

Observed values:

U = 7,42887261670424

D = 0,251217484464214

P-value = 0,0471

Linear	Theta	Index	Ranks	Theta Star
23,9	292,5	1	31	360
		2	25	290,3
		3	6,5	75,5
		4	25	290,3
		5	25	290,3
		6	25	290,3
		7	1	11,6
25,2	225	8	25	290,3
		9	25	290,3
		10	15,5	180
		11	4	46,5
		12	3	34,8
		13	5	58,1
		14	15,5	180
26,7	180	15	9	104,5
26,7	202,5	16	9	104,5
		17	15,5	180

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
 Γενικό Τμήμα
 Εργαστήριο Μαθηματικών & Στατιστικής
 ΠΜΣ "Θετικές Επιστήμες στη Γεωπονία"
 Κλάδος "Γεωπληροφορική"

Μεταπτυχιακή Διατριβή
**Στατιστική Ανάλυση Κυκλικών Δεδομένων:
 Μια Γενική Επισκόπηση και μια Εφαρμογή
 σε Μετεωρολογικά Δεδομένα**
 Λότοτζης Ν. Μαργαρίτης

Επιβλέπων καθηγητής:
 Παπαδόπουλος Γ., Επίκ. Καθηγητής

Calculate Observed Values

Number of Observations: 31
 Number of Permutations: 10000

Simulate

Reset

Exit

2 26 23 5 22 1 6 13 14 30 24 7 11 19 21 9 8 31 18 25 4
 Perm's 1 U= 5,37497428222561

8 7 1 23 21 11 10 4 30 26 14 5 16 12 17 3 28 6 25 27 1
 Perm's 2 U= 9,96263113020397

18 4 25 11 16 10 9 27 8 12 30 14 3 6 26 7 17 13 19 21
 Perm's 3 U= 0,699369697926127

5 1 31 10 23 18 22 2 4 13 19 16 11 17 24 27 7 26 6 3 1
 Perm's 4 U= 3,88125507925796

15 29 8 20 2 18 9 23 14 3 7 22 10 25 28 31 5 17 21 27
 Perm's 5 U= 1,73126125228009

Simulation program that determines the significance level (P-value) of Mardia's non-parametric test for circular-linear correlation (Mardia, 1976; Biometrika, 63: 403-405)

Αθήνα, Αύγουστος 2011

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Γενικό Τμήμα

Εργαστήριο Μαθηματικών & Στατιστικής

ΠΜΣ “Θετικές Επιστήμες στη Γεωπονία”

Κλάδος “Γεωπληροφορική”

Μεταπτυχιακή Διατριβή

Στατιστική Ανάλυση Κυκλικών Δεδομένων: Μια
Γενική Επισκόπηση και μια Εφαρμογή σε
Μετεωρολογικά Δεδομένα

Λότοτζης Ν. Μαργαρίτης

Επιβλέπων καθηγητής:

Παπαδόπουλος Γ., Επίκ. Καθηγητής

Αθήνα, Αύγουστος 2011

Μεταπτυχιακή Διατριβή

Στατιστική Ανάλυση Κυκλικών Δεδομένων: Μια
Γενική Επισκόπηση και μια Εφαρμογή σε
Μετεωρολογικά Δεδομένα

Λότοτζης Ν. Μαργαρίτης

Επιβλέπων καθηγητής:

Παπαδόπουλος Γ., Επίκ. Καθηγητής

Εξεταστική Επιτροπή:

Σωτηράκογλου Κ., Αναπλ. Καθηγήτρια (Μέλος)

Τσίρος Ι., Αναπλ. Καθηγητής (Μέλος)

Περίληψη

Με τον όρο “κυκλικά δεδομένα” εννοούμε δεδομένα διεύθυνσης και κατεύθυνσης. Οι συνήθεις στατιστικές μέθοδοι που εφαρμόζονται σε μη κυκλικά δεδομένα δεν μπορούν να εφαρμοστούν στην παρουσίαση και ανάλυση των κυκλικών δεδομένων παρ’ ότι οι αντίστοιχες έννοιες είναι ανάλογες (συνάρτηση κατανομής, περιγραφικά μέτρα, διαστήματα εμπιστοσύνης, έλεγχοι υποθέσεων κ.λπ.). Επίσης, η σχετική βιβλιογραφία είναι περιορισμένη, καθώς τα συγγράμματα στατιστικής και στατιστικής ανάλυσης σε λίγες περιπτώσεις πραγματεύονται θέματα ανάλυσης κυκλικών δεδομένων.

Η παρούσα μελέτη αποτελεί μια γενική επισκόπηση των στατιστικών μεθόδων παρουσίασης και ανάλυσης των κυκλικών δεδομένων, με έμφαση στους ελέγχους υποθέσεων και τη συσχέτιση μεταξύ κυκλικών και μη κυκλικών μεταβλητών. Για το τελευταίο αναπτύξαμε πρόγραμμα προσομοίωσης σχετικού παραμετρικού ελέγχου. Τέλος, έγινε εφαρμογή αυτών των μεθόδων σε πραγματικά μετεωρολογικά δεδομένα, τα οποία αφορούν στη θερμοκρασία της ατμόσφαιρας και την κατεύθυνση του ανέμου στο σταθμό “Λόφος των Νυμφών”, Θησείο, του Εθνικού Αστεροσκοπείου Αθηνών, για το μήνα Ιούλιο 2007. Διαπιστώθηκε, μεταξύ άλλων, ότι η μέση ημερήσια κατεύθυνση του ανέμου διαφέρει μεταξύ της περιόδου με φυσιολογικές τιμές θερμοκρασίας και της περιόδου καύσωνα και ότι η μέση ωριαία θερμοκρασία συσχετίζεται με την, αντίστοιχη, μέση ωριαία κατεύθυνση του ανέμου.

Λέξεις κλειδιά:

κυκλικά δεδομένα, κυκλική κατανομή, μέση κατεύθυνση, κυκλική διακύμανση, ομοιόμορφη κατανομή, κατανομή von Mises, κυκλική–γραμμική συσχέτιση

Summary

By the term “circular data” we refer to directional and axial data. The common statistical methods which apply to non-circular data cannot be applied to the presentation and analysis of circular data, even though the corresponding concepts are analogous (distribution function, descriptive measures, confidence intervals, hypothesis testing etc.). Furthermore, the relevant bibliography is scant as the statistics and statistical analysis books treat the topics of circular data analysis in only a few cases.

The present thesis constitutes a general survey of the statistical methods of presentation and analysis of circular data, with emphasis to hypothesis testing and correlation between circular and non-circular variables. For the latter we developed a simulation program of a relative non-parametric test. Finally, we applied these methods to real meteorological data, which refer to air temperature and wind direction at “Nymphs’ Hill”, Thiseio, National Observatory of Athens, during July 2007. Analysis mainly suggested that mean-daily wind direction is different between days with normal air temperature and periods of heat-wave and that mean-hourly wind direction correlates with the corresponding mean-hourly air temperature.

Key-words:

circular data; circular distribution; mean direction; circular variance;
uniform distribution; von Mises distribution; circular-linear correlation

Περιεχόμενα

Εισαγωγή	xxiii
1 Γραφική αναπαράσταση και περιγραφικά μέτρα	1
1.1 Γραφική αναπαράσταση	1
1.2 Περιγραφικά μέτρα	6
1.2.1 Μέτρα θέσης	6
1.2.2 Μέτρα διασποράς, λοξότητας και κύρτωσης	16
2 Κυκλικές κατανομές	25
2.1 Εισαγωγή	25
2.2 Η ομοιόμορφη κατανομή	27
2.3 Η κατανομή von Mises	29
2.4 Άλλες κυκλικές κατανομές	32
3 Διαστήματα εμπιστοσύνης και έλεγχοι υποθέσεων	35
3.1 Σημειακές εκτιμήσεις και εκτιμήσεις διαστήματος	35
3.1.1 Σημειακή εκτίμηση παραμέτρων	35
3.1.2 Εκτιμήσεις διαστήματος	36
3.2 Έλεγχοι τυχαιότητας και καλής προσαρμογής	41
3.2.1 Rayleigh test	41
3.2.2 V test	46
3.2.3 Hodges-Ajne test	50
3.2.4 modified Hodges-Ajne test	52
3.2.5 Rao's spacing test	53
3.2.6 range test	55
3.2.7 chi-squared test	57
3.2.8 Kuiper's test	59
3.2.9 Watson's U^2 test	61

3.2.10	Cox's test	62
3.2.11	Σύνοψη	65
3.3	Έλεγχοι δύο δειγμάτων	67
3.3.1	Watson-Williams test	67
3.3.2	Mardia-Watson-Wheeler test	79
3.3.3	chi-squared test	88
3.3.4	Kuiper's test	89
3.3.5	Watson's U^2 test	91
3.3.6	rank-sum test	98
3.3.7	run test	99
3.3.8	Ένα παραμετρικό τεστ για την παράμετρο συγκέντρωσης	100
3.3.9	Έλεγχος μεταβλητότητας γωνιακών αποκλίσεων	101
3.3.10	Ένας απαραμετρικός έλεγχος της γωνιακής διασποράς	102
4	Ανάλυση δεύτερης τάξης	109
4.1	Εισαγωγή	109
4.2	Γενικός μέσος k μέσων γωνιών	111
4.3	Όρια εμπιστοσύνης για τη μέση γωνία και το μέσο μέτρο δεύτε- ρης τάξης	114
4.4	Ένας έλεγχος για την παράμετρο συγκέντρωσης κ	118
4.5	Hotelling's one-sample test	119
4.6	Moore's test	120
4.7	Hotelling's two-sample test	122
4.8	Mardia's two-sample test	123
4.9	Ένας παραμετρικός έλεγχος δύο δεύτερης τάξης δειγμάτων ζευ- γαρωτών παρατηρήσεων για τη μέση κατεύθυνση	126
4.10	Ένας απαραμετρικός έλεγχος δύο δεύτερης τάξης δειγμάτων ζευγαρωτών παρατηρήσεων για τη μέση κατεύθυνση	127
5	Κυκλική συσχέτιση και παλινδρόμηση	129
5.1	Κυκλική–κυκλική συσχέτιση	129
5.2	Κυκλική–γραμμική συσχέτιση	133
5.2.1	Παραμετρική μέθοδος	134
5.2.2	Απαραμετρική μέθοδος	153
5.2.3	Σύγκριση παραμετρικής και απαραμετρικής μεθόδου	170

6 Συμπεράσματα	177
6.1 Αποτελέσματα των ελέγχων τυχαιότητας και κανονικότητας της μεταβλητής της κατεύθυνσης του ανέμου	177
6.1.1 Έλεγχοι τυχαιότητας και κανονικότητας των πρώτης τάξης δειγμάτων	177
6.1.2 Έλεγχοι τυχαιότητας και κανονικότητας των δεύτερης τάξης δειγμάτων	178
6.1.3 Έλεγχοι ισονομίας των κατανομών των δεύτερης τάξης δειγμάτων	179
6.2 Συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου	181
6.2.1 Έλεγχος συσχέτισης των πρώτης τάξης δειγμάτων . . .	182
6.2.2 Έλεγχος συσχέτισης των δεύτερης τάξης δειγμάτων . .	183
Παράρτημα	α΄
Βιβλιογραφία	ιγ΄
Ευρετήριο	ιε΄

Κατάλογος πινάκων

1	Πίνακας σύμβασης για το χαρακτηρισμό της στατιστικής σημαντικότητας των αποτελεσμάτων.	xxv
2	Πίνακας περιγραφής των μεταβλητών της μελέτης	xxviii
3	Δεδομένα της μελέτης — Κατεύθυνση του ανέμου (α)	xxx
4	Δεδομένα της μελέτης — Κατεύθυνση του ανέμου (β)	xxxi
5	Δεδομένα της μελέτης — Θερμοκρασία (α)	xxxii
6	Δεδομένα της μελέτης — Θερμοκρασία (β)	xxxiii
1.1	Αντίστοιχες σχέσεις μεταξύ κυκλικών και γραμμικών στατιστικών	15
3.1	Αντιστοιχίες στατιστικών συναρτήσεων και πληθυσμιακών παραμέτρων.	35
3.2	Εφαρμογή του Rayleigh test στις μεταβλητές WINDIR_D7 και WINDIR_D24	43
3.3	Εφαρμογή του Rayleigh test στη μεταβλητή WINDIR_DAY	44
3.4	Εφαρμογή του Rayleigh test στη μεταβλητή WINDIR_DPREV	45
3.5	Εφαρμογή των Rayleigh test και V test στις μεταβλητές WINDIR_H15 και WINDIR_H17	49
3.6	Εφαρμογή του Rao's spacing test στις μεταβλητές WINDIR_H15 και WINDIR_H17	54
3.7	Έλεγχοι τυχειότητας.	58
3.8	Εφαρμογή των Kuiper's test και Watson's U^2 test στις μεταβλητές WINDIR_D10 και WINDIR_D24	63
3.9	Εφαρμογή των ελέγχων καλής προσαρμογής στις μεταβλητές WINDIR_HOUR και WINDIR_DAY	64
3.10	Έλεγχοι καλής προσαρμογής.	66

3.11 Έλεγχος των συνθηκών για την εφαρμογή του Watson-Williams test στις μεταβλητές WINDIR_DAY [1-17] και WINDIR_DAY [18-31]	71
3.12 Εφαρμογή του Watson-Williams test στις μεταβλητές WINDIR_DAY [1-17] και WINDIR_DAY [18-31]	72
3.13 Έλεγχος των συνθηκών για την εφαρμογή του Watson-Williams test στις μεταβλητές WINDIR_D5, -21, -24, -30	76
3.14 Αποτελέσματα του ελέγχου Watson-Williams test στις μεταβλητές WINDIR_D5, -21, -24, -30	77
3.15 Εφαρμογή του Mardia-Watson-Wheeler test στις μεταβλητές δεύτερης τάξης WINDIR_DAY [1-17] και WINDIR_DAY [18-31]	82
3.16 Εφαρμογή του Mardia-Watson-Wheeler test στις μεταβλητές πρώτης τάξης WINDIR_D5, -21, -24, -30	86
3.17 Πίνακας συχνοτήτων και αθροιστικών συχνοτήτων για το Chi-squared test.	88
3.18 Εφαρμογή του ελέγχου Watson's U^2 test στις μεταβλητές WINDIR_DAY [1-17] και WINDIR_DAY [18-31]	94
3.19 Εφαρμογή του Watson's U^2 test στις μεταβλητές WINDIR_D5, -21, -24, -30	95
3.20 Έλεγχος με το Watson's U^2 test για την ισονομία των πληθυσμών των μεταβλητών WINDIR_DPREV [1-17] και WINDIR_DPREV [18-31]	96
3.21 Έλεγχος για την ισονομία των μεταβλητών WINDIR_DAY και WINDIR_DPREV	97
3.23 Κατευθύνσεις σε μοίρες από το Βορρά και κατά τη φορά των δεικτών του ωρολογίου ομάδας πουλιών ενός συγκεκριμένου είδους.	103
3.24 Κατευθύνσεις σε μοίρες, γωνιακές αποστάσεις (από τις αντίστοιχες μέσες γωνίες, σε μοίρες) και βαθμοί.	104
3.25 Κατευθύνσεις σε μοίρες, γωνιακές αποστάσεις (από τη βόρεια κατεύθυνση, σε μοίρες) και βαθμοί.	105
3.22 Έλεγχοι δύο δειγμάτων για την ισονομία των πληθυσμών.	106
3.26 Άλλοι έλεγχοι δύο ή περισσότερων δειγμάτων.	107
4.1 Σύγκριση του υπολογισμού των μέσων κατά την ανάλυση πρώτης και δεύτερης τάξης.	113

4.2	Εφαρμογή του ελέγχου <i>Mardia's two-sample test</i>	125
4.3	Συνοπτική επισκόπηση των ελέγχων για την ανάλυση δεύτερης τάξης.	128
5.1	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_Hxx</i> και <i>TEMPER_Hxx</i>	138
5.2	Αποτελέσματα των ελέγχων κανονικότητας και καλής προσαρμογής για τις μεταβλητές <i>TEMPER_Dxx</i>	140
5.3	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_Dxx</i> και <i>TEMPER_Dxx</i>	142
5.4	Αποτελέσματα των παλινδρομήσεων <i>WINDIR_Dxx/TEMPER_Dxx</i>	143
5.5	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_HOUR</i> και <i>TEMPER_HOUR</i>	144
5.6	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DAY</i> και <i>TEMPER_DAY</i>	146
5.7	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DAY</i> και <i>TEMPER_DMAX</i>	150
5.8	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DPREV</i> και <i>TEMPER_DAY</i>	151
5.9	Παραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DPREV</i> και <i>TEMPER_DMAX</i>	152
5.10	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_Hxx</i> και <i>TEMPER_Hxx</i>	157
5.11	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_Dxx</i> και <i>TEMPER_Dxx</i>	159
5.12	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_HOUR</i> και <i>TEMPER_HOUR</i>	160
5.13	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DAY</i> και <i>TEMPER_DAY</i>	161
5.14	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DAY</i> και <i>TEMPER_DMAX</i>	163
5.15	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DPREV</i> και <i>TEMPER_DAY</i>	165
5.16	Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών <i>WINDIR_DPREV</i> και <i>TEMPER_DMAX</i>	167

5.17	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx (1)	171
5.18	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx (2)	172
5.19	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx (3)	173
5.20	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Hxx και TEMPER_Hxx (1)	174
5.21	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Hxx και TEMPER_Hxx (2)	175
6.1	Αποτελέσματα των ελέγχων για την ομοιομορφία και την κανονικότητα των κυκλικών μεταβλητών και την κανονικότητα των γραμμικών μεταβλητών της μελέτης.	187
6.2	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών TEMPER_ / WINDIR_ (I).	188
6.3	Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών TEMPER_ / WINDIR_ (II)	189
6.4	Αλγόριθμος PERMUTE-1	β'
6.5	Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-1 για $n = 5$	β'
6.6	Αλγόριθμος PERMUTE-2	γ'
6.7	Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-2 για $n = 5$	γ'
6.8	Αλγόριθμος PERMUTE-3	δ'
6.9	Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-3 για $n = 5$	δ'

Κατάλογος σχημάτων

1	Συνήθεις κυκλικές κλίμακες μετρήσης	xxiv
1.1	Κυκλικά διαγράμματα διασποράς των μεταβλητών WINDIR_HOUR και WINDIR_DAY	2
1.2	Ροδογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY (1)	3
1.3	Ροδογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY (2)	3
1.4	Κυκλικά ιστογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY (1)	4
1.5	Κυκλικά ιστογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY (2)	4
1.6	Γραμμικό ιστόγραμμα των κυκλικών μεταβλητών WINDIR_HOUR και WINDIR_DAY	5
1.7	Κυκλικό διμεταβλητό ιστόγραμμα και ροδόγραμμα της μεταβλητής WINDIR_HOUR	6
1.8	Αριθμητικός μέσος κυκλικών δεδομένων	7
1.9	Το συνισταμένο διάνυσμα $\sum \mathbf{e}_i$ και το μέσο διάνυσμα \mathbf{m}	8
1.10	Αποκλίσεις από τη δειγματική μέση κατεύθυνση $\bar{\phi}$	13
1.11	Το μέσο μέτρο r παίρνει τιμές από 0 έως 1	16
1.12	Μέση γωνιακή απόκλιση s και μέσο μέτρο r	19
1.13	Εύρος ενός κυκλικού δείγματος	21
1.14	Ο παράγοντας επαναπατρισμού v	22
2.1	Γραμμικά διαγράμματα κυκλικής κατανομής	26
2.2	Παράδειγμα συνεχούς κυκλικής κατανομής	27
2.3	Η ομοιόμορφη κατανομή	28
2.4	Η κατανομή von Mises	29

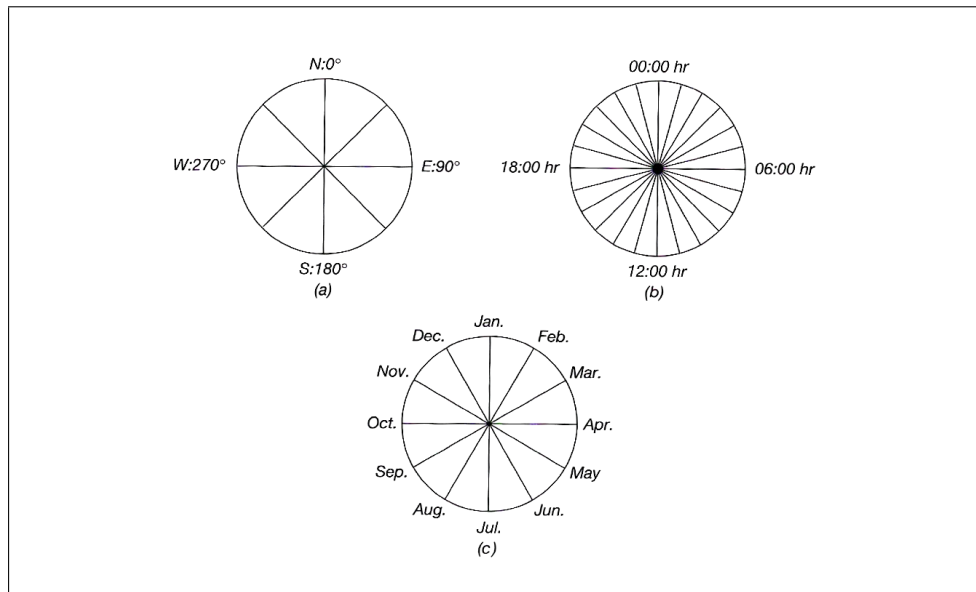
2.5	Η πυκνότητα της κατανομής von Mises	30
2.6	Η συνάρτηση $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$	31
2.7	Τέσσερις γνωστές κυκλικές κατανομές	33
3.1	Διαγράμματα για τον προσδιορισμό διαστημάτων εμπιστοσύνης για το μ_1	37
3.2	Διάγραμμα διασποράς της μεταβλητής WINDIR_DAY	39
3.3	Κυκλικό ιστόγραμμα της μεταβλητής WINDIR_H15 και κυκλικό διάγραμμα διασποράς της μεταβλητής WINDIR_D25	39
3.4	Διαγράμματα για τον προσδιορισμό διαστημάτων εμπιστοσύνης για το κ	40
3.5	Έλεγχος καλής προσαρμογής (ομοιομορφία) των μεταβλητών WINDIR_D7 και WINDIR_D24	46
3.6	Προετοιμασία του V test	47
3.7	Έλεγχος της προσαρμογής των μεταβλητών WINDIR_Hxx	48
3.8	Εφαρμογή του Hodges-Ajne test στη μεταβλητή WINDIR_D24	51
3.9	Εφαρμογή του modified Hodges-Ajne test στη μεταβλητή WINDIR_D10	52
3.10	Εφαρμογή του range test στη μεταβλητή WINDIR_H15	56
3.11	Απόκλιση δύο κυκλικών δειγμάτων	68
3.12	Έλεγχος της ισονομίας των μεταβλητών WINDIR_H16 και WINDIR_H17 με χρήση διαγράμματος Q-Q plot	80
3.13	Έλεγχος της ισονομίας των πληθυσμών των μεταβλητών δεύτερης τάξης WINDIR_DAY [1-17] και WINDIR_DAY [18-31] με χρήση διαγράμματος Q-Q plot	81
3.14	Έλεγχος της ισονομίας των μεταβλητών WINDIR_D19/WINDIR_D22 και WINDIR_D5 / WINDIR_D20 με χρήση διαγράμματος Q-Q plot	91
4.1	Κατανομές συχνοτήτων του r για δείγματα διαφόρου μεγέθους	110
4.2	Προσδιορισμός ορίων εμπιστοσύνης μέσου διανύσματος από δείγμα δεύτερης τάξης	116
4.3	Προσδιορισμός ορίων εμπιστοσύνης μέσης κατεύθυνσης από δείγμα δεύτερης τάξης	116
5.1	Γραμμικό διάγραμμα θετικής συσχέτισης	130
5.2	Κυκλικό διάγραμμα συσχέτισης υπό περιστροφή	131
5.3	Γραμμικό διάγραμμα αρνητικής (:) συσχέτισης	133

5.4	Κατασκευή του συντελεστή κυκλικής-γραμμικής συσχέτισης .	169
6.1	Εφαρμογή του προγράμματος προσομοίωσης (α)	ι'
6.2	Εφαρμογή του προγράμματος προσομοίωσης (β)	ι'
6.3	Εφαρμογή του προγράμματος προσομοίωσης (γ)	ια'

Εισαγωγή

Κυκλικά δεδομένα, δηλαδή δεδομένα διεύθυνσης ή κατεύθυνσης, παρατηρούνται και υπάρχει ανάγκη στατιστικής ανάλυσής τους σε πολλές εφαρμογές στη βιολογία, τη γεωλογία, τη μετεωρολογία, τη φυσική, ακόμα και στα οικονομικά. Πρόκειται για μετρήσεις γωνιών που εκτείνονται από 0 έως 2π ακτίνια ή, ισοδύναμα, από 0° έως 360° . Η άλγεβρα των γωνιών είναι διαφορετική από αυτή που διαχειρίζεται άλλες ποσότητες. Από τη μία, η αρχική (μηδέν) κατεύθυνση είναι εν πολλοίς *αυθαίρετη* — π.χ. ενώ για το αζιμούθιο η αρχική κατεύθυνση είναι ο Βορράς, σε άλλες περιπτώσεις είναι ο θετικός ημιάξονας X του μοναδιαίου κύκλου. Από την άλλη, το άθροισμα ή η διαφορά των γωνιών μπορεί, σε κάποιες περιπτώσεις, να μην ανήκει στο διάστημα $[0^\circ, 360^\circ)$ και έτσι, θα πρέπει να υπολογίζεται πάντοτε modulo 360° . Οι γωνίες, οι οποίες μετρούν τις κατευθύνσεις είναι κυκλικές ποσότητες και γι' αυτό οι αντίστοιχες μεταβλητές καλούνται κυκλικές μεταβλητές. Τέτοιου τύπου μεταβλητές διαφέρουν θεμελιωδώς από μεταβλητές που αναφέρονται σε ποσότητες όπως το μήκος, το βάρος, η θερμοκρασία κ.λπ., οι οποίες έχει επικρατήσει να ονομάζονται γραμμικές, καθώς μετρώνται στην πραγματική ευθεία.

Οι κυκλικές μεταβλητές εμφανίζονται, επίσης, σε προβλήματα μελέτης βιολογικών ρυθμών. Μία περίοδος 24 ωρών, για παράδειγμα, μπορεί να αντιστοιχιστεί σε μια πλήρη στροφή 360° . Όταν ένα γεγονός συμβαίνει επαναλαμβανόμενα κατά την ίδια χρονική στιγμή μιας προκαθορισμένης περιόδου, μπορεί τότε να χαρακτηριστεί από μία γωνία φάσης. Οι στατιστικές μέθοδοι που εφαρμόζονται στο πεδίο των βιολογικών ρυθμών μοιάζουν με τις μεθόδους της ανάλυσης των κατευθύνσεων. Πάντως, στο χώρο των βιολογικών ρυθμών προκύπτουν επιπρόσθετα προβλήματα, καθώς τα δεδομένα συλλέγονται στο ίδιο αντικείμενο σειριακά, πράγμα το οποίο οδηγεί σε *εξαρτημένα* δειγματικά σημεία. Επίσης, η ακριβής περίοδος δεν είναι πάντα γνωστή εκ των προτέρων. Επομένως, για την ανάλυση των βιολογικών ρυθμών απαιτούνται διαφορετι-



Σχήμα 1: Συνήθειες κυκλικές κλίμακες μετρήσης. (α) Κατευθύνσεις πυξίδας (αζιμούθια). (β) Ώρες της ημέρας. (γ) Ημέρες του έτους (σημειώνεται η πρώτη ημέρα κάθε μήνα).

κές στατιστικές μέθοδοι, όπως αυτές της ανάλυσης χρονολογικών σειρών.

Της στατιστικής ανάλυσης των κυκλικών δεδομένων προηγείται η μείωση του όγκου τους, η οποία επιτυγχάνεται με τον υπολογισμό κάποιων μέτρων θέσης και διασποράς. Στη συνέχεια, προσπαθούμε να βρούμε ένα μοντέλο που να τα ερμηνεύει, δηλαδή έναν θεωρητικό πληθυσμό από τον οποίο μπορεί να έχει επιλεγεί το συγκεκριμένο δείγμα. Η προσαρμογή μιας θεωρητικής κατανομής στα δεδομένα απαιτεί τη σημειακή εκτίμηση των παραμέτρων της. Η αποδοχή ή όχι του υποδείγματος αποφασίζεται σύμφωνα με τα αποτελέσματα της χρήσης των στατιστικών ελέγχων καλής προσαρμογής, με βάση τις εκτιμήσεις αυτών των παραμέτρων.

Ένας στατιστικός έλεγχος προϋποθέτει τον εκ των προτέρων καθορισμό ενός επιπέδου σημαντικότητας $\alpha\%$. Κατά κανόνα, θέλουμε να γνωρίζουμε το μικρότερο επίπεδο σημαντικότητας όπου η μηδενική υπόθεση μπορεί να απορριφθεί, το οποίο καλείται *P*-value (*P*) ή κρίσιμο επίπεδο. Στην παρούσα μελέτη, για λόγους συνέπειας της ανάλυσης, σε ό,τι αφορά στη στατιστική σημαντικότητα των αποτελεσμάτων, ακολουθείται η σύμβαση που αναγράφεται στον πίνακα (1).

Πίνακας 1: Πίνακας σύμβασης για το χαρακτηρισμό της στατιστικής σημαντικότητας των αποτελεσμάτων.

<i>P</i> -value	Στατιστική σημαντικότητα
$0.05 \geq P > 0.01$	Σημαντικά αποτελέσματα
$0.01 \geq P > 0.001$	Πολύ σημαντικά αποτελέσματα
$0.001 \geq P$	Πάρα πολύ σημαντικά αποτελέσματα

Οι βασικές έννοιες της θεωρίας πιθανοτήτων και της στατιστικής, δηλαδή οι κατανομές, οι εκτιμήτριες και οι ιδιότητές τους, τα διαστήματα εμπιστοσύνης και οι έλεγχοι υποθέσεων κ.τ.λ. θεωρούνται γνωστές και δεν θα αναλυθούν περαιτέρω στο παρόν.

