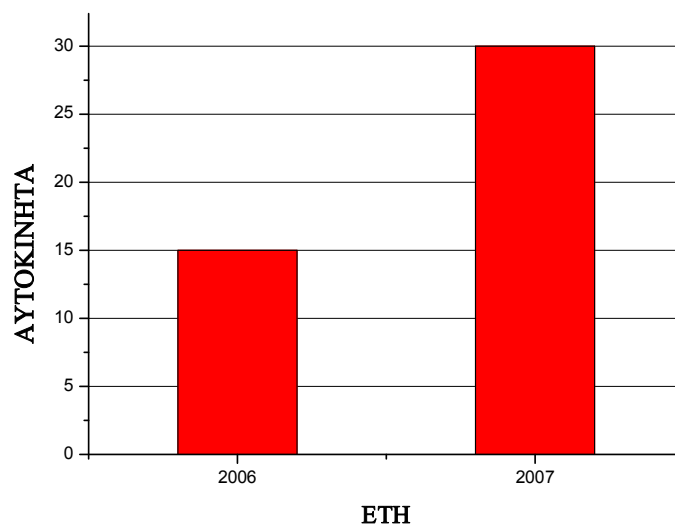


Σχήμα 2.5. Το διάγραμμα του σχήματος 2.4 με τον κάθετο άξονα να αρχίζει από το 0.



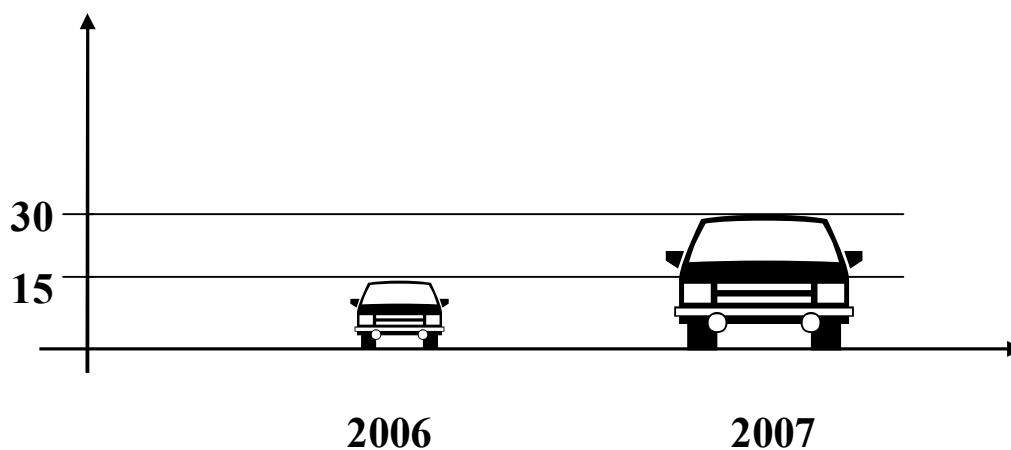
Σχήμα 2.6. Μέσος όρος στα μαθηματικά Α' και Β' Τριμήνου της Α' Γυμνασίου

Το σχήμα 2.7 παριστάνει τις αγορές των νέων αυτοκινήτων σε μια περιοχή το 2006 και το 2007. Τα στοιχεία είναι υποθετικά το ραβδόγραμμα όμως είναι τίμιο αφού οι διπλάσιες πωλήσεις του 2007 αποτυπώνονται με ράβδο διπλασίου ύψους και ίδιου πλάτους. Αν θελήσουμε όμως να παρουσιάσουμε την αύξηση του 2007 μεγαλύτερη από την



Σχήμα 2.7. Οι αγορές νέων αυτοκινήτων (σε χιλιάδες), σε κάποια περιοχή, το 2006 και το 2007.

πραγματικότητα, μπορούμε να παραστήσουμε τα προηγούμενα σχεδιάζοντας για παράδειγμα ένα αυτοκίνητο για τις πωλήσεις του 2006 και με ένα «διπλάσιο» εκείνες του 2007 (σχήμα 2.8). Οι αριθμοί θα είναι οι ίδιοι στο διάγραμμα, αν όμως διπλασιάσουμε τις διαστάσεις του δευτέρου αυτοκινήτου, γνωρίζουμε ότι καταλαμβάνει τετραπλάσια έκταση. Η εντύπωση που δημιουργείται είναι διαφορετική, αφού η τετραπλάσια έκταση σαφώς υπερτιμά τη σχέση των δύο πωλήσεων που είναι 1 προς 2.



