

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΕΛΕΓΧΟΣ ΣΥΜΦΩΝΙΑΣ ΜΕΤΡΗΣΕΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΤΗΝ ΙΑΤΡΙΚΗ

Μιχαήλ Χαραλάμπους Ταλούμης

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος
Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Μάιος 2014

ΑΦΙΕΡΩΣΗ

Στην οικογένεια μου, τους καθηγητές μου και τον κ.Βασδέκη.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ τον κ.Βασδέκη για την ανάθεση της εργασίας, την βιήθεια και υπομονή του και για την απλότητα με την οποία επέλυσε κάθε απορία μου. Ευχαριστώ και την οικογένεια μου για τη στήριξη και όσους φιλους μου με βοήθησαν στην εκπόνηση της εργασίας.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Γεννήθηκα στη αθήνα το 1988, μεγάλωσα στο Ίλιον και αποφοίτησα από το 3^{ο} ενιαίο λύκειο Ιλίου το 2005. Το σεπτέμβρη του 2005 έλαβα μλερος στο προπτυχιακό πρόγραμμα Στατιστική, Αναλογιστικά και Χρηματοοικονομικά μαθηματικά στο πανεπιστήμιο Αιγαίου στη Σάμο και αποφοίτησα το 2010. Τον Οκτώβριο του 2010 έλαβα μέρος στο μεταπτυχιακό πρόγραμμα Στατιστικές Μέθοσοι στην Ιατρική και Φαρμακευτική.

ABSTRACT

Michael Taloumis

TWO DIFFERENT METHODS MESAUREMENTS AGREEMENT TESTS IN MEDICAL

May 2014

How two different cilinical measurement methods agree with each other, can be quantified using the differences between the observations made by each method on the same subject. The 95% limits of agreement estimated by the mean diffenrence ±1.96 tinew the standars deviation of the measurements, give us an interval within which are expected to be 95 % of the differences between the measurements of two methods. We'll see how we can use graphical methods to investigate methods and will also give confidence intervals . We extend the basic approach to data where there is a relationship between the difference and the size of the measurements with a simple logarithmic transformation approach and a new, more general regression approach. We discuss the importance of the repeatability of each method separately and compare an estimate of this to the limits of agreement. We extend the limits of agreement approach to data with repeated measurements, proposing new estimates for equal numbers of replicates by each method on each subject, for unequal numbers of replicates, and for replicated data collected in pairs, where the underlying value of the quantity being measured is changing. We describe a simple nonparametric approach to comparing methods. In our analysis than numerical tests, in many cases we will also have some visual tests with charts. Finally, we will see how we can generate reference centiles, using the absolute residual variance of regression of the data, seperated into groups.

ΠΕΡΙΛΗΨΗ

Μιχαήλ Ταλούμης

ΕΛΕΓΧΟΣ ΣΥΜΦΩΝΙΑΣ ΜΕΤΡΗΣΕΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΤΗΝ ΙΑΤΡΙΚΗ

Μάιος 2014

Το πόσο συμφωνούν δύο μέθοδοι κλινικών μετρήσεων, μπορεί να προσδιοριστεί ποσοτικά, χρησιμοποιώντας τις διαφορές μεταξύ των παρατηρήσεων που έγιναν με τη χρήση των δύο μεθόδων για τα ίδια υποκείμενα. Τα 95% όρια συμφωνίας, εκτιμώμενα από τη μέση διαφορά ± 1.96 επί την τυπική απόκλιση των διαφορών, μας δίνουν ένα διάστημα, εντός του οποίου αναμένονται να βρίσκονται το 95% των διαφορών μεταξύ των μετρήσεων των δύο μεθόδων. Θα δούμε πως μπορούμε να χρησιμοποιήσουμε γραφικές μεθόδους για να διερευνήσουμε τις μεθόδους και θα δώσουμε επίσης διαστήματα εμπιστοσύνης. Επεκτείνουμε τη βασική προσέγγιση στα δεδομένα, όπου υπάρχει μια σχέση μεταξύ της διαφοράς και του μεγέθους των μετρήσεων, με μια απλή λογαριθμική προσέγγιση μετασχηματισμού και με μια νέα, πιο γενική προσέγγιση παλινδρόμησης. Θα συζητήσουμε τη σημασία της επαναληψιμότητας της κάθε μεθόδου ξεχωριστά και θα συγκρίνουμε μια εκτίμηση αυτής στα όρια της συμφωνίας. Έχουμε επεκτείνει τα όρια της συμφωνίας στα δεδομένα με επαναλαμβανόμενες μετρήσεις, προτείνοντας νέες εκτιμήσεις για ίσο αριθμό των επαναλήψεων με κάθε μέθοδο σε κάθε υποκείμενο, για άνισο αριθμό επαναλήψεων, και για αναπαραγωγή δεδομένων που συλλέγονται σε ζεύγη, όπου η ποσότητα των μετρήσεων αλλάζει. Στην ανάλυση μας εκτός από αριθμητικούς ελέγχους, σε πολλά κομμάτια θα δούμε και οπτικούς ελέγχους με τα εκάστοτε διαγράμματα. Ακόμα, θα περιγράψουμε μια μη παραμετρική προσέγγιση για σύγκριση μεθόδων, με αρκετά καλά αποτελέσματα και πολύ απλή στη χρήση της. Τέλος, θα δούμε πως μπορούν να δημιουργηθούν ποσοστιαίες θέσεις αναφοράς, χρησιμοποιώντας τα απόλυτα κατάλοιπα της παλινδρόμησης της διακύμανσης όταν χωρίζουμε τα δεδομένα σε ομάδες ως προς μία μεταβλητή και κάνουμε παλινδρόμηση πάνω σε αυτή.

Περιεχόμενα.....

Εισαγωγή-Συμφωνία μετρήσεων σε μελέτες σύγκρισης μεθόδων 1

1. Συγκρίνοντας δύο μεθόδους..... 3

2. Διάστημα συφωνίας 7

2.1 Γραφική αναπαράσταση της συμφωνίας των μεθόδων 9

2.2 Ακρίβεια της εκτίμησης των ορίων της συμφωνίας 13

3. Σχέση μεταξύ μεγέθους και διαφοράς 15

3.1 Λογαρίθμηση 18

3.2 Μέθοδος μελέτης με παλινδρόμηση για μη ομοιόμορφες διαφορές..... 21

4. Σπουδαιότητα της επανάληψης 29

4.1 Προβλέποντας την επανάληψη 30

5. Μετρώντας την συμφωνία στις επαναλαμβανόμενες μετρήσεις 33

5.1 Ίσος αριθμος επαναλήψεων..... 33

5.2 Άνισος αριθμός επαναλήψεων..... 38

5.3 Επαναλαμβανόμενα δεδομένα σε ζεύγη..... 41

6. Μη παραμετρική προσέγγιση για τη σύγκριση μεθόδων 43

7. Κατασκευή ποσοστιαίων θέσεων αναφοράς, σχετικά με την ηλικία, χρησιμοποιώντας απόλυτα κατάλοιπα 46

8.Βιβλιογραφία.....55

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Μετρήσεις συστολικής πίεσης από 2 ιατρούς(j και r) και από ένα αυτόματο μηχάνημα μέτρησης πίεσης Altman and Bland[1])....2
Πίνακας 2. Paired Sample t-test για το ζεύγος J1-S1.....8
Πίνακας 3 Τεστ κανονικότητας των διαφορών J1-S2.....8
Πίνακας 4 Paired Sample t-test για το ζεύγος J1-S1 έχοντας βγάλει τις 2 ακραίες τιμές.....9
Πίνακας 5 Μη παραμετρικός έλεγχος Spearman's test .....**Error! Bookmark not defined.**1
Πίνακας 6 Περιγραφικά στατιστικά.....13
Πινακας 7. Μετρήσεις όγκου πλάσματος, εκφρασμένο σε ποσοστό της αναμενόμενης τιμής, χρησιμοποιώντας 2 σετ δεδομένων κανονικών τιμών Nadler-Hurley (C.Dore, βλ.Cotes et al[11]). .....**Error! Bookmark not defined.**5
Πίνακας 8.Εκτιμώμενο λίπος στο ανθρώπινο γάλα μετρημένο μέσω των τριγλυκεριδίων και μέσω της κλασικής μεθόδου Gerber[4].....23
Πίνακας 9 Συντελεστές της παλινδρόμησης Gerber-τριγλυκεριδια 25
Πίνακας 10 Όγκος αίματος $(dm^3 / \min)$ που φεύγει από την καρδιά ανά λεπτό. Μετρήσεις με καρδιογράφημα εμπέδησης IC και με ραδιο-κοιλιογραφία RV (δεδομένα απο Dr. LS Bowling[5]).....39
Πίνακας 11 Βλέπουμε τα όρια για να βαθμολογηθεί ένα πιεσόμετρο με A,B,C,D. Οι διαφορές είναι συγκριτικά με ένα σφυγμομανόμετρο[8].....44
Πίνακας 12 Μετρήσεις για το μέγεθος του ποδιού του εμβρύου σε ομάδες μίας εβδομάδας(α) και σε 6 ίσες ομάδες(β).....50

ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

Διάγραμμα 1 Μέτρηση πίεσης από αυτόματο πιεσόμετρο-μέτρηση από γιατρό 10
Διάγραμμα 2 *Διαφορά μέτρησης πίεσης αυτόματου πιεσόμετρου- μέτρησης γιατρού με τη μέση τιμή τους*.....**Error! Bookmark not defined.**1
Διάγραμμα *3 Διαφορά μέτρησης πίεσης αυτόματου πιεσόμετρου- μέτρησης γιατρού με τη μέση τιμή τους (outliers)*.....**Error! Bookmark not defined.**2
Διάγραμμα 4 *Nadler-Hurley*.....**Error! Bookmark not defined.**7
Διάγραμμα 5 *Nadler-Hurley με το μέσο τους*....**Error! Bookmark not defined.**8
Διάγραμμα 6 *Log(Nadler) – Log(Hurley)* .....**Error! Bookmark not defined.**9
Διάγραμμα 7 *Log(Nadler)-Log(Hurley) με το μέσο τους* ..**Error! Bookmark not defined.**9
Διάγραμμα 8 *Nadler/Hurley με το μέσο ποσοστό πλάσματος*.....20
Διάγραμμα 9 *g/100ml (τριγλυκερίδια- Gerber) με το μέσο (τριγλυκερίδια-Gerber)*.....**Error! Bookmark not defined.**4
Διάγραμμα 10 *g/100ml τριγλυκερίδια g/100ml Gerber)*...**Error! Bookmark not defined.**4
Διάγραμμα 11 *Περιεκτικότητα του ανθρώπινου γαλακτος σε λίπος μετα από ενζυμική διαδικασία για τον προσδιορισμο των τριγλυκεριδίων και μετρημένο με τη μέθοδο Gerber* .....**Error! Bookmark not defined.**6
Διάγραμμα 12 *Διάγραμμα διαφοράς των 2 μεθόδων μέτρησης με τον μέσο τους.* .....**Error! Bookmark not defined.**7
Διάγραμμα 13 *Παρουσιάζονται τα διαγράμματα μεταξύ τυπικών αποκλίσεων και μέσων των υποκειμένων για τις δύο μεθόδους (RV, IC)*.....40
Διάγραμμα 14 *Η διαφορά RV-IC προς το μέσο (RV,IC)*.40**Error! Bookmark not defined.**
Διάγραμμα 15 *Από τα δεδομένα του πίνακα 1, βλέπουμε τις διαφορές του σφυγμομανόμετρου με το αυτόματο πιεσόμετρο. Οι γραμμές απεικονίζουν το διάστημα (-15,15).*.....44
Διάγραμμα 16 *Τυπική απόκλιση μήκους ποδιού ως προς το χρόνο κύησης, με ομαδοποίηση ανα εβδομάδα.* 51
Διάγραμμα 17 *Τυπική απόκλιση μήκους ποδιού ως προς το χρόνο κύησης, με ομαδοποίηση σε έξι ομάδες.* 52

Εισαγωγή

Συμφωνία μετρήσεων σε μελέτες σύγκρισης μεθόδων

Σε οποιαδήποτε μέτρηση γίνεται στη φύση, πάντα υπάρχει απόκλιση από την πραγματική τιμή. Σε κάποιες περιπτώσεις αυτό μας επηρεάζει στην απόφαση που πρέπει να πάρουμε και σε κάποιες όχι. Δηλαδή άλλοτε είναι σημαντικό να προσεγγίσουμε οσο το δυνατόν περισσότερο την πραγματική τιμή και άλλοτε οχι.

Στις κλινικές μελέτες, πολλές φορές, υπάρχει πρόβλημα με την ακρίβεια και αξιοπιστία των μετρήσεων που γίνονται, λόγω διαφορετικών οργάνων μέτρησης, διαφορετικών συνθηκών κατα τη διάρκεια των μετρήσεων, διαφορετικών μεθόδων μέτρησης, πολυπλοκότητα μέτρησης, ανθρώπινων λαθών κλπ. Αυτό είνα ένα πρόβλημα που μπορεί να μας επηρεάσει στη μελέτη κάποιου φαινομένου και να μας οδηγήσει σε λάθος συμπεράσματα. Είναι εύκολα κατανοητό, ότι ένα τέτοιο λάθος, σε ένα κλάδο, όπως είναι η ιατρική, μπορεί να έχει πολύ δυσάρεστα αποτελέσμτα, με αντίκτυπο σε μεγάλο πλήθος ανθρώπων. Οπότε το να προσεγγίζουμε όσο το δυνατόν περισσότερο την ακριβή πραγματική τιμή, είναι «ζήτημα ζωής και θανάτου». Για την αποφυγή και ελαχιστοποίηση αυτών των αποκλίσεων των τιμών των μετρήσεων, έχουν βρεθεί διάφοροι τρόποι, για να εκτιμούμε την πραγματική τιμή της μέτρησης που κάνουμε και ως αποτέλεσμα να έχουμε πιο αξιόπιστα και ισχυρά αποτελέσματα.

Οι μέθοδοι αυτές, είναι αποτελέσματα πολλών ετών έρευνας και πολλών πειραμάτων και υπολογισμών, αλλά συνεχώς αναπτύσσονται νέες μέθοδοι, με σκοπό την ευκολότερη, την γρηγορότερη, την πιο αξιόπιστη και την πιο αποτελεσματική μέθοδο.

1. Συγκρίνοντας δύο μεθόδους

Στην κλινική ιατρική, πολύ συχνά, μας ενδιαφέρει να κάνουμε μετρήσεις που αφορούν το σώμα μας, όπως τον όγκο του αίματος σε κάθε παλμό της καρδιάς και την πίεση του αίματος. Αυτό είναι πρακτικά αδύνατο να μετρηθεί με ακρίβεια και γι αυτό οι πραγματικές τιμές είναι άγνωστες. Όμως έχουμε κάποιες μεθόδους μέτρησης και με αυτές συγκρίνουμε κάθε μια καινούρια μέθοδο που παρουσιάζεται. Δεν ξέρουμε όμως ποια μέθοδος είναι πιο κοντά στην πραγματική τιμή που μετράμε και έτσι προσπαθούμε να βρούμε ένα σημείο συμφωνίας μεταξύ των δυο μεθόδων. Η τυπική μέθοδος ονομάζεται «gold standard» αλλά αυτό δεν σημαίνει πως με αυτή τη μέθοδο δεν θα έχουμε σφάλματα.

Η ασυμφωνία μεταξύ των μεθόδων είναι αναπόφευκτη. Αυτό που μας ενδιαφέρει είναι το πόσο διαφέρουν μεταξύ τους. Θέλουμε να μάθουμε κατά πόσο διαφέρει η νέα μέθοδος με την παλιά και αν η διαφορά τους δεν προκαλεί πρόβλημα τότε η παλιά μπορεί να αντικατασταθεί από την καινούρια μέθοδο ή μπορούν να χρησιμοποιούνται και οι 2. Για παράδειγμα αν μελετήσουμε την διαφορά ενός υποθετικού νέου μηχανήματος μέτρησης πίεσης και η διαφορά του από ένα παλιό πιεσόμετρο είναι το πολύ 10mmHg, θεωρούμε ότι δεν δημιουργείται πρόβλημα και το κρατάμε. Αν η διαφορά είναι 30mmHg και πάνω, θα δημιουργηθεί πρόβλημα στη διάγνωση και θεραπεία ενός ασθενή και έτσι δεν θεωρείται το μηχάνημα αξιόπιστο. Τα όρια που θα μας κάνουν να απορρίπτουμε ή να αποδεχόμαστε ένα καινούριο μηχάνημα, εξαρτάται από το που θα χρησιμοποιηθεί αυτή η μέτρηση και είναι και στην κρίση των ιατρών.