Διάρθρωση της μελέτης

- Στο πρώτο κεφάλαιο παρουσιάζονται οι τρόποι γραφικής αναπαράστασης των κυκλικών δεδομένων και τα κύρια περιγραφικά στατιστικά τους, δηλαδή τα μέτρα θέσης και διασποράς.
- Στο δεύτερο κεφάλαιο παρουσιάζονται οι δύο κυριότερες θεωρητικές κυκλικές κατανομές, η ομοιόμορφη κατανομή και η κατανομή von Mises. Η πρώτη αποτελεί το υπόδειγμα ομοιομορφίας, βάσει του οποίου εφαρμόζονται οι έλεγχοι τυχαιότητας του επόμενου κεφαλαίου. Η δεύτερη αποτελεί το υπόδειγμα κανονικότητας των κυκλικών δεδομένων, δηλαδή προσαρμόζεται σε μονοκόρυφες και συμμετρικές κατανομές.
- Στο τρίτο κεφάλαιο δίνονται στατιστικές συναρτήσεις για τη σημειακή εκτίμηση και την εκτίμηση διαστήματος των παραμέτρων του πληθυσμού. Ακολούθως, παρουσιάζονται οι έλεγχοι υποθέσεων με διάκριση ως προς το πλήθος των δειγμάτων (έλεγχοι ενός δείγματος και έλεγχοι δύο ή περισσότερων δειγμάτων) και ως προς το υπόδειγμα (έλεγχοι καλής προσαρμογής της ομοιόμορφης κατανομής, όπου δίνεται βάρος στην ειδικότερη κατηγορία των ελέγχων τυχαιότητας, και της κατανομής von Mises).
- Στο τέταρτο κεφάλαιο παρουσιάζονται οι τεχνικές για την ανάλυση δεδομένων δεύτερης τάξης. Επισημαίνεται η ιδιαιτερότητα των κυκλικών

δεδομένων δεύτερης τάξης σε σχέση με τα αντίστοιχα γραμμικά. Αναφέρονται τα κατάλληλα στατιστικά για τις εκτιμήσεις και τους ελέγχους υποθέσεων.

- Στο πέμπτο κεφάλαιο παρουσιάζεται η ανάλυση συσχέτισης και παλινδρόμησης των κυκλικών μεταβλητών. Γίνεται εκτενής αναφορά στη συσχέτιση κυκλικής με γραμμική μεταβλητή. Παρατίθεται η παραμετρική και η μη παραμετρική μέθοδος. Για τη δεύτερη και για τις περιπτώσεις που δεν καλύπτουν οι πίνακες κρίσιμων τιμών που υπάρχουν στη βιβλιογραφία, δημιουργήθηκε πρόγραμμα, σε γλώσσα προγραμματισμού Visual Basic, με το οποίο υπολογίζεται με προσομοίωση η P -value του αντίστοιχου ελέγχου. Το πρόγραμμα αυτό παρουσιάζεται αναλυτικά στο Παράρτημα.
- Το έκτο κεφάλαιο περιλαμβάνει τα συμπεράσματα της μελέτης.
- Στο Παράρτημα παρουσιάζεται το πρόγραμμα προσομοίωσης για την ύπαρξη ή όχι συσχέτισης κυκλικής με γραμμική μεταβλητή και παρατίθενται οι κύριοι αλγόριθμοι και οι ρουτίνες του.

Μετεωρολογικά δεδομένα Στην παρούσα μελέτη έγινε στατιστική επεξεργασία πραγματικών μετεωρολογικών δεδομένων, τα οποία αφορούν σε μετρήσεις θερμοκρασίας και κατεύθυνσης του ανέμου από το σταθμό “Λόφος των Νυμφών”, Θησείο, του Εθνικού Αστεροσκοπείου Αθηνών,¹ για το χρονικό διάστημα του μηνός Ιουλίου 2007. Κατά τις πρώτες 17 ημέρες του μήνα, τα επίπεδα της θερμοκρασίας ήταν φυσιολογικά,² ενώ στο διάστημα των τελευταίων 14 ημερών υπήρξε κύμα καύσωνα. Εξετάστηκε, κυρίως, η συσχέτιση της θερμοκρασίας με την κατεύθυνση του ανέμου ανά ώρα και ανά ημέρα, σε πρώτο βαθμό (απλές μετρήσεις) και δεύτερο βαθμό (μέσοι), για κάθε διάστημα. Η παρουσίαση των αποτελεσμάτων γίνεται σε διακεκριμένα πλαίσια, σε κάθε κεφάλαιο, κατ’ ακολουθία των αντίστοιχων θεωρητικών προτάσεων.

Περιγραφή των μεταβλητών Για την περιγραφή των μεταβλητών, χρησιμοποιήθηκαν τα προθέματα TEMPER_ για τη θερμοκρασία και WINDIR_ για την κατεύθυνση του ανέμου. Ειδικότερα, για την περιγραφή της μεταβλητής

¹ ΓΠ: 38° 0.00' Β, ΓΜ: 23° 43.48' Α, Υψόμετρο: 110 m

² Με τον όρο “φυσιολογικές τιμές θερμοκρασίας” εννοούμε κανονικές, για την εποχή, θερμοκρασίες

τής θερμοκρασίας *ανά ημέρα*, δηλαδή όπως αυτή καταγράφηκε, για κάθε μία ώρα του εικοσιτετραώρου, καθ' όλη τη διάρκεια της xx ημέρας του μήνα μελέτης (24 τιμές) χρησιμοποιείται η κατάληξη Dxx, οπότε η μεταβλητή προσδιορίζεται ως TEMPER_Dxx. Δηλαδή, η μεταβλητή TEMPER_D12 αποδίδει τις τιμές της θερμοκρασίας κατά τη διάρκεια της 12ης ημέρας του μήνα μελέτης. Η αντίστοιχη μεταβλητή της κατεύθυνσης του ανέμου, προσδιορίζεται ως WINDIR_Dxx.

Με παρόμοιο τρόπο περιγράφονται τα ίδια δεδομένα, θεωρημένα *ανάστροφα* (δηλαδή, λαμβάνοντας υπ' όψη τον ανάστροφο πίνακα τιμών για τα μεγέθη της θερμοκρασίας και της κατεύθυνσης του ανέμου). Έτσι, για την περιγραφή της μεταβλητής της θερμοκρασίας *ανά ώρα*, δηλαδή όπως αυτή καταγράφηκε για όλες τις ημέρες του μήνα μελέτης κατά τη συγκεκριμένη ώρα xx του εικοσιτετραώρου (31 τιμές), χρησιμοποιείται η κατάληξη Hxx, οπότε η μεταβλητή προσδιορίζεται ως TEMPER_Hxx. Δηλαδή, η μεταβλητή TEMPER_H6 αποδίδει τις τιμές της θερμοκρασίας κατά την 6η ώρα (06:00–06:59) του εικοσιτετραώρου για όλες τις ημέρες του μήνα μελέτης. Η αντίστοιχη μεταβλητή της κατεύθυνσης του ανέμου, προσδιορίζεται ως WINDIR_Hxx.

Οι 31 το πλήθος μέσες ημερήσιες θερμοκρασίες περιγράφονται από τη μεταβλητή TEMPER_DAY και, αντίστοιχα, οι 31 μέσες ημερήσιες κατευθύνσεις του ανέμου από τη μεταβλητή WINDIR_DAY. Ακόμα, οι 24 μέσοι όροι των θερμοκρασιών για κάθε ώρα του εικοσιτετραώρου κατά το μήνα μελέτης περιγράφονται από τη μεταβλητή TEMPER_HOUR και οι αντίστοιχες 24 μέσες κατευθύνσεις του ανέμου από τη μεταβλητή WINDIR_HOUR.

Τέλος, οι 31 μέγιστες ημερήσιες θερμοκρασίες περιγράφονται από τη μεταβλητή TEMPER_DMAX, ενώ οι 31 επικρατούσες ημερήσιες κατευθύνσεις του ανέμου περιγράφονται από τη μεταβλητή WINDIR_DPREV.

Η ως άνω ονοματοδοσία των μεγεθών της μελέτης παρουσιάζεται εποπτικότερα στον παρακάτω πίνακα 2, σελ. xxviii.

Βιβλιογραφία Για την επισκόπηση των στατιστικών μεθόδων παρουσίασης και ανάλυσης των κυκλικών δεδομένων βασιστήκαμε στο βιβλίο "Circular Statistics in Biology" του Edward Batschelet (1981). Οι συμβολισμοί και η γενικότερη δομή της παρουσίασης ακολουθούν το συγκεκριμένο εγχειρίδιο. Επίσης, χρησιμοποιήθηκαν τα βιβλία "Biostatistical Analysis" του Jerrold Zar (1999) και "Directional Statistics" των Kanti Mardia & Peter Jupp (2000) και διάφορες εργασίες, με κυριότερες τις "An Application of Circular–Linear Correlation Analysis to

Πίνακας 2: Πίνακας περιγραφής των μεταβλητών της μελέτης

Μεταβλητή	Περιγραφή
TEMPER_Dxx	οι τιμές θερμοκρασίας, όπως καταγράφηκαν κατά το εικοσιτετράωρο της xx ημέρας του μήνα μελέτης
WINDIR_Dxx	οι κατευθύνσεις του ανέμου κατά τη xx ημέρα.
TEMPER_Hxx	οι ωριαίες τιμές της θερμοκρασίας, δηλαδή οι τιμές της θερμοκρασίας σε συγκεκριμένη ώρα (xx) του εικοσιτετράωρου για όλες τις ημέρες του μήνα μελέτης
WINDIR_Hxx	οι ωριαίες κατευθύνσεις του ανέμου, δηλαδή οι κατευθύνσεις του ανέμου σε συγκεκριμένη ώρα (xx) του εικοσιτετράωρου για όλο το μήνα
TEMPER_DAY	οι 31 μέσες ημερήσιες θερμοκρασίες
TEMPER_DAY[1-17]	οι μέσες ημερήσιες θερμοκρασίες κατά τις πρώτες 17 ημέρες του μήνα (φυσιολογικές τιμές θερμοκρασίας)
TEMPER_DAY[18-31]	οι μέσες ημερήσιες θερμοκρασίες κατά τις 14 ημέρες του καύσωνα
WINDIR_DAY	οι 31 μέσες ημερήσιες κατευθύνσεις του ανέμου
WINDIR_DAY[1-17]	οι μέσες ημερήσιες κατευθύνσεις του ανέμου κατά τις πρώτες 17 ημέρες του μήνα
WINDIR_DAY[18-31]	οι μέσες ημερήσιες κατευθύνσεις του ανέμου κατά τις 14 ημέρες του καύσωνα
TEMPER_HOUR	οι 24 μέσοι όροι των θερμοκρασιών για κάθε ώρα του εικοσιτετράωρου κατά το μήνα μελέτης
WINDIR_HOUR	οι 24 μέσες κατευθύνσεις του ανέμου για κάθε ώρα του εικοσιτετράωρου
TEMPER_DMAX	οι 31 μέγιστες ημερήσιες θερμοκρασίες
WINDIR_DPREV	οι 31 επικρατούσες ημερήσιες κατευθύνσεις του ανέμου

the Relationship between Freon Concentration and Wind Direction in Woods Hole Massachusetts” των Andrew Solow, John Bullister & Synthia Nevison (1988) και “Linear–Circular Correlation Coefficients and Rhythmometry” του Kanti Mardia (1976). Τέλος, για τη δημιουργία της ρουτίνας παραγωγής τυχαίων μεταθέσεων του προγράμματος προσομοίωσης του απαραμετρικού ελέγχου συσχέτισης του Mardia χρησιμοποιήθηκε κατάλληλος αλγόριθμος από το βιβλίο “Discrete Algorithmic Mathematics” των Stephen Maurer & Anthony Ralston (2004).

Λογισμικό Η ανάλυση των μετεωρολογικών δεδομένων που χρησιμοποιήθηκαν έγινε με τα στατιστικά προγράμματα *STATGRAPHICS* (εκδόσεις *Plus for Windows* και *Centurion XV.II*), *Minitab* (έκδοση 16) και το στατιστικό πρόγραμμα ανάλυσης κυκλικών δεδομένων *Oriana* (έκδοση 2.0).³ Ακόμα, πολλοί υπολογισμοί έγιναν με το πρόγραμμα *Microsoft Office Excel* (έκδοση 2003). Για τη δημιουργία του προγράμματος προσομοίωσης του απαραμετρικού ελέγχου συσχέτισης κυκλικής με γραμμική μεταβλητή του Mardia χρησιμοποιήθηκε το προγραμματιστικό περιβάλλον *Microsoft Visual Studio* (έκδοση 6.0).

Στοιχειοθεσία Η συγγραφή του παρόντος έγινε με το σύστημα στοιχειοθεσίας \LaTeX (διανομή *TexLive 2010*). Χρησιμοποιήθηκαν οι γραμματοσειρές *Courier New* για την απόδοση των ονομάτων των μεταβλητών, *GFSNeohellenic-Regular* για την απόδοση των πλαισίων της εφαρμογής και *Alkaios* για την απόδοση του κυρίως κειμένου.

³Άλλα προγράμματα ανάλυσης κυκλικών στατιστικών διαθέτουν τα προγραμματιστικά περιβάλλοντα *R* (πακέτο *circular*) και *Matlab* (πακέτο *CircStat*).

Πίνακας 3: Δεδομένα της μελέτης — Κατεύθυνση του ανέμου (α)

JULY 2007		WIND DIRECTION											
DATE	HOURS	1	2	3	4	5	6	7	8	9	10	11	12
	LST												
1	SSE	SW	W	SW	SW	SW	SW	SW	SSW	SSW	SSW	SSW	SW
2	SE	SSE	S	SSW	WNW	WSW	SSW	SSE	S	SW	SSW	SSW	SW
3	SW	SW	W	W	SSW	S	SSW	ENE	ENE	NE	E	ENE	NE
4	C	C	C	C	SE	C	C	S	C	ENE	NE	NE	ENE
5	NNE	E	C	NW	C	C	C	ESE	NE	NE	ENE	NE	ENE
6	ESE	SE	SE	SE	SSE	C	C	SSE	SSE	SSE	S	SW	SW
7	ESE	WNW	S	SSW	WNW	WSW	SSW	SSW	SSW	SW	SW	SW	SW
8	SW	SW	W	C	C	C	S	S	S	S	SW	SSW	SW
9	SSW	SSW	SSW	S	S	S	SSW	SSW	SSE	SSW	SW	SSW	SW
10	C	S	SSE	SE	SSE	SSE	SSE	SSE	ESE	SE	SW	SW	SW
11	C	NNE	NNE	NNE	NNW	NNW	N	E	ENE	NNE	NNE	NNE	ENE
12	SE	ESE	SE	ESE	SE	ESE	SE	SE	SE	SE	SE	SE	ESE
13	SE	SSW	S	SSW	S	SSW	SSW	SSW	SSW	SW	SW	SW	SW
14	SW	SW	SW	SW	SSW	SSW	S	SSW	SSW	SW	SW	SW	SW
15	SW	SW	SW	SW	SW	SW	SW	SSW	SSW	SW	SW	SW	SW
16	SW	SW	SW	SW	SW	SW	SW	SSW	SSW	SW	SW	SW	SW
17	SSW	SW	SW	SW	SW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW
18	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW
19	SSW	SSW	SSW	SSW	C	C	S	SE	SE	SE	SW	SW	SW
20	SSW	SSW	SSW	SSW	S	S	SSW	C	C	SW	SW	SW	SW
21	SSW	SSW	SSW	SSE	SSE	SSE	C	C	E	SE	SW	SW	SW
22	C	S	S	S	SSW	SSW	C	C	C	C	SW	SW	WSW
23	C	SSE	C	SW	SW	WSW	WSW	SSW	SSW	SSW	S	SW	SSW
24	SSW	SSW	SSW	SSW	WSW	WSW	SSW	S	S	SE	ENE	NW	SSW
25	W	W	SSW	SSW	W	SW	SW	NW	ENE	ENE	NNW	NNE	NE
26	N	SW	C	ENE	ENE	N	ENE	NE	NE	SW	SW	SW	SW
27	SSW	SW	C	SE	SE	SE	S	WSW	WSW	SSW	SSE	SSW	SW
28	SSW	SSW	SSW	S	C	C	ENE	SSW	SSW	SSW	SSW	SSW	SSW
29	SW	SSW	SSE	S	SSE	NNE	SSW	SSW	S	SSW	SSW	SSW	SSW
30	SSW	SSW	S	SE	C	E	NE	NE	E	SSW	SSW	SSW	SSW
31	S	S	S	S	S	W	WNW	WNW	C	ENE	SW	SSW	SSW

Πίνακας 4: Δεδομένα της μελέτης — Κατεύθυνση του ανέμου (β)

JULY 2007		WIND DIRECTION																													
DATE	HOURS	LST	13	14	15	16	17	18	19	20	21	22	23	24	PREV																
1	SW	SW	SSW	SW	SSW	SW	SW	SW	SW	SW	SW	SSW	S	SSE	SW																
2	SW	SW	WSW	SW	WSW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW																
3	NNE	NE	NNE	NE	NNE	NE	NE	NE	NE	NE	NE	NE	NNE	C	NE																
4	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NE	NNE	NNE	NE	NNE																
5	ENE	ENE	E	E	E	E	E	E	E	SE	SSE	SE	SE	ESE	E																
6	SW	SW	SW	SW	SW	SW	SW	WSW	WSW	WSW	WSW	SW	SW	C	SW																
7	S	NE	NNE	NNE	NNE	NNE	NNE	NNE	NE	ENE	WNW	SSW	SSW	SSW	SSW																
8	WSW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW																
9	SSW	SW	SSW	SW	SSW	SW	SW	SW	SW	SW	SW	SW	SW	S	SW																
10	SW	WSW	NNE	NNE	NNE	NNE	NNE	NE	E	E	ENE	NE	NNE	C	SSE																
11	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NNE	NE	ESE	SE	NNE																
12	E	E	ENE	NNE	ENE	NNE	NNE	NNE	NE	SW	WSW	SSW	SSW	S	SE																
13	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SSW	SW																
14	SW	SW	SW	SW	SW	SW	SSW	SW	SSW	SSW	SSW	SSW	SSW	SW	SW																
15	SW	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SW	SW	SSW	SW	SW																
16	SSW	SW	SSW	SW	SSW	SW	SW	SW	SSW	SSW	SSW	SSW	SSW	SSW	SW																
17	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SSW	SW	SSW	SW	SW	SSW	SSW	SSW																
18	SSW	SW	SW	SW	SW	SW	SW	SSW	SW	SSW	SSW	SSW	SSW	SSW	SSW																
19	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SSW	SSW	SSW	SSW	SW	SW																
20	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	S	S	S	SW																
21	NNW	NNE	NE	NE	NE	NE	NE	ENE	ENE	NE	NE	E	NE	NE	NE																
22	SW	SW	NE	NE	NE	NE	NE	SW	SW	SW	SW	SSW	WSW	C	SW																
23	SSW	SSW	SW	SW	SW	SW	SW	SW	SW	SW	C	C	SSE	S	SW																
24	NE	NE	NE	NE	NE	NE	NE	NNE	NE	ENE	ESE	E	C	C	NE																
25	NNE	NNE	NE	NNE	NE	NNE	NNE	NE	NNE	C	ENE	NE	NE	NE	NE																
26	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SW																
27	SW	WSW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SSW	SW																
28	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW																
29	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SSW	SW	SW																
30	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NNE	NE	SSW	C	NE																
31	NNE	NE	NE	W	W	WSW	SW	SW	SW	S	S	S	SSW	SW	S																

Πίνακας 5: Δεδομένα της μελέτης — Θερμοκρασίες (α)

JULY 2007		TEMPERATURE														
DATE	HOUR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	24,8	24,6	24,4	24,3	24,2	24,3	25,5	26,9	28,3	29,4	30,6	31,4	31,7	31,7	31,6	31,6
2	24,5	24,5	24	23,9	23,6	23,7	25,9	27,3	28,6	30	31,3	32,3	33	33,1	33,2	33,2
3	25,6	25	24,7	23,7	23,8	23,5	25,1	27,1	29	30,6	31,4	31,5	31,3	31,7	32	32
4	25,3	25	24,7	24,6	24,3	24,5	26,4	28,1	29,8	31	31,4	32	32,5	32,4	32,2	32,2
5	26,3	26,1	25,9	25,6	25	25,5	25,9	27,6	29,2	30,8	33,3	34,8	35,9	36,4	36,9	36,9
6	26,8	26,3	26,2	26,1	25,4	25,8	26,8	27,8	28,8	28,4	30	30,5	31,6	32	32,4	32,4
7	23,9	23,9	23,4	22,8	22,9	23,6	25,5	27,5	29,5	30,6	31,8	32,5	31,9	31,8	31,6	31,6
8	25,2	24,7	24,5	23,7	23,7	23,5	25	28,2	31,2	33,1	34,7	36,1	36,5	36,2	35,8	35,8
9	27,6	26,5	26,3	26,3	26,3	26,6	27,6	29,8	31,9	32,7	33,8	34,2	35,2	35,5	35,7	35,7
10	26,9	26,7	25,8	25,6	25,8	26,6	27,2	29,9	32,4	35	36	36,3	37,8	36,8	35,9	35,9
11	26,9	27,4	26,6	26,3	26,2	26,2	27,2	29,2	31,1	32,8	33,5	34,6	35,2	35,2	35,1	35,1
12	26,7	26	25,3	25,4	24,8	25	26,9	28,5	30	30,9	32,6	33,4	34,3	35,3	36,2	36,2
13	25,6	25,4	24,9	24,9	24,9	24,2	24,9	25,9	26,9	28,2	30	30,4	28,8	29,2	29,6	29,6
14	23,6	24	23,6	23,1	22,6	22,9	23,9	25,8	27,7	28,5	29,3	30,4	30,3	30,5	30,6	30,6
15	25,4	25,3	24,9	24,6	24,3	24,6	25,5	26,7	27,9	29,1	29,5	30,2	30,5	30,8	31,2	31,2
16	25,4	25,2	25,1	24,9	25	25,2	26,5	28,2	29,8	30,1	30,8	31,2	31,3	31,6	31,9	31,9
17	28,3	28,2	28	27,5	27,5	27,6	28,8	30	31,1	32,2	33,4	34,4	34,5	34,6	34,7	34,7
18	28,9	29,1	28,8	27,8	28,2	28,2	29,9	31,2	32,6	34,3	35,2	35,7	36,2	36,4	36,6	36,6
19	29,2	29,3	28,3	27,7	27,5	27,6	28,5	30,8	33,2	35,5	36,7	37,6	38,8	39,1	39,3	39,3
20	29,3	29,1	28,6	28,5	28	28	29,1	31,2	33,4	34,4	36	37,4	38,2	37,6	37	37
21	28,4	27,7	27,2	26,9	26,6	27	27,6	30,5	33,2	34,8	36,4	37,1	35,9	35,6	35,3	35,3
22	27,1	27,1	26,5	26,7	26,6	27,3	28,4	30,8	33,2	35,6	37,2	38,4	39,7	38	36,3	36,3
23	30	28,8	29	29,2	29,4	29,2	30,1	32,9	35,8	37,2	38,7	40	40,1	40,6	41,1	41,1
24	31,3	31,1	31,4	31,5	31,2	30,9	31,6	34,3	36,8	38,5	39,4	39,7	39,4	40,3	41,2	41,2
25	32,9	31,9	31,2	31	31,1	30,6	31	33	35	38,2	40,2	40,8	41,1	40,2	39,3	39,3
26	32,6	32,5	32,5	32,2	31,5	31,4	31,3	32,9	34,4	34,2	35,3	36,9	37,3	37,5	37,6	37,6
27	29,2	28,5	28,1	27,7	27,4	27,3	28	29,9	31,6	32,9	34	35,7	36,8	37	37,1	37,1
28	27,8	27,1	27,2	26,3	25,7	25,8	27,9	29,9	31,8	33,5	34,6	34,7	34,7	35	35,3	35,3
29	27,5	27,1	27,1	26,4	26,3	26,4	27,6	30	32,2	32,6	33,7	35	35	35	34,9	34,9
30	27,1	26,7	26	26	25,8	25,5	26,3	29,1	31,7	33,3	35,5	36,3	35	35,5	35,9	35,9
31	26,9	26,5	26,3	26,1	26	26	26,5	29	31,4	35,4	36,7	37,7	36,8	37,1	37,4	37,4

Πίνακας 6: Δεδομένα της μελέτης — Θερμοκρασία (β)

JULY 2007		TEMPERATURE													
DATE	HOUR	16	17	18	19	20	21	22	23	24	max	min	aver		
1		31,1	30,7	30	28,2	27,3	26,4	26	25,5	24,8	32,4	23,5	27,7		
2		32,7	32,3	31,3	29,8	28,5	27,2	26,7	26,4	25,9	34	23,1	28,3		
3		33,8	32,4	29,7	27,5	26,9	26,2	25,9	25,6	25,5	34,6	23,2	27,9		
4		30,3	30,1	30	29,4	28,5	27,6	27,2	27,3	26,8	33,6	23,9	28,4		
5		36,8	36,4	35	33,5	31,6	29,8	28,9	28	27,3	37,6	24,7	30,5		
6		32	30,8	29,6	28,2	27,1	25,9	25,8	25,4	24,3	33	25,1	28,1		
7		31,3	31,2	30,8	29,7	28	26,2	26,8	26,3	25,5	33,6	22,7	27,9		
8		35,6	34,8	33,5	32	30,9	29,8	29,4	28,8	27,9	37,2	23,4	30,2		
9		36	35	33,5	32,1	30,7	29,4	28,9	28,1	27,8	36,6	26	30,7		
10		36	36	36,1	34,1	31,3	28,6	28,2	27,4	27	39	25,4	31,2		
11		33,8	33,2	32,2	30,6	29,9	29	29	28,3	27,5	36,6	26	30,3		
12		34,7	33	30,9	30,2	28,9	27,6	27,1	26,6	26	36,4	24,6	29,4		
13		30,3	29,8	29,1	27,9	26,6	25,3	24,8	24,5	24,1	30,8	24,2	26,9		
14		30,5	30	29,2	27,9	27,1	26,3	26,2	25,9	25,5	31,6	22,4	26,9		
15		31	30,2	29,2	28,3	27,7	27,1	26,6	26,5	25,9	31,5	23,9	27,6		
16		32,3	31,4	31	30,3	29,6	28,9	28,5	28,4	28,2	32,6	24,7	28,8		
17		34,5	33,5	31,9	30,8	30,4	29,9	29,8	29,6	29,4	35,4	26,8	30,9		
18		36,7	35,8	34,9	33,4	32	30,6	29,8	29,8	29,8	36,8	27,4	32,2		
19		38,4	37,1	35,7	33,9	33	32	31,4	30,2	29,5	39,8	27,4	32,9		
20		36,8	35,8	34,5	33,2	32	30,7	30,1	29,3	28,9	39,3	27,7	32,4		
21		36,7	36	34	31,7	29,9	28	27,7	27,6	27,2	38,2	25,2	31,2		
22		36,7	36,9	36,8	35,4	34,1	32,8	32	30,7	30,5	40,8	26,2	32,7		
23		41	40,6	39,6	38	36,3	34,4	33,4	32,4	31,7	41,4	28,5	35		
24		40,4	40,6	38,9	36,7	35,3	33,8	33,5	33,4	32,3	41,7	30,8	35,6		
25		40,3	40,2	37,3	34,7	34,1	33,4	33,8	34	33,6	41,9	30	35,4		
26		36,9	35,9	33,9	32,3	31,5	30,6	30,2	29,6	29,4	38,6	30,8	33,4		
27		36,4	35,2	32,7	30,7	29,9	28,9	28,4	28,3	28,1	37,7	27	31,2		
28		35,1	34,5	33,2	31,8	30,7	29,6	28,9	28,1	28	35,7	25,6	30,7		
29		35,4	34,6	33,2	31,8	30,6	29,2	28,3	27,8	27,5	35,9	25,9	30,6		
30		35,8	34,6	34,2	32,8	30,2	27,5	26,9	26,5	26,5	37,2	25,4	30,4		
31		37,8	36,5	35,5	34	32,4	30,7	29,7	29,1	29,2	38,8	26,8	31,7		
											36,5	25,8	31,1		

Κεφάλαιο 1

Γραφική αναπαράσταση και περιγραφικά μέτρα

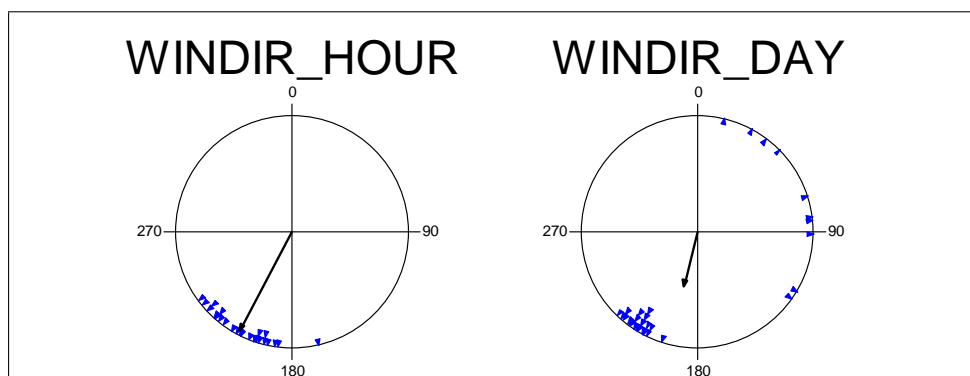
Πριν ορίσουμε τα περιγραφικά στατιστικά μέτρα για κυκλικά δεδομένα, θα παρουσιάσουμε τους τρόπους γραφικής αναπαράστασής τους.

1.1 Γραφική αναπαράσταση

Οι κατευθύνσεις μπορεί να σημειώνονται ως ακτίνες (ημιευθείες) σε έναν κύκλο, όπου η μία από αυτές ορίζεται ως αφετηρία (αρχική ή μηδέν κατεύθυνση) και για την ταξινόμησή τους επιλέγεται η δεξιόστροφη ή αριστερόστροφη φορά. Η παρουσίασή τους γίνεται με τη χρήση των παρακάτω διαγραμμάτων:

Κυκλικό διάγραμμα διασποράς Σχεδιάζουμε έναν κύκλο και τοποθετούμε σημάδια του ίδιου σχήματος (συνήθως τη μύτη βέλους για τα δεδομένα κατεύθυνσης και την τελεία για τα δεδομένα διεύθυνσης) πάνω στην περιφέρεια του κύκλου, στα σημεία τομής της με τις ημιευθείες των κατευθύνσεων. Ο παρατηρητής αποκτά εποπτικά άποψη τόσο για το μέγεθος του δείγματος, όσο και για τη διασπορά των παρατηρήσεων. Αυτή η αναπαράσταση καλείται *κυκλικό διάγραμμα διασποράς* (scatter diagram) και ένα παράδειγμα φαίνεται στο σχήμα 1.1.