Σχήμα 2.8. Οι αγορές νέων αυτοκινήτων (σε χιλιάδες), όπως και στο σχήμα 2.7. Τα μεγέθη όμως των αυτοκινήτων δεν παραπέμπουν στην σωστή σχέση μεταξύ των δύο αγορών.

3. ΟΙ ΑΡΙΘΜΟΔΕΙΚΤΕΣ

Οι αριθμοδείκτες είναι στατιστικά εργαλεία που χρησιμοποιούνται για να δώσουν το αποτέλεσμα της σύγκρισης απλών ή συνθέτων μεγεθών, μιας σύγκρισης που γίνεται μεταξύ δύο ή περισσότερων καταστάσεων. Εφαρμογές βρίσκει η θεωρία των αριθμοδεικτών στην εύρεση του Δείκτη Τιμών Καταναλωτή, στους Χρηματιστηριακούς Δείκτες, στους Δείκτες Εξωτερικού Εμπορίου κλπ, (Τζωρτζόπουλος 2003).

Ας προσπαθήσουμε να τους χρησιμοποιήσουμε όπως μας συμφέρει! Ας υποθέσουμε σε μια απλοποιημένη οικονομία οι τιμές των προϊόντων α και β επηρεάζουν σημαντικά το κόστος ζωής. Ας υποθέσουμε επίσης ότι για το 1 kg αυτών των προϊόντων ισχύει ο πίνακας 3.1, όπου οι αριθμοί εκφράζουν ευρώ. Αυξήθηκε το κόστος ζωής, μειώθηκε ή παρέμεινε σταθερό από πέρυσι μέχρι φέτος;

	Πέρυσι	Φέτος
α	10	20
β	10	5

Πίνακας 3.1

	Πέρυσι	Φέτος
α	1	2
β	1	0,5

Πίνακας 3.2

	Πέρυσι	Φέτος
α	0,5	1
β	2	1

Πίνακας 3.3

	Πέρυσι	Φέτος
α	1	2
β	1	0,5

Πίνακας 3.4

Ας πάρουμε ως βάση σύγκρισης την προηγούμενη χρονιά, καθορίζοντας ως βάση τις τιμές των α , β με 1. Η δεικτοποίηση δίνει τον πίνακα 3.2. Παίρνοντας τον αστάθμητο μέσο αριθμητικό έχουμε ως αποτέλεσμα $(2+0,5)/2=1,25$, άρα οι τιμές ανέβηκαν 25%. Ας πάρουμε τώρα ως βάση σύγκρισης την φετινή χρονιά (πίνακας 3.3). Παίρνοντας τον αστάθμητο μέσο αριθμητικό έχουμε ως αποτέλεσμα $(2+0,5)/2=1,25$, άρα οι περσινές τιμές είναι κατά 25% υψηλότερες των φετιμών. Ας πάρουμε πάλι ως βάση σύγκρισης την προηγούμενη χρονιά, καθορίζοντας ως βάση τις τιμές των α , β με 1. Η δεικτοποίηση δίνει τον πίνακα 3.4. Παίρνοντας όμως τον αστάθμητο μέσο γεωμετρικό, έχουμε ως αποτέλεσμα $\sqrt{2 \cdot 0,5}=1$, άρα οι τιμές έμειναν σταθερές.

4. ΔΕΙΓΜΑΤΟΛΗΨΙΑ

4.1 Το μεροληπτικό δείγμα

Η καταγραφή των τιμών μιας ή περισσοτέρων τυχαίων μεταβλητών ενός πληθυσμού είναι διαδικασία αρκετά δαπανηρή, χρονοβόρα και ενίοτε πρακτικά αδύνατη (απογραφή). Καταφεύγουμε επομένως στην διαδικασία της δειγματοληψίας, όπου σημαντικότατο ρόλο παίζει η επιλογή του δείγματος. Στο *αμερόληπτο* ή *τυχαίο δείγμα* ή *αντιπροσωπευτικό*, πρέπει όλα τα άτομα του πληθυσμού να έχουν την ίδια πιθανότητα να συμπεριληφθούν στο δείγμα.