Στο παρακάτω παράδειγμα θα αναλύσουμε τη σύγκριση 2 μεθόδων

Πίνακας 1. *Μετρήσεις συστολικής πίεσης από 2 ιατρούς(j και r) και από ένα αυτόματο μηχάνημα μέτρησης πίεσης Altman and Bland1)*

Subject	j1	j2	j3	r1	r2	r3	s1	s2	s3
1	100	106	107	98	98	111	122	128	124
2	108	110	108	108	112	110	121	127	128
3	76	84	82	76	88	82	95	94	98
4	108	104	104	110	100	106	127	127	135
5	124	112	112	128	112	114	140	131	124
6	122	140	124	124	140	126	139	142	136

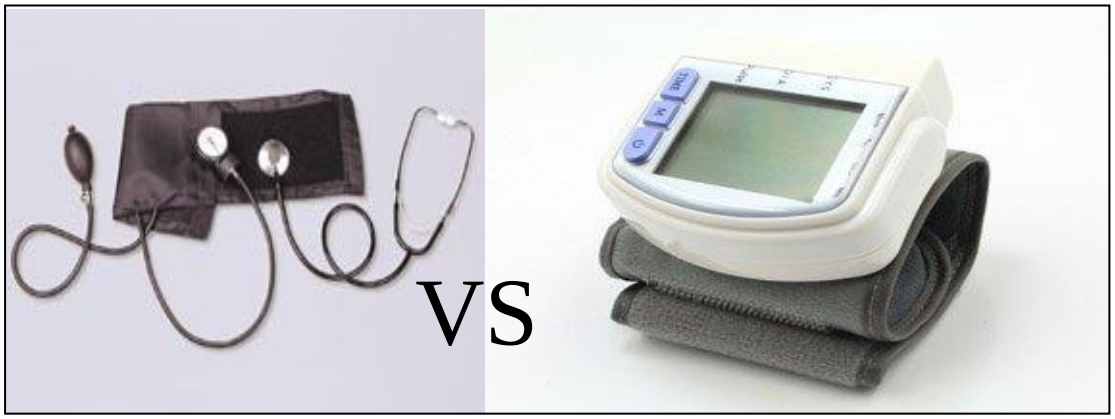
7	116	108	102	118	110	102	122	112	112
8	114	110	112	112	108	112	130	129	135
9	100	108	112	100	106	112	119	122	122
10	108	92	100	108	98	100	126	113	111
11	100	106	104	102	108	106	107	113	111
12	108	112	122	108	116	120	123	125	125
13	112	112	110	114	112	110	131	129	122
14	104	108	104	104	108	104	123	126	114
15	106	108	102	104	106	102	127	119	126
16	122	122	114	118	122	114	142	133	137
17	100	102	102	102	102	100	104	116	115
18	118	118	120	116	118	118	117	113	112
19	140	134	138	138	136	134	139	127	113
20	150	148	144	148	146	144	143	155	133
21	166	154	154	164	154	148	181	170	166
22	148	156	134	136	154	132	149	156	140
23	174	172	166	170	170	164	173	170	154
24	174	166	150	174	166	154	160	155	170
25	140	144	144	140	144	144	158	152	154
26	128	134	130	128	134	130	139	144	141
27	146	138	140	146	138	138	153	150	154
28	146	152	148	146	152	148	138	144	131
29	220	218	220	220	218	220	228	228	226
30	208	200	192	204	200	190	190	183	184
31	94	84	86	94	84	88	103	99	106
32	114	124	116	112	126	118	131	131	124
33	126	120	122	124	120	120	131	123	124
34	124	124	132	126	126	120	126	129	125
35	110	120	128	110	122	126	121	114	125
36	90	90	94	88	88	94	97	94	96
37	106	106	110	106	108	110	116	121	127
38	218	202	208	218	200	206	215	201	207
39	130	128	130	128	126	128	141	133	146
40	136	136	130	136	138	128	153	143	138
41	100	96	88	100	96	86	113	107	102
42	100	98	88	100	98	88	109	105	97
43	124	116	122	126	116	122	145	102	137
44	164	168	154	164	168	154	192	178	171
45	100	102	100	100	104	102	112	116	116
46	136	126	122	136	124	122	152	144	147
47	114	108	122	114	108	122	141	141	137

48	148	120	132	146	130	132	206	188	166
49	160	150	148	160	152	146	151	147	136
50	84	92	98	86	92	98	112	125	124
51	156	162	152	156	158	152	162	165	189
52	110	98	98	108	100	98	117	118	109
53	100	106	106	100	108	108	119	131	124
54	100	102	94	100	102	96	136	116	113
55	86	74	76	88	76	76	112	115	104
56	106	100	110	106	100	108	120	118	132
57	108	110	106	106	118	106	117	118	115
58	168	188	178	170	188	182	194	191	196
59	166	150	154	164	150	154	167	160	161
60	146	142	132	144	142	130	173	161	154
61	204	198	188	206	198	188	228	218	189
62	96	94	86	96	94	84	77	89	101
63	134	126	124	132	126	124	154	156	141
64	138	144	140	140	142	138	154	155	148
65	134	136	142	136	134	140	145	154	166
66	156	160	154	156	162	156	200	180	179
67	124	138	138	122	140	136	188	147	139
68	114	110	114	112	114	114	149	217	192
69	112	116	122	112	114	124	136	132	133
70	112	116	134	114	114	136	128	125	142
71	202	220	228	200	220	226	204	222	224
72	132	136	134	134	136	132	184	187	192
73	158	162	152	158	164	150	163	160	152
74	88	76	88	90	76	86	93	88	88
75	170	174	176	172	174	178	178	181	181
76	182	176	180	184	174	178	202	199	195
77	112	114	124	112	112	126	162	166	148
78	120	118	120	118	116	120	227	227	219
79	110	108	106	110	108	106	133	127	126
80	112	112	106	112	110	106	202	190	213
81	154	134	130	156	136	132	158	121	134
82	116	112	94	118	114	96	124	149	137
83	108	110	114	106	110	114	114	118	126
84	106	98	100	104	100	100	137	135	134
85	122	112	112	122	114	114	121	123	128

Στον πίνακα 1 έχουμε δεδομένα από την μέτρηση της συστολικής πίεσης 85 ανθρώπων. Έγιναν 3 μετρήσεις από 2 πεπειραμένους γιατρούς(j και r) και από ένα αυτόματο πιεσόμετρο(s), σε κάθε άνθρωπο. Δηλαδή 9 μετρήσεις στον καθένα.

2. Διάστημα συφωνίας

Για να μελετήσουμε την συμφωνία των μεθόδων θα πρέπει να συγκρίνουμε τις διαφορές της πίεσης με κάθε μέθοδο, στον ίδιο άνθρωπο. Αν η διαφορά είναι πάντα μονόπλευρη, δηλαδή πάντα μια μέθοδος να δίνει μεγαλύτερα αποτελέσματα από την άλλη, τότε έχουμε μεροληψία και την εκτιμούμε από την μέση διαφορά των μεθόδων στο ίδιο υποκείμενο. Επίσης μας ενδιαφέρει και η διακύμανση της διαφοράς. Αυτές οι εκτιμήσεις έχουν νόημα μόνο αν η μεροληψία και η διακύμανση της διαφοράς είναι ομοιόμορφα σε όλο το δείγμα.



Αρχικά θα εργαστούμε συγκρίνοντας την πρώτη μέτρηση του J γιατρού και την πρώτη μέτρηση του αυτόματου πιεσόμετρου.

Εργαζόμαστε στο SPSS.

Κάνουμε ένα paired samples t-test(πίνακας 2) για να συγκρίνουμε τους μέσους των δύο μεθόδων.

Πίνακας 2. *Paired Sample t-test για το ζεύγος J1-S1.*

Paired Samples Test								
				95% Confidence Interval				
Pair	**Mean**	**Std. Deviation**	**Std. Error Mean**	**Lower**	**Upper**	**t**	**df**	**p-value**
J1-S1	-16.29412	19.61099	2.12711	-20.52411	-12.06412	-7.660	84	0.000

Η μέση διαφορά είναι $\overline{D}$=16.29 και η τυπική απόκλιση είναι s_d=19.61. Υποθέτοντας ότι οι διαφορές ακολουθούν κανονική κατανομή τότε ένα διάστημα 95% των διαφορών θα είναι του τύπου $\overline{D} \pm 1.96 \cdot s_d$. Εδώ έχουμε $16.29 \pm 1.96 \cdot 19.61 \rightarrow (-54.7, 22.1)$. Αν από αυτά τα όρια καταλήγαμε στο ότι οι διαφορές δεν φαίνεται να είναι σημαντικές τότε και οι 2 μέθοδοι θα θεωρούνταν ίδια αποτελεσματικές. Το διάστημα εδώ είναι το διάστημα τιμών της διαφοράς. Δηλαδή το εύρος των τιμών που παίρνει το D. **Δεν** είναι διάστημα εμπιστοσύνης.

Έχουμε κάνει μια υπόθεση κανονικότητας των διαφορών. Φαίνεται λογική εδώ διότι τα δείγμα μας αφορά ανθρώπους με κοντινά χαρακτηριστικά και υπάρχει ένα σφάλμα της μέτρησης όπου από τη φύση του ακολουθεί κανονική κατανομή. Αφαιρώντας 2 μετρήσεις που περιέχουν σφάλματα έχουμε μεγαλύτερη σιγουριά πως ακολουθούν κανονική κατανομή. Μπορούμε να το δούμε όμως και γραφικά κάνοντας ένα ιστόγραμμα αλλά και κάνοντας τα κατάλληλα τεστ. Κάνοντας όμως Kolmogorov-Smirnov(πινακας 3) βλέπουμε πως δεν ακολουθεί κανονική κατανομή. Αυτό συμβαίνει στην περίπτωση που μέσος και διαφορά είναι συσχετισμένα και ο τρόπος απαλοιφής θα περιγραφεί παρακάτω.

Πίνακας 3 *Τεστ κανονικότητας των διαφορών J1-S2*

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	p-value	Statistic	df	p-value
Διαφορές J1-S1	0.170	85	0.000	0.836	85	0.000

Όμως, το να μην ακολουθούν κανονική κατανομή οι διαφορές, στην περίπτωσή μας, δεν είναι και τόσο σημαντικό. Και στις μη-κανονικές κατανομές ισχύει πως το 95% των τιμών θα βρίσκονται στο διάστημα 2 τυπικών αποκλίσεων από τη μέση τιμή με τη διαφορά ότι μπορεί να φεύγουν από το διάστημα από την ίδια κατεύθυνση. Αυτό που μας ενδιαφέρει περισσότερο είναι να έχουμε μικρή τυπική απόκλιση άρα και μικρό διάστημα συμφωνίας.

Στα δεδομένα μας έχουμε κάποιες μεγάλες διαφορές σε μερικά υποκείμενα. Όμως είναι 4 δηλαδή είναι αναμενόμενο καθώς ανήκει στο 5% που θα είναι εκτός διαστήματος. Μπορούμε να κάνουμε ότι κάναμε και πριν βγάζοντας αυτή τη φορά τις 2 ακραίες παρατηρήσεις. Τα αποτελέσματα είναι στον πίνακα 4.

Πίνακας 4 *Paired Sample t-test για το ζεύγος J1-S1 έχοντας βγάλει τις 2 ακραίες τιμές.*

Paired Samples Test								
				95% Confidence Interval				
Pair	**Mean**	**Std. Deviation**	**Std. Error Mean**	**Lower**	**Upper**	**t**	**df**	**p-value**
J1-S1	-14.31325	14.94700	1.64065	-17.57702	-11.04948	-8.724	82	0.000

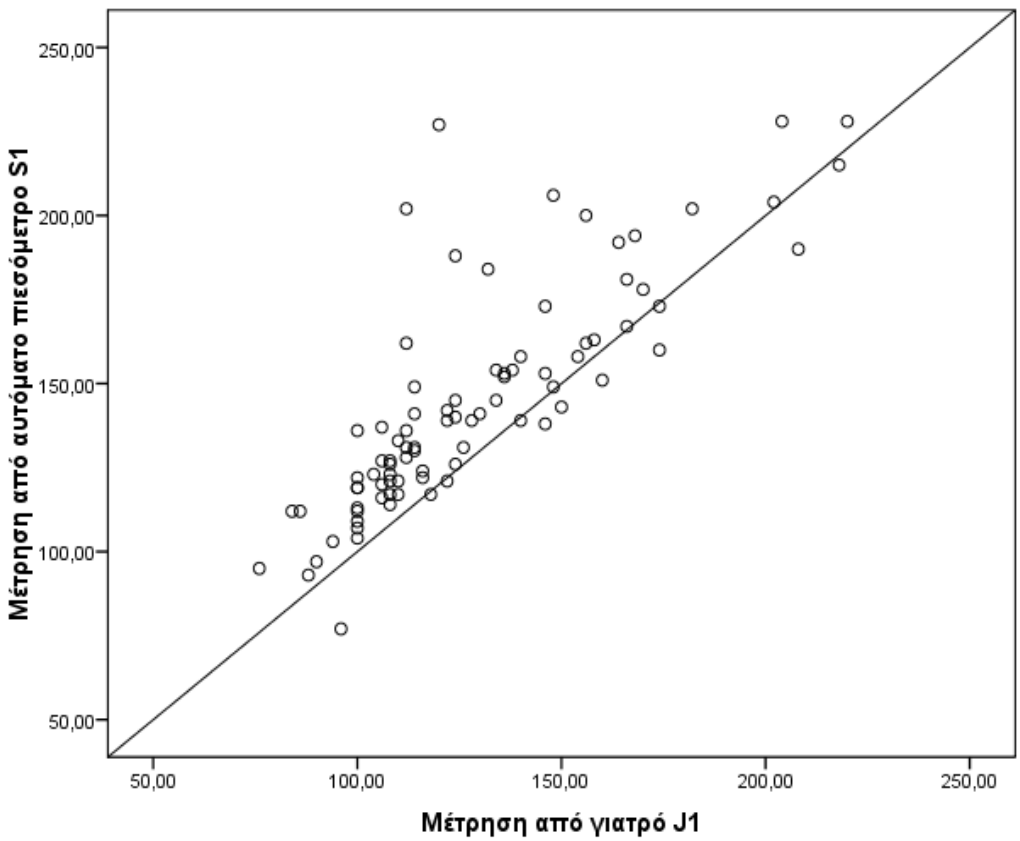
Το διάστημα τιμών εδώ είναι (-43.6,15.0) . Παρατηρούμε ότι το εύρος των τιμών στο $2^{ο}$ διάστημα είναι μικρότερο. Συγκεκριμένα από 77 έγινε 59. Το αν είναι όμως σημαντική ή όχι αυτή η μείωση είναι στην κρίση των ιατρών και δεν μπορούν να απαντηθούν με τη στατιστική.

2.1 Γραφική αναπαράσταση της συμφωνίας των μεθόδων

Στο διάγραμμα 1 βλέπουμε τη σύγκριση της μέτρησης πίεσης από γιατρό και από το αυτόματο πιεσόμετρο, ενώ αναπαρίσταται και η γραμμή

ισότητας, δηλαδή τα σημεία που θα περνούσε η γραφική παράσταση αν οι 2 μέθοδοι έδιναν ίδιο αποτέλεσμα. Δεν μας ενδιαφέρει η παλινδρόμηση των 2 μεθόδων γιατί δεν ψάχνουμε την σχέση τους αλλά αν υπάρχει ισότητα και αν δεν υπάρχει την απόκλιση από αυτήν.

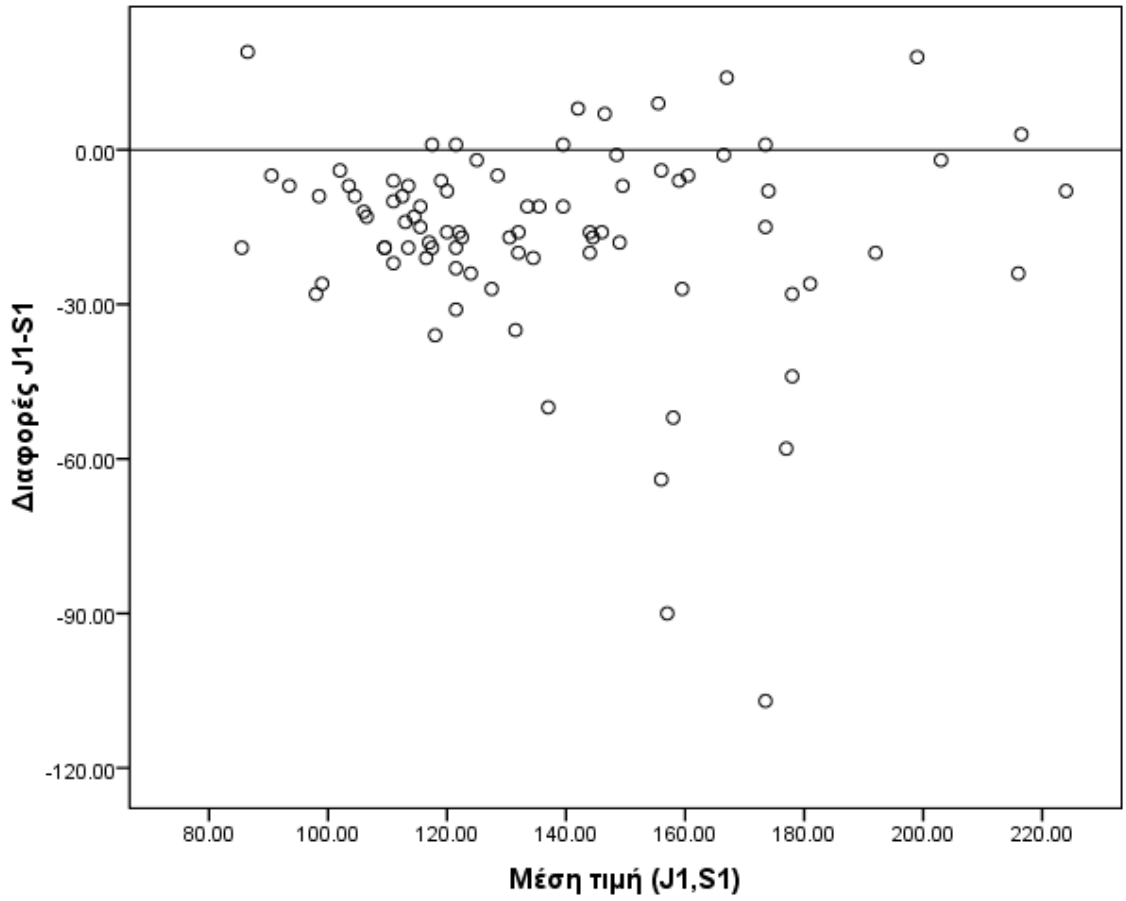
Διάγραμμα 1



Ένας καλύτερος τρόπος αναπαράστασης των δεδομένων είναι να δούμε γραφικά την διαφορά με τον μέσο των μεθόδων(διάγραμμα 2). Σε αυτό το διάγραμμα μπορούμε να δούμε την πιθανή σχέση μεταξύ των διαφορών και της πραγματικής τιμής(πραγματική τιμή θεωρούμε τον μέσο των 2 μετρήσεων (J1, S1) γιατί σε τέτοιες μελέτες που δεν ξέρουμε την ακριβή τιμή αυτού που μετράμε θεωρούμε ως καλύτερη προσέγγιση το μέσο των μετρήσεών μας).

Μπορούμε να υπολογίσουμε την όποια σχέση τους υπολογίζοντας τη συσχέτιση των διαφορών με τον μέσο. Λόγω του ότι οι διαφορές, όπως είδαμε παραπάνω δεν ακολουθούν κανονική κατανομή, θα κάνουμε έναν μη παραμετρικό έλεγχο, το Spearman's test(πίνακας 5).