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

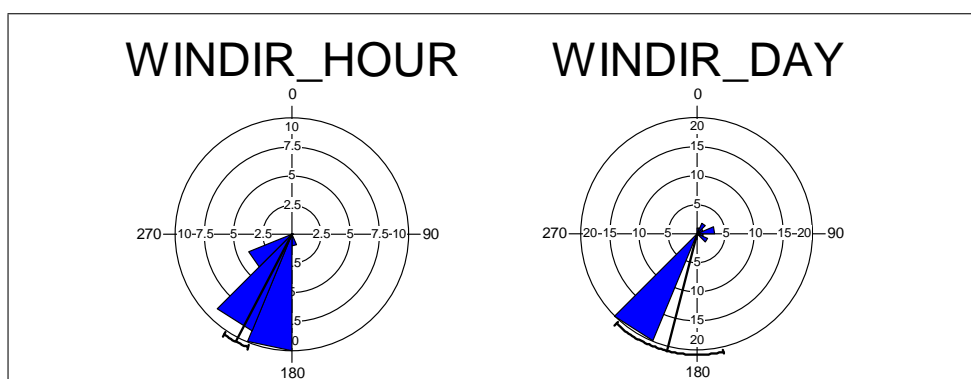


Σχήμα 1.1: Κυκλικά διαγράμματα διασποράς των μεταβλητών WINDIR_HOUR και WINDIR_DAY. Το έντονα μαύρο βέλος σημειώνει το μέσο διάνυσμα των γωνιών του κάθε δείγματος. Το μήκος του βέλους εκφράζει το μήκος του μέσου διανύσματος r . Το μεγαλύτερο βέλος που αντιστοιχεί στη μεταβλητή WINDIR_HOUR δείχνει μεγαλύτερη συγκέντρωση (προσανατολισμό) των δεδομένων.

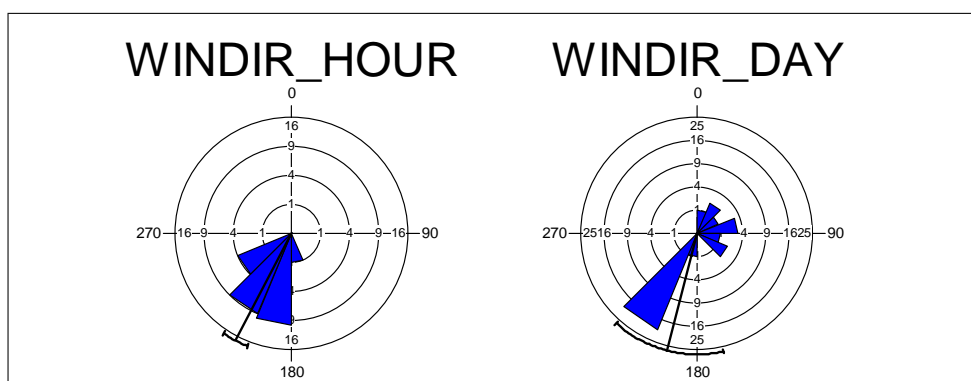
Ροδόγραμμα Εναλλακτικά, μπορούμε να εκφράσουμε τις παρατηρούμενες συχνότητες, ανά εύρος τιμών, με κάποιο *ροδόγραμμα* (rose diagram). Επί της ουσίας, πρόκειται για ένα ιστόγραμμα συχνοτήτων, το οποίο έχει αναδιπλωθεί γύρω από έναν κύκλο. Τα δεδομένα ομαδοποιούνται σε κλάσεις και αντίστοιχα ο κύκλος διαιρείται σε κυκλικούς τομείς. Η συχνότητα της κάθε κλάσης αναπαρίσταται είτε με την ακτίνα (σχήμα 1.2) είτε με το εμβαδόν του αντίστοιχου κυκλικού τομέα (σχήμα 1.3). Όμως, επειδή η οπτική εντύπωση που δημιουργεί ένας κυκλικός τομέας διαμορφώνεται πρωτίστως από το εμβαδό του και δευτερευόντως από την ακτίνα του, το ροδόγραμμα του σχήματος 1.2 μπορεί να παρασύρει σε λάθος συμπεράσματα, αφού υπερτονίζει (οπτικά) τις μεγαλύτερες συχνότητες και υποβαθμίζει τις μικρές. Έτσι, μπορεί να δημιουργηθεί η εντύπωση ότι κάποια κατεύθυνση "ξεχωρίζει" ιδιαίτερα έναντι των άλλων, ενώ τα δεδομένα μπορεί να μην υποστηρίζουν ένα τέτοιο συμπέρασμα. Γι' αυτόν το λόγο, στη βιβλιογραφία προτείνεται οι (απόλυτες ή σχετικές) συχνότητες των κλάσεων να αναπαρίστανται με τα εμβαδά και όχι με τις ακτίνες των αντίστοιχων τομέων. Δηλαδή, η ακτίνα κάθε τομέα προτείνεται να είναι ανάλογη με την τετραγωνική ρίζα της αντίστοιχης συχνότητας και όχι με τη συχνότητα, αφού το εμβαδό του κυκλικού τομέα ισούται με πr^2 , όπου r η ακτίνα του. Στα ιστογράμματα μη κυκλικών δεδομένων και στα κυκλικά ιστογράμματα δεν δημιουργείται ανάλογο πρόβλημα (υπό την προϋπόθεση, πάντα, ότι τα πλάτη των κλάσεων εί-

1.1. Γραφική αναπαράσταση

ναί ίσα), αφού το εμβαδό του κάθε ορθογώνιου παραλληλόγραμμου που αναπαριστά τις συχνότητες ισούται με το γινόμενο της βάσης του (πλάτος κλάσεων) επί το ύψος του (συχνότητα), δηλαδή είναι ανάλογο της συχνότητας. Επίσης, το ροδογράμμο (όπως άλλωστε και το ιστόγραμμα μη κυκλικών δεδομένων) επηρεάζεται δραστικά από το πλάτος των κλάσεων.



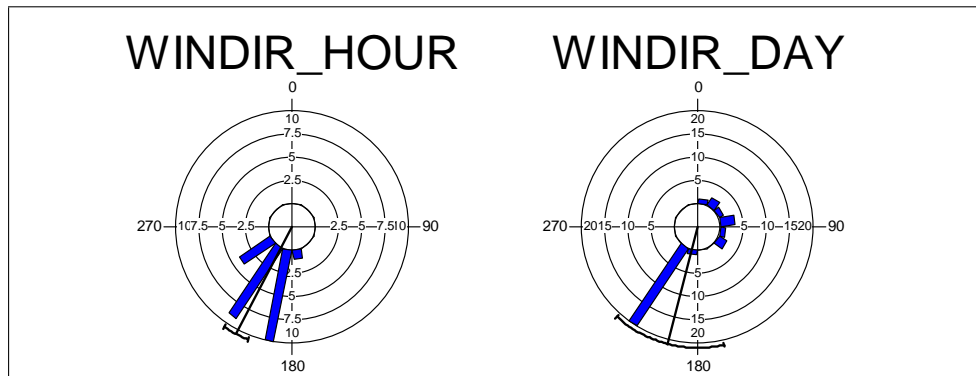
Σχήμα 1.2: Ροδογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY για τα ίδια δεδομένα με αυτά του σχήματος 1.1. Η συχνότητα των παρατηρήσεων εκφράζεται από το μήκος της ακτίνας του τομέα, κάτι που είναι πιθανόν παραπλανητικό. Την αδυναμία αυτή διορθώνουν τα αντίστοιχα ροδογράμματα του σχήματος 1.3), στα οποία η συχνότητα εκφράζεται από το εμβαδόν του του κάθε τομέα. Το έντονα μαύρο ευθύγραμμο τμήμα από το κέντρο του κύκλου έως την περιφέρειά του εκφράζει τη μέση γωνία των δεδομένων και το έντονα μαύρο τόξο κατά μήκος της περιφέρειας του κύκλου σημειώνει το 95% διάστημα εμπιστοσύνης για τη μέση κατεύθυνση του πληθυσμού.



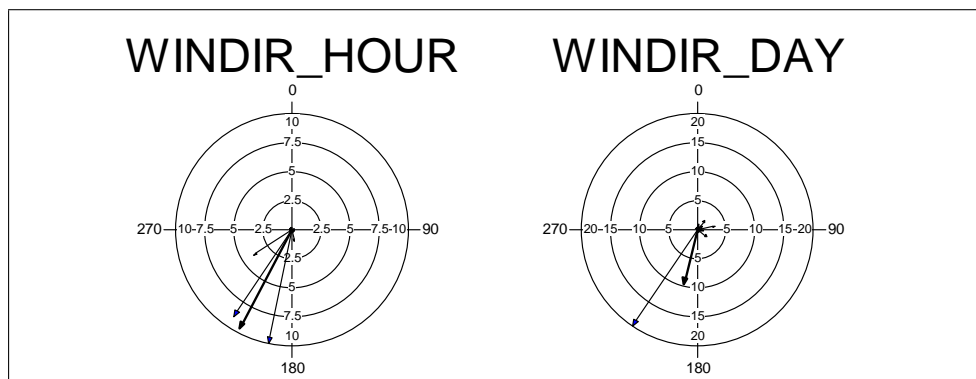
Σχήμα 1.3: Ροδογράμματα των μεταβλητών WINDIR_HOUR (αριστερά) και WINDIR_DAY (δεξιά) για τα ίδια δεδομένα με αυτά του σχήματος 1.2. Στα διαγράμματα αυτά η συχνότητα των παρατηρήσεων εκφράζεται από το εμβαδόν του κάθε τομέα. Η παρουσίαση αυτή των συχνοτήτων των παρατηρήσεων είναι ορθότερη.

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

Κυκλικό ιστόγραμμα Οι παρατηρούμενες συχνότητες μπορεί να αναπαρασταθούν, επίσης, με ένα *κυκλικό ιστόγραμμα* (circular histogram) όπου μπορεί να εκφράζονται είτε, όπως σε ένα γραμμικό ιστόγραμμα, από το μήκος ή το εμβαδό ορθογώνιων παραλληλόγραμμων (βλ. σχήμα 1.4) είτε από το μήκος διανυσμάτων (βλ. σχήμα 1.5). Τα κυκλικά ιστογράμματα χρησιμοποιούνται ευρέως για τη γραφική παρουσίαση της φοράς των ανέμων.

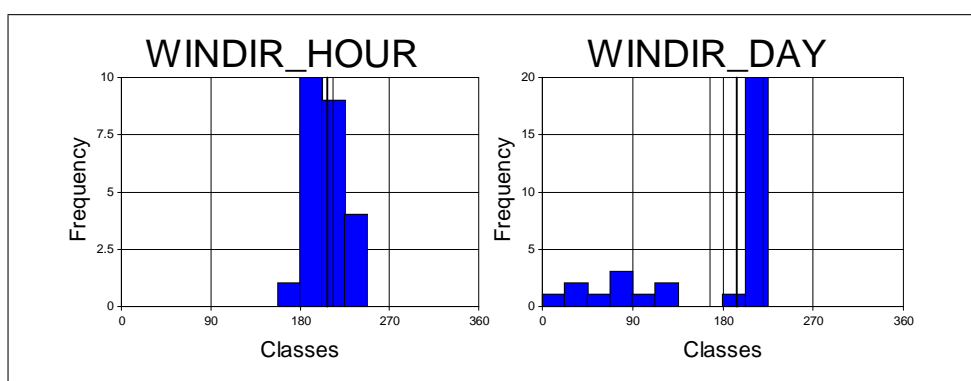


Σχήμα 1.4: Κυκλικά ιστογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY για τα ίδια δεδομένα με αυτά του σχήματος 1.1. Στα διαγράμματα αυτά, όπως και στα ιστογράμματα συχνοτήτων των μη κυκλικών μεταβλητών, η συχνότητα εκφράζεται από το εμβαδό του κάθε ορθογώνιου παραλληλογράμμου.



Σχήμα 1.5: Κυκλικά ιστογράμματα των μεταβλητών WINDIR_HOUR και WINDIR_DAY για τα ίδια ακριβώς δεδομένα με αυτά του κυκλικού ιστογράμματος του σχήματος 1.4. Εδώ, για την αναπαράσταση των συχνοτήτων χρησιμοποιούνται διανύσματα, αντί για ορθογώνια παραλληλόγραμμα, με το μήκος κάθε διανύσματος να εκφράζει τη συχνότητα της κάθε κλάσης. (Τα διανύσματα των κλάσεων συχνοτήτων δεν πρέπει να συγχέονται με το έντονα μαύρο βέλος, το οποίο, βεβαίως, σημειώνει το μέσο διάνυσμα.)

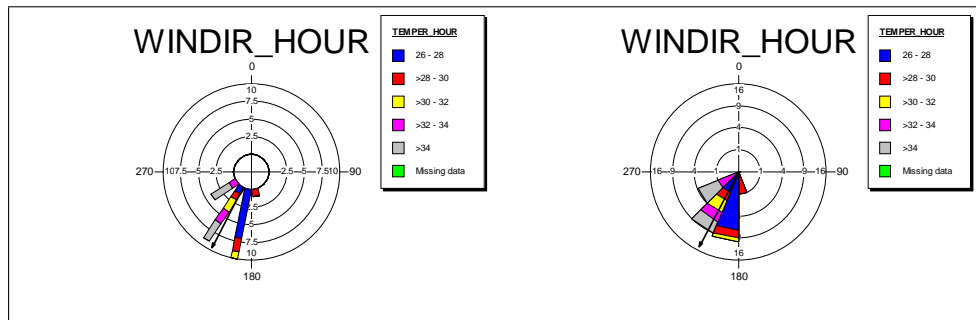
Γραμμικό ιστόγραμμα Πρόκειται για το σύνηθες γραμμικό ιστόγραμμα. Ένα μειονέκτημά του είναι ότι η οπτική εντύπωση που δημιουργεί διαφοροποιείται σημαντικά ανάλογα με την επιλογή της αρχής των αξόνων. Γι' αυτό, όταν το εύρος των τιμών είναι μεγάλο (πάνω από 180°) δεν προτείνεται για την αναπαράσταση κυκλικών δεδομένων.



Σχήμα 1.6: Γραμμικό ιστόγραμμα των κυκλικών μεταβλητών WINDIR_HOUR και WINDIR_DAY. Οι τιμές των μεταβλητών έχουν διαιρεθεί σε κλάσεις, με αυθαίρετο αρχικό σημείο. Το έντονα μαύρο ευθύγραμμο τμήμα σημειώνει το μέσο διάνυσμα και οι διακεκομμένες γραμμές τα 95% όρια εμπιστοσύνης για τη μέση κατεύθυνση.

Κυκλικό διμεταβλητό ιστόγραμμα Στην περίπτωση διμεταβλητών κατανομών, όπου μια κυκλική μεταβλητή σχετίζεται με μια άλλη μεταβλητή, ένας τρόπος παρουσίασης των δεδομένων είναι με τη χρήση ενός *διμεταβλητού ιστογράμματος* (two-variable histogram). Το γράφημα διαιρείται σε κλάσεις συχνοτήτων βάσει των τιμών της άλλης μεταβλητής, για κάθε τομέα (εύρος τιμών) της κυκλικής μεταβλητής και μπορεί να αναπαραστήσει τα δεδομένα είτε με τη μορφή τυπικού κυκλικού ιστογράμματος, είτε με τη μορφή ροδογράμματος. Ένα παράδειγμα τέτοιου διαγράμματος παρουσιάζεται στο σχήμα 1.7.

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ



Σχήμα 1.7: Κυκλικό διμεταβλητό ιστόγραμμα και ροδόγραμμα για τις τιμές της κυκλικής μεταβλητής WINDIR_HOUR ως προς τις αντίστοιχες τιμές της σχετικής γραμμικής μεταβλητής TEMPER_HOUR. Η τιμές της γραμμικής μεταβλητής έχουν διαιρεθεί σε κλάσεις, οι οποίες αποδίδονται στο γράφημα με διαφορετικό χρώμα. Επίσης, σημειώνεται το μέσο διάνυσμα.

1.2 Περιγραφικά μέτρα

Τα μέτρα θέσης και διασποράς των κυκλικών δεδομένων είναι ανάλογα, εννοιολογικά, με τα αντίστοιχα μέτρα κατανομής των μη κυκλικών δεδομένων, όμως τα περισσότερα από αυτά ορίζονται και υπολογίζονται διαφορετικά. Επειδή ο κύκλος είναι κλειστή γραμμή (η αρχή — η όποια αρχή — και το τέλος συμπίπτουν), οι συνήθεις μέθοδοι στατιστικής επεξεργασίας δεδομένων, στην πλειονότητά τους, δεν μπορούν να εφαρμοστούν σε δεδομένα κυκλικής κλίμακας. Για τα τελευταία, απαιτείται άλλη προσέγγιση (διανυσματική).

1.2.1 Μέτρα θέσης

Μέση κατεύθυνση και μέσο διάνυσμα

Έστω το παρακάτω δείγμα δύο κατευθύνσεων ως προς το Βορρά (0°)

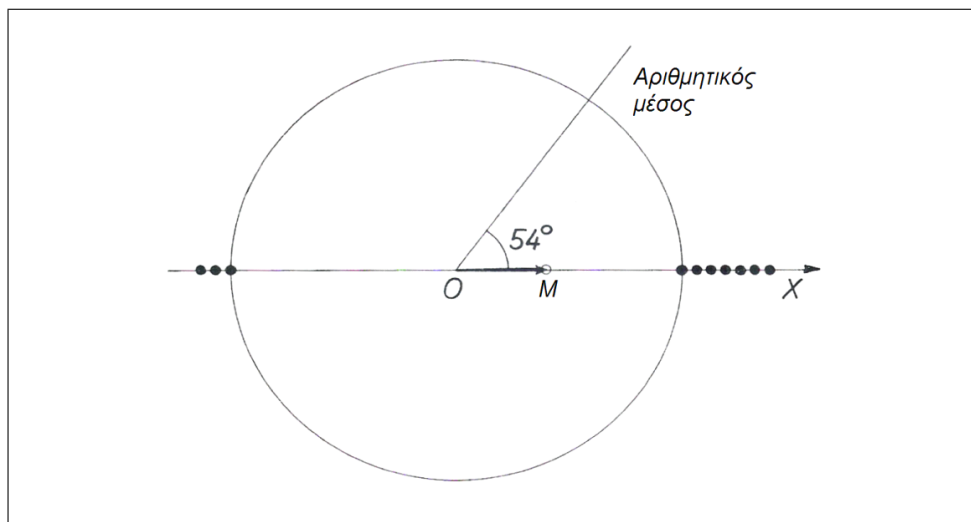
$$\phi_1 = 10^\circ, \quad \phi_2 = 350^\circ$$

Αν θέλουμε να προσδιορίσουμε τη μέση κατεύθυνση αυτού του δείγματος, τότε με την εφαρμογή του *αριθμητικού μέσου* των παραπάνω γωνιών

$$\frac{1}{2}(\phi_1 + \phi_2) = \frac{10^\circ + 350^\circ}{2} = 180^\circ$$

παίρνουμε ως αποτέλεσμα τη νότια κατεύθυνση (180°), κάτι που, βέβαια, είναι απαράδεκτο, καθώς δύο βόρειες κατευθύνσεις δεν είναι δυνατό να έχουν

νότια μέση κατεύθυνση. Ένα ανάλογο παράδειγμα παρουσιάζεται στο σχήμα 1.8.



Σχήμα 1.8: Ο αριθμητικός μέσος ενός δείγματος γωνιών δεν ενδείκνυται για τον προσδιορισμό της μέσης γωνίας. Το υποθετικό δείγμα 10 γωνιών (σημειώνονται με κουκίδες) του σχήματος, με τιμές 0° και 180° , έχει μέση γωνία 0° , η οποία σημειώνεται με το διάνυσμα \vec{OM} — βλ. παρακάτω εξ. (1.5), σ. 9 για τον υπολογισμό. Ο αριθμητικός μέσος των $\frac{1}{10}(7 \times 0^\circ + 3 \times 180^\circ) = 54^\circ$ δίνει ένα άστοχο αποτέλεσμα.

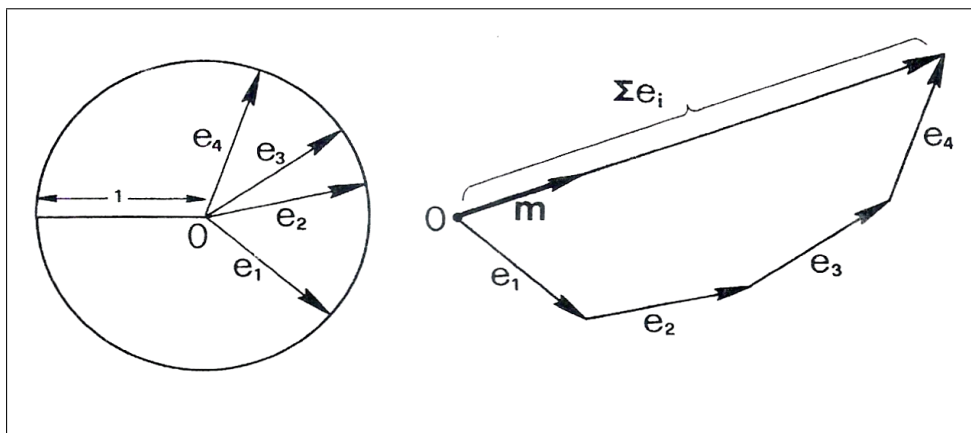
Μία κατάλληλη στατιστική συνάρτηση για τη μέση κατεύθυνση βασίζεται σε μια αρκετά διαφορετική διαδικασία. Έστω τυχαίο δείγμα n κατευθύνσεων που αντιστοιχούν στις γωνίες $\phi_1, \phi_2, \dots, \phi_n$. Αν θεωρήσουμε τις κατευθύνσεις αυτές ως ημιευθείες με αρχή το κέντρο του μοναδιαίου κύκλου, τότε τα σημεία τομής τους με τον κύκλο δίνουν ένα σύνολο από n δειγματικά σημεία. Αναθέτοντας σε κάθε ένα σημείο την ίδια ποσότητα μάζας, υπολογίζουμε το κέντρο ισορροπίας ή κέντρο μάζας ή κέντρο βάρους M του δείγματος. Αν το κέντρο αυτό είναι διαφορετικό από το κέντρο O του μοναδιαίου κύκλου, η ημιευθεία OM προσδιορίζει τη μέση κατεύθυνση του δείγματος.

Ο υπολογισμός αυτού του κέντρου βάρους μπορεί να γίνει με δύο τρόπους, κατά πρώτο με διανυσματική άλγεβρα και, κατά δεύτερο με τριγωνομετρικές συναρτήσεις (Batschelet, 1981, pp.8-11):

Διανυσματική άλγεβρα Αν $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ είναι τα (μοναδιαία) διανύσματα που σημειώνουν τα δειγματικά σημεία, τότε το διάνυσμα

$$\mathbf{m} = \frac{1}{\sum_i M_i} (M_1 \mathbf{e}_1 + M_2 \mathbf{e}_2 + \dots + M_n \mathbf{e}_n)$$

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ



Σχήμα 1.9: Το συνισταμένο διάνυσμα $\sum \mathbf{e}_i$ και το μέσο διάνυσμα \mathbf{m} .

όπου M_i ($i = 1, 2, \dots, n$) οι αντίστοιχες σημειακές μάζες, δείχνει στο κέντρο βάρους. Αφού είναι εξ ορισμού

$$M_1 = M_2 = \dots = M_n = M \text{ δηλαδή, είναι } \sum^i M_i = nM,$$

το διάνυσμα \mathbf{m} απλοποιείται στο

$$\mathbf{m} = \frac{1}{n}(\mathbf{e}_1 + \mathbf{e}_2 + \dots + \mathbf{e}_n),$$

το οποίο ισούται με το συνισταμένο διάνυσμα ή συνισταμένη $\sum^i \mathbf{e}_i$, διαιρεμένο με το πλήθος των σημείων n . Το διάνυσμα \mathbf{m} καλείται μέσο διάνυσμα του δείγματος και το μέτρο του $|\mathbf{m}|$ μήκος του μέσου διανύσματος ή μέσο μέτρο. Αν συμβολίσουμε με R το μήκος της συνισταμένης και με r το μήκος του μέσου διανύσματος, δηλαδή,

$$\left| \sum^i \mathbf{e}_i \right| = R, \quad |\mathbf{m}| = r,$$

τότε είναι

$$r = R/n$$

Το κέντρο βάρους μπορεί να βρίσκεται πάνω στην περιφέρεια του μοναδιαίου κύκλου μόνο στην ειδική περίπτωση κατά την οποία όλες οι σημειακές μάζες συμπίπτουν στο ίδιο σημείο (της περιφέρειας). Σε κάθε άλλη περίπτωση,

το κέντρο βάρους βρίσκεται εντός του μοναδιαίου κύκλου. Επομένως, είναι

$$0 \leq R \leq n \text{ και } 0 \leq r \leq 1$$

Τριγωνομετρικές συναρτήσεις Χρησιμοποιούμε ένα ορθοκανονικό σύστημα συντεταγμένων, με άξονες X και Y και κέντρο O . Αν ϕ_i είναι η i -οστή γωνία του δείγματος και x_i, y_i οι συνιστώσες του αντίστοιχου μοναδιαίου διανύσματος \mathbf{e}_i , τότε είναι

$$x_i = \cos \phi_i, \quad y_i = \sin \phi_i$$

Αν \bar{x} και \bar{y} είναι οι καρτεσιανές συντεταγμένες του κέντρου βάρους και υποθέτοντας ότι είναι $r > 0$, τότε το μέσο διάνυσμα σχηματίζει μια καλώς ορισμένη γωνία πάνω στο θετικό ημιάξονα X . Καλούμε τη γωνία αυτή *μέση γωνία του δείγματος* και τη συμβολίζουμε με $\bar{\phi}$. Τότε είναι

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.1)$$

$$\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n) = \frac{1}{n} \sum_{i=1}^n y_i \quad (1.2)$$

και η μέση γωνία $\bar{\phi}$ αποτελεί τη λύση των εξισώσεων

$$\begin{aligned} \bar{x} &= r \cos \bar{\phi} \\ \bar{y} &= r \sin \bar{\phi} \end{aligned} \quad (1.3)$$

όπου το μέσο μέτρο r δίνεται από την

$$r = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (1.4)$$

Κατά την ειδική περίπτωση όπου είναι $\bar{x} = 0, \bar{y} = 0$ και, ως εκ τούτου, $r = 0$, το μέσο διάνυσμα ισούται με το μηδενικό διάνυσμα.

Η μέση γωνία του δείγματος $\bar{\phi}$ υπολογίζεται από τις

$$\bar{\phi} = \begin{cases} \arctan(\bar{y}/\bar{x}), & \bar{x} > 0 \\ \pi + \arctan(\bar{y}/\bar{x}), & \bar{x} < 0 \end{cases} \quad (1.5)$$

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

Ειδικές περιπτώσεις είναι οι

$$\bar{\phi} = \begin{cases} \pi/2, & \bar{x} = 0 \text{ και } \bar{y} > 0 \\ 3\pi/2, & \bar{x} = 0 \text{ και } \bar{y} < 0 \\ \text{απροσδιόριστο,} & \bar{x} = 0 \text{ και } \bar{y} = 0 \end{cases} \quad (1.6)$$

Μια επαλήθευση είναι η

$$\cos \bar{\phi} = \bar{x}/r, \quad \sin \bar{\phi} = \bar{y}/r$$

Ιδιότητες του μέσου διανύσματος

Παρακάτω, παρουσιάζουμε κάποιες χρήσιμες ιδιότητες του μέσου διανύσματος \mathbf{m} και των πολικών συντεταγμένων του r και $\bar{\phi}$.

1. Έστω ότι περιστρέφουμε την αρχική κατεύθυνση κατά μία γωνία ψ . Τότε, οι δειγματικές γωνίες ϕ_i γίνονται

$$\phi'_i = \phi_i - \psi \quad (i = 1, 2, \dots, n)$$

και οι καρτεσιανές συντεταγμένες του νέου μέσου διανύσματος είναι οι

$$x' = \frac{1}{n} \sum_{i=1}^n \cos \phi'_i \quad y' = \frac{1}{n} \sum_{i=1}^n \sin \phi'_i$$

Τότε, το x' ισούται με

$$\begin{aligned} x' &= \frac{1}{n} \sum_{i=1}^i \cos \phi'_i \\ &= \frac{1}{n} (\cos \phi'_1 + \dots + \cos \phi'_n) \\ &= \frac{1}{n} [\cos(\phi_1 - \psi) + \dots + \cos(\phi_n - \psi)] \\ &= \frac{1}{n} (\cos \phi_1 \cos \psi + \sin \phi_1 \sin \psi + \dots + \cos \phi_n \cos \psi + \sin \phi_n \sin \psi) \\ &= \frac{1}{n} [\cos \psi (\cos \phi_1 + \dots + \cos \phi_n) + \sin \psi (\sin \phi_1 + \dots + \sin \phi_n)] \\ &= \cos \psi \frac{1}{n} \sum_{i=1}^i \cos \phi_i + \sin \psi \frac{1}{n} \sum_{i=1}^i \sin \phi_i \\ &= \cos \psi \bar{x} + \sin \psi \bar{y} \end{aligned} \quad (1.7)$$

οπότε, λόγω των (1.3), είναι

$$\begin{aligned} x' &= \cos \psi r \cos \bar{\phi} + \sin \psi r \sin \bar{\phi} \\ &= r (\cos \psi \cos \bar{\phi} + \sin \psi \sin \bar{\phi}) \\ &= r \cos(\bar{\phi} - \psi) \end{aligned} \quad (1.8)$$

Ανάλογα, αποδεικνύεται ότι είναι

$$y' = r \sin(\bar{\phi} - \psi) \quad (1.9)$$

Διαιρώντας κατά μέλη τις (1.8) και (1.9) έχουμε

$$\tan \bar{\phi}' = \tan(\bar{\phi} - \psi)$$

οπότε είναι

$$\bar{\phi}' = \bar{\phi} - \psi \quad (1.10)$$

Ακόμα, είναι

$$r'^2 = x'^2 + y'^2$$

και, λόγω της (1.8)

$$\begin{aligned} r'^2 &= r^2 \cos^2(\bar{\phi} - \psi) + r^2 \sin^2(\bar{\phi} - \psi) \\ &= r^2 [\cos^2(\bar{\phi} - \psi) + \sin^2(\bar{\phi} - \psi)] \\ &= r^2 \end{aligned}$$

και αφού είναι $r, r' > 0$

$$r = r' \quad (1.11)$$

Επομένως, ισχύει ότι, υπό περιστροφή, η μέση κατεύθυνση είναι αμετάβλητη ως προς την αρχική και το μέσο μέτρο παραμένει σταθερό, δηλαδή το μέσο διάνυσμα δεν εξαρτάται από την αρχική κατεύθυνση.

Η ιδιότητα αυτή είναι η αντίστοιχη της ιδιότητας των γραμμικών στατιστικών, όπου για δείγμα n μετατοπισμένων στον άξονα κατά την ποσότητα α παρατηρήσεων

$$x_1 - \alpha, x_2 - \alpha, \dots, x_n - \alpha$$

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

ο δειγματικός τους μέσος είναι $\bar{x} - \alpha$.