Υπάρχουν πολλές τεχνικές δειγματοληψίας, χωρίζονται όμως σε δύο μεγάλες κατηγορίες : α) Τεχνικές δειγματοληψίας στηριγμένες στην θεωρία πιθανοτήτων, β) Τεχνικές δειγματοληψίας *μη* στηριγμένες στην θεωρία πιθανοτήτων. Στις πρώτες υπάρχει η δυνατότητα προσδιορισμού ορίων εμπιστοσύνης (confidence limits), κάθε παραμέτρου, κάτι που δεν συμβαίνει στις δεύτερες, όπου δεν είναι δυνατόν να υπολογισθούν τυπικά σφάλματα και οι οποίες χρησιμοποιούνται σε περίπτωση αδυναμίας κάλυψης του κόστους ή λόγω ελλείψεως χρόνου ή ... (Kalton 1983). Η προέκταση των αποτελεσμάτων στον συνολικό πληθυσμό στην δεύτερη περίπτωση καθίσταται προβληματική. Επομένως σε κάθε έρευνα είναι απαραίτητο να αναφέρεται η μέθοδος με την οποία έγινε η δειγματοληψία.

Το ποιος πληθυσμός πρόκειται να ερευνηθεί (target population) θέλει προσοχή, αφού η ακριβής του γνώση δεν είναι πάντα εφικτή. Ο προς εξέταση πληθυσμός προσδιορίζεται εκτός από διάφορα στοιχεία του, τις μονάδες που τον απαρτίζουν και από τον χρόνο στον οποίο αναφέρεται.

Οι «αόρατες πηγές αμεροληψίας» πρέπει να μας καθιστούν διστακτικούς απέναντι στα αποτελέσματα μιας έρευνας.

Παραδείγματα:

1. Στέκομαι σε μια πολυσύχναστη πλατεία κάποια πρωινή ώρα και ρωτάω κάθε 10^ο διασταυρούμενο με εμένα άτομο (επίφαση ή διάθεση αμεροληψίας). Αυτόματα αποκλείω ένα σημαντικό κομμάτι του

πληθυσμού της πόλεως, αυτό που δεν κυκλοφορεί έξω το πρωί ή δεν περνά καθόλου από εκείνη την πλατεία.

2. Το μέσο εισόδημα των αποφοίτων του 1924 του πανεπιστημίου του Γέηλ ήταν 25111 \$ (πολύ υψηλό), (Huff Darrel 1983). Ποιους βρήκαν και ρώτησαν; Πως τους βρήκαν; Μόνο οι επιτυχημένοι βρίσκονται ευκολότερα για ευνόητους λόγους.

3. Πείραμα: (Τζωρτζόπουλος 1991). Σε ένα τραπέζι απλώνονται 1200 βότσαλα διαφόρων μεγεθών και μια σειρά από πρόσωπα καλούνται να σχηματίσουν ένα αντιπροσωπευτικό δείγμα. Το πείραμα δείχνει ότι στο δείγμα υπερεκτιμάται το μέσο βάρος του πληθυσμού, πιθανόν για ψυχολογικούς λόγους.

4.2 Το μέγεθος του δείγματος

Το μέγεθος του δείγματος έχει σημασία για την αξιοπιστία μιας δημοσκοπήσης. Για παράδειγμα ρίχνω ένα νόμισμα 10 φορές και έρχονται 8 γράμματα. Αυτό δεν σημαίνει ότι αν ρίξω το ίδιο νόμισμα 100 φορές θα φέρει 80 γράμματα.

Το μέγεθος ενός δείγματος εξαρτάται από παράγοντες όπως το μέγεθος και η ποικιλία του πληθυσμού, καθώς και η μέθοδος της δειγματοληψίας που ακολουθούμε. Γενικά, δείγματα με μεγάλο μέγεθος δίνουν καλύτερα αποτελέσματα από εκείνα με μικρό μέγεθος. Για παράδειγμα αν από ένα άπειρο πληθυσμό διαλέξουμε ένα τυχαίο δείγμα μεγέθους την πρώτη φορά 30 και την δεύτερη 270 ο λόγος των τυπικών σφαλμάτων για την εκτίμηση της μέσης τιμής μια μεταβλητής x , είναι

$\left(\frac{\sigma}{\sqrt{270}}\right) / \left(\frac{\sigma}{\sqrt{30}}\right) = \frac{1}{3}$. Το τυπικό σφάλμα της μέσης τιμής χρησιμοποιείται για

να εκφράσει τη βεβαιότητα των N μετρήσεων μας για τη μέση τιμή του x . Επομένως αν επαναληφθούν οι N μετρήσεις του x θα υπάρχει 68% πιθανότητα η καινούρια μέση τιμή \bar{x} να βρίσκεται μέσα στο διάστημα $(\bar{x} + \sigma_{\bar{x}}, \bar{x} - \sigma_{\bar{x}})$, όπου $\sigma_{\bar{x}} = \sigma / \sqrt{N}$. Σημαίνει επίσης ότι υπάρχει 32% πιθανότητα να είναι έξω από αυτό το διάστημα. Δηλαδή στα 100 τέτοια πειράματα, κατά μέσο όρο τα 32 θα καταλήξουν σε μία τιμή έξω από τα τυπικά σφάλματα.