Διάγραμμα 2

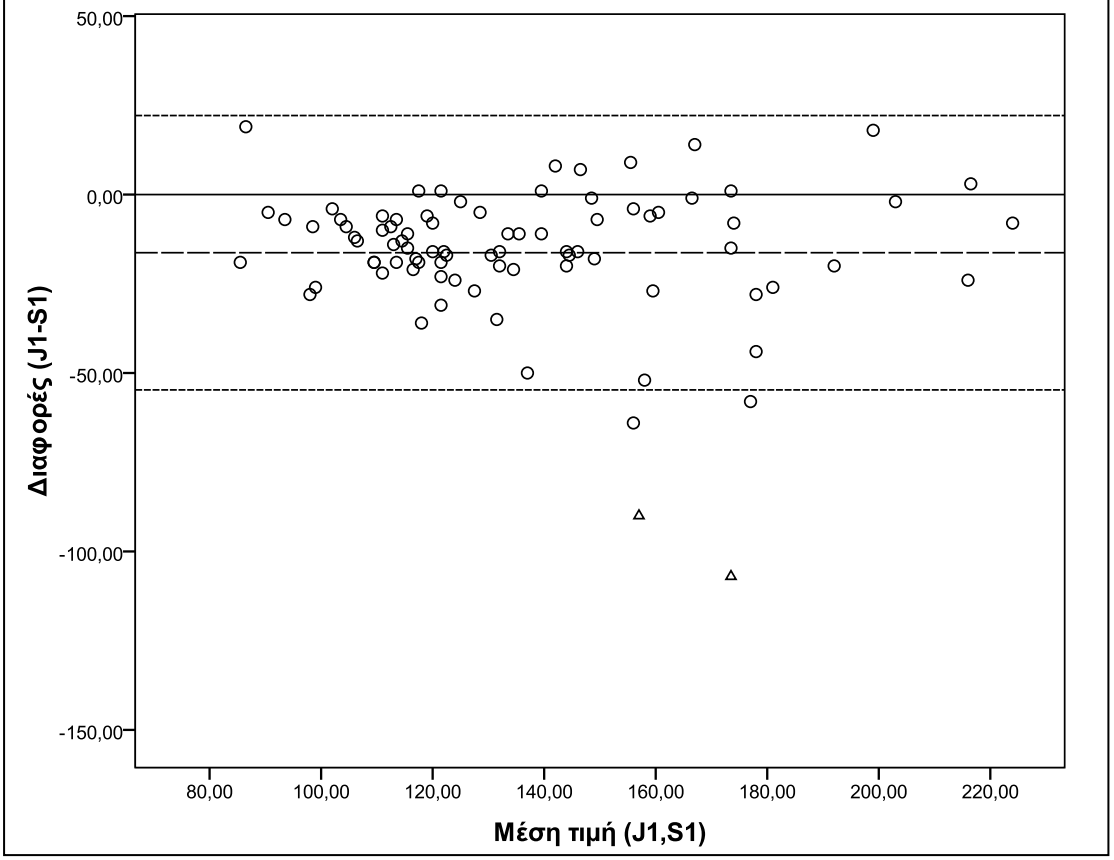


Πίνακας 5 *Μη παραμετρικός έλεγχος Spearman's test*

Correlations				
			Διαφορές J1-S1	Μέση τιμή (J1,S1)
Spearman's rho	Διαφορές J1-S1	Correlation Coeff.	1	0.068
		p-value		0.539
		N	85	85
	Μέση τιμή (J1,S1)	Correlation Coeff.	0.068	1
		p-value	0.539	
		N	85	85

Το αποτέλεσμα του ελέγχου μας είναι 0.068(p-value=0.539), δηλαδή δεν φαίνεται να επηρεάζει η μία τιμή την άλλη, δηλαδή δεν έχουμε συσχέτιση. Για ακόμα καλύτερο οπτικό έλεγχο και παρατήρηση μπορούμε στο διάγραμμα 2 να προσθέσουμε και τα 95% όρια συμφωνίας(διάγραμμα 3).

Διάγραμμα 3



Στα διαγράμματα βλέπουμε και τις ακραίες τιμές. Εδώ είναι 4/85 έξω από το διάστημα συμφωνίας, δηλαδή 4,7%, άρα κανένα πρόβλημα και ακραίες φαίνονται να είναι 2 τιμές μόνο, που συμβολίζονται με τρίγωνο. Ουσιαστικά, στο διάγραμμα παρατηρούμε, πόσο απέχουν οι τιμές μας από τον μέσο του διαστήματος ή ακόμα και από το 0. Όσο πιο μακριά είναι, σημαίνει πως η μέση τιμή, είναι αποτέλεσμα 2 πολύ διαφορετικών τιμών, αφού η διαφορά J1-S1 θα είναι μεγάλη. Αν π.χ. είχαμε μια τιμή στο (100,-100) σημαίνει πως οι 2 τιμές έχουν διαφορά 100 μονάδων(J1=150,S1=50) η οποία είναι πολύ μεγάλη.

2.2 Ακρίβεια της εκτίμησης των ορίων της συμφωνίας

Για τα όρια της συμφωνίας μπορούμε να υπολογίσουμε τυπική απόκλιση , άρα και διαστήματα εμπιστοσύνης, αν υποθέσουμε πως οι διαφορές ακολουθούν κανονική κατανομή. Για να υπολογίσουμε την τυπική απόκλιση s_d του μέσου των διαφορών, διαιρούμε την διακύμανση τους s_d^2 με το μέγεθος του δείγματος n. Αφού ακολουθούν κανονική κατανομή, η διακύμανση της τυπικής απόκλισης εκτιμάται από τον τύπο $\frac{s_d}{2(n-1)}$(αφού s_d^2 έχει κατανομή $x^2 s_d^2(n-1)$ και $\sqrt{x^2}$ έχει διακύμανση ½). Ο μέσος των διαφορών $\overline{D}$ και η τυπική απόκλιση είναι ανεξάρτητα. Άρα για να υπολογίσουμε την διακύμανση των ορίων συμφωνίας:

$$Var(\overline{D} \pm 1.96 s_d) = Var(\overline{d}) + 1.96^2 Var(s_d) = \frac{s_d^2}{n} + 1.96^2 \frac{s_d^2}{2(n-1)}$$
$$= \left(\frac{1}{n} + \frac{1.96^2}{2(n-1)} \right) s_d^2$$

Εκτός και αν το n είναι μικρό όπου ο τύπος θα γίνει:

$$Var(\overline{D} \pm 1.96 s_d) = (1 + \frac{1.96^2}{2}) \frac{s_d^2}{n} = \frac{2.92 s_d^2}{n} = 1.71^2 \frac{s_d^2}{n}$$

Έτσι η τυπική απόκλιση των ορίων θα είναι $\frac{1.71 s_d}{\sqrt{n}} = 1.71 se(\overline{D})$. Για το διάστημα εμπιστοσύνης των ορίων θα χρησιμοποιήσουμε την κατάλληλη τιμή από τον πίνακα της t-student κατανομής με βαθμούς ελευθερίας n-1.

Από την ανάλυση μας έχουμε την τυπική απόκλιση και το μέσο των διαφορών(πίνακας 6).

Πίνακας 6

Περιγραφικά Στατιστικά	
N	85
Μέση τιμή	-16.2941
Τυπική απόκλιση	19.61099

Άρα η τυπική απόκλιση του μέσου των διαφορών $\overline{D}$ είναι $\frac{s_d}{\sqrt{n}}$=2.13. Η τιμή $t_{95,84}$ =1.99. Οπότε το 95% διάστημα εμπιστοσύνης του $\overline{D}$ είναι -16.29±1.99x2.13 → (-20.5,-12.1).

Η τυπική απόκλιση των ορίων του 95% διαστήματος συμφωνίας είναι 1.71se($\overline{D}$)=1.71x2.13= 3.64. Το 95% διάστημα εμπιστοσύνης για τα όρια του διαστήματος συμφωνίας (-54.7,22.1) είναι -54.7 ± 1.99x3.64 → (-61.9,-47.5) για το κάτω όριο και 22.1 ± 1.99x3.64 → (14.9,29.3) για το πάνω όριο.

Βλέπουμε ότι και το πιο στενό διάστημα που προκύπτει αν πάρουμε, το διάστημα θα είναι και πάλι μεγάλο. Έτσι λογικά(θέλουμε και τη γνώμη των γιατρών) καταλήγουμε στο ότι δεν συμφωνούν οι τιμές των 2 μεθόδων.

Τα όρια εμπιστοσύνης είναι κατασκευασμένα έχοντας υπ' όψιν μονό την αβεβαιότητα, λόγω του λάθους της δειγματοληψίας. Θεωρούμε και πως το δείγμα είναι αντιπροσωπευτικό. Ακόμα οι μετρήσεις έγιναν από έναν παρατηρητή μόνο. Η αβεβαιότητα που υπολογίσαμε λοιπόν, για τα όρια της ασυμφωνίας είναι κάπως αισιόδοξη.

3. Σχέση μεταξύ μεγέθους και διαφοράς

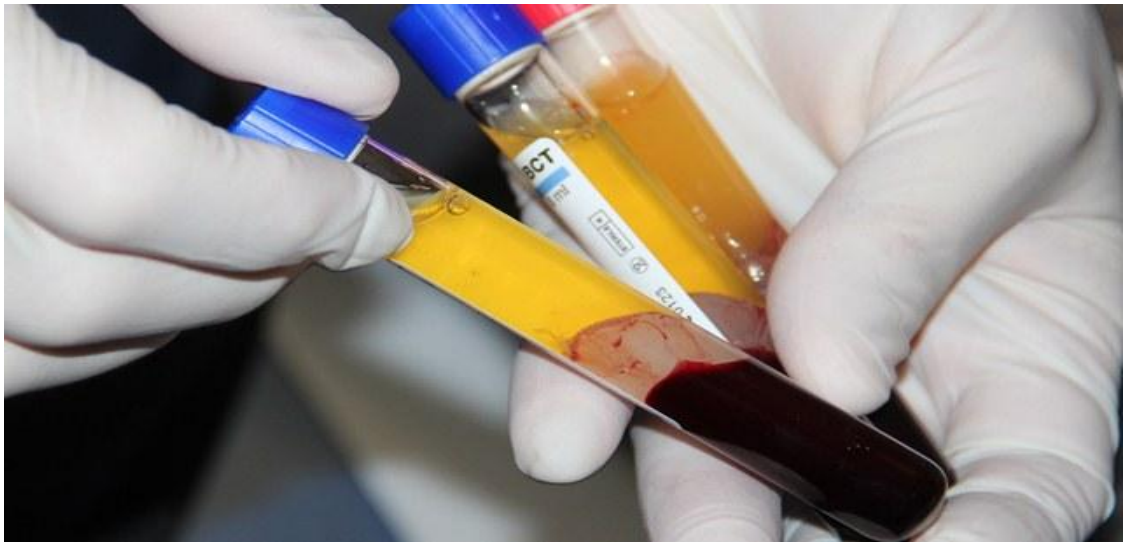
Στην προηγούμενη ενότητα υποθέσαμε πως ο μέσος και η τυπική απόκλιση των διαφορών είναι ίδια σε όλο το εύρος των τιμών. Δηλαδή η διακύμανση και η μέση διαφορά είναι όμοια μεταξύ των μετρήσεων, είτε έχουμε μικρές τιμές είτε μεγάλες. Συνήθως όσο μεγαλύτερες είναι οι τιμές τόσο μεγαλύτερη είναι και η διακύμανση των διαφορών. Για να το μελετήσουμε αυτό κάνουμε ένα διάγραμμα μεταξύ των μεθόδων και θα δούμε όσο μεγαλώνουν οι τιμές, αν έχουμε άπλωμα των δεδομένων. Επίσης η μέση διαφορά μπορεί να είναι ανάλογη με το μέγεθος των τιμών. Για να παρατηρήσουμε αυτό το φαινόμενο κάνουμε διάγραμμα της διαφοράς των μεθόδων ως προς τη μέση τιμή τους.

Θα εξετάσουμε τα δεδομένα του παρακάτω παραδείγματος(Πίνακας 6). Βλέπουμε εύκολα πως η μια μέθοδος(Nadler) δίνει μεγαλύτερες τιμές απ την άλλη(Hurley). Ακόμα φαίνεται πως υπάρχει μια τάση η διαφορά να αυξάνεται όσο αυξάνονται οι τιμές των μετρήσεων(Διάγραμμα 4). Κάνουμε και ένα διάγραμμα μεταξύ της διαφοράς και της μέσης τιμής των μεθόδων για να δούμε αν είναι ανάλογα μεταξύ τους(Διάγραμμα 5). Θα μπορούσαμε να εργαστούμε όπως πριν και να βγάλουμε κάποια διαστήματα συμφωνίας αλλά τα όρια δεν θα είναι σωστά. Για μικρές τιμές το διάστημα θα είναι «φαρδύ» ενώ για μεγάλες θα είναι «στενό».

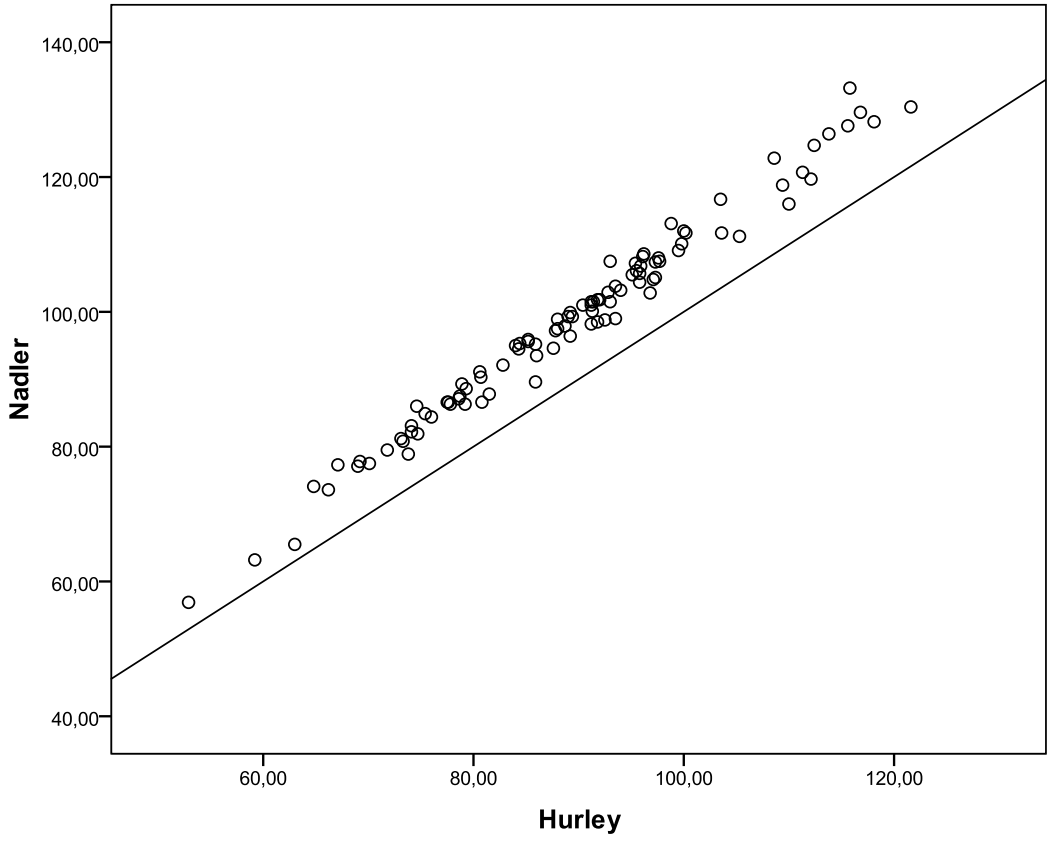
Πινακας 7. *Μετρήσεις όγκου πλάσματος, εκφρασμένο σε ποσοστό της αναμενόμενης τιμής, χρησιμοποιώντας 2 σετ δεδομένων κανονικών τιμών Nadler-Hurley (*C.Dore, βλ.Cotes et al[11]).

Subject	Nadler	Hurley	Subject	Nadler	Hurley
1	56,9	52,9	51	99,3	89
2	63,2	59,2	52	99,3	89,4
3	65,5	63	53	99,9	89,2
4	73,6	66,2	54	100,1	91,3
5	74,1	64,8	55	101	90,4
6	77,1	69	56	101	91,2
7	77,3	67,1	57	101,5	91,4
8	77,5	70,1	58	101,5	93
9	77,8	69,2	59	101,5	91,2
10	78,9	73,8	60	101,8	92

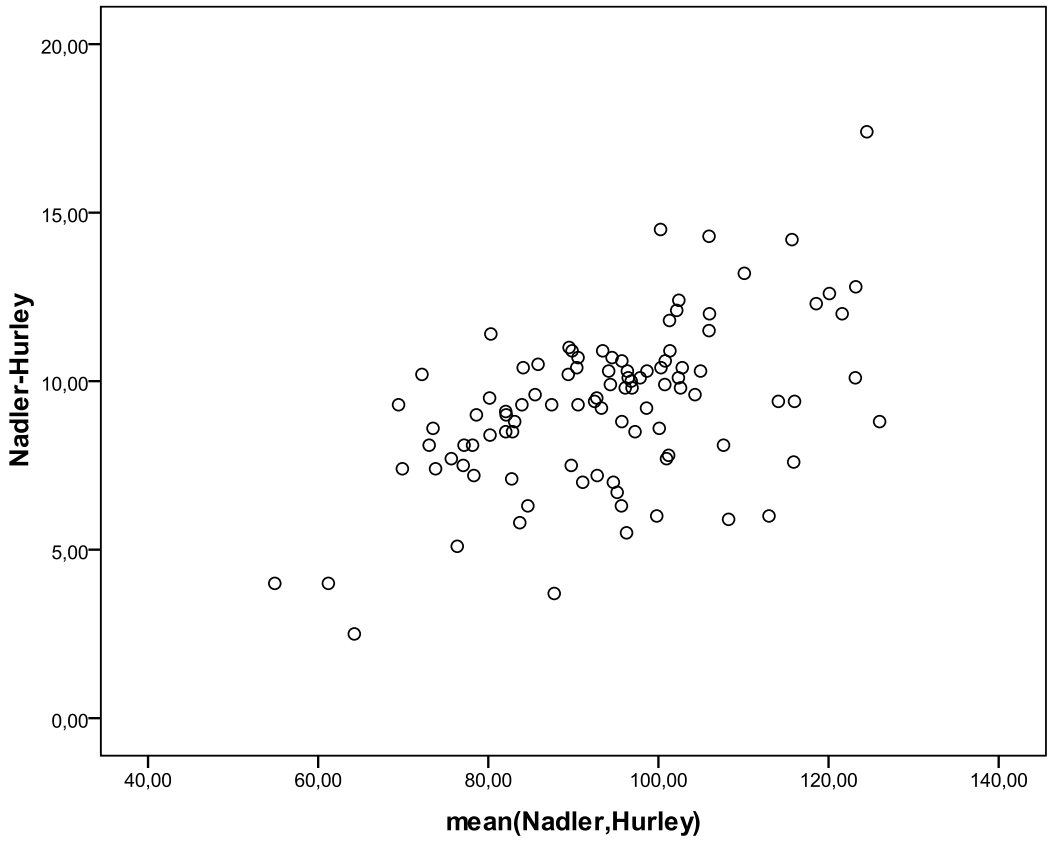
11	79,5	71,8	61	101,8	91,8
12	80,8	73,3	62	102,8	96,8
13	81,2	73,1	63	102,9	92,8
14	81,9	74,7	64	103,2	94
15	82,2	74,1	65	103,8	93,5
16	83,1	74,1	66	104,4	95,8
17	84,4	76	67	104,8	97,1
18	84,9	75,4	68	105,1	97,3
19	86	74,6	69	105,5	95,1
20	86,3	79,2	70	105,7	95,8
21	86,3	77,8	71	106,1	95,5
22	86,6	80,8	72	106,8	95,9
23	86,6	77,6	73	107,2	95,4
24	86,6	77,5	74	107,4	97,3
25	87,1	78,6	75	107,5	97,7
26	87,5	78,7	76	107,5	93
27	87,8	81,5	77	108	97,6
28	88,6	79,3	78	108,2	96,1
29	89,3	78,9	79	108,6	96,2
30	89,6	85,9	80	109,1	99,5
31	90,3	80,7	81	110,1	99,8
32	91,1	80,6	82	111,2	105,3
33	92,1	82,8	83	111,7	103,6
34	93,5	86	84	111,7	100,2
35	94,5	84,3	85	112	100
36	94,6	87,6	86	113,1	98,8
37	95	84	87	116	110
38	95,2	85,9	88	116,7	103,5
39	95,3	84,4	89	118,8	109,4
40	95,6	85,2	90	119,7	112,1
41	95,9	85,2	91	120,7	111,3
42	96,4	89,2	92	122,8	108,6
43	97,2	87,8	93	124,7	112,4
44	97,5	88	94	126,4	113,8
45	97,9	88,7	95	127,6	115,6
46	98,2	91,2	96	128,2	118,1
47	98,5	91,8	97	129,6	116,8
48	98,8	92,5	98	130,4	121,6
49	98,9	88	99	133,2	115,8
50	99	93,5			