2. Για το άθροισμα των ημιτόνων των διαφορών $\phi_i - \bar{\phi}$ έχουμε

$$\begin{aligned}\sum_{i=1}^i \sin(\phi_i - \bar{\phi}) &= \sin \phi_i \cos \bar{\phi} - \sin \bar{\phi} \cos \phi_i \\ &= (\bar{x}/r) n \bar{y} - (\bar{y}/r) n \bar{x} \\ &= \frac{n}{r} (\bar{x} \bar{y} - \bar{y} \bar{x})\end{aligned}$$

$$\text{δηλαδή} \quad \sum_{i=1}^n \sin(\phi_i - \bar{\phi}) = 0 \quad (1.12)$$

Αφού για γωνίες $\psi \approx 0$ ισχύει ότι

$$\sin \psi \approx \psi$$

τότε, για μικρές αποκλίσεις $\phi_i - \bar{\phi} \approx 0$, η εξ. (1.12) δίνει την προσεγγιστική σχέση

$$\sum_{i=1}^i (\phi_i - \bar{\phi}) \approx 0 \quad (1.13)$$

η οποία είναι αντίστοιχη με τη σχέση

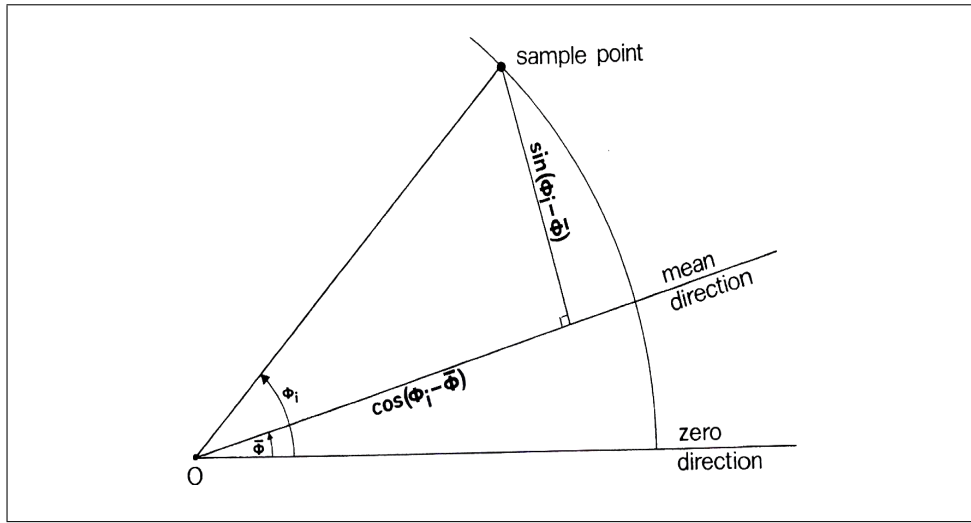
$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

που ισχύει στη γραμμική στατιστική ανάλυση.

3. Για το άθροισμα των συνημιτόνων των διαφορών $\phi_i - \bar{\phi}$ είναι

$$\begin{aligned}\sum_{i=1}^i \cos(\phi_i - \bar{\phi}) &= \cos \bar{\phi} \sum_{i=1}^i \cos \phi_i + \sin \bar{\phi} \sum_{i=1}^i \sin \phi_i \\ &= (\bar{x}/r) n \bar{x} + (\bar{y}/r) n \bar{y} \\ &= \frac{n}{r} (\bar{x}^2 + \bar{y}^2) \\ &= \frac{n}{r} r^2\end{aligned}$$

$$\text{δηλαδή} \quad \sum_{i=1}^n \cos(\phi_i - \bar{\phi}) = nr \quad (1.14)$$



Σχήμα 1.10: Στην ανάλυση κυκλικών δεδομένων, οι αποκλίσεις από τη δειγματική μέση κατεύθυνση $\bar{\phi}$ μετρώνται σε $\sin(\phi_i - \bar{\phi})$ και $\cos(\phi_i - \bar{\phi})$

Ακόμα, είναι

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n 2[1 - \cos(\phi_i - \bar{\phi})] &= \frac{1}{n} \sum_{i=1}^n [2 - 2 \cos(\phi_i - \bar{\phi})] \\
 &= \frac{1}{n} [2n - 2 \sum_{i=1}^n \cos(\phi_i - \bar{\phi})] \\
 &= 2 - \frac{2}{n} \sum_{i=1}^n \cos(\phi_i - \bar{\phi}) \quad (1.15)
 \end{aligned}$$

και, λόγω της εξ. (1.14), έχουμε

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n 2[1 - \cos(\phi_i - \bar{\phi})] &= 2 - \frac{2}{n} nr \\
 &= 2 - 2r \\
 &= 2(1 - r) \quad (1.16)
 \end{aligned}$$

Χρησιμοποιώντας την προσεγγιστική σχέση

$$2(1 - \cos \psi) \approx \psi^2$$

για $\psi \approx 0$, όπου το ψ μετριέται σε ακτίνια, παίρνουμε, για μικρές απο-

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

κλίσεις $\phi_i - \bar{\phi} \approx 0$, τη σχέση

$$2[1 - \cos(\phi_i - \bar{\phi})] \approx (\phi_i - \bar{\phi})^2 \quad (1.17)$$

Έτσι, από τις (1.16) και (1.17), προκύπτει η προσεγγιστική σχέση

$$\frac{1}{n} \sum_{i=1}^n (\phi_i - \bar{\phi})^2 \approx 2(1 - r) \quad (1.18)$$

η οποία είναι αντίστοιχη με την

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

που χρησιμοποιείται στη γραμμική στατιστική ανάλυση ως μέτρο διασποράς.

4. Υπάρχει, επίσης, μια τέταρτη αντιστοιχία μεταξύ κυκλικών και γραμμικών στατιστικών. Σύμφωνα με τον τύπο τού Steiner

$$\sum_{i=1}^n (x_i - u)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - u)^2$$

όπου το u είναι τυχαία ποσότητα, προκύπτει ότι το $\sum_{i=1}^n (x_i - u)^2$ ελαχιστοποιείται, όταν είναι $u = \bar{x}$. Η ανάλογη εξίσωση στα κυκλικά στατιστικά αποδεικνύεται (Batschelet, 1981, p. 17) ότι είναι η

$$\sum_{i=1}^n (\phi_i - \psi) = \sum_{i=1}^n \cos(\phi_i - \bar{\phi}) \cos(\bar{\phi} - \psi) \quad (1.19)$$

για τυχαία γωνία ψ . Εδώ, η ποσότητα $\sum_{i=1}^n (\phi_i - \psi)$ μεγιστοποιείται, όταν είναι $\cos(\bar{\phi} - \psi) = 1$, το οποίο συνεπάγεται $\psi = \bar{\phi} \pmod{360^\circ}$. Σύμφωνα με την εξ. (1.14), το μέγιστο αυτό ισούται με nr .

Στον παρακάτω πίνακα 1.1 συνοψίζουμε τις αντίστοιχες σχέσεις μεταξύ κυκλικών και γραμμικών στατιστικών:

Πίνακας 1.1: Αντίστοιχες σχέσεις μεταξύ κυκλικών και γραμμικών στατιστικών

Γραμμικά στατιστικά	Κυκλικά στατιστικά
$x_i - \bar{x}$	$\sin(\phi_i - \bar{\phi})$
$\sum^i (x_i - \bar{x}) = 0$	$\sum^i \sin(\phi_i - \bar{\phi}) = 0$
$(x_i - \bar{x})^2$	$2[1 - \cos(\phi_i - \bar{\phi})]$
$\frac{1}{n} \sum^i (x_i - \bar{x})^2 = s^2$	$\frac{1}{n} \sum_{i=1}^n (\phi_i - \bar{\phi})^2 \approx 2(1 - r)$

Διάμεση κατεύθυνση

Σε κάποιες περιπτώσεις, είναι προτιμότερη η χρησιμοποίηση ενός μέτρου θέσης, διαφορετικού από τη μέση γωνία. Για το σκοπό αυτό, διαιρούμε το κυκλικό δείγμα με μία διάμετρο έτσι, ώστε τα μισά δειγματικά σημεία να βρίσκονται στη μία πλευρά της διαμέτρου και τα άλλα μισά στην άλλη. Αν το δείγμα είναι μονοκόρυφο και το πλήθος των σημείων περιττός αριθμός, η διάμετρος αυτή ορίζεται μονοσήμαντα (διατρέχει το κεντρικό σημείο). Αν το πλήθος των δειγματικών σημείων είναι άρτιος αριθμός, η διάμετρος αυτή τέμνει το τόξο ανάμεσα στα δύο κεντρικά σημεία. Η γωνία που σχηματίζει η διάμετρος αυτή καλείται *διάμεση γωνία* και συμβολίζεται συνήθως με $\tilde{\phi}$.

Πρέπει να επισημανθεί, ότι κάποιο μέτρο μέσης θέσης (προτιμώμενη κατεύθυνση, ώρα αιχμής) έχει πρακτική αξία μόνο όταν τα δεδομένα είναι συγκεντρωμένα ή ομαδοποιημένα γύρω από το μέσο. Για παράδειγμα, αν υποθέσουμε ότι μια ομάδα ποντικών παρουσιάζει δραστηριότητα κατά τα διαστήματα ωρών 21:00–22:30 και 01:20–02:40, ενώ κατά τον υπόλοιπο χρόνο της ημέρας είναι αδρανής, τότε μια “μέση ώρα” δραστηριότητας θα έπεφτε σε διάστημα αδράνειας, κάτι το οποίο δεν αποδίδει την πραγματική εικόνα (Batschelet, 1981, p. 19).

Πολυκόρυφα δείγματα

Σε διάφορες εφαρμογές, τα δείγματα μπορεί να είναι δικόρυφα ή τετρακόρυφα, γενικά πολυκόρυφα. Ένας λόγος μπορεί να είναι ότι, για παράδειγμα, μια ομάδα ζωντανών οργανισμών καλείται να διαλέξει μια κατεύθυνση μεταξύ δύο ή περισσότερων προτιμώμενων, κατά μικρό ή μεγάλο βαθμό, κατευθύνσεων. Μπορεί να θεωρηθεί, ότι δείγματα τέτοιου τύπου προέρχονται από μια κατανομή, η οποία παράγεται από δύο ή περισσότερες άλλες επικαλυπτόμενες (overlapping) κατανομές. Τότε, αναφερόμαστε σε μια *μείξη κατανομών*.

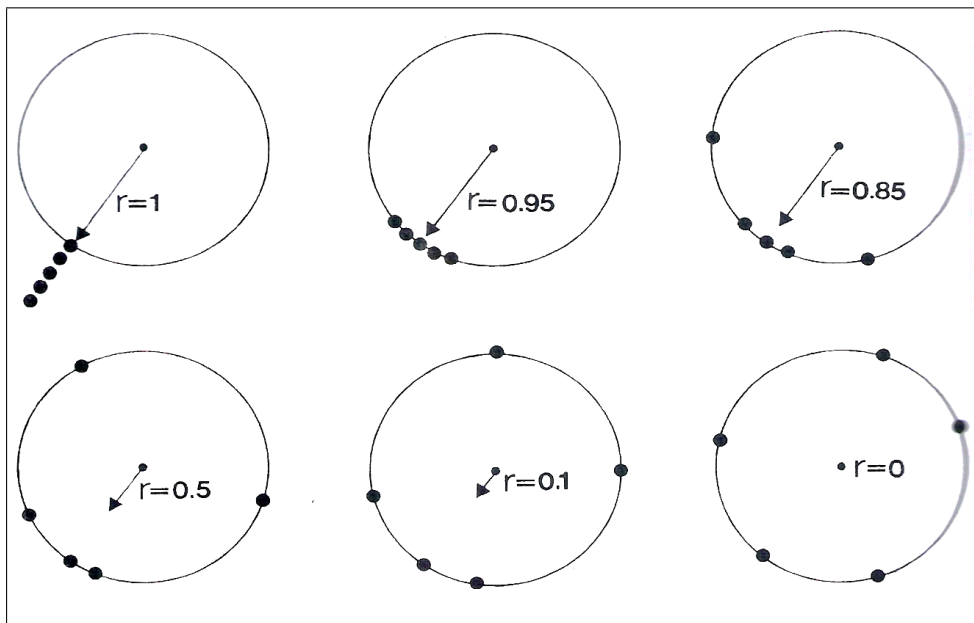
ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

Δεδομένα διεύθυνσης Υπάρχει, πάντως, ένας σημαντικός ειδικός τύπος δι-κόρυφου δείγματος, στην ανάλυση του οποίου επιτρέπονται οι περισσότερες στατιστικές τεχνικές: η περίπτωση *δύο ίσων και αντίθετων κορυφών*. Μια παρόμοια κατάσταση προκύπτει κατά την παρατήρηση *μη κατευθυνόμενων ευθειών ή αξόνων*. Στην τελευταία περίπτωση, δεν μπορεί να γίνει διάκριση μεταξύ δύο διαμετρικά αντίθετων σημείων. Κατά τη γραφική αναπαράσταση μιας τέτοιας δειγματικής κατανομής, προκύπτει ένα *κεντρικά συμμετρικό σχήμα*. Στις περιπτώσεις αυτές, αναφερόμαστε σε *δεδομένα διεύθυνσης*.

1.2.2 Μέτρα διασποράς, λοξότητας και κύρτωσης

Μέτρο της συγκέντρωσης

Η μέση γωνία ενός δείγματος, το οποίο προέρχεται από πληθυσμό με μονοκόρυφη κατανομή, υποδεικνύει μια προτιμώμενη κατεύθυνση, αλλά δεν περιέχει πληροφορία σχετικά με το βαθμό κατά τον οποίο οι δειγματικές τιμές συγκεντρώνονται γύρω από τη γωνία αυτή. Όπως στα γραμμικά στατιστικά, έτσι και στην ανάλυση των κυκλικών δεδομένων ένα μέτρο διασποράς είναι εξ ίσου σημαντικό με ένα μέτρο θέσης (Batschelet, 1981, p. 31).



Σχήμα 1.11: Το μήκος του μέσου διανύσματος r παίρνει τιμές από 0 έως 1. Ως μέγεθος, είναι ανάλογο με τη συγκέντρωση των δειγματικών σημείων γύρω από τη μέση κατεύθυνση — στη σειρά των γραφημάτων αυτών, μειώνεται, όσο τα σημεία διασπείρονται στην περιφέρεια του μοναδιαίου κύκλου.

Βλέποντας τα γραφήματα του σχήματος 1.11, παρατηρούμε ότι, μειούμενης της συγκέντρωσης των σημείων του δείγματος γύρω από μια συγκεκριμένη (μέση) κατεύθυνση, μειώνεται το μέσο μέτρο r . Ξεκινώντας από την περίπτωση της μέγιστης συγκέντρωσης ($r = 1$), όπου όλα τα σημεία συμπίπτουν, καταλήγουμε στην περίπτωση όπου τα σημεία ισαπέχουν πάνω στην περιφέρεια του μοναδιαίου κύκλου ($r = 0$). Συνεπώς, συμπεραίνουμε ότι σε μονοκόρυφα δείγματα, το μήκος του μέσου διανύσματος r μπορεί να χρησιμοποιηθεί ως μέτρο της συγκέντρωσης.

Γωνιακή διακύμανση και γωνιακή απόκλιση

Το μέσο μέτρο r , εκτός από μέτρο της συγκέντρωσης, μπορεί, προφανώς, να χρησιμοποιηθεί ως μέτρο της διασποράς. Μάλιστα, συχνά, είναι πιο επιθυμητή η χρήση ενός μέτρου διασποράς απ' ό,τι ενός μέτρου συγκέντρωσης. Προφανώς, τα δύο αυτά μέτρα σχετίζονται κατά τέτοιο τρόπο, ώστε το ένα απ' αυτά να αυξάνεται, ενόσω το άλλο μειώνεται.

Αφού το r κινείται στο διάστημα $1 \rightarrow 0$ καθώς η διασπορά των σημείων αυξάνεται, είναι φυσικό το να θεωρήσουμε το $1 - r$ ως μέτρο της διασποράς των κατευθύνσεων γύρω από τη μέση κατεύθυνσή τους και, έτσι, να ορίσουμε την *κυκλική διακύμανση* (circular variance)

$$S^2 = 1 - r \quad (\text{τετρ. ακτίνα}) \quad (1.20)$$

Λαμβάνοντας υπόψη τη σχέση (1.18), σ. 12, από την (3) ιδιότητα του μέσου διανύσματος, ο Batschelet (1965) προτείνει ως μέτρο της διασποράς τη *γωνιακή διακύμανση* (angular variance)

$$s^2 = 2(1 - r), \quad (\text{τετρ. ακτίνα}) \quad (1.21)$$

η οποία είναι ασυμπτωτικά ισοδύναμη με τη στατιστική συνάρτηση

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

της διακύμανσης στα γραμμικά στατιστικά, και ορίζει την τετραγωνική ρίζα της αυτής της ποσότητας ως *μέση γωνιακή απόκλιση* ή *γωνιακή απόκλιση* (angular deviation)

$$s = \sqrt{2(1 - r)} \quad (\text{ακτίνα}) \quad (1.22)$$

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

Στον Batschelet (1981, p. 317) δίνεται πίνακας τιμών για τον προσδιορισμό της s .

Ο Mardia (1972) προτείνει ως μέτρο της διασποράς την ποσότητα

$$s_0^2 = -2 \ln r \quad (\text{τετρ. ακτίνα}) \quad (1.23)$$

και ορίζει την τετραγωνική ρίζα της αυτής της ποσότητας ως *κυκλική τυπική απόκλιση* (circular standard deviation)

$$s_0 = \sqrt{-2 \ln r} \quad (\text{ακτίνα}) \quad (1.24)$$

η οποία, επίσης, προσεγγίζει ασυμπτωτικά τη γνωστή τυπική απόκλιση.

Για $r \approx 1$, τα s και s_0 παίρνουν προσεγγιστικά την ίδια τιμή. Όταν είναι $r \rightarrow 1$, και τα δύο γίνονται ασυμπτωτικά ίσα με τη γνωστή τυπική απόκλιση των γραμμικών στατιστικών. Μόνο για μικρές τιμές του r , δηλαδή σε περιπτώσεις με μεγάλη διασπορά, τα s και s_0 διαφέρουν ουσιαστικά μεταξύ τους. Έτσι, όταν το r τείνει στο μηδέν, ενώ το s παραμένει πεπερασμένο, το s_0 απειρίζεται.

Αφού το μήκος του μέσου διανύσματος r εξαρτάται από το μέγεθος του δείγματος, κατά συνέπεια, όλα τα μέτρα που παράγονται από το r εξαρτώνται, επίσης, από το μέγεθος του δείγματος.

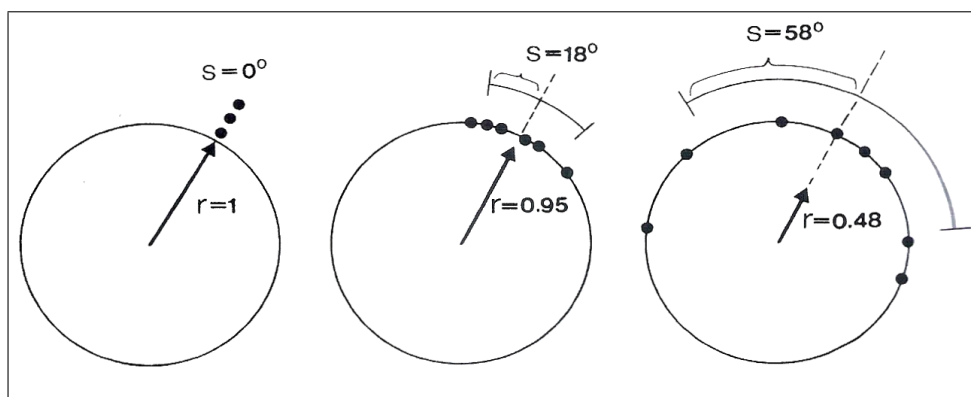
Για κάθε δείγμα της μορφής

$$\phi_1, \phi_2, \dots, \phi_n, \phi_1 + \pi, \phi_2 + \pi, \dots, \phi_n + \pi$$

είναι $r = 0$. Αυτό σημαίνει ότι μηδέν μέσο μέτρο δεν συνεπάγεται ότι οι παρατηρήσεις κατανέμονται ισομερώς στην περιφέρεια του κύκλου (Mardia and Jupp, 2000, p. 18).

Πρέπει να τονιστεί ότι τα S^2 , s και s_0 , όπως και στη στατιστική των γραμμικών μεταβλητών, έχουν νόημα μόνο για μονοκόρυφες κατανομές.

Δικόρυφα δείγματα Προκειμένου να υπολογίσουμε τη γωνιακή απόκλιση ενός δικόρυφου δείγματος, εφαρμόζουμε τη μέθοδο του *διπλασιασμού των γωνιών*. Για το τροποποιημένο δείγμα υπολογίζουμε το μήκος του νέου μέσου διανύσματος r_2 και μέσω της εξ. (1.22) τη νέα μέση γωνιακή απόκλιση s_2 . Τελικά, για να επιστρέψουμε στα αρχικά δεδομένα, ακυρώνουμε την επίδραση του διπλασιασμού, υποδιπλασιάζοντας το s_2 , οπότε η τελική τιμή της γωνια-



Σχήμα 1.12: Η μέση γωνιακή απόκλιση s αυξάνεται καθώς το μέσο μέτρο r ελαττώνεται.

κής απόκλισης του δείγματος είναι

$$s_1 = s_2/2$$

Διόρθωση για ομαδοποιημένα δεδομένα

Κατά την ομαδοποίηση των γωνιακών δεδομένων, ο κύκλος μπορεί να διαιρεθεί σε ισομήκη διαστήματα (τόξα) και να αριθμηθούν οι παρατηρήσεις κάθε διαστήματος, οι οποίες, τότε, παίρνουν ως τιμή το μέσο του διαστήματος αυτού. Αν k είναι ο αριθμός των τόξων, τότε το μήκος κάθε τόξου ισούται με

$$\lambda = 2\pi/k \quad (\text{ακτίνια})$$

Το λ καλείται *μήκος κλάσης*. Αν $\phi_1, \phi_2, \dots, \phi_n$ είναι οι γωνίες που μετρούντα μέσα σημεία των κλάσεων και n_1, n_2, \dots, n_k είναι οι συχνότητες των δειγματικών σημείων σε κάθε κλάση, με $n = n_1 + n_2 + \dots + n_k$ το μέγεθος του δείγματος, τότε οι ποσότητες

$$\begin{aligned} \bar{x} &= \frac{1}{n}(n_1 \cos \phi_1 + n_2 \cos \phi_2 + \dots + n_k \cos \phi_k) \\ \bar{y} &= \frac{1}{n}(n_1 \sin \phi_1 + n_2 \sin \phi_2 + \dots + n_k \sin \phi_k) \end{aligned} \quad (1.25)$$

είναι οι συνιστώσες του μέσου διανύσματος \mathbf{m} .

Χωρίς διόρθωση, ενώ η μέση γωνία $\bar{\phi}$ δεν επηρεάζεται, η εκτίμηση για το μέσο μέτρο r είναι μεροληπτική και τείνει να υποεκτιμά την αντίστοιχη παράμετρο του πληθυσμού ρ_1 . Γι' αυτό, πολλαπλασιάζουμε το r με ένα διορθωτικό

ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

παράγοντα $c > 1$. Τότε, η διορθωμένη τιμή του r είναι

$$r_c = cr \quad (1.26)$$

Αν το μήκος των κλάσεων μετριέται σε ακτίνια, αποδεικνύεται [Greenwood (1959), Gilroy (1965) και Mardia (1972) — τρεις διαφορετικές αποδείξεις] ότι ο παράγοντας διόρθωσης c είναι μεγάλους

$$c = \frac{\lambda/2}{\sin \lambda/2} \quad (1.27)$$

Η διόρθωση των ομαδοποιημένων δεδομένων επηρεάζει έμμεσα τη μέση γωνιακή απόκλιση s . Αν με s_c συμβολίσουμε τη διορθωμένη τιμή του s , υπολογίζουμε από την εξ. (1.22) ότι είναι

$$s_c = \sqrt{2(1 - r_c)} \quad (1.28)$$

όπου το s_c μετριέται σε ακτίνια.

Η παραπάνω διόρθωση για ομαδοποιημένα δεδομένα δίνει ικανοποιητικά αποτελέσματα μόνο για δείγματα από μονοκόρυφες και αρκετά συμμετρικές κατανομές (Batschelet, 1981, p. 38). Επίσης, αν ο αριθμός των ομάδων είναι μεγαλύτερος από 8, δηλαδή τα διαστήματα ομαδοποίησης είναι $\leq 45^\circ$, είναι $c = \frac{\lambda/2}{\sin \lambda/2} \leq 1.03$ και τότε η διόρθωση έχει αμελητέα επίδραση, οπότε παραβλέπεται (Mardia and Jupp, 2000, p. 23).

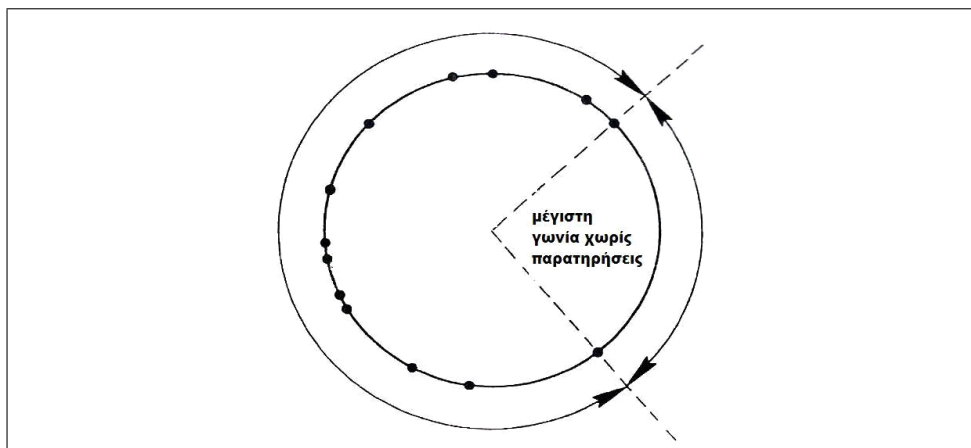
Άλλα μέτρα διασποράς

Δειγματική κυκλική διασπορά Οι Mardia and Jupp (2000, p. 19) ορίζουν ως *δειγματική κυκλική διασπορά* $\hat{\delta}$ την ποσότητα

$$\hat{\delta} = \frac{1 - r_2}{2r^2} \quad (1.29)$$

όπου r_2 και r το μέσο μέτρο των διπλασιασμένων και των αρχικών γωνιών του δείγματος, αντίστοιχα.

Κυκλικό εύρος Πρακτική αξία παρουσιάζει το *κυκλικό εύρος* (circular range) ή *εύρος*, το οποίο ισούται με *το μήκος του μικρότερου τόξου που περιέχει όλα τα δειγματικά σημεία* (βλ. σχήμα 1.13). Πάντως, το εύρος είναι μάλλον ένα χονδροειδές μέτρο της διασποράς (Batschelet, 1981, p. 41).



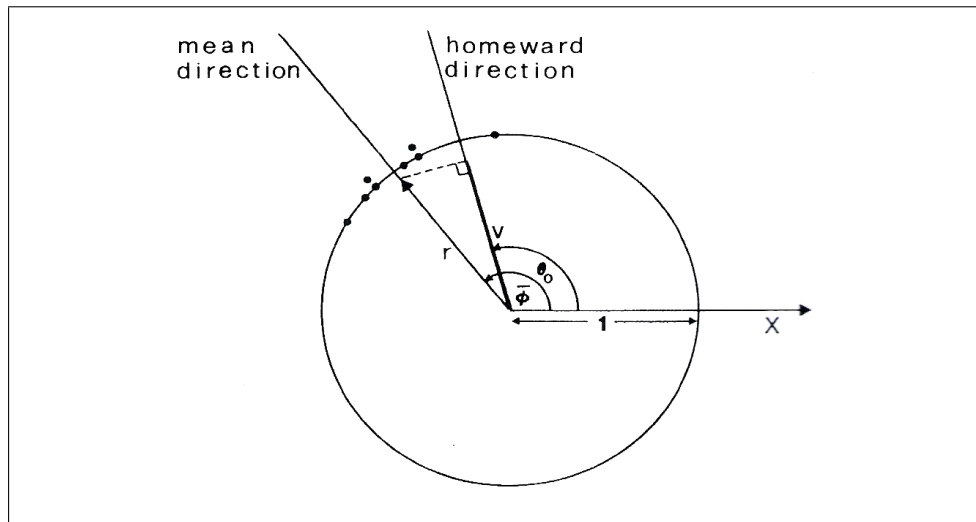
Σχήμα 1.13: Το εύρος ενός κυκλικού δείγματος ισούται με το μήκος του μικρότερου τόξου που περιέχει όλα τα δειγματικά σημεία

Παράγοντας επαναπατριsmού Σε πειράματα επαναπατριsmού (homing experiments) δεν ενδιαφέρει μόνο το κατά πόσο διαφέρουν οι κατευθύνσεις επιστροφής (vanishing directions) από τη μέση κατεύθυνση, αλλά και το κατά πόσο η μέση αυτή κατεύθυνση είναι κοντά στην κατεύθυνση επαναπατριsmού (homeward direction). Προκειμένου να αποκτήσουμε ένα κατάλληλο μέτρο γι' αυτού του τύπου τη συμπεριφορά επαναπατριsmού (homing behaviour), μπορούμε να συνδυάσουμε το μέτρο της συγκέντρωσης r με τη γωνία μεταξύ της μέσης κατεύθυνσης και της κατεύθυνσης επαναπατριsmού. Για το λόγο αυτό, υποθέτουμε ότι η κατεύθυνση επαναπατριsmού σχηματίζει με το θετικό x -ημιάξονα γωνία θ_0 . Τότε, για μέσο μέτρο r και μέση γωνία $\bar{\phi}$, η ποσότητα

$$v = r \cos(\bar{\phi} - \theta_0) \quad (1.30)$$

αποτελεί τη συνιστώσα του μέσου διανύσματος με αναλογία προς την κατεύθυνση επαναπατριsmού. Για το λόγο αυτό, το v καλείται παράγοντας επαναπατριsmού (homeward component). Το v παίρνει τη μέγιστη τιμή του, $v = 1$, αν, για παράδειγμα, όλα τα ζώα κινούνται ακριβώς κατά την κατεύθυνση επαναπατριsmού. Ανάλογα, όσο μεγαλύτερη είναι η διασπορά και όσο περισσότερο η μέση κατεύθυνση αποκλίνει από την κατεύθυνση επαναπατριsmού, τόσο μικρότερο γίνεται το v . Ο παράγοντας επαναπατριsmού v γίνεται αρνητικός, όταν η διαφορά $\bar{\phi} - \theta_0$ υπερβεί τις 90° . Επομένως, ο παράγοντας επαναπατριsmού μπορεί να χρησιμοποιηθεί ως μέτρο της συμπεριφοράς επαναπατριsmού.

Όπως το μέσο μέτρο r , έτσι και ο παράγοντας επαναπατριsmού v εξαρτά-



Σχήμα 1.14: Ο παράγοντας επαναπατρισμού v ως μέτρο της συμπεριφοράς επαναπατρισμού.