4.3 Λάθη κατά την δειγματοληψία

Υπάρχουν λάθη που οφείλονται στις απαντήσεις των ερωτώμενων, σκόπιμα ή μη, αλλά υπάρχουν και λάθη από την μεριά του συνεντευκτή. (Τζωρτζόπουλος 1991).

Τα σφάλματα των απαντήσεων μπορεί να: οφείλονται σε αδυναμία της μνήμης, ή να είναι σκόπιμα. Για λόγους ματαιοδοξίας, παρουσιαζόμαστε περισσότερο μορφωμένοι ή φιλόπρωτοι, αποκρύπτουμε τα εισοδήματά μας, λόγω του φόβου της Οικονομικής Εφορίας, δεν αποκαλύπτουμε μη κοινωνικά αποδεκτές συνήθειες και αντιλήψεις, κλπ.

Από την μεριά των συνεντευκτών είναι δυνατόν να συναντήσουμε λάθη που οφείλονται στην βιασύνη συμπλήρωσης του ερωτηματολογίου.

Προσοχή θέλει η διατύπωση των ερωτήσεων.

Παράδειγμα:

A) Ποιες είναι οι χώρες της Ενωμένης Ευρώπης;

B) Γνωρίζετε ποιες είναι οι χώρες της Ενωμένης Ευρώπης;

Γ) Σίγουρα γνωρίζετε ποιες είναι οι χώρες της Ενωμένης Ευρώπης.

Δ) Γνωρίζετε ποιες είναι οι χώρες της Ενωμένης Ευρώπης. Έτσι δεν είναι;

Η λέξη «σίγουρα» πιέζει τον ερωτώμενο να απαντήσει ναι, καθώς και η θετική διατύπωση «έτσι δεν είναι;».

5. ΜΕΤΡΑ ΘΕΣΗΣ

Ως γνωστό τα σημαντικότερα μέτρα θέσης είναι η Μέση Τιμή, η Διάμεσος και η Επικρατούσα Τιμή. Η χρήση, μόνο του καθενός από αυτά δεν δίνει πάντα πολλές πληροφορίες, αντίθετα είναι πιθανόν να συσκοτίζει. Για παράδειγμα η διάμεσος είναι το καλλίτερο μέτρο θέσεως για ασύμμετρες κατανομές (οικογενειακό εισόδημα, κατά κεφαλήν εισόδημα κλπ) ή όταν υπάρχουν τιμές σε μεγάλη απόσταση από τον κύριο όγκο των δεδομένων (έκτροπες), ενώ η μέση τιμή αποτυγχάνει να δώσει σωστές πληροφορίες.

Παράδειγμα: Μια μέση οικογένεια αποτελείται από 3,6 άτομα. Αυτό σημαίνει 3 ή 4 άτομα. Οι κατασκευαστές οικιών φτιάχνουν κυρίως κατοικίες με 2 ή 3 κρεβατοκάμαρες. Όμως η κατηγορία με 3 ή 4 άτομα είναι η μικρότερη αριθμητικά.

6. ΟΙ ΑΡΙΘΜΟΙ ΠΟΥ ΛΕΙΠΟΥΝ

6.1 Επίπεδο σημαντικότητας

Όταν κάποιοι αριθμοί λείπουν από μια στατιστική έρευνα, εξαπατούν ή και ταλαιπωρούν τον κόσμο.

Παράδειγμα: Πριν 5 χρόνια βρέθηκε ότι το μέσο βάρος των μαθητών της Γ' Γυμνασίου της πόλης Α ακολουθούσε την κανονική κατανομή με μέσο βάρος 70,4 kg και διακύμανση 7,29. Σε ένα τυχαίο δείγμα φέτος βρέθηκε μέσο βάρος 70,8. Μπορούμε να ισχυριστούμε ότι το μέσο βάρος των μαθητών της Γ' τάξης αυξήθηκε; Πρέπει να απαντήσουμε αναφέροντας και την πιθανότητα σφάλματος (επίπεδο σημαντικότητας). Π.χ αποδεικνύεται ότι αν το μέσο βάρος ενός δείγματος είναι μεγαλύτερο ή ίσο του 71,18, τότε το μέσο βάρος των μαθητών έχει αυξηθεί με πιθανότητα σφάλματος 2%. (Κιόχος 1998).