Διάγραμμα 4


140,00
120,00
100,00
80,00
60,00
40,00
Nadler
60,00
80,00
100,00
120,00
Hurley

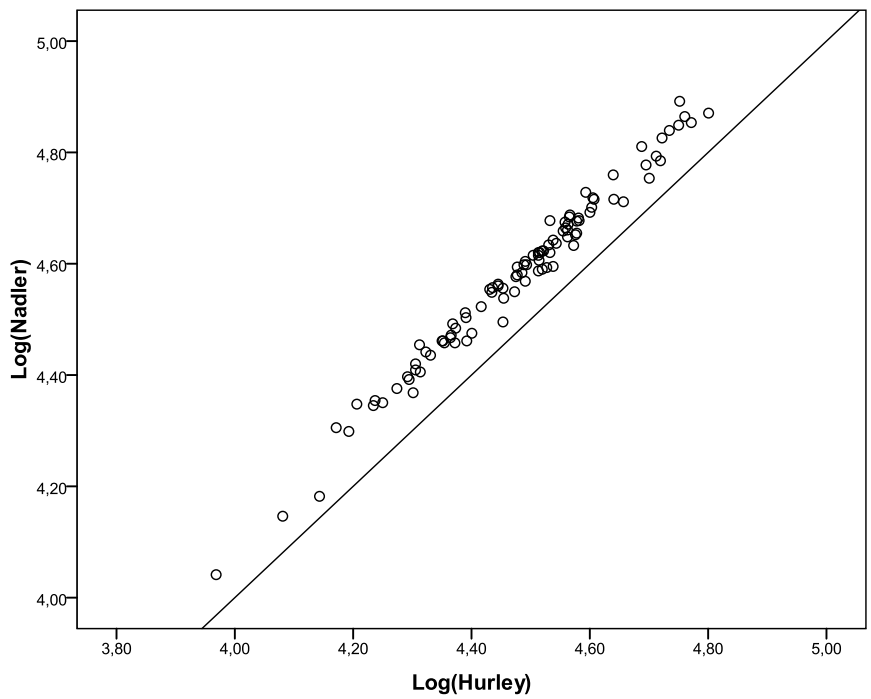
Διάγραμμα 5



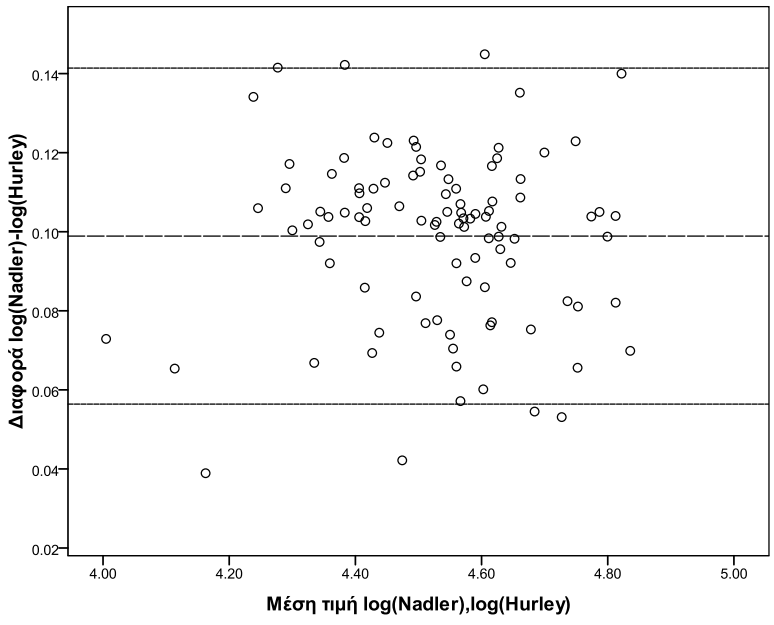
3.1 Λογαρίθμηση

Σε αυτήν την περίπτωση, λογαριθμίζουμε τις μετρήσεις και των 2 μεθόδων πριν την ανάλυση και έτσι ισχύει η κανονικότητα. Έτσι θα βγάλουμε τα όρια του διαστήματος συμφωνίας και απολογαριθμίζοντας θα βρούμε τα όρια συμφωνίας για την αναλογία Nadler/Hurley^2.

Διάγραμμα 6



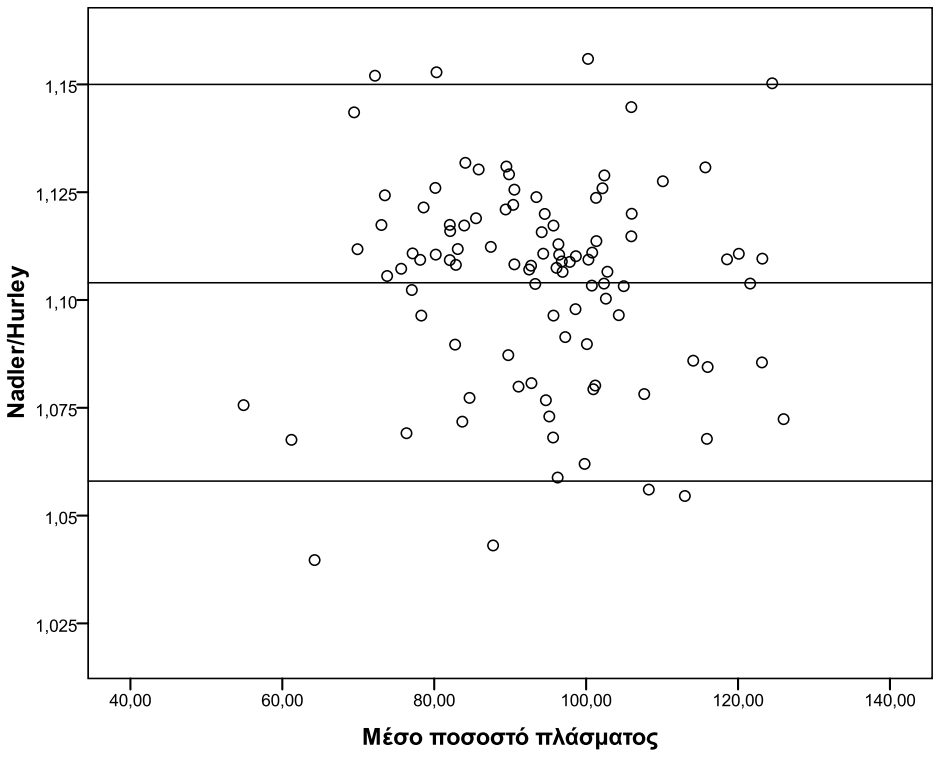
Διάγραμμα 7



Στα δεδομένα του πίνακα 6, η λογαρίθμηση φαίνεται να δίνει πολύ καλά αποτελέσματα. Από το διάγραμμα 7 βλέπουμε πως η διαφορά των μεθόδων δεν είναι ανάλογη της μέσης τιμής τους αφού οι τελείες είναι διάσπαρτες. Επίσης στο διάγραμμα 7 έχουμε και το 95% διάστημα συμφωνίας. Μέση τιμή Log(Nadler)-Log(Hurley) είναι ίση με 0.0989 και η τυπική της απόκλιση είναι 0,0217 οπότε το διάστημα συμφωνίας είναι 0.0989±1.96*0.0217→(0.0564,0.1414). Για να βρούμε το δ.ε. των ορίων του διαστήματος συμφωνίας πρέπει να βρούμε αρχικά την τυπική απόκλιση της μέσης διαφοράς. Άρα $\frac{s_d}{\sqrt{n}} = \frac{0.0217}{\sqrt{98}} = 0.00401, t_{0.025,98} = 1.98$. Άρα το δ.ε. του κάτω ορίου είναι 0.0564±1.97x0.00401→(0.0485,0.064).

Όμως τα αποτελέσματα αναφέρονται σε διαφορές μεταξύ λογαριθμισμένων τιμών και δεν γίνεται να ερμηνευθούν εύκολα. Οπότε όπως είπαμε και πριν, θα απολογαριθμίσουμε τα αποτελέσματα για να δούμε τις αναλογίες μεταξύ των δύο μεθόδων. Ο μέσος και το 95% δ.ε. του μέσου είναι 1.104 και (1.058,1.15). Άρα η μέθοδος Nadler δίνει κατά 95%, από 6% έως 15%, μεγαλύτερες τιμές από τη Hurley. Όλα τα παραπάνω φαίνονται γραφικά στο διάγραμμα 8.

Διάγραμμα 8



Βλέπουμε πως είναι σημαντικά και η διαφορά D και η τυπική απόκλιση μεταξύ των μεθόδων s_d. Εδώ η διαφορά είναι στην πραγματικότητα η μεροληψία αφού όλες οι διαφορές μεταξύ των μεθόδων είναι μονόπλευρες. Δηλαδή η μέθοδος Nadler πάντα δίνει μεγαλύτερες τιμές. Λόγω όμως της μικρής διασποράς γύρω από τη μέση τιμή της διαφοράς θα μπορούσαμε διαιρώντας την μέθοδο Nadler με 1.104 ή πολλαπλασιάζοντας τη μέθοδο Hurley με 1.104 να έχουμε συμφωνία των 2 μεθόδων.

Θα μπορούσαμε κατευθείαν να έχουμε δουλέψει με τον τελευταίο τρόπο στο διάγραμμα 8. Χωρίς να λογαριθμίσουμε και χρησιμοποιώντας τον μέσο και την τυπική απόκλιση του πηλίκου Nadler/Hurley.

3.2 Μέθοδος μελέτης με παλινδρόμηση για μη ομοιόμορφες διαφορές

Κάποιες φορές οι διαφορές μεταξύ 2 μεθόδων είναι πολύπλοκες και χρειαζόμαστε άλλες μεθόδους για να τις μελετήσουμε. Για παράδειγμα, μπορεί στη σύγκριση 2 μεθόδων A-B, για χαμηλές τιμές των μετρήσεων η μέθοδος A να δίνει μεγαλύτερες τιμές, ενώ για μεγάλες τιμές των μετρήσεων η B να δίνει μεγαλύτερες τιμές. Σε αυτές τις περιπτώσεις, ένα διάγραμμα σαν το διάγραμμα 7 είναι χρήσιμο για να συγκρίνουμε τις 2 μεθόδους, αλλά αν προσπαθήσουμε να το αναλύσουμε με τον προηγούμενο τρόπο, τα όρια συμφωνίας θα είναι πολύ μακριά μεταξύ τους.

Ο τρόπος μελέτης σε αυτήν την περίπτωση είναι να μοντελοποιήσουμε την διακύμανση της τυπικής απόκλισης της διαφοράς των μεθόδων d_i συναρτήσει του μεγέθους της μέτρησης, χρησιμοποιώντας μια μέθοδο που βασίζεται στα απόλυτα κατάλοιπα από την γραμμή παλινδρόμησης. Αν αλλάζει και η μέση διαφορά ανάλογα με το μέγεθος της μέτρησης, μπορούμε να μοντελοποιήσουμε αυτή τη σχέση και μετά την τυπική απόκλιση των καταλοίπων.

Αρχικά εξετάζουμε τη σχέση της διαφοράς με το μέγεθος της μέτρησης. Θα λάβουμε υπ’ όψιν μας ότι διαφορές των μεθόδων αλλάζουν ανάλογα με το μέγεθος των μετρήσεων αλλά και η τυπική απόκλιση αλλάζει ανάλογα με τον μέσο των 2 μεθόδων(A).

Ο τύπος της απλής παλινδρόμησης που χρειαζόμαστε είναι:

$$D = b_0 + b_1 A \qquad (1)$$

Αν ο συντελεστής b_1 δεν είναι στατιστικά σημαντικός, δηλαδή ίσος με 0, ισχύει η ισότητα $D = \bar{d}$. Αν είναι στατιστικά σημαντικό, εκτιμούμε τη διαφορά των δύο μεθόδων από την εξίσωση (1). Το τι είναι στατιστικά σημαντικό εξαρτάται κάθε φορά από την κρίση των γιατρών σε κάθε έρευνα.

Θα κάνουμε ένα παράδειγμα για την εφαρμογή αυτής της μεθόδου.

Στον πίνακα 8 παρακάτω βλέπουμε το εκτιμώμενο λίπος του ανθρώπινου γάλακτος (g/100ml) μετρημένο μέσω της γλυκερόλης που απελευθερώνεται από την υδρόλυση των τριγλυκεριδίων και με την κλασική μέθοδο Gerber[4].

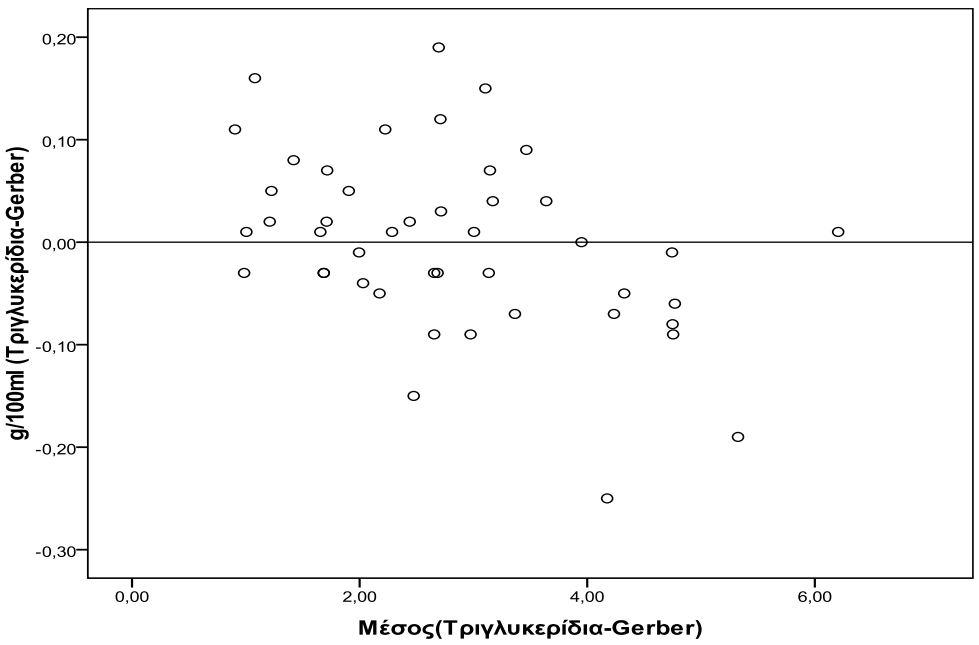


Πίνακας 8.*Εκτιμώμενο λίπος στο ανθρώπινο γάλα μετρημένο μέσω των τριγλυκεριδίων και μέσω της κλασικής μεθόδου Gerber*[4].

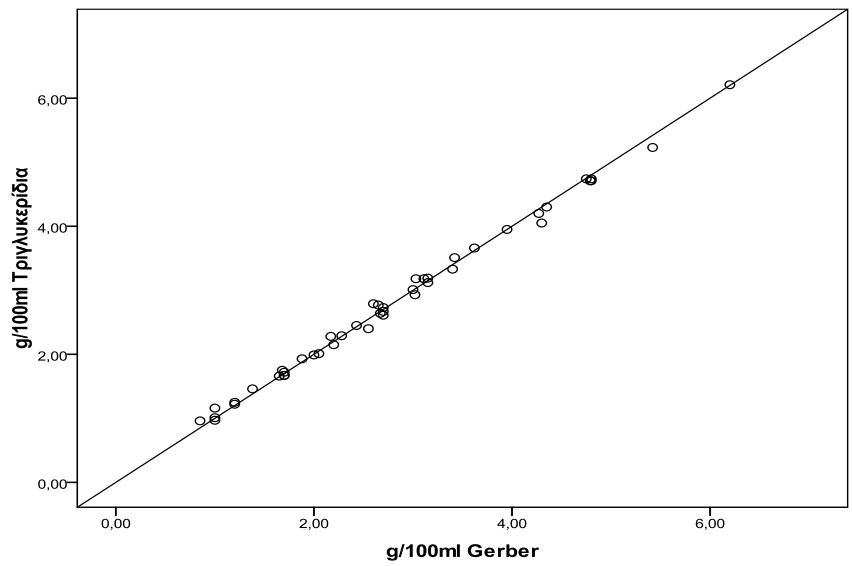
Τριγλυκερίδια	Gerber		Τριγλυκερίδια	Gerber
0.96	0.85		2.73	2.70
1.16	1.00		2.67	2.70
0.97	1.00		2.61	2.70
1.01	1.00		3.01	3.00
1.25	1.20		2.93	3.02
1.22	1.20		3.18	3.03
1.46	1.38		3.18	3.11
1.66	1.65		3.19	3.15
1.75	1.68		3.12	3.15
1.72	1.70		3.33	3.40
1.67	1.70		3.51	3.42
1.67	1.70		3.66	3.62
1.93	1.88		3.95	3.95
1.99	2.00		4.20	4.27
2.01	2.05		4.05	4.30
2.28	2.17		4.30	4.35
2.15	2.20		4.74	4.75
2.29	2.28		4.71	4.79
2.45	2.43		4.71	4.80
2.40	2.55		4.74	4.80
2.79	2.60		5.23	5.42
2.77	2.65		6.21	6.20
2.64	2.67			

Θα δούμε μέσω 2 διαγραμμάτων τη σχέση μεταξύ των δύο μεθόδων. Το ένα διάγραμμα(διάγραμμα 9) θα είναι μεταξύ των τιμών των 2 μεθόδων και το δεύτερο (διάγραμμα 10) θα είναι μεταξύ των διαφορών των μεθόδων και του μέσου τους.

Διάγραμμα 9



Διάγραμμα 10



Στο διάγραμμα 9 βλέπουμε ότι οι 2 μέθοδοι είναι κοντά. Όμως στο διάγραμμα 10 φαίνεται να υπάρχει αλλαγή στο πρόσημο της διαφοράς τριγλυκερίδια-Gerber. Ενώ αρχικά τα τριγλυκερίδια δίνουν μεγαλύτερες τιμές, όσο αυξάνονται οι τιμές των μετρήσεων, η μέθοδος Gerber δίνει

μεγαλύτερες. Κάνοντας παλινδρόμηση όπως προαναφέραμε παίρνουμε το εξής μοντέλο (από πίνακα 9)

$$Var(Y) = \sigma_t^2 + \sigma_{yI}^2 + \sigma_{yw}^2 \quad g/100ml$$

Πίνακας 9

Coefficients				
Μοντέλο	Τιμή	Τυπιό σφάλμα	t	p-value
βο	0.079	0.029	2.72	0.009
β1	-0.028	0.009	-2.993	0.005

Η τυπική απόκλιση των καταλοίπων είναι 0.08033.