ται από το μέγεθος του δείγματος. Σχετικά μικρά δειγματικά μεγέθη ευνοούν μεγάλες τιμές του v . Για το λόγο αυτό, δεν μπορούμε να συγκρίνουμε τους παράγοντες επαναπατρισμού από δείγματα διαφορετικού μεγέθους.

Τα παραπάνω μέτρα έχουν νόημα μόνο για δείγματα που προέρχονται από μονοκόρυφους πληθυσμούς.

Μέτρα λοξότητας και κύρτωσης

Ο Batschelet (1981, p. 43) αναφέρει ότι ο βασικός όρος που προσδιορίζει τη *λοξότητα* ενός δείγματος κυκλικών δεδομένων είναι ο

$$r_2 \sin(\bar{\phi}_2 - 2\bar{\phi}) \quad (1.31)$$

όπου r_2 και $\bar{\phi}_2$ το μέσο μέτρο και η μέση γωνία που προκύπτουν από τη μέθοδο του διπλασιασμού των γωνιών του δείγματος και $\bar{\phi}$ είναι η μέση γωνία του αρχικού δείγματος. Σε ένα συμμετρικό δείγμα, η παραπάνω ποσότητα μηδενίζεται.

Οι Mardia and Jupp (2000, p. 23) δίνουν το παραπάνω στατιστικό ελαφρώς τροποποιημένο

$$\frac{r_2 \sin(\bar{\phi}_2 - 2\bar{\phi})}{(1 - r)^{3/2}} \quad (1.32)$$

όπου r το αρχικό μέσο μέτρο.

Ακόμα, ο Mardia (1972) προτείνει ως μέτρο λοξότητας το

$$g_1 = \frac{r_2 \sin(\bar{\phi}_2 - 2\bar{\phi})}{s^3} \quad (1.33)$$

όπου τα r_2 , $\bar{\phi}_2$ και $\bar{\phi}$ ορίζονται όπως στην (1.31) και s είναι η γωνιακή απόκλιση της (1.22). Ο παρονομαστής στην εξ. (1.33) χρησιμοποιείται για να ελαχιστοποιήσει πιθανές επιδράσεις της διασποράς.

Μια ποσότητα που υποδεικνύει κατάλληλα την *κύρτωση*, δηλαδή το βαθμό κατά τον οποίο η κατανομή του δείγματος παρουσιάζει κορυφή σχετικά "επίπεδη" (flat) ή "απότομη" (sharply peaked), κατά τον Batschelet (1981, p. 43), είναι η

$$\frac{r_2 \cos(\bar{\phi}_2 - 2\bar{\phi}) - r^4}{(1 - r)^2} \quad (1.34)$$

με τα r_2 , $\bar{\phi}_2$ και $\bar{\phi}$ να ορίζονται όπως στην (1.31).

Για την εκτίμηση της κύρτωσης δειγμάτων από *αναδιπλούμενες* (wrapped) κανονικές κατανομές, οι Mardia and Jupp (2000, p. 23) προτείνουν την ποσότητα

$$\frac{r_2 \sin(\bar{\phi}_2 - 2\bar{\phi})}{(1 - r)^{3/2}} \quad (1.35)$$

με τα r_2 , $\bar{\phi}_2$, $\bar{\phi}$ και r να ορίζονται όπως στην (1.32).

Τα μέτρα λοξότητας και κύρτωσης έχουν νόημα μόνο για μονοκόρυφες κατανομές.

*ΚΕΦΑΛΑΙΟ 1. ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΑ
ΜΕΤΡΑ*

Κεφάλαιο 2

Κυκλικές κατανομές

2.1 Εισαγωγή

Οι κυκλικές κατανομές είναι κατανομές που εκτείνονται από 0 έως 2π ακτίνα (ισοδύναμα, από 0 έως 360 μοίρες), ενώ οι κατανομές γραμμικών μεταβλητών μπορεί είτε να έχουν πεπερασμένο εύρος (διωνυμική κατανομή) είτε να επεκτείνονται στο $+\infty$ (κατανομή Poisson) είτε να εκτείνονται καθ' όλο το μήκος του πραγματικού άξονα (κανονική κατανομή).

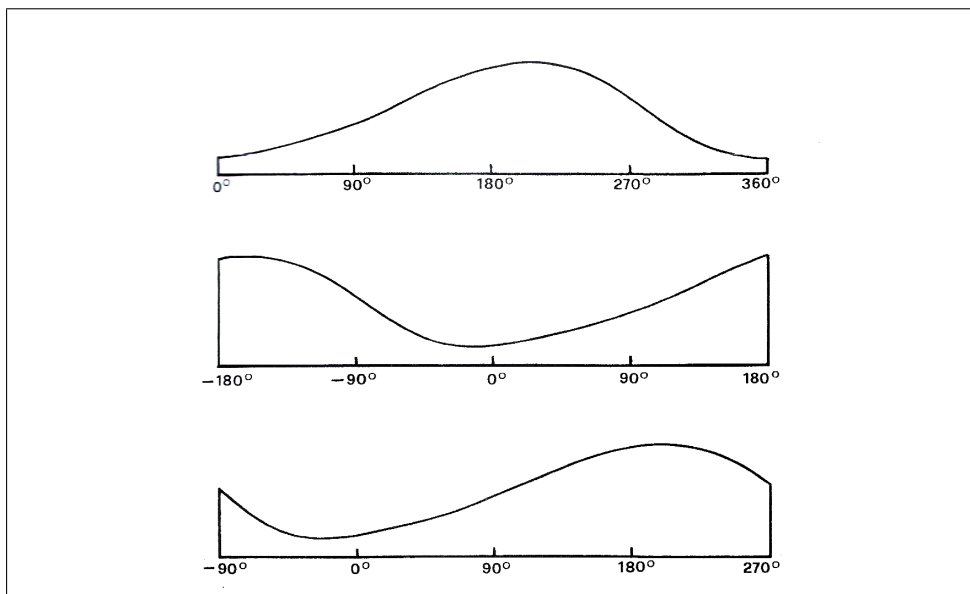
Κατά κανόνα, θεωρούμε τις κυκλικές κατανομές *συνεχείς*: υπάρχει μία *πυκνότητα πιθανότητας*, δηλαδή μία συνεχής συνάρτηση της γωνίας ϕ ,

$$f(\phi), \quad 0 \leq \phi < 2\pi$$

Η πυκνότητα αυτή είναι πάντοτε μη αρνητική και το ολοκλήρωμά της σε όλο το εύρος της ισούται με την ολική πιθανότητα 1:

$$\int_0^{2\pi} f(\phi) d\phi = 1 \quad (2.1)$$

Η κατανομή πιθανότητας μιας κυκλικής μεταβλητής μπορεί να αναπαρασταθεί με τη χρησιμοποίηση ενός γραμμικού διαγράμματος, όπου η πυκνότητα $f(\phi)$ εκφράζεται από ζεύγη καρτεσιανών συντεταγμένων. Σοβαρό μειονέκτημα σ' αυτή την περίπτωση αποτελεί το γεγονός ότι το (αυθαίρετο πάντοτε στις κυκλικές κατανομές) αρχικό σημείο που θα επιλεγεί μπορεί να προκαλέσει συγχύσεις. Αυτό παρουσιάζεται εμφανώς στο παρακάτω σχήμα 2.1, όπου αναπαρίσταται σε γραμμικό διάγραμμα η ίδια κυκλική κατανομή, υπό τρία διαφορετικά αρχικά σημεία.



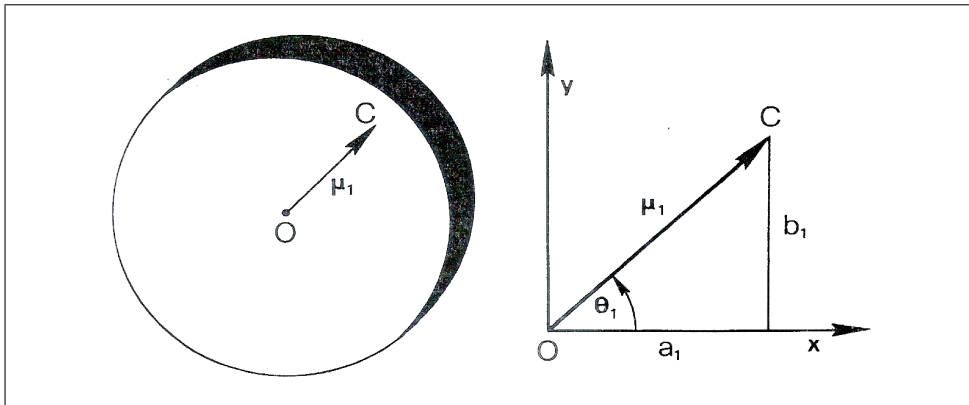
Σχήμα 2.1: Τρία διαφορετικά γραμμικά διαγράμματα της ίδιας κυκλικής κατανομής. Η συγκεκριμένη κατανομή είναι μονοκόρυφη, με την κορυφή της να βρίσκεται στη γωνία των 200° . Έχουν επιλεγεί τρία διαφορετικά αρχικά σημεία, τα 0° , -180° και -90° . Είναι προφανής η σύγχυση ως προς τη μορφή της κατανομής.

Μία διαισθητικότερη άποψη της κατανομής πιθανότητας μιας κυκλικής κατανομής μπορεί να δώσει ένα κυκλικό διάγραμμα, όπως αυτό του σχήματος 2.2, όπου από έναν κύκλο αυθαίρετης ακτίνας, η πυκνότητα $f(\phi)$ σχεδιάζεται προς την κατεύθυνση του μέσου διανύσματος μ_1 . Καθώς οι κατευθύνσεις εκλαμβάνονται και ως σημεία πάνω στην περιφέρεια του μοναδιαίου κύκλου, μία κυκλική κατανομή μπορεί να εννοηθεί ως μία διμεταβλητή κατανομή υποκειμένη στον περιορισμό του να διασπείρεται η ολική πιθανότητα (ολική μάζα) στο μοναδιαίο κύκλο, όπου, όπως και κατά τη στατιστική των συνήθων διμεταβλητών κατανομών, υπάρχει ένα μέσο διάνυσμα που σημειώνει το κέντρο βάρους.

Έστω μ_1 το μέσο διάνυσμα μιας κυκλικής κατανομής και ρ_1 το μήκος του έτσι, ώστε:

$$\rho_1 = |\mu_1|, \quad 0 \leq \rho_1 \leq 1 \quad (2.2)$$

Ακόμα, έστω θ_1 η πολική γωνία του μέσου διανύσματος μ_1 . Τότε, αν είναι $\rho_1 > 0$, καλούμε τη θ_1 μέση κατεύθυνση της κατανομής. Μπορούμε ακόμα να εισάγουμε τις καρτεσιανές συντεταγμένες του μ_1 , έστω τις a_1 και b_1 κατά τη διεύθυνση του x -άξονα και του y -άξονα, αντίστοιχα. Τότε, κατ' αντιστοιχία με τις εξ. (1.3), (1.4), (1.5) και (1.6), σελ. 9 για τα δειγματικά μεγέθη, είναι



Σχήμα 2.2: Παράδειγμα συνεχούς κυκλικής κατανομής. Αριστερά, φαίνεται η πιθανότητα, η οποία μπορεί να εκληφθεί ως μάζα μοναδιαίου μεγέθους διεσπαρμένη γύρω από την περιφέρεια του μοναδιαίου κύκλου και, ακόμα, το μέσο διάνυσμα μ_1 , το οποίο δείχνει στο κέντρο βάρους. Δεξιά, φαίνονται οι καρτεσιανές συντεταγμένες του μ_1 είναι οι a_1 και b_1 , που ορίζουν τη μέση γωνία θ_1 .

για τις παραμέτρους του πληθυσμού:

$$a_1 = \rho_1 \cos \theta_1, \quad b_1 = \rho_1 \sin \theta_1 \quad (2.3)$$

Η επίλυση των παραπάνω εξισώσεων ως προς ρ_1 και θ_1 δίνει:

$$\rho_1 = \sqrt{a_1^2 + b_1^2} \quad (2.4)$$

$$\theta_1 = \begin{cases} \arctan(b_1/a_1), & a_1 > 0 \\ \pi + \arctan(b_1/a_1), & a_1 < 0 \end{cases} \quad (2.5)$$

με τις ειδικότερες περιπτώσεις

$$\theta_1 = \begin{cases} \pi/2, & a_1 = 0 \text{ και } b_1 > 0 \\ 3\pi/2, & a_1 = 0 \text{ και } b_1 < 0 \\ \text{δεν προσδιορίζεται,} & a_1 = 0 \text{ και } b_1 = 0 \end{cases} \quad (2.6)$$

2.2 Η ομοιόμορφη κατανομή

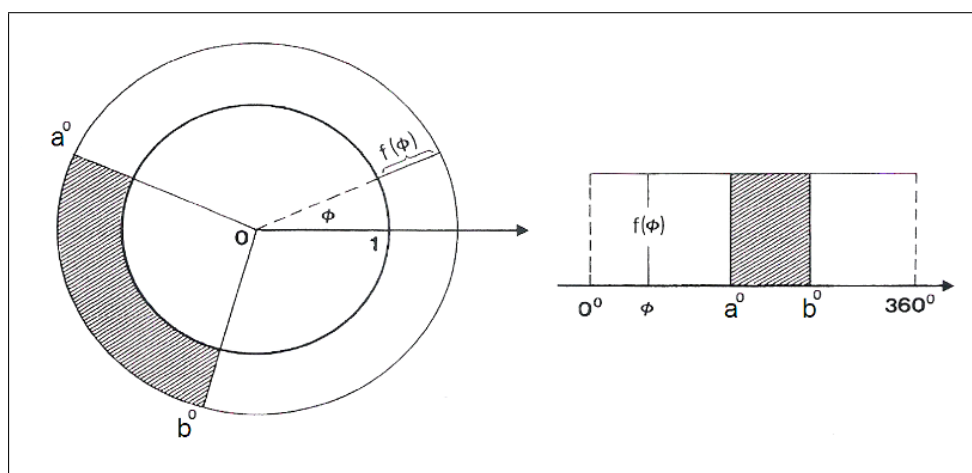
Αν όλες οι κατευθύνσεις πάνω σε ένα επίπεδο μπορεί να επιλέγονται με ίση πιθανότητα, δηλαδή τυχαία, τότε η κατανομή καλείται *ομοιόμορφη* ή *ισοτροπική*. Επίσης, ευρέως χρησιμοποιούμενος είναι για την περίπτωση αυτή ο όρος *τυχαία*, κυρίως από τους φυσικούς επιστήμονες. Πολύ συχνά, χρησιμοποιείται ως το μηδενικό υπόδειγμα (null model).

ΚΕΦΑΛΑΙΟ 2. ΚΥΚΛΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

Για μια τυχαία κυκλική μεταβλητή ϕ , η οποία κατανέμεται ομοιόμορφα και μετρείται σε ακτίνια, η πυκνότητα πιθανότητας $f(\phi)$ είναι σταθερή γύρω από την περιφέρεια του μοναδιαίου κύκλου και ισούται με

$$f(\phi) = \frac{1}{2\pi} \quad (2.7)$$

δηλαδή, η πιθανότητα είναι ανάλογη του μήκους του τόξου.



Σχήμα 2.3: Η ομοιόμορφη κατανομή ως υπόδειγμα της τυχαιότητας. Αριστερά, κυκλικό διάγραμμα της πυκνότητας πιθανότητας και δεξιά, το αντίστοιχο γραμμικό διάγραμμα. Η γραμμοσκιασμένη περιοχή σημειώνει την πιθανότητα του τομέα από a° έως b° , $a \leq b$, δηλαδή την $P(a < \phi \leq b) = \frac{b-a}{2\pi}$.

Λόγω συμμετρίας, το κέντρο της κατανομής συμπίπτει με το αρχικό σημείο O . Ως εκ τούτου, είναι $\rho_1 = 0$, δηλαδή δεν υπάρχει μέση κατεύθυνση.

Η ομοιόμορφη κατανομή στον κύκλο έχει την αξιοσημείωτη ιδιότητα ότι είναι η μόνη κυκλική κατανομή, για την οποία, για τυχαία δείγματα σταθερού μεγέθους, η μέση γωνία είναι *στοχαστικά ανεξάρτητη* από το μήκος του μέσου διανύσματος (Batschelet, 1981, p. 279). Επίσης, είναι αμετάβλητη υπό περιστροφή και αντικατοπτρισμό (Mardia and Jupp, 2000, p. 35).

Για κάθε σύνολο ανεξάρτητων και ισόνομων τυχαίων κυκλικών μεταβλητών $\phi_1, \phi_2, \dots, \phi_n$, το άθροισμά τους $S_n = \phi_1 + \phi_2 + \dots + \phi_n$ τείνει να κατανέμεται ομοιόμορφα, καθώς το $n \rightarrow \infty$, υπό τη συνθήκη ότι δεν προέρχονται από πληθυσμό που ακολουθεί τη *δικτυωτή κατανομή* (lattice distribution) (Mardia and Jupp, 2000, p. 36).

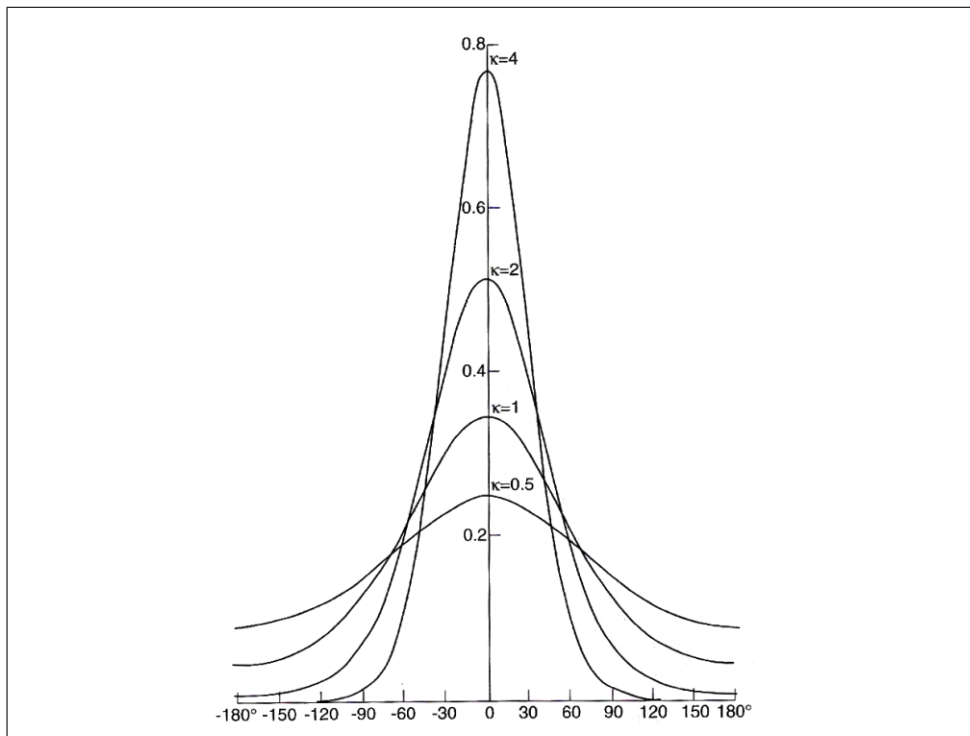
2.3 Η κατανομή von Mises

Η κατανομή von Mises αποτελεί μια μονοκόρυφη κυκλική κατανομή με πυκνότητα πιθανότητας

$$f(\phi) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\phi - \theta_1)} \quad (2.8)$$

όπου ϕ τυχαία κυκλική μεταβλητή. Η κατανομή προσδιορίζεται από δύο παραμέτρους, την *παραμέτρο συγκέντρωσης* (concentration parameter) κ ($\kappa \geq 0$) και τη μέση κατεύθυνση θ_1 . Η ποσότητα $I_0(\kappa)$ εκφράζει την τιμή της τροποποιημένης συνάρτησης Bessel

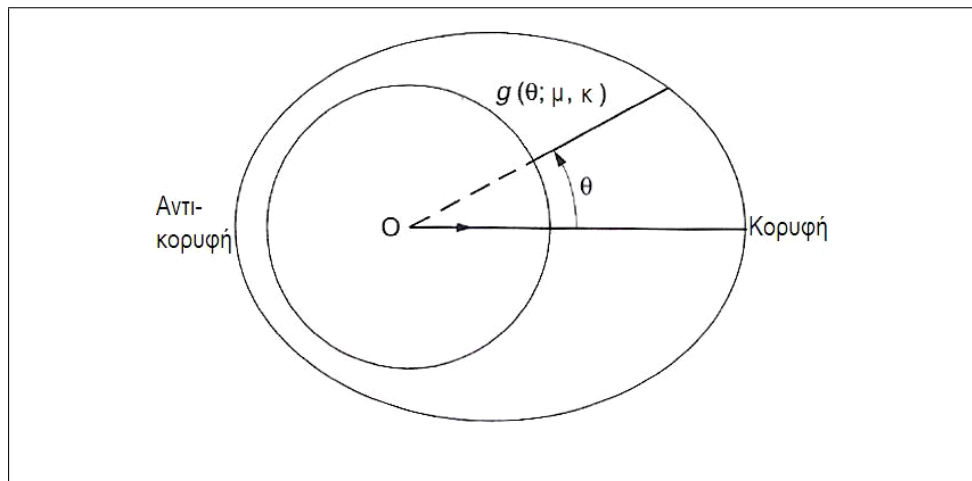
$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \phi} d\phi \quad (2.9)$$



Σχήμα 2.4: Η πυκνότητα της κατανομής von Mises $VM(0, \kappa)$, για $\kappa = 0.5, 1, 2, 4$.

Η συνάρτηση (2.8) παίρνει τη μέγιστη τιμή της στο $\phi = \theta_1$, δηλαδή η θ_1 είναι η κορυφή (mode). Η κατανομή είναι μονοκόρυφη και συμμετρική ως προς την κορυφή θ_1 , δηλαδή τη μέση κατεύθυνση. Η αντικορυφή (antimode) βρίσκεται

ται στο σημείο $\theta_1 + \pi$.



Σχήμα 2.5: Πολική αναπαράσταση της πυκνότητας $g(\theta; \mu, \kappa)$ της κατανομής von Mises $VM(\mu, \kappa)$.

Όσο μεγαλύτερο είναι το κ , τόσο περισσότερο η κατανομή συγκεντρώνεται γύρω από τη μέση κατεύθυνση θ_1 . Αυτό εκφράζεται, επίσης, από το λόγο της πυκνότητας της κορυφής προς την πυκνότητα της αντικορυφής, ο οποίος ισούται με $e^{2\kappa}$ (Mardia and Jupp, 2000, p. 36). Για $\kappa = 0$, η κατανομή von Mises εκφυλίζεται στην ομοιόμορφη κατανομή.

Το κέντρο βάρους της κατανομής σημειώνεται από το μέσο διάνυσμα μ_1 . Η πολική του γωνία είναι η θ_1 και το μήκος του ισούται με

$$|\mu_1| = \rho_1 = A(\kappa)$$

με τη συνάρτηση $A(\kappa)$ να προσδιορίζεται ως

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} \quad (\kappa \geq 0) \quad (2.10)$$

όπου το $I_0(\kappa)$ δίνεται από την εξ. (2.9) και το $I_1(\kappa)$ αποτελεί την πρώτη παράγωγο του

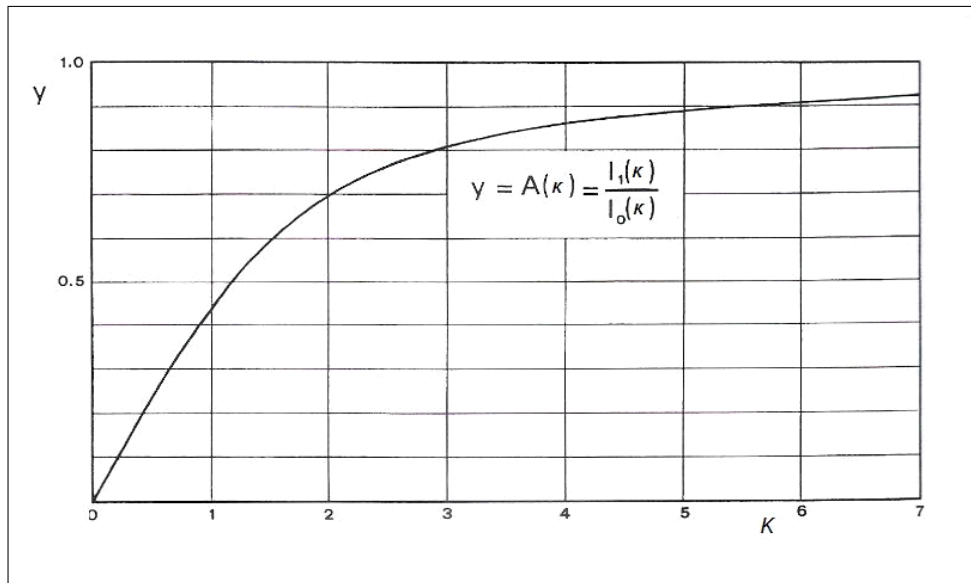
$$I_1(\kappa) = \frac{d}{d\kappa} I_0(\kappa)$$

Η συνάρτηση (2.10) είναι γνησίως αύξουσα, με ελάχιστη τιμή $A(0) = 0$ για $\kappa = 0$, η οποία προσεγγίζει ασυμπτωτικά το $+\infty$, καθώς το $\kappa \rightarrow +\infty$ (βλ. σχήμα 2.6). Αφού η $y = A(\kappa)$ είναι γνησίως μονότονη συνάρτηση, υπάρχει η

αντίστροφή της

$$\kappa = A^{-1}(y) \quad (2.11)$$

η οποία είναι, επίσης, γνησίως μονότονη. Ο Batschelet (1981, pp. 317, 318) δίνει πίνακες τιμών για τις συναρτήσεις (2.10) και (2.11), δηλαδή, για τη μετατροπή της παραμέτρου κ σε ρ_1 και, αντίστροφα, για τη μετατροπή της παραμέτρου ρ_1 σε κ .



Σχήμα 2.6: Γράφημα της συνάρτησης $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$, ($\kappa \geq 0$). Καθώς αυξάνεται η τιμή της παραμέτρου συγκέντρωσης κ (οριζόντιος άξονας), αυξάνεται το μέσο μέτρο $\rho_1 = A(\kappa)$ (κάθετος άξονας).

Οι κατανομές $VM(\theta_1 + \pi, \kappa)$ και $VM(\theta_1, -\kappa)$ είναι ταυτόνομες. Για την αποφυγή συγχύσεων, συνήθως ορίζουμε $\kappa \geq 0$ (Mardia and Jupp, 2000, p. 36).

Η συνάρτηση κατανομής της $VM(0, \kappa)$ δίνεται από την

$$F(\phi; 0, \kappa) = \frac{1}{I_0(\kappa)} \int_0^\phi e^{\kappa \cos u} du \quad (2.12)$$

Ο υπολογισμός της (2.12) είναι δυσχερής. Οι Mardia and Jupp (2000, pp. 354–360) δίνουν πίνακα τιμών της $F(\phi; 0, \kappa)$ για $0 < \kappa \leq 10$.

Η συγκεκριμένη κατανομή προτάθηκε από τον von Mises (1918). Στην ανάλυση των κατευθύνσεων, η κατανομή von Mises αποτελεί την πιο συχνά χρησιμοποιούμενη κατανομή και παίζει έναν παρόμοιο ρόλο με αυτόν της κανονικής κατανομής στη γραμμική στατιστική ανάλυση. Για το λόγο αυτό, οι Gumbel,

Greenwood και Durand (1953) την ονόμασαν *κυκλική κανονική κατανομή*. Πάντως, τα τελευταία χρόνια ο όρος "κυκλική κανονική" είναι λιγότερο δημοφιλής, εν μέρει διότι μπορεί να προκαλέσει σύγχυση σε σχέση με τη διμεταβλητή κανονική κατανομή με $\sigma_1 = \sigma_2, \rho = 0$. Στη δεύτερη περίπτωση, οι ισοσταθμικές καμπύλες της πυκνότητας πιθανότητας είναι κύκλοι.

2.4 Άλλες κυκλικές κατανομές

Εκτός από τις δύο βασικές κυκλικές κατανομές, δηλαδή την ομοιόμορφη (ενότ. 2.2) και τη von Mises (ενότ. 2.3), υπάρχουν και άλλες, συχνά χρησιμοποιούμενες, κυκλικές κατανομές, μερικές από τις οποίες παρουσιάζονται εν συντομία παρακάτω.

Αναδιπλούμενες (wrapped) κατανομές Μια δεδομένη γραμμική κατανομή στην πραγματική ευθεία, μπορεί να αναδιπλωθεί κατά την περιφέρεια του μοναδιαίου κύκλου. Έτσι, αν x είναι μια τυχαία γραμμική μεταβλητή, η αντίστοιχη κυκλική μεταβλητή x_w της αναδιπλούμενης κατανομής δίνεται από την

$$x_w = x \pmod{2\pi}^1 \quad (2.13)$$

Αναδιπλούμενη κανονική κατανομή Η αναδιπλούμενη κανονική κατανομή (wrapped normal distribution) $WN(\mu_1, \rho_1)$ προκύπτει από την αναδίπλωση στον κύκλο της κανονικής κατανομής $N(\mu_1, \sigma^2)$, όπου είναι

$$\sigma^2 = -2 \ln \rho_1 \Leftrightarrow \rho_1 = e^{-\sigma^2/2} \quad (2.14)$$

με συνάρτηση πυκνότητας πιθανότητας

$$f(\phi; \mu_1, \rho_1) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{k=-\infty}^{+\infty} \exp \left[-\frac{(\phi - \mu_1 + 2\pi k)^2}{2\sigma^2} \right] \quad (2.15)$$

Η κατανομή $WN(\mu_1, \rho_1)$ είναι μονοκόρυφη και συμμετρική γύρω από την κορυφή μ_1 . Τείνει στην ομοιόμορφη κατανομή, καθώς το $\rho_1 \rightarrow 0$, ενώ τείνει στη σημειακή κατανομή μ_1 , καθώς το $\rho_1 \rightarrow 1$ (Mardia and Jupp, 2000, p. 50).