Λόγω της στατιστικής αβεβαιότητας, δεν μπορούμε να αποδείξουμε μια υπόθεση βασισμένη σε ένα δείγμα. Εξάγουμε συμπεράσματα που συνοδεύονται από μια πιθανότητα λάθους, η οποία πρέπει να αναφέρεται. (Εκτός και το δείγμα μας ταυτιστεί με τον πληθυσμό).

6.2 Μέτρα Διασποράς

Η διακύμανση μας καθορίζει την απόκλιση των τιμών από τον μέσο όρο, ο οποίος μόνος του, ενίοτε δεν λέει πολλά πράγματα και ίσως είναι άχρηστος.

Παράδειγμα 1^ο: Σύμφωνα με έρευνες ένα μικρό παιδί αρχίζει να στέκεται όρθιο στην ηλικία των x μηνών. Αν δεν συμβεί αυτό για το παιδί του ένας γονιός αρχίζει να ανησυχεί. Όμως λόγω του κεντρικού

οριακού θεωρήματος, ο πραγματικός μέσος θα βρίσκεται στο διάστημα x συν πλην δύο τυπικές αποκλίσεις του δείγματος, με πιθανότητα 95%.

Παράδειγμα 2^ο: Η περιοχή Α έχει μέση ετήσια θερμοκρασία 20^ο C. Θαυμάσια μέρος για μόνιμη διαμονή συνταξιούχων, αν η συγκεκριμένη περιοχή δεν είχε εύρος τιμών από -20^ο C έως 48^ο C.

7. ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΑΙΤΙΟΤΗΤΑ

Συσχέτιση δεν σημαίνει και αιτιότητα. Είναι εύκολο να δείξει κανείς ότι υπάρχει συσχέτιση σε ένα σωρό ζευγάρια καταστάσεων. Το να χαρακτηρίζει ένα από αυτά σαν αιτία του άλλου είναι ενίοτε ανόητο.

Παράδειγμα 1^ο: Οι φωλιές των πελαργών στην στέγη και ο αριθμός παιδιών σε μια δανέζικη οικογένεια σχετίζονται θετικά. Οι δύο αυτές καταστάσεις είναι αποτελέσματα των περισσότερων δωματίων (άρα και καμινάδων σε μια πολυμελή οικογένεια).

Παράδειγμα 2^ο: Έχει επισημανθεί ότι ο μισθός των καθηγητών και η τιμή του καφέ έχουν μια θετική συσχέτιση. Ποιο είναι το αίτιο και ποιο το αποτέλεσμα; Η αύξηση της τιμής του καφέ θα επιφέρει και αυξήσεις στους μισθούς μας ή θα πρέπει να κρατήσουμε σταθερούς τους μισθούς των καθηγητών για να απολαμβάνουμε φτηνό καφέ;

8. ΠΑΡΕΡΜΗΝΕΙΑ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Η παρερμηνεία των στατιστικών πληροφοριών είναι ένα συνηθισμένο πρόβλημα. Μπορεί να είναι αποτέλεσμα πολλών παραγόντων. Οι σημαντικότεροι είναι:

- Παρανόηση των στοιχείων
- Χρήση εννοιών κατά λανθασμένο ή διαφορετικό τρόπο.
- Σκόπιμη παρερμηνεία της πληροφορίας

Κάθε μέρα βομβαρδιζόμαστε από αριθμούς στα ΜΜΕ. Αν και οι πληροφορίες μπορεί να είναι αξιόπιστες, έχει ενδιαφέρον να παρατηρήσουμε πως για παράδειγμα οι εφημερίδες, αναλύουν με τον δικό τους ανορθόδοξο τρόπο τα δεδομένα.

Το περιοδικό «The Daily Statistics Canada» είναι το επίσημο περιοδικό της στατιστικής υπηρεσίας του Καναδά το οποίο αναρτάται και στο διαδίκτυο (<http://www.statcan.ca>).



Σχήμα 8.1. Το περιοδικό The Daily Statistics Canada ανεβασμένο στο διαδίκτυο.