Τώρα πρέπει να να μοντελοποιήσουμε τη διασπορά των καταλοίπων του μοντέλου μας συναρτήσει του μεγέθους των μετρήσεων (A). Η μοντελοποίηση είναι απλή με την υπόθεση ότι τα κατάλοιπα ακολουθούν κανονική κατανομή ασχέτως του μεγέθους των μετρήσεων. Έτσι μοντελοποιούμε τις απόλυτες τιμές των καταλοίπων R με το μέσο των μετρήσεων A:

$$R = c_0 + c_1 A \qquad (2)$$

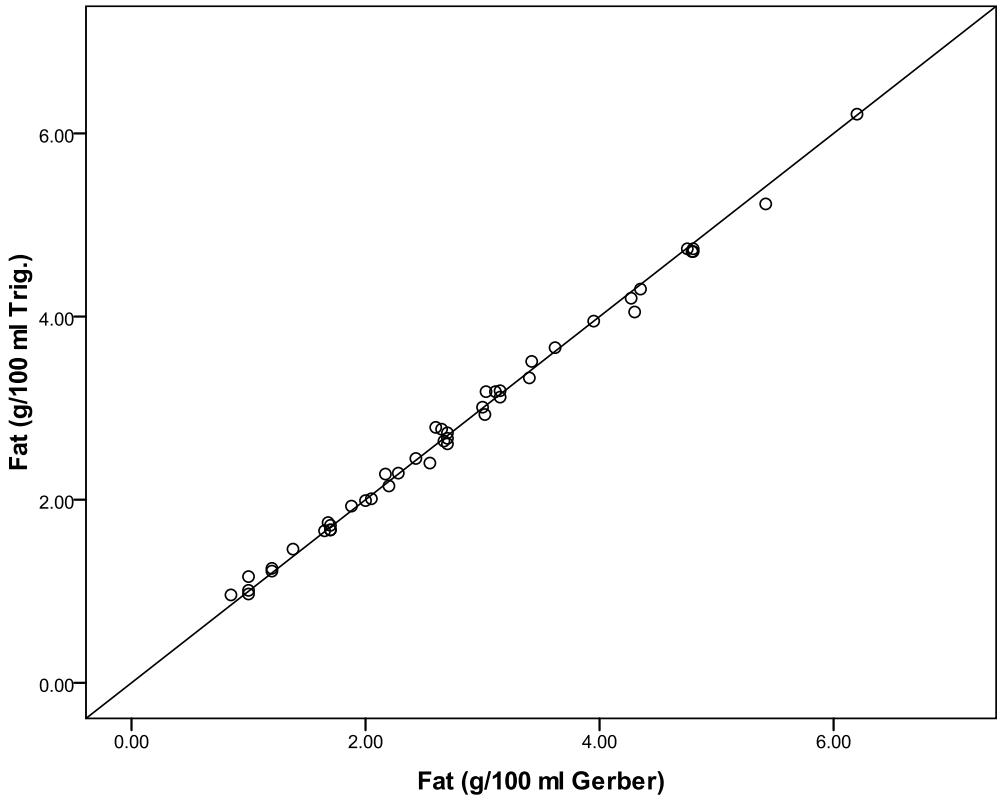
Αν έχουμε μία κανονική κατανομή με μέσο 0 και διακύμανση σ^2, οι απόλυτες τιμές των καταλοίπων ακολουθούν την μισή κανονική κατανομή με μέσο $\frac{\sigma}{\sqrt{2\pi}}$. Η τυπική απόκλιση τους προκύπτει πολλαπλασιάζοντας τις προβλεπόμενες τιμές με $\sqrt{\frac{\pi}{2}}$. Τα όρια συμφωνίας προκύπτουν από τον συνδυασμό των 2 εξισώσεων(Altman[3]).

Αν και αρχικα κάθε μοντέλο παλινδρόμησης θα μπορούσε να προσαρμοστεί, αν η τυπική απόκλιση δεν ειναι σταθερή, η γραμμική παλινρόμηση είναι επαρκης για να περιγράψει τη σχέση. Αν δεν υπάρχει σημαντική σχέση μετξύ R και A, η εκτιμώμενη τυπική απόκλιση είναι απλά η τυπική απόκλιση των προσαρμοσμένων διαφορών, τα κατάλοιπα της εξίσωσης.

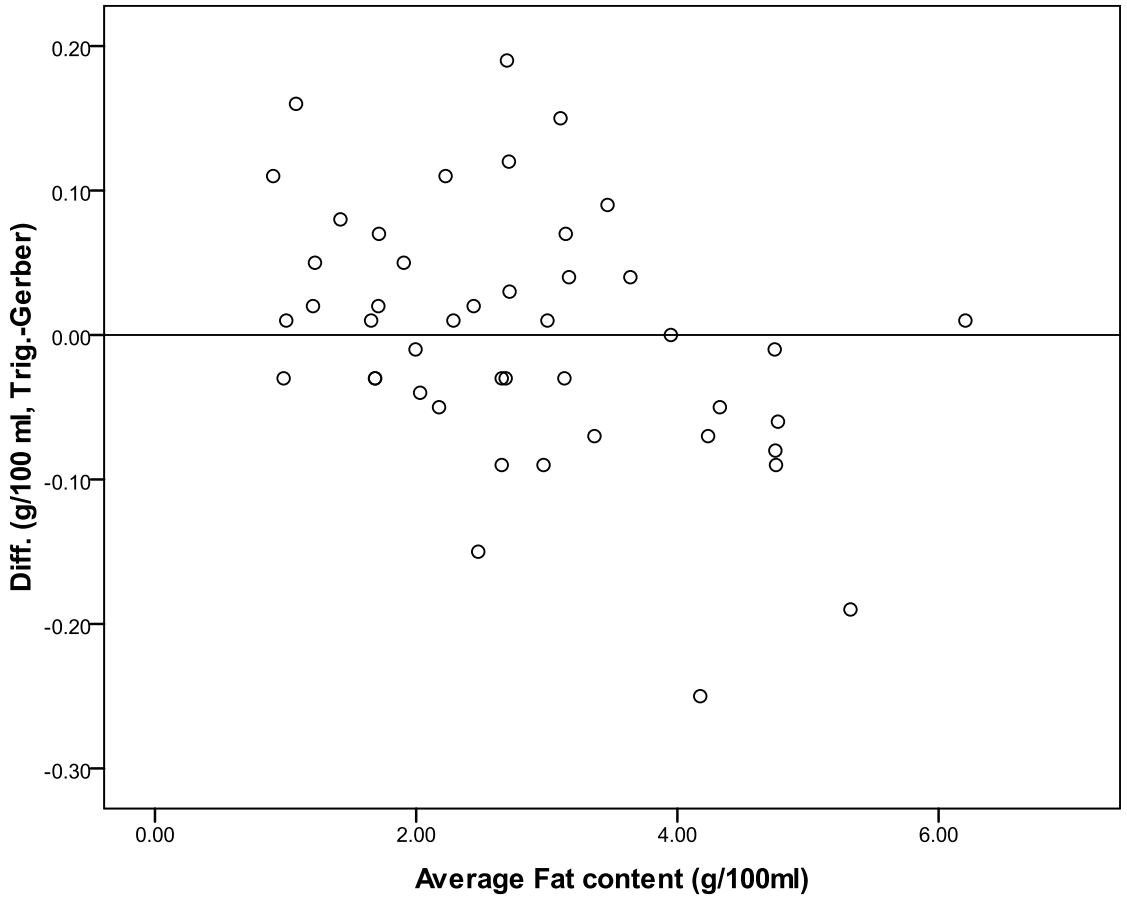
Στη γενική περίπτωση που και τα 2 μοντέλα (1),(2) χρησιμοποιούνται. Η προβλεπόμενη τιμή της διαφοράς των μεθόδων δίνεται από τον τύπο $D = b_0 + b_1 A$ και τα 95% όρια συμφωνίας δίνονται από τον τύπο

$$D \pm 1.96\sqrt{\frac{\pi}{2}}R = D \pm 2.46R \quad \text{ή} \quad b_0 + b_1 A \pm 2.46\{c_0 + c_1 A\}$$

Διάγραμμα 11 Περιεκτικότητα του ανθρώπινου γαλακτος σε λίπος μετα από ενζυμική διαδικασία για τον προσδιορισμο των τριγλυκεριδίων και μετρημένο με τη μέθοδο Gerber



Διάγραμμα 12 Διάγραμμα διαφοράς των 2 μεθόδων μέτρησης με τον μέσο τους.



Γυρνώντας στο παράδειγμα βλέπουμε πως δεν υπαρχει σχέση μεταξύ του μοντέλου παλινδρόμησης και του Α, επομένως η τυπική απόκλιση των προσαρμοσμένων διαφορών, ειναί τα κατάλοιπα απο την παλινδρόμηση, άρα s_d=0.08033. Μπορούμε να υπολογίσουμε τα 95% όρια συμφωνίας βάσει της παλινδρόμησης : 0,079-0,283α±1,96x0,08033 g/100ml όπου α το ο μέσος των δύο μεθόδων μέτρησης του λίπους. Τα όρια τα βλέπουμε και γραφικα στο διάγραμμα 12. Όταν χρησιμοποιούμε μία μέθοδο το α ειναι απλά η τιμη της μέτρησης και όχι καποιος μέσος.

4. Σπουδαιότητα της επανάληψης

Η σύγκριση της επανάληψης κάθε μεθόδου είναι σχετική με τη σύγκριση μεθόδων διότι οι επανλήψεις των δύο μεθόδων μέτρησης περιορίζουν το ποσοστό συμφωνίας το οποίο είναι πιθανο. Περιέργως, επαναλαμβανόμενες μετρήσεις σπανίως γίνονται σε μελέτες σύγκρισης μεθόδων και έτσι χάνεται ενας πολύ σημαντικός τρόπος σύγκρισης. Αν έχουμε μόνο μια μέτρηση απο κάθε μέθοδο για κάθε υποκείμενο, δεν μπορούμε να αποφανθούμε για το ποια μέθοδος είναι πιο ακριβής. Η έλλειψη επαναληψιμότητας παρεμβαίνει στη σύγκριση των μεθόδων διότι αν μια μέθοδος δεν είναι ακριβής, δηλαδη έχει σημαντικές διαφορές στις επαναλαμβανόμενες μετρήσεις της, τότε η συμφωνία μεταξύ των μεθόδων θα είναι κακη. Ακόμα και αν οι μέσες τιμές των μεθόδων έδειχναν συμφωνία, μια μέθοδος που θα είχε κακή επαναληψιμότητα, θα οδηγούσε σε κακή συμφωνία μεταξυ των μεθόδων. Αν μια παλιά μέθοδος έχει κακη επαναληψιμότητα και τη συγκρίνουμε μια μια νέα πολύ καλύτερη, δεν θα βρίσκαμε συμφωνία μεταξύ των μεθόδων αυτών. Γενικά έλλειψη συμφωνίας με μεθόδους που δεν έχουν επαναλαμβανόμενες μετρήσεις μπορεί να μας δείχνει πως η νέα μέθοδος δεν είναι καλή, αλλά αυτό μπορεί να οφείλεται στην κακή επαναληψιμότητα της παλιάς μεθόδου. Αν και οι δύο μέθοδοι έχουν κακή επαναληψιμότητα, η μη συμφωνία τους είναι σχεδόν σίγουρη. Γι αυτό θα πρέπει να εστιάζουμε στη συνεχή εκτίμηση των επαναλήψεων και της συμφωνίας συλλέγοντας δεδομένα με επαναλαμβανόμενες μετρήσεις.

Επαναλαβανόμενες μετρήσεις είναι οι μετρήσεις που γίνονται στο ίδιο υποκείμενο και σε ίδιες συνθηκες. Πιο γενικά ειναι μετρήσεις που λαμβάνονται σε γρήγορη διαδοχη.

Κάτι σημαντικό στις επαναλαμβανόμενες μετρήσεις είναι ότι θα πρέπει να είναι ανεξάρτητες μεταξύ τους. Αυτό μπορεί να επιτευχθεί αν ο παρατηρητής κάνοντας την μέτρηση, ανεξάρτητα απο την ήδη γνωστή προηγούμενη μέτρηση. Αυτό είναι πρακτικά δύσκολο.

4.1 Προβλέποντας την επανάληψη

Μια παρόμοια ανάλυση με τα όρια συμφωνίας μπορεί να εφαρμοστεί για να ποσοτικοποιήσουμε την επαναληψιμότητα μιας μεθόδου από δεδομένα που μετρήθηκαν σε ίδιες μεθόδους. Κάνοντας one-way ANOVA, με τα υποκείμενα ως παράγοντα, μπορούμε να προβλέψουμε την μεταξύ των υποκειμένων τυπική απόκλιση, από την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος. Μπορούμε να συγκρίνουμε τις τυπικές αποκλίσεις των διαφορετικών μεθόδων για να ελέγξουμε ποια είναι πιο ακριβής.Κάθε τυπική απόκλιση μπορεί επίσης να χρησιμοποιηθεί για να υπολογιστούν τα όρια μεταξύ των οποίων περιμένουμε οι διαφορά μεταξύ δυο μετρήσεων να είναι ψευδής. Εκτός του ότι είναι κατατοπιστική από μόνη της, η επαναληψιμότητα είναι ένα σημείο αναφοράς βάσει του οποίου θα δούμε την μεταξύ μέθοδου μεταβλητότητα.

Η ανάλυση είναι απλή γιατί δεν περιμένουμε να υπάρχει συστηματική διαφορά στις μετρήσεις του ίδιου δείγματος και άρα περιμένουμε η μέση διαφορά των επαναλήψεων να είναι μηδέν. Αν υπάρχει συστηματική διαφορά μπορούμε να πούμε πως οι τιμές δεν είναι επαναλήψεις. Ένα διάγραμμα μπορεί να μας δείξει αν υπόθεση μας είναι λογική και αν οι διαφορές είναι ανεξάρτητες του μέσου. Αν η ακρίβεια χειροτερεύει όσο μεγαλώνουν οι τιμές των μετρήσεων μπορεί να χρειαστεί να λογαριθμήσουμε τα δεδομένα ακριβώς όπως κάναμε για να συγκρίνουμε μεθόδους μεταξύ τους.

Επιστρέφοντας στα δεδομένα της πίεσης του αίματος του πίνακα 1, μπορούμε να εκτιμήσουμε την ακρίβεια κάθε μεθόδου. Για τον J ιατρό, η within subject διακύμανση είναι $s_w^2 = 37.408$ και για τον R ιατρό είναι $s_w^2 = 37.98$,ενώ για το αυτόματο πιεσόμετρο η διακύμανση είναι $s_w^2 = 83.141$. Οπότε βλέπουμε ότι οι 2 ιατροί είναι πιο ακριβείς και πιο πανομοιότυποι από το αυτόματο πιεσόμετρο.

Η επαναληψιμότητα της μεθόδου είναι $1.96\sqrt{2}s_w$ ή $2.77s_w$ για το 95% των υποκειμένων. Αυτός είναι ο συντελεστής επαναληψιμότητας. Για τον ιατρό J το $s_w = \sqrt{37.408} = 6.116$ και ο συντελεστής επαναληψιμότητας είναι 2,77x6,116= 16,95 mmHg. Για το μηχάνημα S έχουμε $s_w = \sqrt{83.141} = 9.118$ και ο συντελεστής επαναληψιμότητας είναι 2,77x9,118= 25,27mmHg. Άρα η

επαναληψιμότητα του μηχανήματος είναι 50% μεγαλύτερη από του ιατρού. Μπορούμε να συγκρίνουμε τους 95% συντελεστές επαναληψιμότητας με τα 95% διαστήματα συμφωνίας. Αυτά είναι ($-2,77s_w$, $2,77s_w$). Αν αυτά είναι ίδια τότε η έλλειψη συμφωνίας των δύο μεθόδων οφείλεται στην έλλειψη επαναληψιμότητας. Αν τα όρια συμφωνίας είναι σημαντικά μεγαλύτερα από όσο θα έπρεπε σύμφωνα με την επαναληψιμότητα, τότε υπάρχει κάποιος άλλος παράγοντας που ελλαττώνει την συμφωνία των μεθόδων.

5. Μετρώντας την συμφωνία στις επαναλαμβανόμενες μετρήσεις

Όταν έχουμε επαναλαμβανόμενες μετρήσεις δύο μεθόδων στο ίδιο υποκείμενο είναι προφανές ότι χρειαζεται να χρησιμοποιήσουμε όλα τα δεδομένα που έχουμε. Ένα λογικό πρώτο βήμα είναι να βρούμε τους μέσους όρους των επαναλήψεων κάθε μεθόδου σε κάθε υποκείμενο και να τους χρησιμοποιήσουμε για να συγκρίνουμε τις μεθόδους χρησιμοποιώντας τα διαστήματα συμφωνίας. Η εκτίμηση της μεροληψίας δεν θα επηρεαστεί από το μέσο όρο, αλλά η εκτίμηση της τπικής απόκλισης των διαφορών θα είναι μικρή, επειδή ένα μέρος της επίδρασης των σφαλμάτων μετρήσεων έχει εξαλειφθεί. Θέλουμε την τυπική απόκληση των μέσων μεταξύ μονών μετρήσεων και όχι μεταξύ μέσων όρων πολλών επαναλήψεων.

Τέτοια δεδομένα μπορεί να έχουν ίσο αριθμό επαναλήψεων ή και άνισο αριθμο επαναληψεων.