¹Με τη χρησιμοποίηση μοναδιαίων μιγαδικών αριθμών, η σχέση αναδίπλωσης $x \mapsto x_w$ μπορεί να γραφτεί στη μορφή $x \mapsto e^{2\pi i x}$ (Mardia and Jupp, 2000, p. 47)

Αναδιπλούμενη Cauchy κατανομή Μια γραμμική κατανομή Cauchy στην πραγματική ευθεία, αναδιπλούμενη δίνει μια μονοκόρυφη και συμμετρική κυκλική κατανομή, την *αναδιπλούμενη Cauchy κατανομή* (wrapped Cauchy distribution). Με μέση γωνία θ_1 και μέσο μέτρο ρ_1 , η πυκνότητα πιθανότητας ισούται με

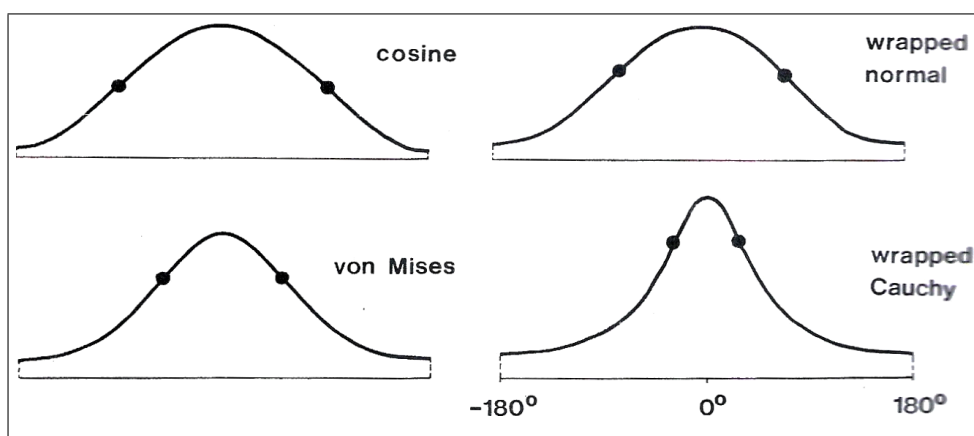
$$f(\phi) = \frac{1}{2\pi} \frac{1 - \rho_1^2}{1 + \rho_1^2 - 2\rho_1 \cos(\phi - \theta_1)}. \quad (2.16)$$

Για $\rho_1 = 0$, η wrapped Cauchy κατανομή εκφυλίζεται στην ομοιόμορφη κυκλική κατανομή (Batschelet, 1981, p. 282).

Κατανομή συνημίτονο Η κατανομή *συνημίτονο* (sine-wave distribution) παράγεται από μια συνημιτονοειδή καμπύλη πάνω στο διάστημα $[-180^\circ, +180^\circ]$. Η πυκνότητα πιθανότητάς της, για μέση γωνία θ_1 και μέσο μέτρο ρ_1 είναι

$$f(\phi) = \frac{1}{2\pi} + \frac{\rho_1}{\pi} \cos(\phi - \theta_1) \quad (2.17)$$

όπου, προκειμένου να αποφύγουμε αρνητικές τιμές για την $f(\phi)$, υποθέτουμε ότι είναι $\rho_1 \leq 1/2$. Συνεπώς, η κατανομή *συνημίτονο* δεν μπορεί να χρησιμοποιηθεί ως μοντέλο για την καλή προσαρμογή δείγματος μέσου μέτρου $r_1 > 1/2$.



Σχήμα 2.7: Τέσσερις κυκλικές κατανομές με παρόμοιο σχήμα και διαφορετική σε κάθε περίπτωση απόσταση των σημείων καμπής (του αριστερού και του δεξιού σκέλους της κάθε κατανομής). Ακραίες περιπτώσεις αποτελούν η κατανομή *συνημίτονο* με τη μεγαλύτερη από τις τέσσερις κατανομές απόσταση των σημείων καμπής (ίση με 180°) και η *αναδιπλούμενη Cauchy* κατανομή με τη μικρότερη τέτοια απόσταση.

Πολυκόρυφες κατανομές Πολλές περιπτώσεις ανάλυσης δεδομένων κατεύθυνσης οδηγούν σε πολυκόρυφες κυκλικές κατανομές. Για παράδειγμα, αν ενδιαφέρει μόνο η παρατήρηση της θέσης στο επίπεδο (δηλαδή της διεύθυνσης) των ευθειών που προκύπτουν από κάποια δεδομένα, τότε η δειγματική κατανομή αποκτά δύο κορυφές, μία κατά την πρώτη κατεύθυνση (όποια από τις δύο ορίζουμε ως πρώτη) και μία κατά την αντίθετη κατεύθυνση. Ειδικότερα, κατά την περίπτωση πολυκόρυφων κυκλικών κατανομών, τύπου von Mises η κάθε μία, αν με ν δηλώσουμε το πλήθος των κορυφών, η συνάρτηση πυκνότητας αυτής της *πολυκόρυφης von Mises κατανομής* δίνεται από την

$$f(\phi) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos \nu(\phi - \theta)] \quad (2.18)$$

όπου το θ δηλώνει κάποια από τις ν κορυφές. Η συνάρτηση αυτή διαφέρει ελάχιστα (μόνο κατά τον παράγοντα ν) από τη συνάρτηση πυκνότητας (2.8) της (μονοκόρυφης) von Mises κατανομής. Επισημαίνεται, ότι σε μια τέτοια πολυκόρυφη κατανομή το κέντρο βάρους της βρίσκεται στο κέντρο του μοναδιαίου κύκλου. Συνεπώς, το μήκος του μέσου διανύσματος ρ_1 είναι μηδέν.

Κεφάλαιο 3

Διαστήματα εμπιστοσύνης και έλεγχοι υποθέσεων

3.1 Σημειακές εκτιμήσεις και εκτιμήσεις διαστήματος

3.1.1 Σημειακή εκτίμηση παραμέτρων

Η προσαρμογή μιας θεωρητικής κατανομής στα δεδομένα ενός δείγματος απαιτεί τη *σημειακή* εκτίμηση των παραμέτρων της. Η αποδοχή ή όχι της προσαρμογής αποφασίζεται σύμφωνα με τα αποτελέσματα κατάλληλων στατιστικών ελέγχων. Η εκτίμηση των παραμέτρων του μελετούμενου πληθυσμού γίνεται με τον υπολογισμό αντίστοιχων ποσοτήτων από το δείγμα, με τη βοήθεια κατάλληλων στατιστικών συναρτήσεων. Στον παρακάτω πίνακα παρουσιάζουμε την αντιστοίχισή τους:

Πίνακας 3.1: Αντιστοιχίες στατιστικών συναρτήσεων και πληθυσμιακών παραμέτρων.

	Στατιστική συνάρτηση (Δείγμα)	Παράμετρος (Πληθυσμός)	Εκτίμηση
Μέσο διάνυσμα	\mathbf{m}	$\boldsymbol{\mu}_1$	$\hat{\boldsymbol{\mu}}_1$
Μήκος του μέσου διανύσματος	r	ρ_1	$\hat{\rho}_1$
Μέση γωνία	$\bar{\phi}$	θ_1	$\hat{\theta}_1$

3.1.2 Εκτιμήσεις διαστήματος

Για κάθε άγνωστη παράμετρο ενός πληθυσμού, μπορεί να γίνει εκτίμηση του διαστήματος, το οποίο με προκαθορισμένη πιθανότητα περιέχει αυτή την παράμετρο. Προκειμένου να προσδιορίσουμε αυτά τα όρια εμπιστοσύνης πρέπει, κατά πρώτο, να επιλέξουμε ένα συντελεστή εμπιστοσύνης Q (συνήθως $Q = 0.95$ ή $Q = 0.99$). Ο συντελεστής αυτός συνδέεται με το επίπεδο σημαντικότητας α με τη σχέση

$$Q = 1 - \alpha$$

Διαστήματα εμπιστοσύνης για τη μέση κατεύθυνση

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου, γνωστής περιόδου. Υποθέτουμε ότι το συγκεκριμένο δείγμα επιλέχθηκε από πληθυσμό με κατανομή von Mises. Θέλουμε να προσδιορίσουμε όρια εμπιστοσύνης $\bar{\phi} - \delta$ και $\bar{\phi} + \delta$ για τη μέση κατεύθυνση θ_1 .

Για το σκοπό αυτό, υπολογίζουμε το μήκος r του μέσου διανύσματος \mathbf{m} και τη μέση γωνία $\bar{\phi}$ του δείγματος. Κατόπιν, αποφασίζουμε για το συντελεστή εμπιστοσύνης Q . Ο Batschelet (1981, p. 86) δίνει διαγράμματα για τον υπολογισμό της γωνίας απόκλισης δ , κατά περίπτωση, από τους Stephens (1962) και Brown and Mewaldt (1968) (βλ. σχήματα 3.1). Η γωνία απόκλισης προσδιορίζεται από το μέσο μέτρο r και το μέγεθος του δείγματος n .

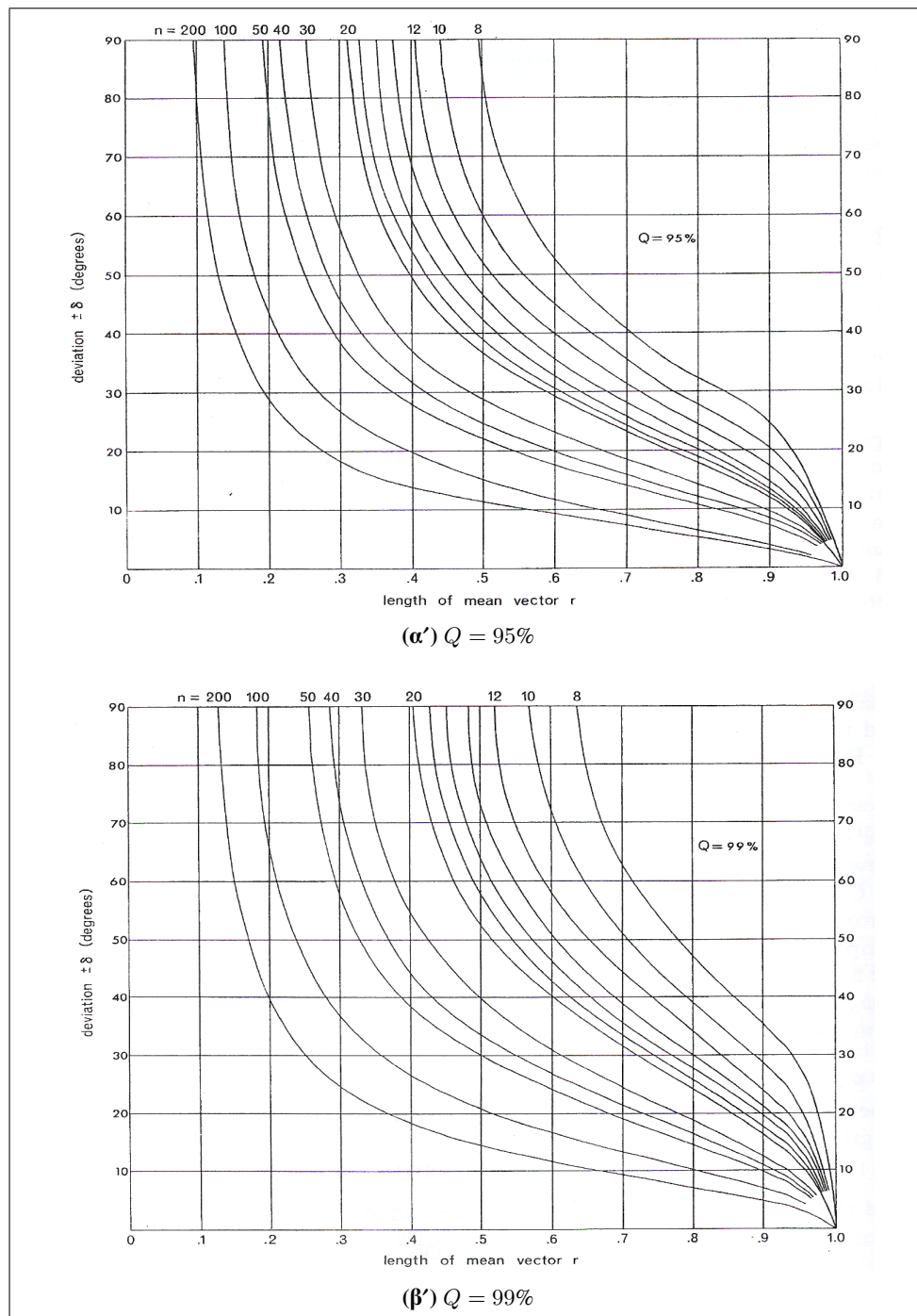
Ο Zar (1999, pp. 605-606) αναφέρει μια μέθοδο, η οποία έχει προταθεί από τον Upton (1986), για τον υπολογισμό του δ , όταν το μέγεθος του δείγματος είναι $n \leq 8$. Τότε, για τιμές του μήκους του μέσου διανύσματος $r \leq 0.9$ και $r > \sqrt{\chi_{\alpha,1}^2/2n}$, το δ δίνεται από τον τύπο

$$\delta = \arccos \left[\frac{\sqrt{\frac{2n(2R^2 - n\chi_{\alpha,1}^2)}{4n - \chi_{\alpha,1}^2}}}{R} \right] \quad (3.1)$$

και για $r \geq 0.9$, το δ δίνεται από τον τύπο

$$\delta = \arccos \left[\frac{\sqrt{n^2 - (n^2 - R^2) \exp(\chi_{\alpha,1}^2/n)}}{R} \right] \quad (3.2)$$

3.1. Σημειακές εκτιμήσεις και εκτιμήσεις διαστήματος



Σχήμα 3.1: Διαγράμματα για τον προσδιορισμό διαστήματος εμπιστοσύνης για τη μέση κατεύθυνση, με συντελεστή εμπιστοσύνης $Q = 95\%$ και $Q = 99\%$, για διάφορα δειγματικά μεγέθη. (Batschelet, 1981).

όπου είναι $R = nr$.

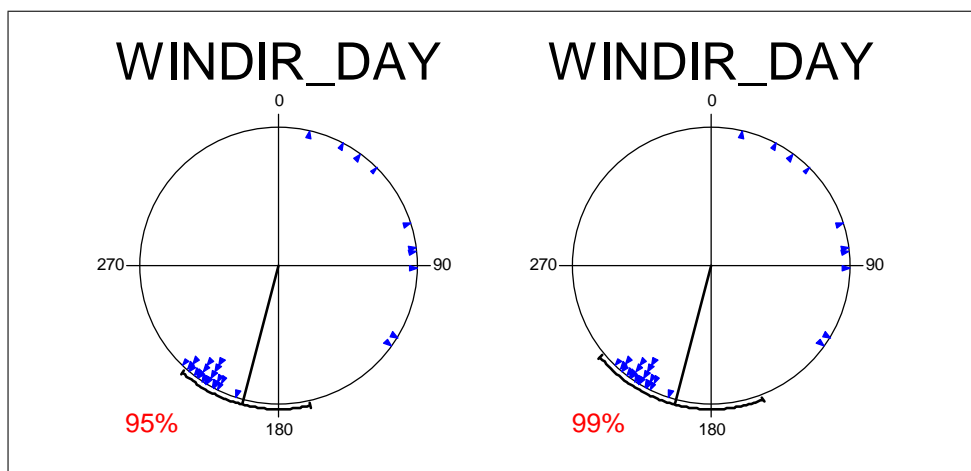
Τα διαστήματα εμπιστοσύνης μπορεί να χρησιμοποιούνται και για τους ελέγχους υποθέσεων,¹ όπου εμπλέκεται μια παράμετρος. Αν, για παράδειγμα, θέλουμε να ελέγξουμε το κατά πόσο η δειγματική μέση γωνία $\bar{\phi}$ διαφέρει σημαντικά από το αζιμούθιο της κατεύθυνσης επαναπατρισμού ϕ_0 , τότε μπορούμε, προσδιορίζοντας τα όρια $\bar{\phi} \pm \delta$, να εξετάσουμε αν το ϕ_0 βρίσκεται ανάμεσα στα όρια αυτά. Αν το ϕ_0 δεν ανήκει στο παραπάνω διάστημα εμπιστοσύνης, έχουμε σημαντικά αποτελέσματα και συμπεραίνουμε ότι η μέση κατεύθυνση του πληθυσμού διαφέρει σημαντικά από τη ϕ_0 .

Διαστήματα εμπιστοσύνης για την παράμετρο συγκέντρωσης κ

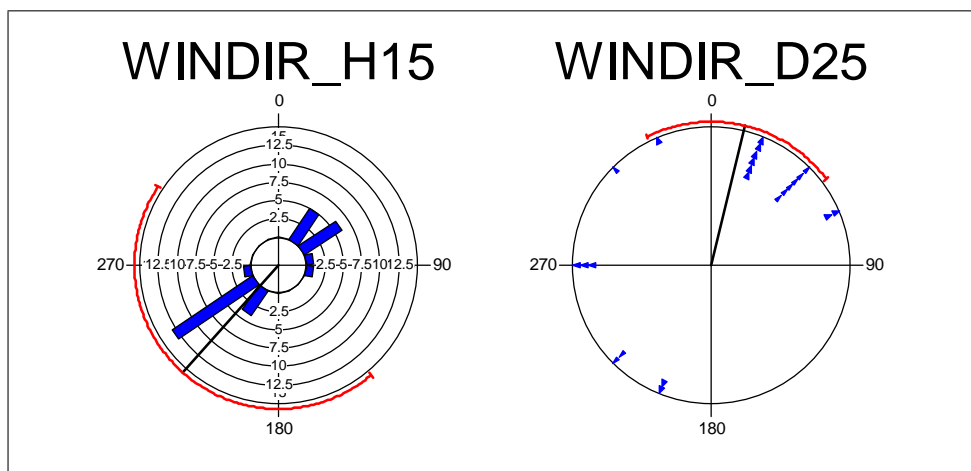
Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Υπό την υπόθεση ότι το δείγμα επιλέχθηκε από πληθυσμό με von Mises κατανομή, θέλουμε να εκτιμήσουμε την παράμετρο κ (οι τιμές του κ κυμαίνονται από 0 έως $+\infty$).

Για το σκοπό αυτό, υπολογίζουμε το μήκος του μέσου διανύσματος r του δείγματος και κατόπιν επιλέγουμε το συντελεστή εμπιστοσύνης Q . Ο Batschelet (1981, pp. 91-92) δίνει διαγράμματα από τα οποία μπορούμε να πάρουμε τα όρια εμπιστοσύνης των κ , κατά περίπτωση ($Q = 90\%$ και $Q = 98\%$). Τα διαγράμματα αυτά βασίζονται στις εργασίες των Stephens (1969d) και Papakonstantinou (1979).

¹ Πρέπει να επισημανθεί, πάντως, ότι, όταν το μέγεθος του δείγματος και η συγκέντρωση έχουν χαμηλές τιμές, τα όρια εμπιστοσύνης μπορεί να είναι αναξιόπιστα — βλ. σχήμα (3.3), σελ.39.

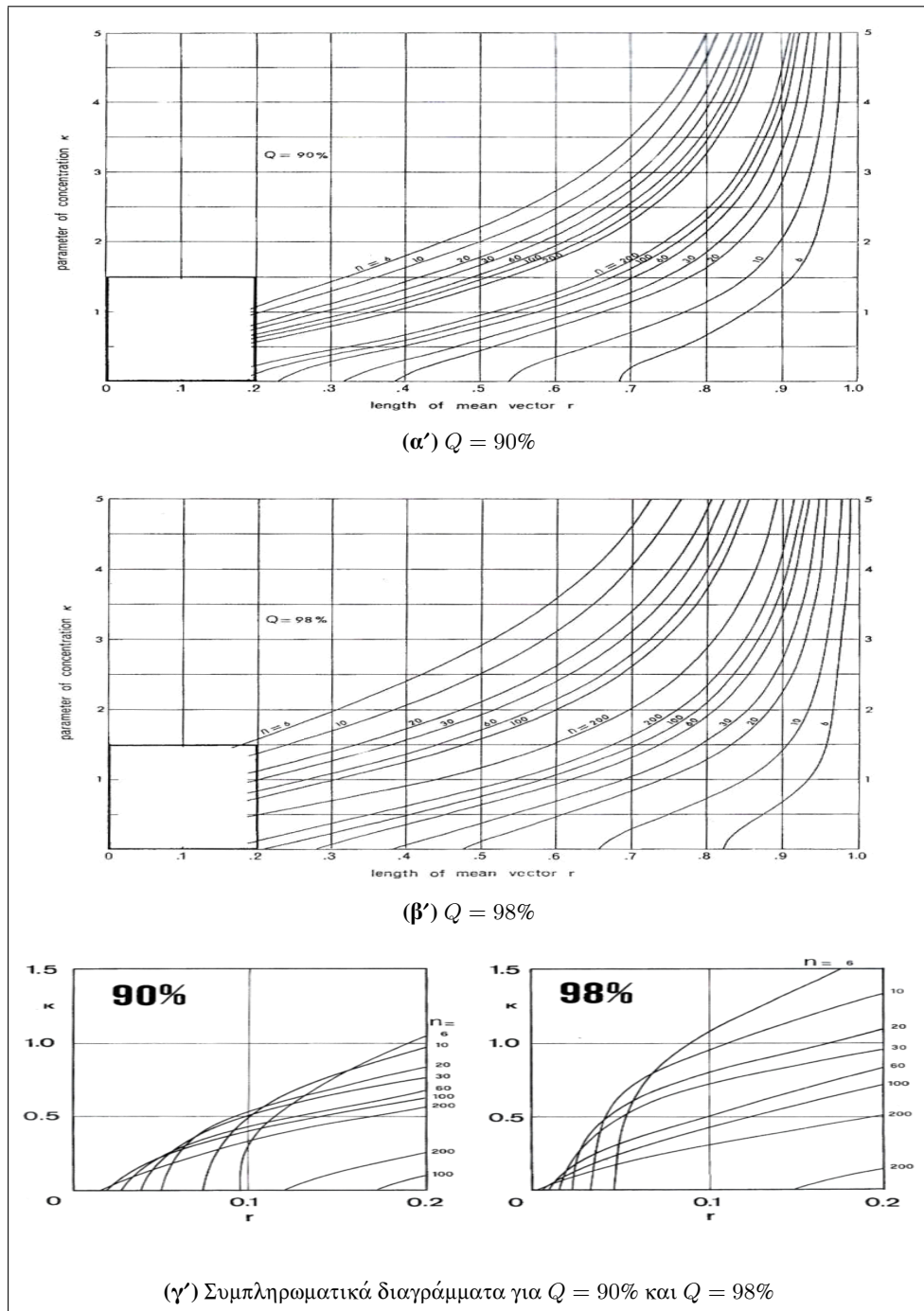


Σχήμα 3.2: Διάγραμμα διασποράς της μεταβλητής δεύτερης τάξης WINDIR_DAY. Η έντονα μαύρη ακτίνα σημειώνει τη μέση γωνία των δεδομένων και τα έντονα μαύρα τόξα τα 95% (αριστερά) και 99% (δεξιά) διαστήματα (τόξα) εμπιστοσύνης για τη μέση κατεύθυνση του πληθυσμού. Το 95% διάστημα εμπιστοσύνης είναι, βεβαίως, μικρότερο.



Σχήμα 3.3: Κυκλικό ιστόγραμμα της μεταβλητής WINDIR_H15 και κυκλικό διάγραμμα διασποράς της μεταβλητής WINDIR_D25. Το κόκκινο τόξο αναπαριστά και στις δύο περιπτώσεις το 95% διάστημα εμπιστοσύνης για τη μέση κατεύθυνση του πληθυσμού, όμως ο κόκκινος χρωματισμός δηλώνει την αναξιοπιστία της εκτίμησης, λόγω του ότι ο συνδυασμός του μεγέθους του δείγματος n και της συγκέντρωσης *Concentration* είναι χαμηλός. Αυτό δεν είναι πάντα ξεκάθαρο: στο συγκεκριμένο σχήμα, η μεταβλητή WINDIR_H15 έχει πράγματι σχετικά χαμηλό συνδυασμό $n = 31$ και *Concentration* = 0.352, αλλά δεν μπορούμε να ισχυριστούμε το ίδιο και για τη μεταβλητή WINDIR_D25, με συνδυασμό $n = 31$ και *Concentration* = 0.869.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ



Σχήμα 3.4: Διαγράμματα για τον προσδιορισμό διαστημάτων εμπιστοσύνης για την παράμετρο συγκέντρωσης κ , για διάφορα δειγματικά μεγέθη, με συντελεστή εμπιστοσύνης $Q = 90\%$ και $Q = 98\%$. Η τομή των καμπυλών με τον πάνω άξονα του κάθε διαγράμματος δίνει το άνω όριο εμπιστοσύνης κ_u και, αντίστοιχα, η τομή τους με τον κάτω άξονα δίνει το κάτω όριο εμπιστοσύνης κ_l . Για τιμές του $r < 0.2$ και του $\kappa < 1.5$ (κενά παραλληλόγραμμα) δίνονται αντίστοιχα συμπληρωματικά διαγράμματα. (Batschelet, 1981).

3.2 Έλεγχοι τυχαιότητας και, γενικότερα, καλής προσαρμογής

Μια από τις σημαντικότερες υποθέσεις για την κατανομή ενός πληθυσμού που συνήθως ελέγχεται είναι αυτή της ομοιομορφίας ή, ισοδύναμα, τυχαιότητας. Στην ενότητα αυτή παρουσιάζονται κάποια τεστ για τον έλεγχο του αν η κατανομή του πληθυσμού από τον οποίο προέρχεται ένα συγκεκριμένο δείγμα ακολουθεί τη θεωρητική ομοιόμορφη κατανομή ή όχι.

ΕΛΕΓΧΟΙ ΤΥΧΑΙΟΤΗΤΑΣ

3.2.1 Rayleigh test

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Θέλουμε να εξετάσουμε το αν ο πληθυσμός από τον οποίο προέρχεται το τυχαίο αυτό δείγμα διαφέρει σημαντικά από το τυχαίο πρότυπο, δηλαδή αν υπάρχει στατιστική ένδειξη προσανατολισμού.

Το μέσο μέτρο r δίνει κάποια ένδειξη προσανατολισμού. Για δείγματα που προέρχονται από ομοιόμορφους πληθυσμούς, το r είναι σχετικά μικρό. Αν το r είναι επαρκώς μεγάλο, η μηδενική υπόθεση της τυχαιότητας μπορεί να απορριφθεί υπέρ της εναλλακτικής υπόθεσης του προσανατολισμού. Επομένως, η ποσότητα r μπορεί να χρησιμοποιηθεί ως το στατιστικό του ελέγχου. Ο Batschelet (1981, p. 334-335) δίνει πίνακες που περιέχουν το κρίσιμο επίπεδο P σε σχέση με το μέγεθος του δείγματος n και το μήκος του μέσου διανύσματος r . Η κατανομή της πιθανότητας του r δόθηκε από τον Mardia (1972).

Το Rayleigh test εξετάζει το πόσο μεγάλο πρέπει να είναι το r , προκειμένου να υποδείξει με εγκυρότητα μια μη ομοιόμορφη κατανομή πληθυσμού. Ο Zar (1999, p. 617) αναφέρει ως "Rayleigh's R " την ποσότητα

$$R = nr \quad (3.3)$$

και παρουσιάζει μια προσέγγιση της πιθανότητας P του R , η οποία δόθηκε αρχικά από τους Greenwood and Durand (1955)

$$P = \exp[\sqrt{1 + 4n + 4(n^2 - R^2)} - (1 + 2n)] \quad (3.4)$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Η παραπάνω προσέγγιση του P είναι ακριβής στα τρία δεκαδικά ψηφία, για $n \leq 10$ και στα δύο δεκαδικά ψηφία, για $n \leq 5$. Ακόμα, ο Mardia (1972a) έχει προτείνει μια λιγότερο ακριβή προσέγγιση του P , θεωρώντας ότι το στατιστικό $z = nr^2$ ακολουθεί τη χ^2 κατανομή με δύο βαθμούς ελευθερίας — η προσέγγιση αυτή είναι ακριβής στα δύο δεκαδικά ψηφία, για $n \leq 15$. Για μεγάλες τιμές του r , οι οποίες δεν καλύπτονται από τους πίνακες, μπορούμε να χρησιμοποιήσουμε το στατιστικό

$$z = nr^2, \quad (3.5)$$

το οποίο ακολουθεί την κανονική κατανομή. Εδώ βεβαίως η σημαντικότητα (δηλαδή, η μη τυχαιότητα) υπάρχει, όταν είναι $z > z(\alpha)$.

Οι Mardia and Jupp (2000, p. 95) αναφέρουν ότι, υπό τη μηδενική υπόθεση της ομοιομορφίας και για μεγάλα δείγματα, η ασυμπτωτική κατανομή της ποσότητας $2nr^2$ είναι

$$2nr^2 \sim \chi_2^2, \quad (3.6)$$

με σφάλμα της τάξης του $O(n^{-1/2})$. Ακόμα, η τροποποίηση S^* του παραπάνω στατιστικού, όπου

$$S^* = \left(1 - \frac{1}{2n}\right) 2nr^2 + \frac{nr^4}{2}$$

κατανέμεται σύμφωνα με τη χ_2^2 κατανομή, με σφάλμα της τάξης του $O(n^{-1})$ (Jupp, 1999).

Αν το κρίσιμο επίπεδο P , όπως προκύπτει από τους πίνακες, είναι μικρότερο από το προεπιλεγέν επίπεδο α , τότε η μηδενική υπόθεση της τυχαιότητας μπορεί να απορριφθεί. Στην περίπτωση αυτή, το μήκος του μέσου διανύσματος r θα διαφέρει σημαντικά από το 0. Όταν είναι $P > \alpha$, η ομοιομορφία της κατανομής του πληθυσμού δεν μπορεί να απορριφθεί.

Αν υπάρχουν επαρκείς λόγοι για να υποθέσουμε ότι η κατανομή του πληθυσμού είναι μονοκόρυφη, τότε το Rayleigh test αποδεικνύει όχι μόνο τον προσανατολισμό αλλά και τη συγκέντρωση των κατευθύνσεων γύρω από τη μέση κατεύθυνση. Την κατεύθυνση αυτή αποκαλούμε, επίσης, προτιμώμενη ή επικρατούσα.

Αν η εναλλακτική υπόθεση για την μηδενική υπόθεση είναι η κατανομή von Mises με συγκεκριμένο $\kappa > 0$, τότε το Rayleigh test είναι *συνεπές* και, μάλιστα, *ομοιόμορφα* πιο ισχυρό. Δεν είναι όμως *συνεπές* για εναλλακτικές υποθέσεις όπου είναι $\rho = 0$ και, ειδικότερα, κατανομές με διαμετρική συμ-

3.2. Έλεγχοι τυχαιότητας και καλής προσαρμογής

Πίνακας 3.2: Εφαρμογή του Rayleigh test στις μεταβλητές WINDIR_D7 και WINDIR_D24. Ενώ η εφαρμογή του Rayleigh test έδωσε για την πλειονότητα των μεταβλητών WINDIR_Dxx πάρα πολύ σημαντικά αποτελέσματα, υπήρξαν διαφοροποιήσεις, με κύριες εξαιρέσεις τις WINDIR_D7 και WINDIR_D24, οι οποίες έδωσαν υψηλά *P*-value (0.221 και 0.283, αντίστοιχα — αυτό σημαίνει ότι το Rayleigh test δεν απέριψε την ομοιομορφία της κατανομής τους). Ακόμα, από τα στοιχεία του πίνακα φαίνεται ότι, για την WINDIR_D24, οι έλεγχοι καλής προσαρμογής (της ομοιόμορφης κατανομής) Kuiper's test και Watson's U^2 test (παρουσιάζονται παρακάτω, υποενότ. 3.2.8, σελ. 59 και υποενότ. 3.2.9, σελ. 61) δεν απορρίπτονται, επίσης, τη μηδενική υπόθεση της τυχαιότητας, αλλά με μικρότερες τιμές *P*-value. Βλέπε, επίσης, σχήμα 3.5, σελ. 46.