Στο τεύχος της 18^{ης} Ιουλίου 2000 (σχ. 8.1), υπάρχει μια στατιστική αναφορά στην εγκληματικότητα αυτής της χώρας. Αξίζει να παρατηρήσει κανείς πως η εικόνα που δίνει το περιοδικό μπορεί να υπεραπλουστευθεί και να παρανοηθεί από τα ΜΜΕ. Το περιοδικό σε υπότιτλους αναφέρει (<http://www.statcan.ca/english/κλπ>):

- Μείωση των εγκλημάτων βίας σε όλες τις κατηγορίες.
- Μεγάλη μείωση των διαζυγίων.
- Αύξηση των παραβάσεων που σχετίζονται με ναρκωτικά, στασιμότητα στις οδικές παραβάσεις.

Βασισμένες στις παραπάνω πληροφορίες κάποιες εφημερίδες με μεγάλους τίτλους ανακοινώνουν:

- Η αυστηρή οικονομική πολιτική σταματά την αύξηση της εγκληματικότητας.
- Ο μειούμενος ρυθμός των ανθρωποκτονιών επηρεάζει την προεκλογική κίνηση.
- Στα χαμηλότερα επίπεδα στα 32 τελευταία χρόνια οι ανθρωποκτονίες.

Εστιάζοντας μονόπλευρα στα δεδομένα, οι εφημερίδες αγνόησαν κάτι σημαντικό, ότι η εγκληματικότητα μειώνεται για συνεχή όγδοη χρονιά. Τα δεδομένα αποκαλύπτουν ότι τα εγκλήματα βίας μειώθηκαν κατά 2,4% μέσα στο 1999, πρόκειται για την 7^η συνεχή μείωση, μετά από 15 χρόνια αυξήσεων. Όλες οι κύριες κατηγορίες των εγκλημάτων βίας εμφανίζουν μείωση, όπως οι ανθρωποκτονίες (-4,7%), απόπειρες φόνων (-8,8%), βιαιοπραγίες (-2,0%), σεξουαλικές βιαιοπραγίες (-7,3%) και κλοπές (-1,5%).

Αν και είναι δυνατόν να εικάσουμε ότι η αυξανόμενη ευημερία, έχει συνεισφέρει στο ρυθμό μείωσης της παραβατικότητας, τα δεδομένα που παρείχε η Στατιστική Υπηρεσία του Καναδά, δεν συσχετιζαν την οικονομία με την εγκληματικότητα. Και αν συνέβαινε κάτι τέτοιο, η παρατήρηση θα ήταν ότι η μείωση της εγκληματικότητας άρχισε το 1991, χρονιά που άρχισε η ύφεση στον Καναδά. Η Daily, πράγματι ανέφερε αρκετούς παράγοντες που συνεισφέρουν στην μεταβολή της εγκληματικότητας. Κάποιοι δημοσιογράφοι το εξέλαβαν ως γεγονός. Οι συνδέσεις που έγιναν στα πρωτοσέλιδα ήταν καθαρά εικασίες και δεν προέκυπταν από γεγονότα. Οι δημοσιογράφοι παρερμήνευσαν τα δεδομένα πιθανόν, γιατί παρανόησαν τις αιτίες που βρίσκονται πίσω από τις εγκληματικές πράξεις. Οι εργαζόμενοι στην Στατιστική Υπηρεσία καθημερινά προσπαθούσαν να ανασκευάσουν τις ανακρίβειες των δημοσιογράφων.

Είναι σημαντικό να καταλαβαίνουμε τους στατιστικούς ορισμούς και έννοιες που βρίσκονται πίσω από την πληροφορία που χρησιμοποιούμε. Αν για παράδειγμα εξετάζουμε εργατικά θέματα, πρέπει να μας είναι οικείοι οι όροι όπως ανεργία, απασχόληση, μερική απασχόληση κλπ.

Ένα μεγάλο πλεονέκτημα της στατιστικής είναι ότι μπορεί να κάνει συγκρίσεις, ανακαλύπτοντας τάσεις και χαρακτηριστικά του υπό εξέτασιν πληθυσμού. Όμως προβλήματα εγείρονται στην σύγκριση

δεδομένων όταν οι ορισμοί, η ταξινόμηση και η μέθοδος συλλογής των δεδομένων είναι διαφορετικοί. Ιδιαίτερα για δεδομένα από διαφορετικές πηγές.

Το ακόλουθο παράδειγμα είναι από μετρήσεις κοινωνικού περιεχομένου και αφορά τον Καναδά.