5.1 Ίσος αριθμος επαναλήψεων

Όταν πραγματοποιούμε επαναλαμβανόμενες μετρήσεις του ίδιου υποκειμένου από δύο μεθόδους, οι μετρήσεις από κάθε μέθοδο θα είναι καταναμημένες γύρω από την αναμενόμενη τιμή της κάθε μεθόδου γι αυτό το υποκείμενο. Αυτοί οι μέσοι των μεθόδων δεν θα είναι απαραίτητα ίδιοι διότι μπορεί να διαφέρουν από υποκείμενο σε υποκείμενο. Αυτή η μεταβλητότητα αποτελεί την αλληλεπίδραση των υποκειμένων και της μεθόδου. Εμείς θέλουμε να μελετήσουμε την διακύμανση των διαφορών μεταξύ μετρήσεων των διαφορετικών μεθόδων D= X – Y όπου X και Y οι μετρήσεις των δυο διαφορετικών μεθόδων. Για κάθε μέθοδο η διακύμανση είναι

$$Var(X) = \sigma_t^2 + \sigma_{xI}^2 + \sigma_{xw}^2$$

$$Var(Y) = \sigma_t^2 + \sigma_{yI}^2 + \sigma_{yw}^2$$

Όπου σ_t^2 η διακύμανση των μετρήσεων, σ_{xI}^2 και σ_{yI}^2 είναι η αλληλεπίδραση μεθόδου και υποκειμένου και σ_{xw}^2 και σ_{yw}^2 είναι η within subject διακύμανση των μετρήσεων της ίδιας μεθόδου, για την X και την Y αντίστοιχα. Έτσι η between subject διακύμανση των διαφορών των μετρήσεων είναι

$$Var(X-Y)=\sigma_D^2=\sigma_{xI}^2+\sigma_{yI}^2+\sigma_{xw}^2+\sigma_{yw}^2 \quad (3)$$

Σκοπός είναι να εκτιμήσουμε τη διακύμανση από την ανάλυση των μέσων των μετρήσεων για κάθε υποκείμενο, $\overline{D}=\overline{X}-\overline{Y}$ και ο τυπος είναι $Var(\overline{X}-\overline{Y})$. Χρησιμοποιώντας τον μέσο όρο των επαναλήψεων θα ελλαττωθεί η within subject διακύμανση αλλα δεν θα επηρεαστούν οι αλληλεπιδράσεις. Έχουμε

$$Var(\overline{X})=\sigma_t^2+\sigma_{xI}^2+\frac{\sigma_{xw}^2}{m_x} \quad (4)$$

όπου m_x είναι ο αριθμός των μετρήσεων που έγιναν σε κάθε υποκείμενο με τη X μέθοδο, αφού μόνο για τα within subject και within method κατάλοιπα έχουμε μέσους.

$$\sigma_d^2=358.492+(1-\frac{1}{3})37.408+(1-\frac{1}{3})83.141=438.858 \quad (5)$$

Από (4),(5)

$$Var(\overline{X}-\overline{Y})=Var(\overline{D})=\sigma_{xI}^2+\frac{\sigma_{xw}^2}{m_x}+\sigma_{yI}^2+\frac{\sigma_{yw}^2}{m_y}$$

Η κατανομή του $\overline{D}$ εξαρτάται μόνο από τα κατάλοιπα και τις αλληλεπιδράσεις, γιατί οι πραγματικές τιμές βρίσκονται στα X και Y τα οποία αφαιρέσαμε. Από (3) έχουμε

$$Var(X-Y)=\sigma_{xI}^2+\sigma_{xw}^2+\sigma_{yI}^2+\sigma_{yw}^2+\frac{\sigma_{xw}^2}{m_x}+\frac{\sigma_{yw}^2}{m_y}-\frac{\sigma_{xw}^2}{m_x}-\frac{\sigma_{yw}^2}{m_y}$$

$$=\sigma_{xI}^2+\frac{\sigma_{xw}^2}{m_x}+\sigma_{yI}^2+\frac{\sigma_{yw}^2}{m_y}+\sigma_{xw}^2-\frac{\sigma_{xw}^2}{m_x}+\sigma_{yw}^2-\frac{\sigma_{yw}^2}{m_y}$$

$$=Var(\overline{D})+\sigma_{xw}^2(1-\frac{1}{m_x})+\sigma_{yw}^2(1-\frac{1}{m_y}) \qquad (6)$$

Έστω ότι η παρατηρούμενη διακύμανση των διαφορών των μέσων όρων των παρατηρήσεων κάθε υποκειμένου είναι s_d^2, τότε το $Var(X-Y)=\sigma_d^2$ εκτιμάται από

$$\sigma_d^2=s_{\bar{d}}^2+(1-\frac{1}{m_x})s_{wx}^2+(1-\frac{1}{m_y})s_{wy}^2 \quad (7)$$

Αν έχουμε μόνο δύο επαναλήψεις τότε ο τύπος γίνεται

$$\sigma_d^2=s_{\bar{d}}^2+\frac{s_{xw}^2}{2}+\frac{s_{yw}^2}{2} \quad \text{(Bland and Altman}^2\text{)}$$

Αυτός ο τύπος λειτουργεί ακόμα και όταν η μία μέθοδος είναι επαναλαμβανόμενη και η άλλη όχι.

Ας συγκρίνουμε για παράδειγμα από τον πίνακα 1 τον γιατρό J και το αυτόματο πιεσόμετρο S. Όπως είδαμε πριν, η within-subject διακύμανση για τον J είναι 37,408 και για το αυτόματο πιεσόμετρο S είναι 83,141. Η μέση διαφορά είναι -15,62mmHg. Έχουμε 3 μετρήσεις για κάθε μέθοδο οπότε $m_j=m_s=3$ και η διακύμανση της διαφοράς των μέσων των παρατηρήσεων κάθε υποκειμένου είναι $s_d^2=358.492$. Άρα έχουμε

$$\sigma_d^2=358.492+(1-\frac{1}{3})37.408+(1-\frac{1}{3})83.141=438.858$$

και $s_d=\sqrt{438.858}=20.95mmHg$. Τα όρια του διαστήματος συμφωνίας είναι $-15.62\pm1.96x20.95=(-56.68,25.44)$ το οποίο διάστημα είναι κοντινό με αυτό που βρήκαμε αρχικά {*σελ. 4* $(-54.7,22.1)$} λαμβάνοντας υπ' όψην μόνο μία παρατήρηση.

Για να βρούμε την τυπική απόκλιση και τα 95% διαστήματα εμπιστοσύνης των ορίων της συμφωνίας υποθέτουμε πως τα κατάλοιπα ακολουθούν κανονική κατανομή και είναι ανεξάρτητα. Για n υποκείμενα το $n(m_x-1)\frac{s_{xw}^2}{\sigma_{xw}^2}$ ακολουθεί X^2 με $n(m_x-1)$ βαθμούς ελευθερίας και $2n(m_x-1)$ διακύμανση. Άρα

$$Var(s_{xw}^2)=\frac{2\sigma_{xw}^2}{n(m_x-1)} \quad \text{και} \quad Var(s_{yw}^2)=\frac{2\sigma_{yw}^2}{n(m_y-1)} \quad (8)$$

Και από τον τύπο (7)

$$Var\left(\left(1-\frac{1}{m_x}\right)s_{xw}^2+\left(1-\frac{1}{m_y}\right)s_{yw}^2\right)= \left(1-\frac{1}{m_x}\right)^2\frac{2\sigma_{xw}^4}{n(m_x-1)}+\left(1-\frac{1}{m_y}\right)^2\frac{2\sigma_{yw}^4}{n(m_y-1)}= \frac{2(m_x-1)\sigma_{xw}^4}{nm_x^2}+\frac{2(m_y-1)\sigma_{yw}^4}{nm_y^2} \quad (9)$$

(10)

Ομοίως η διακύμανση του s_d^2 δίνεται από

$$Var\left(s_{\bar{d}}^2\right)=\frac{2\sigma_{\bar{d}}^4}{n-1} \quad (11)$$

Αντικαθστώντας στην (7) τις (10) και (11), έχουμε

$$Var(\sigma_d^2)=\frac{2\sigma_{\bar{d}}^2}{n-1}+\frac{2(m_x-1)\sigma_{xw}^4}{nm_x^2}+\frac{2(m_y-1)\sigma_{yw}^4}{nm_y^2} \quad (12)$$

Χρησιμοιποιώντας τα

$$Var(f(z)) \approx \left(\frac{df(z)}{dz} \right)^2_{z=E(z)} Var(z)$$

Επομένως

$$Var(\sqrt{z}) \approx \left(\frac{1}{2\sqrt{z}} \right)^2_{z=E(z)} Var(z) = \frac{1}{4E(z)} Var(z)$$

Και

$$Var(\sigma_d) = \frac{1}{4\sigma_d^2} \left(\frac{2\sigma_{\bar{d}}^2}{n-1} + \frac{2(m_x-1)\sigma_{xw}^4}{nm_x^2} + \frac{2(m_y-1)\sigma_{yw}^4}{nm_y^2} \right) =$$
$$\frac{1}{2\sigma_d^2} \left(\frac{\sigma_{\bar{d}}^2}{n-1} + \frac{(m_x-1)\sigma_{xw}^4}{nm_x^2} + \frac{(m_y-1)\sigma_{yw}^4}{nm_y^2} \right) \qquad (13)$$

Η διακύμανση της μέσης διαφοράς $\bar{d}$ εκτιμάται από το $\frac{\sigma_d^2}{n}$ και η μέση τιμή και η τυπική απόκλιση των διαφορών είναι ανεξάρτητα. Αντικαθιστώντας στον τύπο (13) τις εκτιμήσεις των διακυμάνσεων, η διακύμανση των ορίων συμφωνίας $\bar{d} \pm 1.96\sigma_d^2$ δίνεται από τον τύπο

$$Var(\bar{d} \pm 1.96\sigma_d^2) = \frac{\sigma_d^2}{n} + \frac{1.96^2}{2\sigma_d^2} \left(\frac{\sigma_{\bar{d}}^4}{n-1} + \frac{(m_x-1)s_{xw}^4}{nm_x^2} + \frac{(m_y-1)s_{yw}^4}{nm_y^2} \right) \quad (14)$$

Για $m_x = m_y = 2$ η (14) γίνεται

$$Var(\bar{d} \pm 1.96\sigma_d^2) = \frac{\sigma_d^2}{n} + \frac{1.96^2}{2\sigma_d^2} \left(\frac{\sigma_{\bar{d}}^4}{n-1} + \frac{s_{xw}^4}{4n} + \frac{s_{yw}^4}{4n} \right)$$

Και για μη επαναλαμβανόμενες μετρήσεις με $m_x = m_y = 1$, η τυπική απόκλιση σ_d αντικαθίσταται από τον εκτιμητή s_d και έχουμε

$$Var(\bar{d} \pm 1.96\sigma_d^2) = \frac{s_d^2}{n} + \frac{1.96^2}{2s_d^2}\frac{s_d^4}{n-1} = s_d^2\left(\frac{1}{n} + \frac{1.96^2}{2(n-1)}\right)$$

όπως στο 2.2. Αυτός ο τύπος μπορεί να χρησιμοποιηθεί για να εκτιμήσουμε τα 95% διαστήματα συμφωνίας.

Για τα δεδομένα μέτρησης της πίεσης το αίματος, η διακύμανσης των ορίων του διαστήματος συμφωνίας είναι

$$Var(\bar{D} \pm 1.96s_D) = \frac{438.859}{85} + \frac{1.96^2}{2\times 438.859}\left(\frac{358.493^2}{84} + \frac{2\times 37.4078^2}{9\times 85}\frac{2\times 83.1412}{9\times 85}\right) =$$
$$= 11.9941$$

Επομένως η τυπική απόκλιση είναι $\sqrt{11.9941} = 3.463 mmHg$. Το 95% διάστημα εμπιστοσύνης για το κάτω όριο του διαστήματος συμφωνίας είναι $-56.68 - 1.96\times 3.463$ έως $m_{yi} = m_y$ δηλαδή $(-63.5, -49.9) mmHg$ και για το πάνω όριο αντίστοιχα είναι $(18.7, 32.2) mmHg$.

Η τυπική απόκλιση είναι σχεδόν ίση με την τυπική απόκλιση που βρήκαμε για μόνο μία επανάληψη (3.64mmHg) στην παράγραφο 2.2. Η χρησιμότητα των επαναλήψεων είναι να ελαττώσει τις διαφορές λόγω της έλλειψης ακρίβειας που έχουν οι μέθοδοι, ενώ η αλληλεπίδραση επαναλήψεων-υποκειμένου παραμένει. Αν υπάρχουν μεγάλες διαφορές μεταξύ των υποκειμένων σε όλες τις επαναλήψεις, τότε οι επαναλαμβανόμενες μετρήσεις δεν θα βοηθήσουν και πολύ στην ακρίβεια των ορίων συμφωνιας. Οι δύο επαναλήψεις όμως είναι καλό να γίνονται, για να μελετάται η ακρίβεια των μεθόδων.

5.2 Άνισος αριθμός επαναλήψεων

Υπάρχουν και περιπτώσεις που δεν έχουμε ίσο αριθμο παρατηρήσεων ανά υποκείμενο, m_{xi} και m_{yi} για κάθε μέθοδο X και Y για το υποκείμενο i. Τέτοια δεδομένα μπορεί να προκύψουν όταν μετράμε κάποια υποκείμενα σε τακτα χρονικά διαστήματα κατά τη διάρκεια κάποιας διαδικασίας, η οποία δεν έχει συγκεκριμένη διάρκεια, οπως πχ. οι μετρήσεις κατά τη διάρκεια μιας

εγχείρησης. Ο πίνακας 4 δείχνει μετρήσεις του όγκου του αίματος που φεύγει από την καρδιά σε χρόνο ενός λεπτού $(dm^3/\min)$ κάνοντας μετρήσεις με καρδιογράφημα εμπέδησης IC και μα ρεαδιο-κοιλιογραφία RV.

Πίνακας 10 Όγκος αίματος $(dm^3/\min)$ που φεύγει από την καρδιά ανά λεπτό. Μετρήσεις με καρδιογράφημα εμπέδησης IC και μα ρεαδιο-κοιλιογραφία RV (δεδομένα απο Dr. LS Bowling[5])

Sub	RV	IC
1	7.83	6.57
1	7.42	5.62
1	7.89	6.9
1	7.12	6.57
1	7.88	6.35
2	6.16	4.06
2	7.26	4.29
2	6.71	4.26
2	6.54	4.09
3	4.75	4.71
3	5.24	5.5
3	4.86	5.08
3	4.78	5.02
3	6.05	6.01
3	5.42	5.67
4	4.21	4.14
4	3.61	4.2
4	3.72	4.61
4	3.87	4.68
4	3.92	5.04

Sub	RV	IC
5	3.13	3.03
5	2.98	2.86
5	2.85	2.77
5	3.17	2.46
5	3.09	2.32
5	3.12	2.43
6	5.92	5.9
6	6.42	5.81
6	5.92	5.7
6	6.27	5.76
7	7.13	5.09
7	6.62	4.63
7	6.58	4.61
7	6.93	5.09
8	4.54	4.72
8	4.81	4.61
8	5.11	4.36
8	5.29	4.2
8	5.39	4.36
8	5.57	4.2

Sub	RV	IC
9	4.48	3.17
9	4.92	3.12
9	3.97	2.96
10	4.22	4.35
10	4.65	4.62
10	4.74	3.16
10	4.44	3.53
10	4.5	3.53
11	6.78	7.2
11	6.07	6.09
11	6.52	7
11	6.42	7.1
11	6.41	7.4
11	5.76	6.8
12	5.06	4.5
12	4.72	4.2
12	4.9	3.8
12	4.8	3.8
12	4.9	4.2
12	5.1	4.5

Για τη μελέτη δεδομένων με ίσες επαναλήψεις (τύπος (6)) χρησιμοποιούμε το ότι η διακύμανση του μέσου n ανεξάρτητων μεταβλητών που έχουν ίδια μέση τιμή και διακύμανση $s_{yw}^2 = 0.138$, είναι σ^2/n. Αν W_i είναι η μέση τιμή m_i παρατηρήσεων με μέση τιμή μ και διακύμανση σ^2, έχει διακύμανση σ^2/m_i, τότε η αναμενόμενη διακύμανση των μέσων θα είναι

$$Var(Wi) = \frac{1}{n}\left(\sum \frac{1}{m_i}\right)\sigma^2 \qquad (15)$$

Για το υποκείμενο i έχουμε m_{xi} παρατηρήσεις με τη μέθοδο X και m_{yi} παρατηρήσεις με τη μέθοδο Y. Για κάθε υποκείμενο, υπολογίζουμε τις

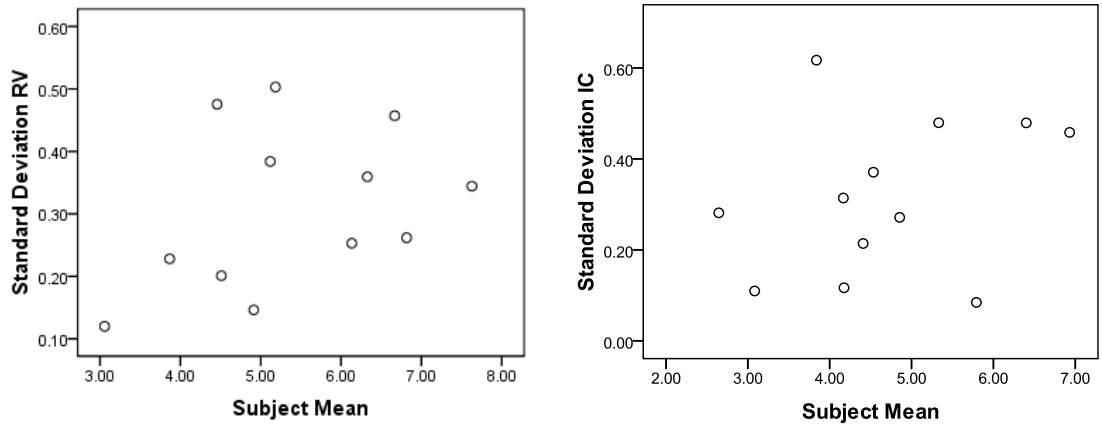
διαφορές μεταξύ των μέσων των μετρήσεων των δύο μεθόδων και υπολογίζουμε τη διακύμανση των διαφορών αυτών. Η αναμενόμενη διακύμανση βρίσκεται από τον τύπο

$$Var(\overline{D}) = \sigma_{xI}^2 + \frac{1}{n}\left(\sum \frac{1}{m_{xi}}\right)\sigma_{xw}^2 + \sigma_{yI}^2 + \frac{1}{n}\left(\sum \frac{1}{m_{yi}}\right)\sigma_{yw}^2 \quad (16)$$

Επομένως έχουμε

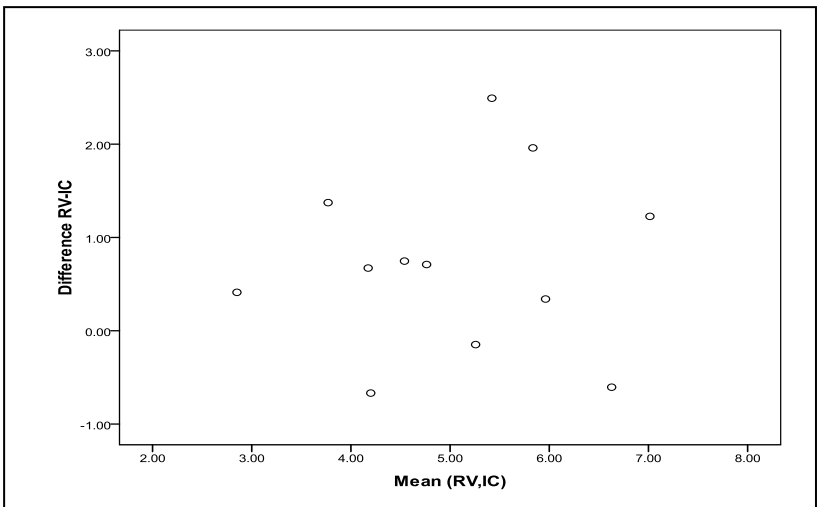
$$Var(D) = Var(\overline{D}) + \left(1 - \frac{1}{n}\left(\sum \frac{1}{m_{xi}}\right)\right)\sigma_{xw}^2 + \left(1 - \frac{1}{n}\left(\sum \frac{1}{m_{yi}}\right)\right)\sigma_{yw}^2 \quad (17)$$

όπου, αν $m_{xi} = m_x$ και $m_{yi} = m_y$, καταλήγουμε στον τύπο (6).



Διάγραμμα 13 Παρουσιάζονται τα διαγράμματα μεταξύ τυπικών αποκλίσεων και μέσων των υποκειμένων για τις δύο μεθόδους (RV, IC)

Στη συνέχεια κάνουμε ένα διάγραμμα των διαφορών των μέσων των δύο μεθόδων για κάθε υποκείμενο, με την μέση τιμή τους (διάγραμμα 14).