<i>Variable</i>	<i>WINDIR_D7</i>	<i>WINDIR_D24</i>
Data Type	Angles	Angles
Number of Observations	24	22
Mean Vector (μ)	221.849°	91.033°
Length of Mean Vector (r)	0,252	0,241
Median	225°	78.75°
Concentration	0,52	0,496
Circular Variance	0,748	0,759
Circular Standard Deviation	95.195°	96.706°
Standard Error of Mean	32.346°	35.36°
95% Confidence Interval (-/+ for μ)	158.438°	21.713°
	285.26°	160.353°
99% Confidence Interval (-/+ for μ)	138.52°	359.939°
	305.178°	182.127°
Rayleigh Test (Z)	1,518	1,274
Rayleigh Test (p)	0,221	0,283
Rao's Spacing Test (U)	225	180
Rao's Spacing Test (p)	< 0.01	< 0.01
Watson's U^2 Test (Uniform, U^2)	0,211	0,173
Watson's U^2 Test (p)	< 0.05	0.1 > p > 0.05
Kuiper's Test (Uniform, V)	1,914	1,697
Kuiper's Test (p)	< 0.025	0.10 > p > 0.05
Watson's U^2 Test (von Mises, U^2)	0,136	0,122
Watson's U^2 Test (p)	< 0.005	< 0.005
Kuiper's Test (von Mises, V)	1,629	1,417
Kuiper's Test (p)	0.10 > p > 0.05	> 0.15

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.3: Εφαρμογή του Rayleigh test στη μεταβλητή WINDIR_DAY. Όπως φαίνεται από τα στοιχεία του πίνακα, η μέση ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ομοιόμορφα ούτε κατά τις πρώτες 17 ημέρες του μήνα (φυσιολογικές τιμές θερμοκρασίας) (μεταβλητή WINDIR_DAY [1-17]), ούτε κατά τις υπόλοιπες 14 ημέρες του μήνα (περίοδος του καύσωνα) (μεταβλητή WINDIR_DAY [18-31]), ούτε κατά τη μικτή περίοδο (σύνολο των ημερών του μήνα) (μεταβλητή WINDIR_DAY). Τα πολύ σημαντικά αποτελέσματα των υπολοίπων ελέγχων ομοιομορφίας επιβεβαιώνουν αυτό τον ισχυρισμό. Επίσης, οι παραπάνω μεταβλητές δεν κατανέμονται ούτε κανονικά (von Mises).

<i>Variable</i>	<i>WINDIR_</i> <i>DAY</i>	<i>WINDIR_</i> <i>DAY [1-17]</i>	<i>WINDIR_</i> <i>DAY [18-31]</i>
Data Type	Angles	Angles	Angles
Number of Observations	31	17	14
Mean Vector (μ)	194,672°	192,444°	197,051°
Length of Mean Vector (r)	0,49	0,462	0,525
Median	211,579°	211,897°	211,29°
Concentration	1,121	1,039	1,112
Circular Variance	0,51	0,538	0,475
Circular Standard Deviation	68,447°	71,238°	65,037°
Standard Error of Mean	13,887°	20,063°	20,042°
95% Confidence Interval (-/+) for μ	167,447°	153,112°	157,761°
	221,896°	231,776°	236,341°
99% Confidence Interval (-/+) for μ	158,896°	140,757°	145,419°
	230,448°	244,131°	248,683°
Rayleigh Test (Z)	7,44	3,623	3,86
Rayleigh Test (p)	4,03E-04	0,024	0,018
Rao's Spacing Test (U)	237,807	232,462	245,933
Rao's Spacing Test (p)	< 0.01	< 0.01	< 0.01
Watson's U ² Test (Uniform, U ²)	0,769	0,384	0,421
Watson's U ² Test (p)	< 0.005	< 0.005	< 0.005
Kuiper's Test (Uniform, V)	3,49	2,524	2,629
Kuiper's Test (p)	< 0.01	< 0.01	< 0.01
Watson's U ² Test (von Mises, U ²)	0,595	0,283	0,341
Watson's U ² Test (p)	< 0.005	< 0.005	< 0.005
Kuiper's Test (von Mises, V)	3,101	2,225	2,399
Kuiper's Test (p)	< 0.01	< 0.01	< 0.01

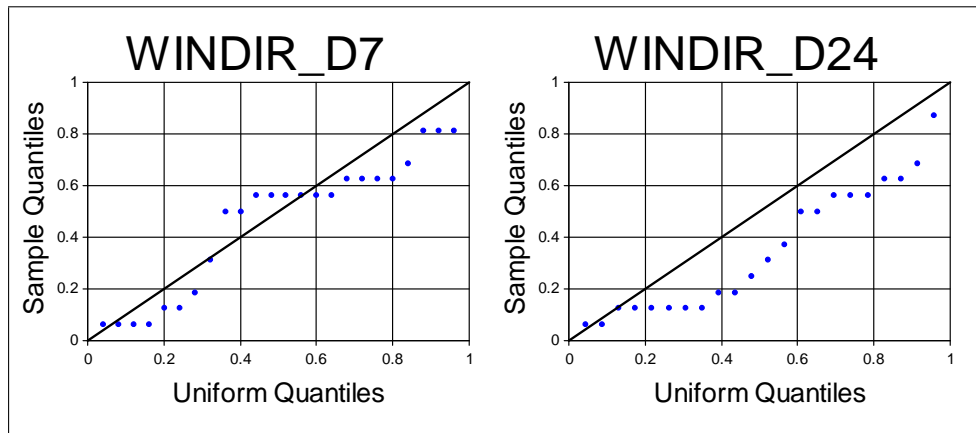
\underline{x} = αναξιόπιστο αποτέλεσμα λόγω χαμηλής συγκέντρωσης
(ομοιόμορφη κατανομή)

3.2. Έλεγχοι τυχαιότητας και καλής προσαρμογής

Πίνακας 3.4: Εφαρμογή του Rayleigh test στη μεταβλητή WINDIR_DPREV. Όπως φαίνεται από τα στοιχεία του πίνακα, η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ομοιόμορφα ούτε κατά τις πρώτες 17 ημέρες του μήνα (φυσιολογικές τιμές θερμοκρασίας) (μεταβλητή WINDIR_DPREV[1-17]), ούτε κατά τη μικτή περίοδο (σύνολο των ημερών του μήνα) (μεταβλητή WINDIR_DPREV). Όμως, κατά τις τελευταίες 14 ημέρες του μήνα (περίοδος του καύσωνα) (μεταβλητή WINDIR_DPREV[18-31]) φαίνεται να μην μπορεί να απορριφθεί η μηδενική υπόθεση της ομοιομορφίας ($P = 0.094$). Πάντως, οι υπόλοιποι έλεγχοι τυχαιότητας δίνουν πολύ σημαντικά αποτελέσματα, σε κάθε περίπτωση. Επίσης, οι παραπάνω μεταβλητές δεν κατανέμονται ούτε κανονικά (von Mises).

<i>Variable</i>	<i>WINDIR_ DPREV</i>	<i>WINDIR_ DPREV [1-17]</i>	<i>WINDIR_ DPREV [18-31]</i>
Data Type	Angles	Angles	Angles
Number of Observations	31	17	14
Mean Vector (μ)	209,378°	206,08°	214,047°
Length of Mean Vector (r)	0,446	0,477	0,41
Median	225°	225°	225°
Concentration	0,995	1,084	0,739
Circular Variance	0,554	0,523	0,59
Circular Standard Deviation	72,84°	69,686°	76,545°
Standard Error of Mean	15,456°	19,32°	27,838°
95% Confidence Interval (-/+) for μ	179,079°	168,205°	159,474°
	239,677°	243,954°	268,62°
99% Confidence Interval (-/+) for μ	169,561°	156,308°	142,331°
	249,195°	255,851°	285,762°
Rayleigh Test (Z)	6,158	3,873	2,35
Rayleigh Test (p)	0,002	0,018	0,094
Rao's Spacing Test (U)	267,097	211,765	263,571
Rao's Spacing Test (p)	< 0.01	< 0.01	< 0.01
Watson's U ² Test (Uniform, U ²)	0,737	0,4	0,371
Watson's U ² Test (p)	< 0.005	< 0.005	< 0.005
Kuiper's Test (Uniform, V)	3,36	2,535	2,334
Kuiper's Test (p)	< 0.01	< 0.01	< 0.01
Watson's U ² Test (von Mises, U ²)	0,562	0,28	0,304
Watson's U ² Test (p)	< 0.005	< 0.005	< 0.005
Kuiper's Test (von Mises, V)	3,162	2,296	2,263
Kuiper's Test (p)	< 0.01	< 0.01	< 0.01

\bar{x} = αναξιόπιστο αποτέλεσμα λόγω χαμηλής συγκέντρωσης
(ομοιόμορφη κατανομή)



Σχήμα 3.5: Διαγράμματα για τον έλεγχο της προσαρμογής της ομοιόμορφης κατανομής στα δεδομένα των μεταβλητών WINDIR_D7 και WINDIR_D24. Ενώ το Rayleigh test δεν απορρίπτει την ομοιομορφία τους, η προσαρμογή τους στα διαγράμματα, μάλλον, δεν κρίνεται ικανοποιητική. Μάλιστα, η WINDIR_D7, η οποία δείχνει να προσαρμόζεται λίγο καλύτερα από την WINDIR_D24, έδωσε πολύ σημαντικά αποτελέσματα ($P < 0.01$) στα Watson's U^2 test και Kuiper's test, με την WINDIR_D24 να δίνει λιγότερο σημαντικά αποτελέσματα ($0.1 > P > 0.05$). Αυτό ίσως είναι ενδεικτικό της μικρής ισχύος του Rayleigh test, σε σχέση με άλλους ελέγχους, όταν η εναλλακτική υπόθεση δεν είναι η κατανομή von Mises (όπως φαίνεται παρακάτω, οι μεταβλητές WINDIR_Dxx δεν ακολουθούν ούτε τη von Mises κατανομή).

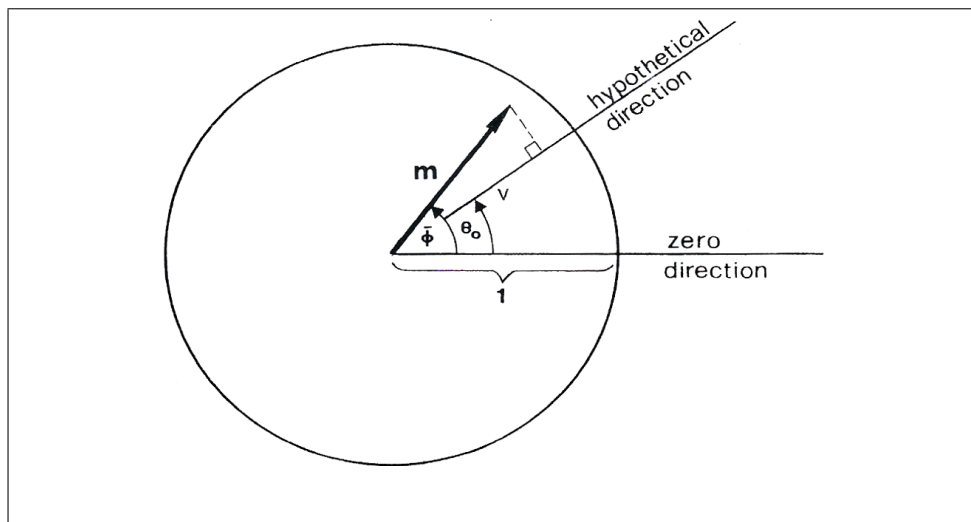
μετρία (antipodal symmetry). Έλεγχοι ομοιομορφίας που είναι συνεπείς έναντι κάθε εναλλακτικής είναι οι Kuiper's test και Watson's U^2 test (Mardia and Jupp, 2000, p. 98).

Για ομαδοποιημένες παρατηρήσεις χρησιμοποιείται το προσαρμοσμένο r , $r_c = cr$, όπου ο παράγοντας διόρθωσης c δίνεται από την εξ. (1.27). Για δεδομένα διεύθυνσης, γίνεται η σχετική προσαρμογή με το διπλασιασμό των τιμών.

Η στατιστική σημαντικότητα του μέσου μέτρου r αρχικά μελετήθηκε από τον λόρδο Rayleigh το 1880. Ο Pearson γενίκευσε το πρόβλημα το 1905 και την ίδια χρονιά ο Kluyver ανακάλυψε την ακριβή δειγματική κατανομή του r , υπό την μηδενική υπόθεση της τυχαιότητας.

3.2.2 V test

Σε αρκετές περιπτώσεις, πριν από την εκτέλεση ενός πειράματος, αναμένεται μια συγκεκριμένη μέση κατεύθυνση. Τότε, η υπόθεση που θέλουμε να ελέγξουμε είναι η τυχαιότητα έναντι του προσανατολισμού, όμως παρ' όλ' αυτά θα αποτελούσε απώλεια πληροφορίας, αν εγκαταλειπόταν η γνώση αυτής της αναμενόμενης κατεύθυνσης.



Σχήμα 3.6: Προετοιμασία του V test. Το μέσο διάνυσμα \mathbf{m} προβάλλεται στην υποθετική κατεύθυνση. Η προβολή αυτή v παίρνει τιμές από -1 έως +1. Όσο μεγαλύτερο είναι το v , τόσο περισσότερο τα δειγματικά σημεία συγκεντρώνονται προς την υποθετική κατεύθυνση.

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Με το συγκεκριμένο τεστ θέλουμε να ελέγξουμε το αν οι παρατηρούμενες γωνίες έχουν την τάση να ομαδοποιούνται γύρω από μια δεδομένη γωνία θ_0 και ως εκ τούτου η κατανομή διαφέρει σημαντικά από την τυχαία.

Έστω \mathbf{m} το μέσο διάνυσμα ενός δείγματος, r το μήκος του και $\bar{\phi}$ η πολική γωνία του. Αν V είναι η προβολή του \mathbf{m} πάνω στη δεδομένη κατεύθυνση θ_0 , τότε η ποσότητα

$$V = r \cos(\bar{\phi} - \theta_0)$$

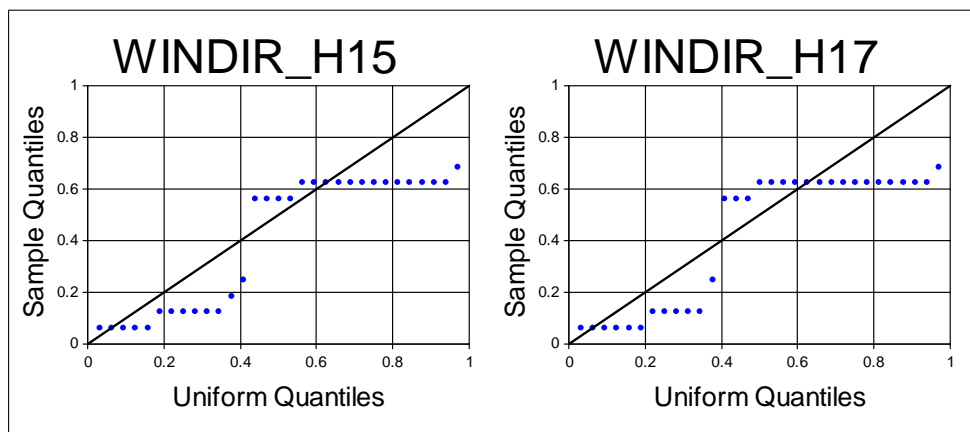
κυμαίνεται μεταξύ -1 και +1. Από τον παραπάνω τύπο προκύπτει ότι αν οι παρατηρούμενες γωνίες ϕ_i δεν διαφέρουν πολύ από τη θ_0 , τότε το V πλησιάζει τη μονάδα. Σε άλλη περίπτωση το V είναι σημαντικά μικρότερο του 1. Επομένως, το V ή κάποιος μετασχηματισμός του μπορεί να χρησιμοποιηθεί ως στατιστικό ελέγχου. Το στατιστικό που χρησιμοποιείται (και δίνει τη σημαντικότητα του V) είναι το

$$u = (2n)^{1/2}V \quad (3.7)$$

Η μηδενική υπόθεση της τυχειότητας απορρίπτεται, όταν $u \geq u(\alpha)$. Στην περίπτωση αυτή, το μήκος του μέσου διανύσματος r διαφέρει σημαντικά από το 0 και οι δεδομένες κατευθύνσεις συγκεντρώνονται γύρω από τη συγκεκριμένη

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

υποθετική κατεύθυνση. Αν είναι $u < u(\alpha)$, η ομοιομορφία των δεδομένων δεν μπορεί να απορριφθεί. Οι Batschelet (1981) και Zar (1999) δίνουν πίνακες των κρίσιμων τιμών $u(\alpha)$ για διάφορα επίπεδα σημαντικότητας α .



Σχήμα 3.7: Διαγράμματα για τον έλεγχο της προσαρμογής της ομοιόμορφης κατανομής στα δεδομένα των μεταβλητών WINDIR_H15 και WINDIR_H17. Ενώ το Rayleigh test και περισσότερο το V test υποδεικνύουν την ομοιομορφία τους, η προσαρμογή τους στα διαγράμματα, μάλλον, δεν κρίνεται ικανοποιητική. Γενικά, οι δύο αυτοί έλεγχοι δείχνουν να απορρίπτουν πιο δύσκολα τη μηδενική υπόθεση της τυχαιότητας. Αυτό ίσως οφείλεται στο ότι η εναλλακτική υπόθεση δεν αφορά στην κατανομή von Mises.

Για ομαδοποιημένες παρατηρήσεις χρησιμοποιείται το προσαρμοσμένο r , $r_c = cr$, όπου ο παράγοντας διόρθωσης c δίνεται από την εξ. (1.27). Για δεδομένα διεύθυνσης, γίνεται η σχετική προσαρμογή με το διπλασιασμό των τιμών.

Αν οι κατευθύνσεις τείνουν να ομαδοποιούνται γύρω από μια υποθετική κατεύθυνση, το V test είναι *ισχυρότερο* στο να απορρίψει την τυχαιότητα απ' ό,τι το Rayleigh test. Στην ειδική περίπτωση όπου η εναλλακτική υπόθεση είναι η κατανομή von Mises με $\kappa > 0$, το V test είναι *ομοιόμορφα* πιο ισχυρό (Batschelet, 1981, p. 60).

Το συγκεκριμένο τεστ πρέπει να χρησιμοποιείται με μηδενική υπόθεση *μόνο* την υπόθεση της τυχαιότητας. Έτσι, στην περίπτωση όπου τα δεδομένα είναι σαφώς προσανατολισμένα και είναι επιθυμητό να γίνει έλεγχος για το αν η μέση κατεύθυνση του δείγματος αποκλίνει σημαντικά από μια θεωρητική κατεύθυνση, το V test δεν μπορεί να εφαρμοστεί. Για την περίπτωση αυτή θα πρέπει να χρησιμοποιηθεί ένα διάστημα εμπιστοσύνης για τη μέση κατεύθυνση.

Πρέπει, επίσης, να επισημανθεί ότι αν το V test δεν απορρίψει τη μηδενική υπόθεση της ομοιομορφίας, δεν μπορούμε να συμπεράνουμε αν ο πληθυσμός κατανέμεται ομοιόμορφα ή αν έχει μια μέση κατεύθυνση διαφορετική από τη

3.2. Έλεγχοι τυχειότητας και καλής προσαρμογής

Πίνακας 3.5: Εφαρμογή των Rayleigh test και V test στις μεταβλητές WINDIR_H15 και WINDIR_H17 κατά το μήνα μελέτης. Η εφαρμογή των ελέγχων ομοιομορφίας έδωσε για το σύνολο, σχεδόν, των μεταβλητών WINDIR_Hxx πάρα πολύ σημαντικά αποτελέσματα. Ως εξαιρέσεις, επιλέξαμε τις WINDIR_H15 και WINDIR_H17, οι οποίες έδωσαν υψηλά *P*-value (0.396 και 0.164, αντίστοιχα) μόνο για το Rayleigh test, ενώ στους υπόλοιπους ελέγχους τυχειότητας έδωσαν σημαντικά έως πολύ σημαντικά αποτελέσματα. Το V test, εφαρμοζόμενο ως προς τη μέση γωνία (*expected mean*, 221,542° και 231,313°, αντίστοιχα) κάθε δείγματος, δεν απορρίπτει την τυχειότητα της WINDIR_H15, δίνοντας *P*-value=0,086, ενώ απορρίπτει την τυχειότητα της WINDIR_H17, δίνοντας *P*-value=0,028. Παρατηρούμε ότι, για την WINDIR_H17, όπου φαίνεται να υπάρχει ένας σαφέστερος προσανατολισμός των κατευθύνσεων (με μέσο μέτρο $r = 0,174$ και συγκέντρωση *Concentration* = 0,352), σε σχέση με την WINDIR_H15, το V test απορρίπτει την τυχειότητα, όταν το Rayleigh test δεν την απορρίπτει (με *P*-value=0,164). Αυτό ίσως είναι ενδεικτικό του ότι το V test είναι ισχυρότερο στο να απορρίψει την τυχειότητα απ' ό,τι το Rayleigh test, όταν οι κατευθύνσεις τείνουν να ομαδοποιούνται γύρω από μια συγκεκριμένη κατεύθυνση. Σημειώνουμε, πάντως, ότι η κατανομή των WINDIR_Hxx δεν είναι ούτε von Mises, ώστε να είναι ομοιόμορφα πιο ισχυρός ο συγκεκριμένος έλεγχος.

<i>Variable</i>	<i>WINDIR_H15</i>	<i>WINDIR_H17</i>
Data Type	Angles	Angles
Number of Observations	31	31
Mean Vector (μ)	221,542°	231,313°
Length of Mean Vector (r)	0,174	0,242
Median	225°	225°
Concentration	0,352	0,498
Circular Variance	0,826	0,758
Circular Standard Deviation	107,237°	96,568°
Standard Error of Mean	41,618°	29,664°
95% Confidence Interval (-/+ for μ)	139,955°	173,16°
	303,129°	289,466°
99% Confidence Interval (-/+ for μ)	114,327°	154,893°
	328,757°	307,733°
Rayleigh Test (Z)	0,933	1,81
Rayleigh Test (p)	0,396	0,164
Rao's Spacing Test (U)	278,71	290,323
Rao's Spacing Test (p)	< 0.01	< 0.01
Watson's U ² Test (Uniform, U ²)	0,464	0,556
Watson's U ² Test (p)	< 0.005	< 0.005
Kuiper's Test (Uniform, V)	2,802	2,988
Kuiper's Test (p)	< 0.01	< 0.01
V Test (V; <i>expected mean</i>)	0,174	0,242
V Test (u)	1,366	1,903
V Test (p)	0,086	0,028

\underline{x} = αναξιόπιστη τιμή, λόγω χαμηλής συγκέντρωσης (ομοιόμορφη κατανομή)

θ_0 (Zar, 1999, p. 620).

Το V test προτάθηκε αρχικά από τους Greenwood and Durand (1955) και αναπτύχθηκε περαιτέρω από τους Durand and Greenwood (1958).

3.2.3 Hodges-Ajne test

Με το συγκεκριμένο παραμετρικό τεστ θέλουμε να ελέγξουμε το αν ο πληθυσμός από τον οποίο προέρχεται ένα τυχαίο δείγμα κατανέμεται ακολουθώντας το τυχαίο πρότυπο. Μάλιστα, σε αντίθεση με το Rayleigh test, δεν απαιτείται να κάνουμε κάποια παραδοχή για την κατανομή του πληθυσμού από τον οποίο δειγματοληπτούμε. Γι' αυτό, ο συγκεκριμένος έλεγχος καλείται "omnibus", καθώς λειτουργεί ικανοποιητικά για μονοκόρυφες, δικόρυφες και γενικά πολυκόρυφες δειγματικές κατανομές. Πάντως, αν δεχόμαστε ότι ο πληθυσμός είναι μονοκόρυφος, τότε το Rayleigh test είναι ισχυρότερο.

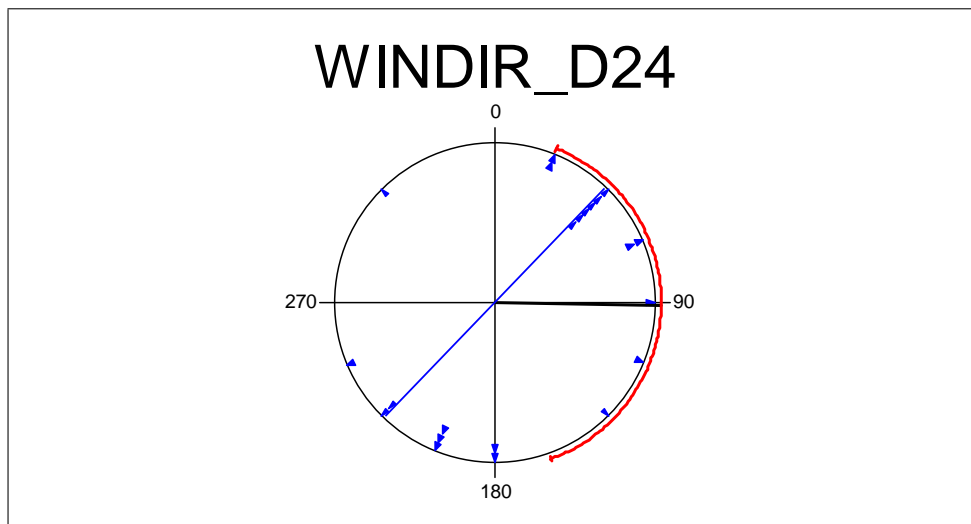
Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Σχεδιάζουμε το δείγμα σε έναν κύκλο και φέρουμε μια διάμετρο l . Κατόπιν περιτρέφουμε τη διάμετρο γύρω από το κέντρο του κύκλου μέχρι να καταφέρουμε να συγκεντρώσουμε το μεγαλύτερο δυνατό αριθμό παρατηρήσεων από τη μια πλευρά της διαμέτρου (δηλαδή, σε ένα εύρος 180°). Ένας μικρός αριθμός παρατηρήσεων από την άλλη πλευρά της διαμέτρου υποδεικνύει απόκλιση από το τυχαίο πρότυπο. Έτσι, το στατιστικό του ελέγχου είναι το

$K =$ ο ελάχιστος αριθμός παρατηρήσεων από τη μία πλευρά της περιστρεφόμενης διαμέτρου

Οι Batschelet (1981, p. 337) και Zar (1999, App.191-192) δίνουν πίνακες με τα κρίσιμα επίπεδα P -value, συναρτήσει του μεγέθους του δείγματος n και του επιπέδου σημαντικότητας α . Αν το κρίσιμο επίπεδο P -value είναι μικρότερο από το προεπιλεγέν επίπεδο σημαντικότητας α , η μηδενική υπόθεση απορρίπτεται υπέρ της εναλλακτικής υπόθεσης του προσανατολισμού.

Τα δεδομένα δεν πρέπει να είναι ομαδοποιημένα ή αν είναι ομαδοποιημένα, ο αριθμός των ομάδων να είναι μεγάλος σε σχέση με το μέγεθος του δείγματος, ενώ σε δεδομένα διεύθυνσης απαιτείται η σχετική προσαρμογή τους.

Αν υπάρχουν ενδείξεις ότι ο πληθυσμός κατανέμεται σύμφωνα με την κατανομή von Mises με $\kappa > 0$, τότε το Hodges-Ajne test δεν είναι τόσο ισχυρό, όσο το Rayleigh test (Batschelet, 1981, p. 63).



Σχήμα 3.8: Στο σχήμα φαίνεται η εφαρμογή του Hodges-Ajne test στις τιμές της μεταβλητής WINDIR_D24. Η περιστρεφόμενη διάμετρος αναπαρίσταται με μπλε χρώμα. Το κόκκινο τόξο εκφράζει το 95% διάστημα εμπιστοσύνης για τη μέση κατεύθυνση — ο κόκκινος χρωματισμός δηλώνει αναξιοπιστία της εκτίμησης, καθώς ο συνδυασμός του μεγέθους του δείγματος ($n = 29$) και της συγκέντρωσης ($Concentration = 0.496$) είναι χαμηλός. Ο έλεγχος δίνει $K = 6$ και για $n = 29$ από τους πίνακες δεν παίρνουμε σημαντικά αποτελέσματα ($P = 0.30$). Με εξαίρεση το Rao's spacing test ($P < 0.01$), κανένας άλλος έλεγχος δεν έδωσε σημαντικά αποτελέσματα για τα δεδομένα αυτά — βλ. τον παρακάτω πίνακα 3.8, σελ. 63.

Ο Zar (1999, p. 621) δίνει την πιθανότητα P για το K , υπό τη μηδενική υπόθεση της κυκλικής ομοιομορφίας, ώστε να μην είναι απαραίτητη η χρήση των πινάκων των κρίσιμων τιμών του K :

$$P = \binom{n}{K} 2^{1-n} (n - 2K) \quad (3.8)$$

καθώς και μια προσέγγιση του P , για $n > 50$

$$P \approx \frac{\sqrt{2\pi}}{A} \exp -\pi^2 / (8A^2)$$

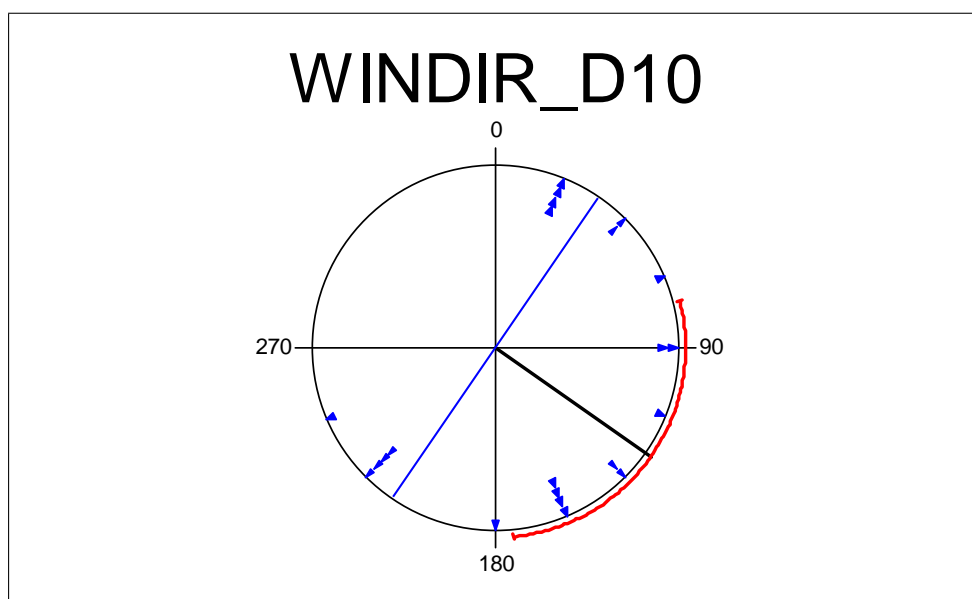
όπου

$$A = \frac{\pi\sqrt{n}}{2(n - 2K)}$$

Το κυκλικό αυτό τεστ δημοσιεύθηκε από τον Anje (1968). Σύντομα αποδείχθηκε από τον Rao (1969) και ανεξάρτητα από τους Bhattacharyya and Johnson (1969) ότι το τεστ του Anje αποτελεί ειδική περίπτωση του διμεταβλητού *sign test* του Hodges (1955).