Table 1. Married people by sex, 1998 to 2002

Married	1998	1999	2000	2001	2002
Number of persons					
Both sexes	14,630,173	14,711,793	14,806,694	14,913,766	15,018,130
Males	7,299,132	7,337,226	7,381,266	7,431,522	7,476,537
Females	7,331,041	7,374,567	7,425,428	7,482,244	7,541,593

Source: CANSIM table 051-0010

Πίνακας 8.2. Παντρεμένα άτομα κατά φύλο από το 1998 έως το 2002.

Σύμφωνα με την Υπηρεσία Στατιστικής του Καναδά, ο όρος «παντρεμένος» περιέχει εκείνους που έχουν κάνει νόμιμο γάμο και ζουν μαζί, εκείνους που μετά από ένα νόμιμο γάμο χώρισαν και αυτούς που ζουν με σύμφωνο ελεύθερης συμβίωσης. Σε άλλη έρευνα αν στον ορισμό του παντρεμένου δεν περιέχεται η τελευταία κατηγορία τα αποτελέσματα θα είναι διαφορετικά. Μια άλλη σημαντική παρατήρηση είναι ότι ο αριθμός των παντρεμένων γυναικών είναι ελαφρά μεγαλύτερος, ενώ θα ανέμενε κανείς ισότητα στις δύο κατηγορίες. Εκτός των άλλων φαίνεται ότι ο όρος παντρεμένος έχει διαφορετικό περιεχόμενο σε κάποιους ανθρώπους!

9. ΈΛΕΓΧΟΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΕΚΤΙΜΗΣΕΩΝ ΚΑΙ ΕΡΕΥΝΩΝ

« Η τακτική των πολιτικών δεν αναζητά συστηματικά την αλήθεια η ακόμη την απλή ακρίβεια. Ψάχνουν να βρουν πολεμοφόδια για να τα χρησιμοποιήσουν στον πόλεμο της πληροφορίας. Δεδομένα, πληροφορίες και γνώση δεν είναι ανάγκη να είναι αληθινά, προκειμένου να εξουδετερωθεί ένας αντίπαλος». Alvin Toffler⁴

Ο Toffler μπορεί να φαίνεται αρκετά κυνικός, αλλά στην πραγματικότητα, άνθρωποι και οργανισμοί «μαγειρεύουν» τις πληροφορίες σύμφωνα με το συμφέρον τους. Γι' αυτό τον λόγο πρέπει να είμαστε πάντα κριτικοί στις πληροφορίες που μας παρέχονται. Η τεκμηριωμένη πληροφορία πρέπει να είναι ακέραια, να έχει αντικειμενικότητα και ακρίβεια. Είναι όμως πιθανόν να παρερμηνευθεί από έλλειψη προσωπικής αμεροληψίας, μη ακριβή δεδομένα και ακόμη από την προσθήκη κατασκευασμένων στοιχείων.

Μελετώντας τα αποτελέσματα μιας έρευνας πρέπει να μας απασχολούν τα παρακάτω ερωτήματα:

α) Ποιος το λέει αυτό;

Βεβαιωθείτε από που προέρχεται η πληροφορία και εξετάστε αν η πηγή είναι αξιόπιστη. Έχει σημασία η αυθεντία να βρίσκεται πίσω από την πληροφορία και όχι δίπλα σε αυτή. Έχει άλλη βαρύτητα το ότι η έρευνα έγινε για παράδειγμα στο πανεπιστήμιο του Stanford και άλλη ότι την διενήργησε το ίδιο το πανεπιστήμιο.

Δεδομένα από διαφορετικές πηγές δεν είναι πάντοτε συγκρίσιμα.

β) Πως το ξέρει;

Το δείγμα ήταν τυχαίο; Ποιες μέθοδοι χρησιμοποιήθηκαν για την συλλογή των στοιχείων, καθώς και για την επεξεργασία τους. Τέλος είναι σημαντικό να γνωρίζετε πόσο ακριβείς είναι οι δειγματοληπτικές έρευνες, αφού υπόκεινται σε δύο είδη λαθών, τα δειγματοληπτικά σφάλματα, λόγω του ότι χρησιμοποιείται δείγμα αντί του πληθυσμού και σε σφάλματα που οφείλονται στον συνεντεύκτη ή και του

⁴ Alvin Toffler (1928-) Αμερικανός συγγραφέας και μελλοντολόγος.

ερωτώμενου. Τζωρτζόπουλος (1991). Τα τελευταία είναι παρόντα είτε σε απογραφές είτε σε δειγματοληπτικές έρευνες (βλ. §4).