Διάγραμμα 14 Η διαφορά RV-IC προς το μέσο (RV,IC)

Και στα 3 διαγράμματα βλέπουμε πως ο ισχυρισμός της ανεξαρτησίας φαίνεται να είναι λογικός. Μετά υπολογίζουμε τα σ_{xw}^2 και σ_{yw}^2 κάνοντας one way ANOVA για κάθε μέθοδο ξεχωριστά. Τα αποτελέσματα είναι $s_{xw}^2 = 0.1072$ και $s_{yw}^2 = 0.1379$.

Σε αυτήν την περίπτωση τα $\frac{1}{n}\left(\sum \frac{1}{m_{xi}}\right)$ και $\frac{1}{n}\left(\sum \frac{1}{m_{yi}}\right)$ είναι ίσα με 0.2097, γιατί τα δεδομένα σε κάθε υποκείμενο είναι ισορροπημένα. Η διακύμανση της μέσης διαφοράς μεταξύ των μεθόδων για κάθε υποκείμενο είναι 0.9123. Για να υπολογίσουμε τη διακύμανση της διαφοράς μιας μέτρησης μιας μεθόδου και μιας μέτρησης μιας άλλης μεθόδου, χρησιμοποιούμε τον τύπο (17).

$$Var(IC - RV) = 0.9123 + (1 - 0.2097) \times 0.1072 + (1 - 0.2097) \times 0.1379 = 1.106$$

Η τυπική απόκλιση των διαφορών μεταξυ δύο μοναδικών παρατηρήσεων, μια απο κάθε μέθοδο, εκτιμάται από τον τύπο $\sigma_d = \sqrt{1.106} = 1.0517$. Η μέση διαφορά είναι 0.7092, οπότε το 95% διάστημα συμφωνίας για τη διαφορά των δυό μεθόδων είναι $\sigma_d^2 = \sigma_{dI}^2 + \sigma_{dw}^2$.

5.3 Επαναλαμβανόμενα δεδομένα σε ζεύγη

Οι μεθοδοι στις ενότητες 5.1 και 5.2 υποθέτουν πως οι πραγματικές τιμές δεν αλλάζουν καθώς κάνουμε επαναλαμβανόμενες μετρήσεις. Μερικές φορές θέλουμε να πετρήσουμε τις στιγμιαίες τιμές κάποιας συνεχόμενα μεταβαλλόμενης ουσίας. Δημιουργούμε πολλά ζευγάρια μετρήσεων των δύο μεθόδων για κάθε υποκείμενο, καθώς η πραγματική τιμή είναι διαφορετική για κάθε ζευγάρι. Μπορούμε να εκτιμήσουμε τα όρια συμφωνίας κάνοντας ανάλυση διακύμανσης για επαναλαμβανόμενες μετρήσεις. Βρίσκουμε τις διαφορές για κάθε ζευγάρι. Η διαφορά του j ζευγαριού του i υποκειμένου μοντελοποιείται ως εξής

$$D_{ij} = B + I_i + E_{ij}$$

Όπου B είναι η σταθερή μεροληψία, I_i η αλληλεπίδραση υποκειμένου χρόνου μεθόδου και E_{ij} το τυχαίο σφάλμα του i υποκειμένου για το j ζεύγος μετρήσεων. Οπότε το η διακύμανση του D_{ij} είναι

$$\sigma_d = \sigma_{dI}^2 + \sigma_{dw}^2$$

Μπορούμε να εκτιμήσουμε τα σ_{dI}^2 και σ_{dw}^2 ανάλυση διακύμανσης κατά έναν παράγοντα[6]. Έστω ότι για το υποκείμενο i έχουμε m_i ζεύγη μετρήσεων και ότι έχουμε n υποκείμενα. Έχουμε το μέσο τετραγωνικό σφάλμα MS_w και το between subject μέσο τετραγωνικό σφάλμα MS_b. Τα μέρη της διακύμανσης μπορούν να εκτιμηθούν από τους τύπους $\hat{\sigma}_{dw}^2 = MS_w$ και

$$\hat{\sigma}_{dI}^2 = \frac{\left(\sum m_i\right)^2 - \sum m_i^2}{(n-1)\sum m_i}(MS_b - MS_w)$$

Το άθροισμα αυτών των εκτιμήσεων μας δίνει το $\hat{\sigma}_d^2$. Η μέση μεροληψία εκτιμάται από τον τύπο $\frac{\sum m_i \bar{d}_i}{\sum m_i}$, όπου $\bar{d}_i$ είναι ο μέσος των διαφορών για το i υποκείμενο. Και έτσι βρίσκουμε το 95% διάστημα συμφωνίας. Μέθοδοι υπολογισμού σιαστ΄ματος συμφωνίας για συνδυασμούς ανάλυση διακύμανσης κατά έναν παράγοντα δίνονται από τους Burdick και Graybill[7].

6. Μη παραμετρική προσέγγιση για τη σύγκριση μεθόδων

Οι μεταξύ των υποκειμένων διαφορές δεν ακολουθούν πάντα κανονική κατανομή, όμως όπως αναφέρθηκε στην ενότητα 2, αυτό δεν θα μας επηρεάσει πολύ τα διάστημα συμφωνίας μας. Αν όμως υπάρχουν μία ή περισσότερες μεγάλες διαφορές μεταξύ των μεθόδων, μια μη παραμετρική προσέγγιση είναι πιο θεμιτή. Μια τέτοια περίπτωση συναντάμε στην αξιολόγηση των ηλεκτρονικών αυτόματων πιεσόμετρων, όπως είδαμε στον διάγραμμα 1. Γι αυτό το λόγο, το πρωτόκολο της βρετανικής εταιρίας υπέρτασης πρότεινε μία απλή μη παραμετρική μέθοδο για την αξιολόγηση αυτών των πιεσόμετρων[8].

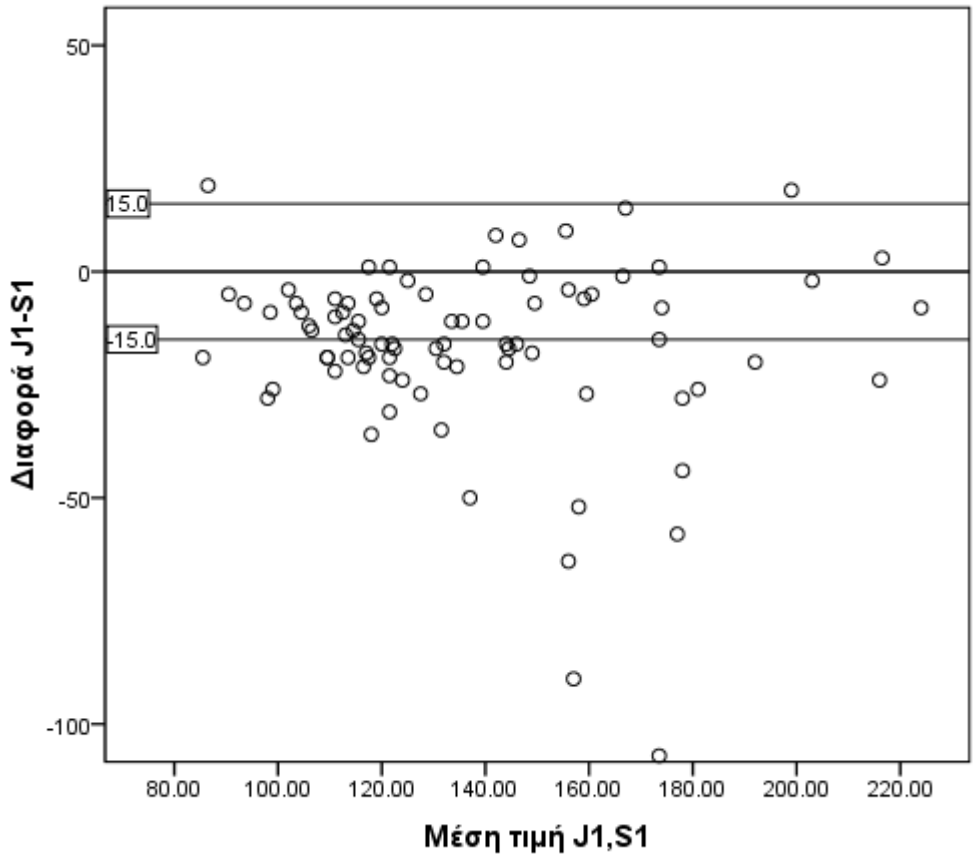
Μπορούμε να διατηρήσουμε την προσέγγιση που κάναμε στην ενότητα 2.1 μέχρι και το σημείο του διαγράμματος των διαφορών έναντι των μέσων των δύο μεθόδων. Έχουμε δύο τρόπους να περιγράψουμε τέτοια δεδομένα χωρίς να υποθέτουμε κανονική κατανομή των διαφορών. Πρώτον, μπορούμε να υπολογίσουμε το ποσοστό των διαφορών που υπερβαίνουν κάποιο όριο (πχ. 10mmHg) και μπορούμε να τα επισημάνουμε στο διάγραμμα διαφορών-μέσων. Δεύερον υπολογίζουμε τις τιμές που βρίσκονται σε κάποιο ποσοστό (πχ. 10%) των πιο ακραίων παρατηρήσεων. Για να το κάνουμε βάζουμε τα δεδομένα σε αύξουσα σειρά και διαγράφουμε από κάθε άκρη το μισό του ποσοστού που θέλουμε (πχ. εδω 5%). Και αυτά μπορούν να επισημανθούν στο διάγραμμα μας. Η δεύτερη μέθοδος είναι ενα μη παραμετρικό είδος διαστήματος συμφωνίας. Οι 2 μη παραμετρικές μέθοδοι είναι λιγότερο αξιόπιστες από αυτές που χρησιμοποιούμε την κανονικη κατανομή ως υπόθεση, ειδικά αν έχουμε μικρό δείγμα. Διαστήματα εμπιστοσύνης μπορούν να κατασκευαστούν χρησιμοποιώντας τη μέθοδο των διωνυμικών ποσοστών ή τα τυπικά σφάλματα του ποσοστημορίου.

Η βρεατνική εταιρία υπέρτασης, πρότεινε να χρησιμοποιήσουν τις διαφορές των ποσοστών βάζοντας όρια για την αξιολόγηση των αυτόματων πιεσόμετρων. Τα όρια αυτά παρουσιάζονται στον πίνακα 5.

Πίνακας 11 Βλέπουμε τα όρια για να βαθμολογηθεί ένα πιεσόμετρο με A,B,C,D. Οι διαφορές είναι συγκριτικά με ένα σφυγμομανόμετρο[8].

	Διαφορά(mmHg)		
Βαθμος	≤5	≤10	≤15
A	60	85	95
B	50	75	90
C	40	65	85
D	*Δεν πληρεί τα παραπάνω*		

Θα χρησιμοποιήσουμε τα δεδομένα του πίνακα 1 όπως στο διάγραμμα 1 με τη διαφορά ότι τώρα απεικονίζεται στο διάγραμμα το διάστημα (-15,15) για να μπορέσουμε εύκολα να διακρίνουμε το πόσο "απλώνονται" οι διαφορές γύρω από αυτό.



Διάγραμμα 15 Από τα δεδομένα του πίνακα 1, βλέπουμε τις διαφορές του σφυγμομανόμετρου με το αυτόματο πιεσόμετρο. Οι γραμμές απεικονίζουν το διάστημα (-15,15).

Τα ποσοστά που πετυχαίνει το πιεσόμετρο S του πρώτου πίνακα είναι ≤5→15.3%, ≤10→36.5% και ≤15→49.4%. Οπότε ο βαθμός του σύμφωνα με τον πίνακα 11 είναι D.

Η μη παραμετρική μέθοδος, παρ όλη την απλότητά της, μας δίνει αρκετα καλά αποτελέσματα. Χρησιμοποιούνταν και παλιότερα[9,10] αλλά ελάχιστα, γιατί ίσως το ότι είναι αρκετά απλή μέθοδος, δημιουργεί την εντύπωση ότι δεν κάνει σωστή ανάλυση στα δεδομένα.

7. Κατασκευή ποσοστιαίων θέσεων αναφοράς, σχετικά με την ηλικία, χρησιμοποιώντας απόλυτα κατάλοιπα

Για μια μεταβλητή που ακολουθεί κανονική κατανομή με τυπική απόκλιση (SD) s, τα παραμετρικά εκατοστιαία σημεία αναφοράς υπολογιζονται ως εξης: μέση τιμή ± ks, όπου k είναι συνήθως 1.645 ή 1.96 για 90% και 95% διαστήματα αναφοράς αντίστοιχα. Όταν με το πέρασμα του χρόνου η μεταβλητή αλλάζει, χρησιμοποιούμε πολυωνυμική παλινδρόμηση για να μοντελοποιήσουμε την μέση τιμή συναρτήσει του χρόνου. Όμως η τυπική απόκλιση, παρόλο που αλλάζει και αυτή με το χρόνο, δεν μοντελοποιείται συχνά. Σε πολλές περιπτώσεις πολλά δημοσιευμένα πεδία τιμών αναφοράς που σχετίχονται με το χρόνο υπολογίζονται ως μέση τιμή ± k x τυπική απόκλιση SD ενώ το SD δεν είναι σταθερό[12-14]. Ο Goldstein[15] πρότεινε να μοντελοποιηθεί η διακύμανση ως προς την ηλικία χωρίζοντας το χρόνο σε μικρά κομμάτια και κάνοντας παλινδρόμηση της διακύμανσης πάνω στο χρόνο. Ο Isaacs et al.[16] πρότεινε την ίδια διαδικασία χρησιμοποιώντας την τυπική απόκλιση SD. Ο Royston[17] πρότεινε μία απλούστερη προσέγγιση, όπου χωρίζει το χρόνο σε 3 κομμάτια και συγκρίνει την τυπική απόκλιση των παρατηρήσεων στο χαμηλότερο και στο υψηλότερο τρίτο του χρόνου. Εδώ θα δούμε έναν απλούστερο τρόπο να μοντελοποιούμε την τυπική απόκλιση, ο οποίος έχει πολλά πλεονεκτήματα σε σχέση με άλλες μεθόδους όταν τα δεδομένα ακολουθούν κανονική κατανομή ή μπορουν να μετασχηματιστούν ώστε να ακολουθουν κανονικη κατανομή.

ΜΕΘΟΔΟΣ

Αν η μέση τιμή μιας μεταβλητής Y μοντελοποιείται επαρκώς συναρτήσει του χρόνου, από κάποια γραμμική, τετραγωνική ή κυβική καμπύλη, όλη η πληροφορία για την τυπική απόκλιση των μετρήσεων βρίσκεται γύρω από την καμπύλη αυτή. Η παραμετρική μέθοδος, για να

βρούμε ένα πεδίο τιμών αναφοράς, βασίζεται στην υπόθεση ότι η μεταβλήτή μας ακολουθεί κανονική κατανομή σε όλους τους χρόνους, όπως και τα κατάλοιπα. Έτσι τα απόλυτα κατάλοιπα ακολουθούν μισή κανονική κατανομή. Η μέση τιμή μιας μισής τυπικής κανονικής κατανομής είναι $\sqrt{\pi/2}$ [7]. Έτσι η μέση τιμή των απόλυτων καταλοίπων πολλαπλασιασμένη με $\sqrt{\pi/2}$ είναι μια εκτίμηση της τυπικής απόκλισης των καταλοίπων. Από αυτό συνεπάγεται, αν η τυπική απόκλιση δεν είναι σταθερή με την πάροδο του χρόνου, οι προβλεπόμενες τιμές από την παλινδρόμηση των απόλυτων καταλοίπων συναρτήσει του χρόνου πολλαπλασιασμένα με $\sqrt{\pi/2}$ είναι εκτιμήσεις της τυπικής απόκλισης των καταλοίπων ως προς την ηλικία, άρα και της Y. Δεν θα χρειαστεί πιο πολύπλοκη από κυβική καμπύλη για να προσαρμοστεί ιακνοποιητικά στην τυπική απόκλιση.

Λόγω του ότι τα απόλυτα κατάλοιπα έχουν ασύμμετρη κατανομή, τα τεστ που βασίζονται στα τυπικά σφάλματα των παραμέτρων της παλινδρόμησης δεν είναι ενταλώς σωστά. Όμως δεν είναι τόσο σημαντικό πρόβλημα καθώς η επιλογή του καταλληλότερου μοντέλου δεν πρέπει να βασίζεται μονο στα p-values, αλλά και στους οπτικούς ελέγχους για την αποδοχή του διαστήματος αναφορας και της καλής προσαρμογής. Ο Royston[17] πρότεινε να κάνουμε διάγραμμα με τα κατάλοιπα έναντι του χρόνου και να ελέγχουμε την κανονικότητα τους. Αν η τυπική αποκλιση δεν είναι σταθερή, οι διαδικασίες πρέπει να γίνονται χρησιμοποιώντας τα τυποποιημένα κατάλοιπα. Αυτά προσδιορίζονται ως εξής

$$\frac{y_i - y_{pred}(x_i)}{s_{pred}(x_i)}$$

για το υποκείμενο i με την μέτρηση y_i στο χρόνο x_i, όπου y_{pred} και s_{pred} είναι οι εκτιμώμενες ως προς το χρόνο τυπικές αποκλίσεις SD του Y. Πιθανό είναι, να μην είναι τόσο εύκολο να δούμε με το μάτι την μη-γραμμικότητα, όταν το Y έχει ασύμμετρη κατανομή, απ' όσο όταν έχει συμμετρική κατανομη. Μπορεί να μας βοηθήσει όμως μια μέθοδος εξομάλυνσης όπως η μέθοδος του Cleveland[18] ως βοήθημα για να προσδιοριστεί το σχήμα του του πιο κατάλληλου μοντέλου.

Η προτεινόμενη μέθοδος δεν συνεπάγεται την κατασκευή αυθαίρετων ομάδων χρόνου και δεν γίνονται επιπλέον υποθέσεις, αλλά είναι πιο

ελκυστική ίσως επειδή είναι απλή υπολογιστικά. Οι μέθοδοι που στηρίζονται στην ομαδοποίηση ως προς το χρόνο, περιλαμβάνουν ξεχωριστές αναλύσεις συνοπτικών στατιστικών στοιχείων, για να πάρουμε μια εξίσωση παλινδρόμησης η οποία θα πρέπει να εισαχθεί στην κύρια αναλυσή μας. Αυτή η ανάλυση δύο σταδίων είναι κουραστική και υπάρχει κίνδυνος σφάλματος λόγω αντιγραφής. Είναι ιδιαίτερα κουραστικό όταν η ανάλυση πρέπει να επαναληφθεί πολλές φορές για την εύρεση του καταλληλότερου μετασχηματισμού. Ένα ακόμα πλεονέκτημα της μεθόδου που στηρίζεται στα απόλυτα κατάλοιπα είναι ότι οδηγούμαστε σε ένα χρήσιμο γράφημα για να βοηθηθούμε στην αξιολόγηση της καλής προσαρμογής.