γ) Τι λείπει;

Π.χ οι ανεπεξέργαστοι αριθμοί. Έτσι συχνά δίνονται τα ποσοστά και όχι οι ανεπεξέργαστοι αριθμοί. Όταν το πανεπιστήμιο John Hopkins άρχισε να δέχεται φοιτήτριες, κάποιος ανέφερε ότι το 33,3% των φοιτητριών είχαν παντρευτεί καθηγητές του πανεπιστημίου. Τα απόλυτα νούμερα και οι πληροφορίες έλεγαν ότι ήταν μόνο τρεις φοιτήτριες και η μία ήταν ήδη παντρεμένη με καθηγητή. (Huff 1983 Κεφ.10).

Η επικαιρότητα των στοιχείων είναι απαραίτητη. Πρέπει λοιπόν να αναφέρεται η χρονολογία συλλογής των στοιχείων.

δ) Μήπως άλλαξε κάποιος το θέμα;

Σε μια έρευνα διαπιστώθηκε πως οι Βρετανοί κάνουν μπάνιο συχνότερα από τις γυναίκες τους. Με δεδομένη την στάση των Βρετανών σε θέματα σωματικής καθαριότητας, είναι πιθανόν οι Βρετανοί να ισχυρίζονται ότι κάνουν μπάνιο συχνότερα από τις γυναίκες τους! (Huff 1983 Κεφ.10).

ε) Έχει νόημα;

Π.χ η εξέλιξη ενός πληθυσμού μιας χώρας. Υπάρχει μια τάση, θα διατηρηθεί και στο μέλλον;

10 ΣΥΜΠΕΡΑΣΜΑ

Από τα ανωτέρω είναι φανερό ότι η παραπλάνηση ηθελημένη ή μη πλήττει τον ανυποψίαστο μη ενημερωμένο ή βιαστικό αναγνώστη, ο οποίος εύκολα πέφτει θύμα των «παραπλανητικών» διαγραμμάτων και γενικότερα των στατιστικών δεδομένων.

Η λανθασμένη χρήση στατιστικών δεδομένων είναι δυνατόν να έχει απρόβλεπτες συνέπειες. Η χρήση της μέσης τιμής του ΑΕΠ της Στερεάς Ελλάδος με τον συνυπολογισμό των πλουσίων βιομηχανιών, κυρίως της περιοχής των Οиноφύτων, αποστέρεισε την συγκεκριμένη περιφέρεια από αναγκαίους κοινοτικούς πόρους (βλέπε <http://viotia.isol-servers.org/portal/> κλπ, την διαμαρτυρία του νομάρχη Βοιωτίας).

Η ενημέρωση του κόσμου γύρω από θέματα στατιστικής, καθιστά και τις εταιρείες δημοσκοπήσεων προσεκτικότερες. Όλο και συχνότερα εκτός των αποτελεσμάτων των δημοσκοπήσεων, αναφέρουν και το στατιστικό σφάλμα. Ιστολόγια ή ιστότοποι που διενεργούν δημοσκοπήσεις αναφέρουν (ορθά), ότι δεν έχουν επιστημονικό χαρακτήρα, αφού το δείγμα δεν είναι τυχαίο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξένη

Huff Darrel. (1983). How to lie with statistics. W.W. Norton & Company, Inc N.Y 10110

<http://jdasovic.com/2008/01/lies-damned-lies-and-excel-charts/>

<http://www.statcan.ca>

<http://www.statcan.ca/english/edu/power/ch6/misinterpretation/misinterpretation.htm>

Kalton, G., (1983). Introduction to Survey Sampling. Sage, Beverly Hills.

Ελληνική

Κιόχος Π., (1998), Επαγωγική Στατιστική, Εκδόσεις INTERBOOKS

Τζωρτζόπουλος, Π., (1991). Οργάνωση και Διεξαγωγή Δειγματοληπτικών Ερευνών, Αθήνα, Ο.Π.Α.

Τζωρτζόπουλος, Π., Λειβαδά Α. (2003). Αριθμοδείκτες: Θεωρία και Εφαρμογές, Αθήνα, Ο.Π.Α.

[http://viotia.isol-](http://viotia.isol-servers.org/portal/index.php?option=com_content&task=view&id=24&Itemid=46)

[servers.org/portal/index.php?option=com_content&task=view&id=24&Itemid=46](http://viotia.isol-servers.org/portal/index.php?option=com_content&task=view&id=24&Itemid=46) Επίσημος ιστοχώρος της νομαρχίας Βοιωτίας.