Η εκτιμώμενη τυπική απόκλιση ως προς το χρόνο SD, s_{pred}, θα μπορούσε να χρησιμοποιηθεί για να επανεκτημηθεί το μοντέλο για τη μέση τιμή σταθμίζοντας με το $1/s^2_{pred}(x_i)$, όπως προτάθηκε από τον Aitkin[19]. Το αποτέλεσμα αυτής της επαναλαμβανόμενης διαδικασίας εξαρτάται από τη διακύμανση των δεδομένων και το μέγεθος του δείγματος, αλλά είναι απίθανο να κάνει μεγάλη διαφορά αν η διακύμανση δεν αλλάξει σημαντικά.

Παράδειγμα

Τα παρακάτω δεδομένα προέρχονται από μια έρευνα που ως σκοπό έχει να αναπτύξει ολοκληρωμένες ποσοστημόρια αναφοράς για το εμβρυακό μέγεθος σε σχέση με το χρόνο κύησης. Τα ποσοστημόρια αυτά χρησιμοποιούνται ευρέως για την αξιολόγηση της ανάπτηξης του εμβρύου. Για να γίνουν αυτές οι μετρήσεις χρησιμοποιήθηκαν υπερηχογραφήματα σε διάφορες γυναίκες με εγκυμοσύνη. Μια μέτρηση πραγματοποιήθηκε από κάθε υποκείμενο για τη δημιουργία του ποσοστημορίου αναφορας. Εδώ (πινακας 6) έχουμε τις μέσες τιμές, τις τυπικες αποκλίσεις και τον αρθμο των υποκειμένων για κάθε γκρουπ που προέρχονται από 450 μετρήσεις για το μέγεθος του ποδιού του εμβρύου.

Πίνακας 12 *Μετρήσεις για το μέγεθος του ποδιού του εμβρύου σε ομάδες μίας εβδομάδας(α) και σε 6 ίσες ομάδες(β).*

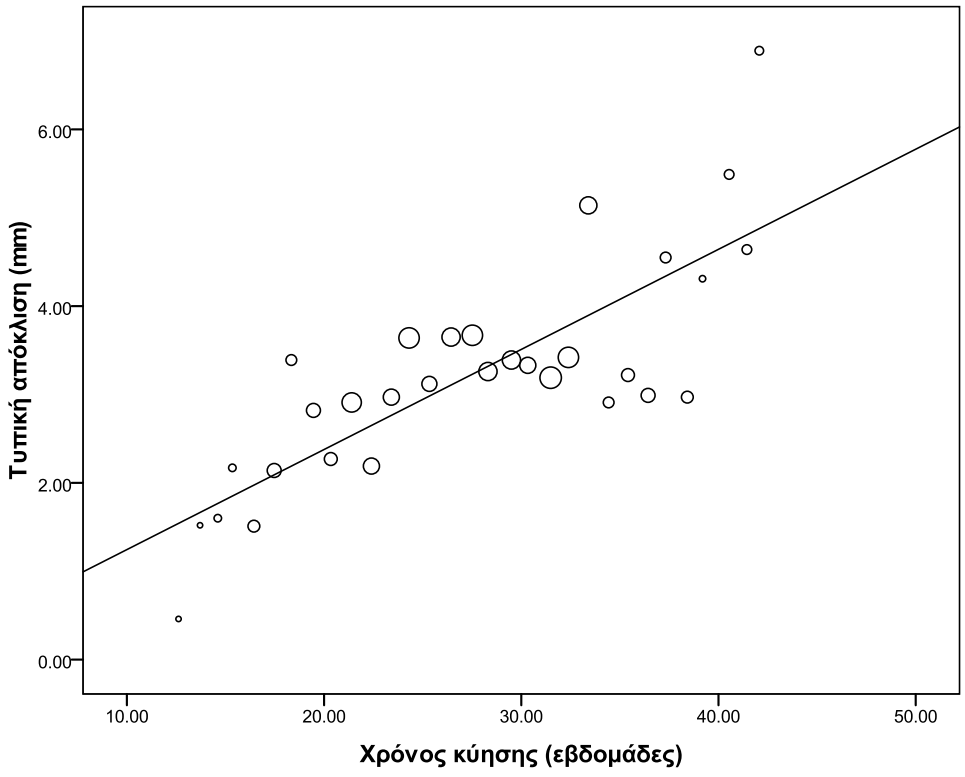
α) 31 ομάδες ανά μια εβδομάδα κύησης			β) 6 ομάδες ως προς το χρόνο κύησης		
Μέσος χρόνος κύησης	Τυπική απόκλιση μήκους ποδιού	N	Μέσος χρόνος κύησης	Τυπική απόκλιση μήκους ποδιού	n
12.62	0.46	3	16.03	1.86	50
13.71	1.52	3	20.78	2.84	82
14.61	1.60	7	25.62	3.43	98
15.35	2.17	7	30.63	3.33	106
16.44	1.51	12	35.42	4.02	69
17.47	2.14	16	40.32	4.82	45
18.34	3.39	11			
19.46	2.82	16			
20.34	2.27	15			
21.40	2.91	21			
22.40	2.19	18			
23.41	2.97	18			
24.31	3.64	22			
25.34	3.12	17			
26.44	3.65	20			
27.52	3.67	22			
28.31	3.26	20			
29.50	3.39	20			
30.33	3.33	18			
31.49	3.19	24			
32.39	3.42	22			
33.40	5.14	19			
34.43	2.91	11			
35.41	3.22	15			
36.43	2.99	16			
37.32	4.55	11			
38.42	2.97	12			
39.19	4.31	6			
40.54	5.49	10			
41.44	4.64	10			
42.07	6.89	8			

Η μέση τιμή του μήκους του ποδιού μοντελοποιείται με πολυωνυμική παλινδρόμηση $2^{ου}$ βαθμού δίνοντας μας πολύ καλη προσαρμογή στα δεδομένα. Όμως κατά την ανάλυση οπτικά φαίνεται να έχουν καλύτερη προσαρμογή με παλινδρόμηση $3^{ου}$ βαθμού $y=a+bx+cx^3$, παρόλα αυτά στην πράξη η διαφορά είναι αμελητέα, εκτός των ακραίων τιμών του X. Το μοντέλο $3^{ου}$ βαθμού για το μήκος των ποδιών των εμβρύων είναι

$$y_{pred} = 35.08 + 3.574x - 0.0004406x^3 \qquad (18)$$

Η τυπική απόκλιση των καταλοίπων αυτού του μοντέλου για κάθε εβδομάδα κύησης δίνεται στον πίνακα 12α και στο διάγραμμα 16 όπου το μέγεθος των κύκλων είναι ανάλογο του μεγέθους του δείγματος.

Διάγραμμα 16 Τυπική απόκλιση μήκους ποδιού ως προς το χρόνο κύησης, με ομαδοποίηση ανα εβδομάδα.



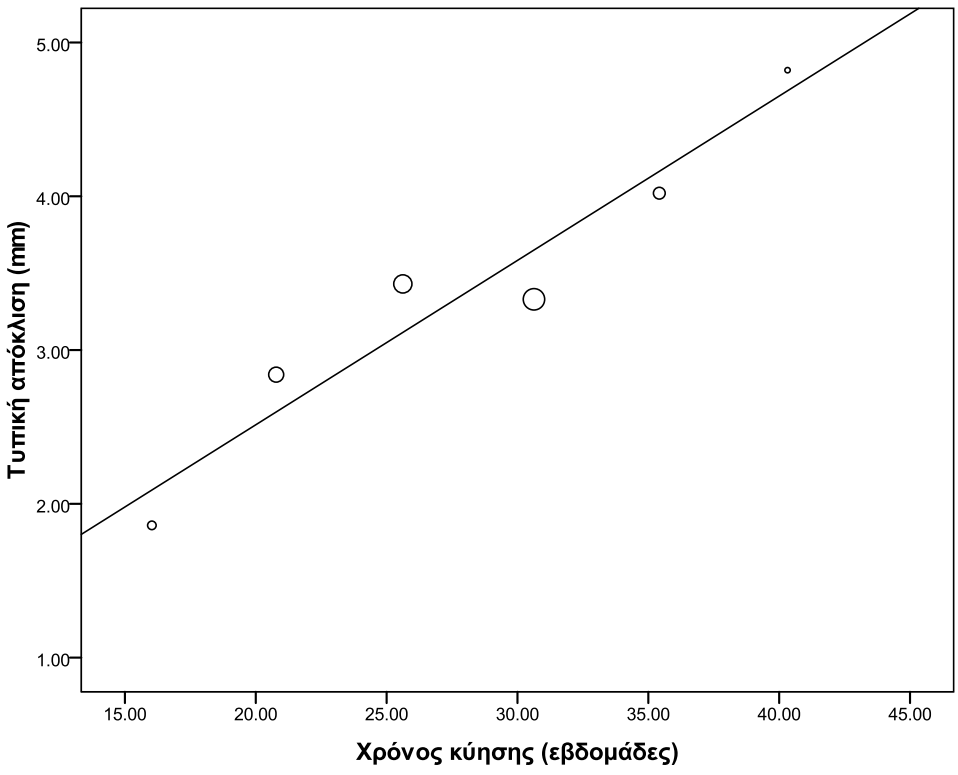
Ο τύπος της γραμμικής παλινδρόμησης της τυπικής απόκλισης σταθμισμένος στο χρόνο κύησης ειναι

$$s_{pred} = 0.566 + 0.0979x$$

Στο διάγραμμα 16 φαίνεται πως μάλλον το μοντέλο 3^{ου} βαθμού είναι το καταλληλότερο και πράγματι το μοντέλο 3^{ου} βαθμού είναι στατιστικά σημαντικό. Όμως έναντι της ευθείας γραμμης (μοντέλο 1^{ου} βαθμού) η βελτίωση που παρουσιάζει το μοντέλο 3^{ου} βαθμού είναι είναι πολύ μικρή. Το μοντέλο 1^{ου} βαθμού είναι επαρκές για την περιγραφή των δεδομένων. Στον πίνακα 12β και στο διάγραμμα 17 βλέπουμε τα αποτελέσματα για τις 6 ομάδες του χρόνου κύησης. Η σταθμισμένη ανάλυση παλινδρόμησης έδωσε τον τύπο της τυπικής απόκλισης

$$s_{pred} = 0.576 + 0.0997x$$

Διάγραμμα 17 Τυπική απόκλιση μήκους ποδιού ως προς το χρόνο κύησης, με ομαδοποίηση σε έξι ομάδες.



Η απλή γραμμική παλινδρόμηση μας δίνει τον τύπο 0.52+0.07691x. Η εκτιμώμενη τυπική απόκλιση βρίσκεται από αυτόν τον τύπο πολλαπλασιασμένο με $\sqrt{\pi/2}$.

$$s_{pred} = 0.651 + 0.0964x$$

Παρατηρούμε ότι ο τύπος αυτός είναι αρκετά κοντα στους τύπους που βρήκαμε έχοντας κάνει χρονικές ομάδες κύησης.

Ο Royston[17] περιέγραψε μία προσέγγιση για τον υπολογισμό εκατοστιαίων σημείων αναφοράς που σχετίζονται με το χρόνο. Δεν έδωσε ιδιαίτερη προσοχή, στην πιθανότητα να αλλάζει η τυπική απόκλιση στο χρόνο, παρόλο που σε κάποιες περιπτώσεις, όπως η μέτρηση του μεγέθους των εμβρύων, η τυπική απόκλιση αλλάζει. Χώρισε τα δεδομένα σε 3 ίσες χρονικές ομάδες και σύγκρινε τις τυπικές αποκλίσεις στο μικρότερο και μεγαλύτερο τρίτο των δεδομένων το οποιο είναι ένα τεστ γραμμικής τάσης της τυπικής απόκλισης. Ενώ αυτή η μέθοδος λειτουργεί καλά όταν έχουμε γραμμική αλλαγή της τυπικής απόκλισης, δεν μπορεί να εντοπίσει μη-γραμμικές(συμπεριλαμβανομένων των μη-μονότονων) αλλαγές της. Μία παρόμοια προσέγγιση, αλλά με περισσότερες χρονικές ομάδες, μας δίνει τη δυνατότητα να μοντελοποιήσουμε την τυπική απόκλιση ή διακύμανση, συναρτήσει του χρόνου.

Αν και οι 2 μέθοδοι εμπεριέχουν αυθαίρετη ομαδοποίηση των δεδομένων, λειτουργούν καλά σε πολλές περιπτώσεις. Παρόλα αυτά, η προτεινόμενη μέθοδος βασίζεται στη χρήση των απόλυτων καταλοίπων και δεν χρειάζεται ομαδοποίηση στο χρόνο και έτσι είναι ευκολότερη στην εφαρμοφή της. Επίσης είναι καλύτερο να λειτουργούμε με τα υποκείμενα ξεχωριστά παρά με ομάδες. Στην πραγματικότητα κάθε παρατήρηση μας δίνει μία εκτίμηση της τυπικής απόκλισης στο χρόνο. Μπορούμε επίσης εύκολα να χρησιμοποιήσουμε αυτή τη μέθοδο όταν η τυπική απόκλιση δεν αλλάζει γραμμικα με την πάροδο του χρόνου.

Η προτεινόμενη προσέγγιση σχεδιάστηκε για να αποφευχθεί η ομαδοποίηση. Επίσης προτάθηκε και από τον Efron[20] ως μέθοδος για τον έλεχχο του πόσο σημαντική είναι η διακύμανση στην εξάπλωση των ποσοστημορίων της παλινδρόμησης. Η μέθοδος του Efron περιλαμβάνει την

εκτίμηση ποσοστημορίων ασυμμετρικά σταθμισμένα τετράγωνα καταλοίπων, πάνω και κάτω από την γραμμή παλινδρόμησης, όπως και την πρόταση των Koenker και Bassett[21] για τη χρήση των ασυμμετρικά σταθμισμένων απόλυτων καταλοίπων. Ο Efron παρατήρησε πως οποιαδήποτε εκατοστιαία θέση μπορεί να εκτιμηθεί με τον τρόπο που παρουσιάστηκε εδώ και ενώ λειτουργεί καλα στο παράδειγμα του, προτείνει μια πιο γενική προσέγγιση. Επίσης, σημειώνει πως η αναλογία μεταξύ της τυπικής απόκλισης και της μέσης τιμής των απολύτων καταλοίπων μπορεί να συγκριθεί με την τιμή $\sqrt{\pi/2} \approx 1,25$ για να έχουμε μια ιδέα για την καταλληλότητα της προσέγγισης. Η ιδέα της παλινδρόμησης των απολύτων καταλοίπων με την επεξηγηματική μεταβλητή εμφανίζεται και σε άρθρο του Glejser[22], ως η βάση για τον έλεγχο της ετεροσκεδαστικότητας.

Η υπόθεση του ελέγχου των συντελεστών της πολυωνυμικής παλινδρόμησης χρησιμοποιώντας τα απόλυτα κατάλοιπα δεν είναι απολύτως σωστή λόγω του ότι να απόλυτα κατάλοιπα δεν ακολουθούν κανονική κατανομή το οποίο όμως σε κάποιες περιπτώσεις δεν είναι πρόβλημα.

8. Βιβλιογραφία

1. **Altman DG, Bland JM.** The analysis of blood pressure data. In O'Brien E, O'Malley K eds. Blood pressure measurement. Amsterdam: Elsevier, 1991: 287±314.
2. **Bland JM, Altman DG.** *Statistical methods for assessing agreement between two methods of clinical measurement*. Lancet 1986; i: 307±10.
3. **Altman DG.** *Calculating age-related reference centiles using absolute residuals. Statistics in Medicine* 1993; 12: 917±24.
4. **Lucas A, Hudson GJ, Simpson P, Cole TJ, Baker BA.** *An automated enzymic micromethod for the measurement of fat in human milk.* Journal of Dairy Research 1987; 54: 487±92.
5. **Bowling LS, Sageman WS, O'Connor SM, Cole R, Amundson DE.** *Lack of agreement between measurement of ejection fraction by impedance cardiography versus radionuclide ventriculography.* Critical Care Medicine 1993; 21: 1523±27.
6. **Searle SR, Cassela G, McCulloch CE.** *Variance components*. New York: New York, 1992.
7. **Burdick RK, Graybill FA.** *Confidence intervals on variance components*. New York: Dekker, 1992.
8. **O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman DG, Bland M, Coats A, Atkins N.** *The British Hypertension Society protocol for the evaluation of blood pressure measuring devices.* Journal of Hypertension 1993; 11(suppl. 2): S43±S62.
9. **Polk BF, Rosner B, Feudo F, Vandenburgh M.** *An evaluation of the Vita-Stat automatic blood pressure measuring device*. Hypertension 1980; 2: 221±27.
10. **Latis GO, Simionato L, Ferraris G.** *Clinical assessment of gestational age in the newborn infant.* Early Human Development 1981; 5: 29-37.
11. **Cotes PM, DoreÂ CJ, Liu Yin JA, Lewis SM, Messinezy M, Pearson TC, Reid C.** *Determination of serum immunoreactive*

erythropoietin in the investigation of erythrocytosis. New England Journal of Medicine 1986; 315: 283±87.

12. **Walther, F. J., Siassi, B. and Wu, P. Y. K.** '*Echocardiographic measurement of left ventricular stroke volume in newborn infants: a correlative study with pulsed Doppler and M-mode echocardiography'*, Journal of Clinical Ultrasound, 14, 37-41 (1986).
13. **Blair, J. I., Carachi, R., Gupta, R., Sim, F. G., McAllister, E. J. and Weston, R.** '*Plasma CI fetoprotein reference ranges in infancy: effect of prematurity'*, Archives of Disease in Childhood, 62, 362-369 (1987).
14. **Platt, L. D., Medearis, A. L., DeVore, G. R., Horenstein, J. M., Carlson, D. E. and Brar, H. S.** '*Fetal foot length: relationship to menstrual age and fetal measurements in the second trimester'*, Obstetrics and Gynecology, 71, 526-531 (1988).
15. **Goldstein, H.** '*The construction of standards for measurements subject to growth'*, Human Biology, 44, 255-261 (1972).
16. **Isaacs, D., Altman, D. G., Tidmarsh, C. E., Valman, H. B. and Webster, A. D. B.** '*Serum immunoglobulin concentration in preschool children measured by laser nephelometry: reference ranges for IgG, IgA, IgM'*, Journal of Clinical Pathology, 36, 1193-1 196 (1983).
17. **Royston, P.** '*Constructing time-specific reference ranges'*, Statistics in Medicine, 10, 675-690 (1991).
18. **Cleveland, W. S.** '*Robust locally-weighted regression and smoothing scatter plots'*, Journal of the American Statistical Association, 74, 829-836 (1 979).
19. **Aitkin, M. A.** *'Modelling variance heterogeneity in normal regression using GLIM*', Applied Statistics, 36, 332-339 (1987).
20. **Efron, B.** '*Regression percentiles using asymmetric squared error loss'*, Statistica Sinica, 1,93-125 (1991).
21. **Koenker, R. and Bassett, G.** *'Regression quantiles'*, Econometrica, 46, 33-50 (1978).
22. **Glejser, H.** '*A new test for heteroskedasticity'*, Journal of the American Statistical Association, 64, 316-323 (1969)