3.2.4 modified Hodges-Ajne test

Με το συγκεκριμένο απαραμετρικό τεστ θέλουμε να ελέγξουμε το αν οι παρατηρούμενες γωνίες έχουν την τάση να ομαδοποιούνται γύρω από μια δεδομένη γωνία και, ως εκ τούτου, η κατανομή διαφέρει σημαντικά από την τυχαία. Το τεστ αυτό αποτελεί μια τροποποίηση του Hodges-Ajne test, ανάλογη με αυτή του V test ως προς το Rayleigh test.



Σχήμα 3.9: Εφαρμογή του modified Hodges-Ajne test στις τιμές της μεταβλητής WINDIR_D10. Η κάθετη στην υποτιθέμενη μέση κατεύθυνση (τη μέση γωνία των δεδομένων — έντονα μαύρη ακτίνα) διάμετρος αναπαρίσταται με μπλε χρώμα. Το κόκκινο τόξο εκφράζει το 95% διάστημα εμπιστοσύνης για τη μέση κατεύθυνση — και εδώ, ο κόκκινος χρωματισμός δηλώνει αναξιοπιστία της εκτίμησης, παρ' όλο που ο συνδυασμός του μεγέθους του δείγματος ($n = 29$) και της συγκέντρωσης ($Concentration = 0.706$) δεν είναι ιδιαίτερα χαμηλός. Για $k = 9$ και $n = 29$, οι πίνακες δεν δίνουν σημαντικά αποτελέσματα ($P = 0.61$). Απ' ό,τι φαίνεται και στον παρακάτω πίνακα 3.8, σελ. 63, ο συγκεκριμένος έλεγχος συμβαδίζει με την πλειονότητα των υπολοίπων ελέγχων ως προς τη μη σημαντικότητα των αποτελεσμάτων.

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου και μια υποθετική μέση κατεύθυνση θ_0 . Χρησιμοποιούμε πάλι μια διάμετρο, όμως αυτή τη φορά κάθετη προς τη δεδομένη κατεύθυνση θ_0 . Έστω

$k =$ ο αριθμός των παρατηρήσεων που συγκεντρώνονται στην άλλη πλευρά του κύκλου σε σχέση με την κατεύθυνση θ_0

Αν το k είναι επαρκώς μικρό σε σχέση με το μέγεθος του δείγματος, η μηδενική υπόθεση της τυχαιότητας μπορεί να απορριφθεί.

Ο Batschelet (1981, p. 338) δίνει πίνακα των κρίσιμων επιπέδων P , συναρτήσει του μεγέθους του δείγματος n και του k . Αν το κρίσιμο επίπεδο P -value είναι μικρότερο από το προεπιλεγέν επίπεδο σημαντικότητας α , απορρίπτουμε την μηδενική υπόθεση της τυχαίας κατανομής και συμπεραίνουμε ότι η κατανομή του πληθυσμού συγκεντρώνεται γύρω από την κατεύθυνση θ_0 . Αν είναι $P > \alpha$, δεν μπορούμε να απορρίψουμε την ομοιομορφία των δεδομένων.

Αν το δείγμα περιέχει δεδομένα διεύθυνσης, απαιτείται η σχετική προσαρμογή τους.

Το συγκεκριμένο τεστ προτάθηκε από τον Batschelet (1981).

3.2.5 Rao's spacing test

Με το συγκεκριμένο απαραμετρικό τεστ θέλουμε να ελέγξουμε το αν ο πληθυσμός από τον οποίο προέρχεται ένα τυχαίο δείγμα κατανέμεται ακολουθώντας το τυχαίο πρότυπο. Υπό την υπόθεση της τυχαιότητας, τα δειγματικά σημεία αναμένεται να είναι ομοιόμορφα διατεταγμένα (κατ' ίσα διαστήματα) στον κύκλο. Αν n είναι το μέγεθος του δείγματος, το τόξο μεταξύ δύο διαδοχικών παρατηρήσεων αναμένεται να ισούται με $2\pi/n$. Μικρές αποκλίσεις είναι ανεκτές λόγω των τυχαίων επιδράσεων, όμως μεγάλες αποκλίσεις από το $2\pi/n$ υποδεικνύουν μη τυχαιότητα.

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Υποθέτουμε ότι οι γωνίες είναι διατεταγμένες σε αύξουσα σειρά έτσι, ώστε να είναι

$$\phi_1 \leq \phi_2 \leq \dots \leq \phi_n$$

Κατόπιν υπολογίζουμε τα μήκη των τόξων μεταξύ των διαδοχικών δειγματικών σημείων

$$T_1 = \phi_2 - \phi_1, T_2 = \phi_3 - \phi_2, \dots, T_{n-1} = \phi_n - \phi_{n-1}, T_n = 2\pi + \phi_1 - \phi_n$$

Βεβαίως είναι

$$\sum_{i=1}^n T_i = 2\pi$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Η απόκλιση των T_i από την αναμενόμενη τιμή τους $E(T_i) = 2\pi/n$ είναι

$$|T_i - (2\pi/n)|, \quad i = 1, 2, \dots, n$$

Υπό την παραδοχή ότι τα δεδομένα δεν είναι ομαδοποιημένα, χρησιμοποιούμε το στατιστικό

$$U = \frac{1}{2} \sum_{i=1}^n |T_i - (2\pi/n)| \quad (3.9)$$

Οι Rao (1969, 1976), Batschelet (1981) και Mardia and Jupp (2000) δίνουν πίνακες κρίσιμων τιμών του U σε μοίρες. Αν η τιμή U του στατιστικού είναι μεγαλύτερη από την κρίσιμη τιμή $U(\alpha)$ των πινάκων, η μηδενική υπόθεση της ομοιομορφίας απορρίπτεται. Όταν είναι $U < U(\alpha)$, δεν μπορεί να απορριφθεί η ομοιομορφία των δεδομένων.

Πίνακας 3.6: Εφαρμογή του Rao's spacing test στις τιμές των μεταβλητών WINDIR_H15 και WINDIR_H17. Ο έλεγχος αυτός απορρίπτει τη μηδενική υπόθεση της τυχαιότητας των δεδομένων, δίνοντας πολύ σημαντικά αποτελέσματα. Οι έλεγχοι Watson's U^2 και Kuiper's συμφωνούν, δίνοντας, επίσης, πολύ σημαντικά αποτελέσματα. Για τις ίδιες μεταβλητές, το Rayleigh test είχε αποτύχει να απορρίψει την τυχαιότητα (βλ. πίνακα 3.5, σελ. 49).

<i>Variable</i>	<i>WINDIR_H15</i>	<i>WINDIR_H17</i>
Data Type	Angles	Angles
Number of Observations	31	31
Length of Mean Vector (r)	0,174	0,242
Concentration	0,352	0,498
Circular Variance	0,826	0,758
Circular Standard Deviation	107,237°	96,568°
Rayleigh Test (Z)	0,933	1,81
Rayleigh Test (p)	0,396	0,164
Rao's Spacing Test (U)	278,71	290,323
Rao's Spacing Test (p)	< 0.01	< 0.01
Watson's U^2 Test (Uniform, U^2)	0,464	0,556
Watson's U^2 Test (p)	< 0.005	< 0.005
Kuiper's Test (Uniform, V)	2,802	2,988
Kuiper's Test (p)	< 0.01	< 0.01

Η κατανομή ενός πληθυσμού δεν είναι πάντοτε μονοκόρυφη. Σε κάποιες περιπτώσεις, συμβαίνει να προβάλλουν δύο ή περισσότερες κύριες κατευθύν-

σεις. Ως λύση, η μέθοδος του διπλασιασμού των γωνιών μπορεί να εφαρμοστεί μόνο όταν τα δεδομένα είναι διεύθυνσης. Στις υπόλοιπες περιπτώσεις πολυκόρυφων κατανομών, ούτε το Rayleigh test, ούτε το Hodges-Ajne test είναι ισχυρά — μπορεί να αποτύχουν να απορρίψουν την τυχαιότητα, όταν στην πραγματικότητα υπάρχουν δύο ή περισσότερες κορυφές. Σ' αυτό το σημείο υπερέχει το Rao's spacing test: χειρίζεται το ίδιο καλά τα δείγματα τόσο από μονοκόρυφες, όσο και από πολυκόρυφες κατανομές. Αυτό φαίνεται χαρακτηριστικά στο παράδειγμα του πίνακα 3.6.

Άλλο ένα πλεονέκτημα του Rao's spacing test είναι η ευκολία υπολογισμού του στατιστικού. Το άθροισμα των θετικών αποκλίσεων ισούται με αυτό των αρνητικών αποκλίσεων και έτσι μπορούμε να πάρουμε την απλοποιημένη φόρμουλα

$$U = \sum_{i=1}^n [T_i - (2\pi/n)]$$

όπου το άθροισμα εκτείνεται μόνο στις θετικές αποκλίσεις.

Το συγκεκριμένο τεστ παρουσιάστηκε από τον Rao (1969, 1976).

3.2.6 range test

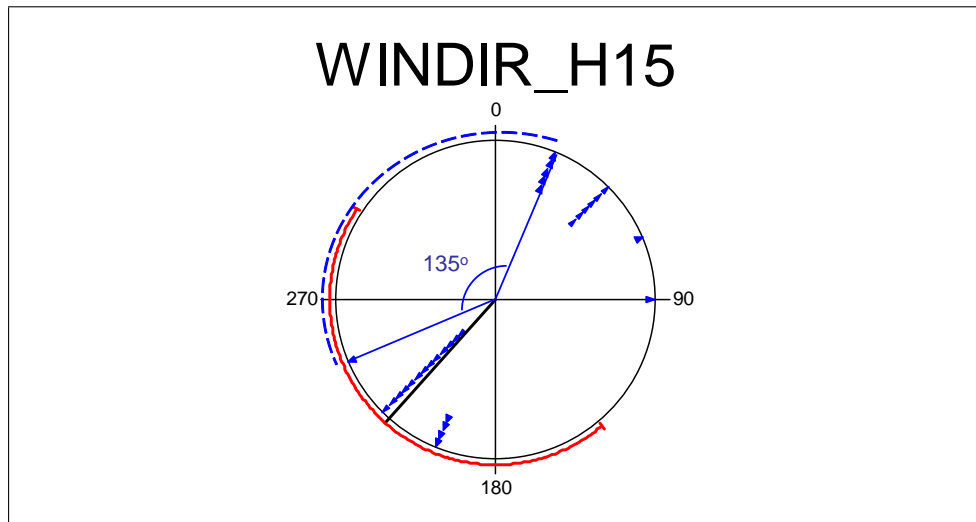
Με το συγκεκριμένο παραμετρικό τεστ θέλουμε να ελέγξουμε το αν ο πληθυσμός από τον οποίο προέρχεται ένα τυχαίο δείγμα κατανέμεται ακολουθώντας το τυχαίο πρότυπο.

Έστω τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου. Επιπρόσθετα, θεωρούμε ότι το δείγμα αυτό προέρχεται από πληθυσμό με μονοκόρυφη κατανομή. Έστω

w = το μήκος του μικρότερου τόξου του κύκλου που περιλαμβάνει όλα τα δειγματικά σημεία

Τότε το μήκος αυτό καλείται *εύρος* (range) του δείγματος. Μικρό εύρος υποδεικνύει ομαδοποίηση των δεδομένων και τότε μπορούμε να απορρίψουμε την ομοιομορφία.

Οι Laubsher and Rudolph (1968), Batschelet (1981, p. 340) και Mardia and Jupp (2000, p. 367) δίνουν πίνακες κρίσιμων τιμών (σε μοίρες) του εύρους $w(\alpha)$. Αν το δειγματικό εύρος w είναι μικρότερο από την τιμή των πινάκων για το προεπιλεγέν επίπεδο σημαντικότητας α , η μηδενική υπόθεση της ομοιομορφίας



Σχήμα 3.10: Εφαρμογή του range test στις τιμές της μεταβλητής WINDIR_H15. Η διακεκομμένη μπλε καμπύλη ορίζει το μέγιστο τόξο χωρίς παρατηρήσεις (135°). Η έντονα μαύρη ακτίνα αναπαριστά τη μέση γωνία των δεδομένων. Το κόκκινο τόξο εκφράζει το 95% διάστημα εμπιστοσύνης για τη μέση κατεύθυνση — εδώ, επίσης, ο κόκκινος χρωματισμός δηλώνει αναξιοπιστία της εκτίμησης, καθώς ο συνδυασμός του μεγέθους του δείγματος ($n = 31$) και της συγκέντρωσης ($Concentration = 0.352$) είναι χαμηλός. Σύμφωνα με τους πίνακες, για $n = 31$, ο έλεγχος δίνει πολύ σημαντικά αποτελέσματα, καθώς είναι $w = 135^\circ < w(0.005) = 266.7^\circ +$ (δίνονται τιμές του $w(\alpha)$ έως $n = 30$ παρατηρήσεις). Ακριβώς το ίδιο αποτέλεσμα παίρνουμε και για τη μεταβλητή WINDIR_H17. Το Rayleigh test δεν είχε απορρίψει την τυχαιότητα των μεταβλητών (βλ. πίνακα 3.5, σελ. 49).

απορρίπτεται υπέρ της εναλλακτικής υπόθεσης του προσανατολισμού.

Οι Mardia and Jupp (2000, p. 107) δίνουν τη συνάρτηση κατανομής του κυκλικού εύρους w , υπό τη συνθήκη της ομοιομορφίας

$$Pr(w \leq r) = \sum_{k=1}^{\infty} (-1)^{k-1} \binom{n}{k} \left[1 - k \left(1 - \frac{r}{2\pi} \right) \right]^{n-1} \quad (3.10)$$

όπου το άθροισμα εκτείνεται στις τιμές του k , για τις οποίες είναι

$$1 - k(1 - r/2\pi) > 0.$$

Το range test λειτουργεί ικανοποιητικά υπό τις παραδοχές ότι τα δεδομένα δεν είναι ομαδοποιημένα και ότι για δεδομένα διεύθυνσης γίνεται η σχετική προσαρμογή. Ακόμα, το συγκεκριμένο τεστ είναι ισχυρό μόνο αν η κατανομή του πληθυσμού έχει σχετικά μικρή γωνιακή απόκλιση, ειδικά αν η κατανομή αυτή είναι και μονοκόρυφη (Batschelet, 1981, p. 70).

Το range test προτάθηκε και μελετήθηκε ανεξάρτητα από τους Laubshier and Rudolph (1968) και Rao (1969).

ΕΛΕΓΧΟΙ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

3.2.7 chi-squared test

Έστω ένα τυχαίο δείγμα από n γωνίες $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αντιπροσωπεύουν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές κάποιου κυκλικού φαινομένου και μια κυκλική κατανομή (για παράδειγμα, η ομοιόμορφη κατανομή ή η von Mises κατανομή με $\kappa > 0$). Θέλουμε να εξετάσουμε το κατά πόσο η θεωρητική αυτή κατανομή ταιριάζει στα δεδομένα του δείγματος.

Η προετοιμασία για την εφαρμογή του chi-squared test περιλαμβάνει τη διαίρεση του κύκλου σε έναν επαρκή αριθμό τόξων, όχι απαραίτητα ισομήκων, για την κατάλληλη ομαδοποίηση των δεδομένων. Σε κάθε τέτοιο τόξο καταγράφουμε την (παρατηρούμενη) συχνότητα n_i των παρατηρήσεων και υπολογίζουμε την αναμενόμενη συχνότητα e_i διαιρώντας το πλήθος των παρατηρήσεων n με το λόγο του μήκους του τόξου κάθε κλάσης προς τον κύκλο. Η εφαρμογή της θεωρητικής κατανομής κρίνεται ικανοποιητική, όταν οι παρατηρούμενες συχνότητες δεν αποκλίνουν πολύ από τις αναμενόμενες, δηλαδή οι διαφορές $n_i - e_i$ είναι σχετικά μικρές. Τα δεδομένα ομαδοποιούνται έτσι, ώστε η αναμενόμενη συχνότητα σε κάθε ομάδα να ισούται τουλάχιστον με $e_i = 5$. Επίσης, η ελευθερία στην επιλογή των διαστημάτων πρέπει να χρησιμοποιείται με ορθό στατιστικό τρόπο.

Αν k είναι το πλήθος των ομάδων (κλάσεων), τότε το στατιστικό του ελέγχου είναι το

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i} \quad (3.11)$$

Από την παραπάνω συνάρτηση, φαίνεται ότι μεγάλες διαφορές μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων ($n_i - e_i$), δίνουν μεγάλες τιμές στο χ^2 . Γενικά, το χ^2 μπορεί να πάρει τιμές από 0 (τέλεια προσαρμογή) και πλέον — δεν μπορεί να πάρει αρνητικές τιμές.

Πίνακες με κρίσιμες τιμές της χ^2 συναντώνται στα περισσότερα βιβλία στατιστικής. Η μηδενική υπόθεση της καλής προσαρμογής της θεωρητικής κατανομής στα δεδομένα του δείγματος απορρίπτεται, όταν το κρίσιμο επίπεδο P -value, όπως προκύπτει από τους πίνακες της χ^2 κατανομής, είναι μικρότερο

Πίνακας 3.7: Έλεγχοι τυχασιότητας.

Έλεγχος	Είδος ελέγχου	Παραδοχές	Στατιστικό	Κανόνας απόρριψης	Παρατηρήσεις
Rayleigh test	Παραμετρικός	Μονοκόρυφες κατανομές. Για ομαδοποιημένες παρατηρήσεις, απαιτείται προσαρμογή. Για δεδομένα διεύθυνσης, εφαρμόζεται η τεχνική του διτλάστιστου	$r = \text{Μέσο μέτρο}$	$P < \alpha$	Για μεγάλα n χρησιμοποιούμε το στατ. $z = n \cdot r^2$. Απόρριψη της H_0 , όταν $z > z(\alpha)$
V test	Παραμετρικός	Μονοκόρυφες κατανομές. Για ομαδοποιημένες παρατηρήσεις, απαιτείται προσαρμογή. Για δεδομένα διεύθυνσης, εφαρμόζεται η τεχνική του διτλάστιστου	$V = r \cos(\bar{\phi} - \theta_0), u = (2n)^{1/2}V$	$u > u(\alpha)$	Αν υπάρχει μια τάση προς κάποιο θ_0 , είναι ισχυρότερο από το Rayleigh test. Αν η H_1 είναι η κατανομή von Mises με $\kappa > 0$, είναι ομοιόμορφα πιο ισχυρό. Για πλήθος von Mises με $\kappa > 0$, δεν είναι τόσο ισχυρό, όσο το Rayleigh test
Hodges-Ajne test	Μη παραμετρικός	Μη ομαδοποιημένα δεδομένα. Για δεδομένα διεύθυνσης, εφαρμόζεται η τεχνική του διτλάστιστου	$K = \text{ο ελάχιστος αριθμός παρατηρήσεων από τη μία πλευρά της περιστρεφόμενης διαμέτρου}$	$P < \alpha$	
modified Hodges-Ajne test	Μη παραμετρικός	Για δεδομένα διεύθυνσης, εφαρμόζεται η τεχνική του διτλάστιστου	$k = \text{ο αριθμός των παρατηρήσεων που συγκεντρώνονται στην άλλη πλευρά του κύκλου σε σχέση με την κατεύθυνση } \theta_0$	$P < \alpha$	
Rao's spacing test	Μη παραμετρικός	Μη ομαδοποιημένα δεδομένα. Για δεδομένα διεύθυνσης, εφαρμόζεται η τεχνική του διτλάστιστου	$U = \frac{1}{2} \sum_{i=1}^n T_i - (2\pi/n) , T_i \text{ διαφορές μεταξύ διαδοχικών παρατηρήσεων}$	$U > U(\alpha)$	Υπερέχει των Rayleigh test και Hodges-Ajne test διότι χειρίζεται καλύτερα τις πολυκόρυφες κατανομές
range test	Παραμετρικός	Μονοκόρυφες κατανομές. Μη ομαδοποιημένα δεδομένα	$w = \text{το μήκος του μικρότερου τόξου του κύκλου που περιλαμβάνει όλα τα δειγματικά σημεία}$	$P < \alpha$	Είναι ισχυρό μόνο αν η κατανομή του πλήθους του έχει σχετικά μικρή γωνιακή απόκλιση

3.2. Έλεγχοι τυχειότητας και καλής προσαρμογής

από το προεπιλεγέν επίπεδο σημαντικότητας α . Στην αντίθετη περίπτωση, η προσαρμογή κρίνεται ικανοποιητική. Οι βαθμοί ελευθερίας ισούνται με τον αριθμό των ομάδων μείον ένα. Στην περίπτωση όπου πρέπει να υπολογιστούν από το δείγμα κάποιες παράμετροι, οι βαθμοί ελευθερίας μειώνονται με τον αριθμό των παραμέτρων αυτών.

Το chi-squared test προτάθηκε από τον Pearson, περίπου το 1900 και αναπτύχθηκε από τον ίδιο και τον Fisher. Ο Rao (1972) προσέφερε μια βελτίωση για τα κυκλικά στατιστικά: καθώς το αποτέλεσμα του chi-squared test εξαρτάται σε μεγάλο βαθμό από την επιλογή του αρχικού σημείου και το μήκος των τόξων που ορίζουν τις ομάδες, ο Rao πρότεινε ένα μέσο *chi-squared* προκειμένου να εξαλείψει την επίδραση του αρχικού σημείου.

Το συγκεκριμένο τεστ δεν μπορεί να εφαρμοστεί στα δεδομένα της μελέτης αυτής, καθώς δεν πληρείται η συνθήκη $e_i \geq 5$: ακόμα και με πλάτος κλάσης 90° , καθώς και απόρριψη των κλάσεων χωρίς παρατηρήσεις ($n_i = 0$), πάνω από το 20% των ομάδων περιείχαν λιγότερες από 5 παρατηρήσεις.

3.2.8 Kuiper's test

Διάφορα τεστ γραμμικών κατανομών βασίζονται στη γενική απόκλιση της εμπειρικής συνάρτησης κατανομής από την υποθετική αθροιστική συνάρτηση κατανομής. Για την κατασκευή ανάλογων ελέγχων των κυκλικών κατανομών, θα πρέπει να καθοριστεί η εμπειρική συνάρτηση κατανομής για τα κυκλικά δεδομένα. Επιλέγοντας αρχική κατεύθυνση και φορά στον κύκλο, ταξινομούμε τις παρατηρήσεις κατ' αύξουσα σειρά και προσθέτουμε επιπλέον τις βοηθητικές παρατηρήσεις $\phi_{(0)}$ και $\phi_{(n+1)}$ έτσι, ώστε να είναι

$$\phi_{(0)} = 0 \leq \phi_{(1)} \leq \dots \leq \phi_{(n)} \leq \phi_{(n+1)} = 2\pi$$

Τότε, η εμπειρική συνάρτηση κατανομής S_n ορίζεται ως (Mardia and Jupp, 2000, p. 100):

$$S_n(\phi) = i/n, \quad \text{αν } \phi_{(i)} \leq \phi < \phi_{(i+1)}, \quad i = 0, 1, \dots, n.$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Επισημαίνεται ότι τα S_n και $F(\phi)$ εξαρτώνται από την επιλογή του αρχικού σημείου και της φοράς. Για την ομοιόμορφη κατανομή είναι:

$$F(\phi_i) = \frac{\phi_i}{2\pi}$$

Έστω ένα τυχαίο δείγμα από n γωνίες $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αντιπροσωπεύουν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές κάποιου κυκλικού φαινομένου και μια δεδομένη κυκλική κατανομή (για παράδειγμα, η ομοιόμορφη κατανομή ή η von Mises κατανομή με $\kappa > 0$). Θέλουμε να εξετάσουμε το κατά πόσο η θεωρητική αυτή κατανομή ταιριάζει στα δεδομένα του δείγματος. Σχεδιάζουμε σε ένα ορθογώνιο σύστημα αξόνων, με τον οριζόντιο άξονα διαιρεμένο σε 360° μοίρες (ανάπτυξη του κύκλου) και τον κάθετο άξονα ως διάστημα πραγματικών αριθμών $([0, 1])$, τη θεωρητική συνάρτηση κατανομής, η οποία είναι μια συνεχής καμπύλη, και την δειγματική (εμπειρική) συνάρτηση κατανομής, η οποία είναι μια βαθμωτή συνάρτηση. Το αρχικό σημείο του οριζόντιου άξονα είναι παρ' όλ' αυτά αυθαίρετο (μπορεί να είναι για παράδειγμα από -180° έως $(+)180^\circ$). Θέλουμε να εξετάσουμε το πόσο διαφέρουν μεταξύ τους τα δύο γραφήματα.

Έστω D^+ η μεγαλύτερη και D^- η μικρότερη κάθετη απόσταση της βαθμωτής συνάρτησης της δειγματικής κατανομής από τη συνεχή καμπύλη, αντίστοιχα. Ορίζουμε τότε ως

$$V_n = D^+ + D^- \quad (3.12)$$

το άθροισμα των δύο αποκλίσεων. Το V_n , σε αντίθεση με τα D^+ και D^- δεν εξαρτάται από την επιλογή του αρχικού σημείου ή της κατεύθυνσης της φοράς. Οι ποσότητες V_n, D^+ και D^- είναι όλες μεταξύ 0 και 1. Μικρή τιμή του V_n υποδηλώνει καλή προσαρμογή, ενώ μεγάλη σχετικά τιμή του V_n οδηγεί πιθανώς σε σημαντικά αποτελέσματα. Το στατιστικό του ελέγχου είναι το

$$K = n^{1/2} V_n \quad (3.13)$$

Ο Batschelet (1981, p. 341) δίνει πίνακα των κρίσιμων τιμών $K(\alpha)$ του στατιστικού K . Η μηδενική υπόθεση της καλής προσαρμογής απορρίπτεται, όταν η τιμή του στατιστικού υπερβαίνει την κρίσιμη τιμή των πινάκων. Στην περίπτωση αυτή, η θεωρητική κατανομή δεν εφαρμόζει στα δεδομένα του δείγματος. Στην αντίθετη περίπτωση, η προσαρμογή κρίνεται ικανοποιητική.

Αν πρόκειται για ομαδοποιημένα δεδομένα, θα πρέπει οι κλάσεις να μην

είναι μεγαλύτερες από 5°.

Ως κανόνας, το Kuiper's test είναι πιο ισχυρό από το chi-squared test τόσο για τις μονοκόρυφες, όσο και για τις δικόρυφες κατανομές. Είναι συνεπής έλεγχος έναντι κάθε εναλλακτικής ως προς την ομοιομορφία. Επίσης, όπως και κάθε άλλο τεστ καλής προσαρμογής υποδείγματος, το Kuiper's test μπορεί να χρησιμοποιηθεί ως τεστ ελέγχου της τυχειότητας (έλεγχος της προσαρμογής της ομοιόμορφης κατανομής στα δεδομένα) (Batschelet, 1981, p. 78).

Ο Kolmogorov συνέλαβε την ιδέα της χρησιμοποίησης της απόκλισης της εμπειρικής από τη θεωρητική κατανομή για τον έλεγχο της καλής προσαρμογής. Ο Kuiper (1960) προσαρμοσε το τεστ στις ανάγκες των κυκλικών κατανομών.

3.2.9 Watson's U^2 test

Αντί του υπολογισμού της γενικής διαφοράς μεταξύ της εμπειρικής συνάρτησης κατανομής και της αθροιστικής συνάρτησης κατανομής του υποδείγματος, μέσω μιας μεταβλητής μέγιστης απόκλισης, όπως στο παραπάνω Kuiper's test, μπορούμε να χρησιμοποιήσουμε μια (διορθωμένη) μέση τετραγωνική απόκλιση.

Έστω ένα τυχαίο δείγμα από n γωνίες $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αντιπροσωπεύουν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές κάποιου κυκλικού φαινομένου και μια κυκλική κατανομή (για παράδειγμα, η ομοιόμορφη κατανομή ή η von Mises κατανομή με $\kappa > 0$). Θέλουμε να εξετάσουμε το κατά πόσο η θεωρητική αυτή κατανομή ταιριάζει στα δεδομένα του δείγματος.

Εδώ, η βασική ιδέα είναι η χρησιμοποίηση μιας μέσης τετραγωνικής απόκλισης. Μεγάλη τιμή της απόκλισης αυτής οδηγεί στο συμπέρασμα ότι η θεωρητική κατανομή δεν είναι κατάλληλη για την περιγραφή του πληθυσμού. Για την προετοιμασία του τεστ, παίρνουμε τα δεδομένα όχι ομαδοποιημένα, αλλά διατεταγμένα σε αύξουσα σειρά:

$$\phi_1 \leq \phi_2 \leq \dots \leq \phi_n.$$

Αν είναι $u_i = F(\phi_i)$ οι τιμές της συνάρτησης κατανομής $F(\cdot)$ της θεωρητικής κατανομής για τις παρατηρήσεις του δείγματος ϕ_i , i οι δείκτες ταξινόμησης

και c_i οι αριθμοί $c_i = 2i-1$, τότε το στατιστικό του ελέγχου είναι το

$$U_n^2 = \sum_{i=1}^n u_i^2 - \sum_{i=1}^n \frac{c_i u_i}{n} + n[1/3 - (\bar{u} - 1/2)^2] \quad (3.14)$$

όπου είναι

$$\bar{u} = \frac{u_1 + u_2 + \dots + u_n}{n}$$

Ο Batschelet (1981, p. 342) δίνει πίνακα των κρίσιμων τιμών $U(\alpha)$ του στατιστικού U . Αν οι τιμή U του στατιστικού υπερέρχει της κρίσιμης τιμής $U(\alpha)$ του πίνακα, τότε η μηδενική υπόθεση της καλής προσαρμογής απορρίπτεται. Αλλιώς, η προσαρμογή κρίνεται ικανοποιητική.

Το Watson's U^2 test είναι κατάλληλο τόσο για μονοκόρυφες, όσο και για πολυκόρυφες κατανομές. Είναι μάλιστα ιδιαίτερα ισχυρό στις περιπτώσεις των μικρών δειγμάτων. Επίσης, μπορεί να χρησιμοποιηθεί ως τεστ ελέγχου της τυχαιότητας.

Η μέθοδος της χρησιμοποίησης μέσης τετραγωνικής απόκλισης στις γραμμικές μεταβλητές προτάθηκε από τους Cramér, von Mises, Smirnov. Ο Watson (1961) έκανε την προσαρμογή της μεθόδου στα κυκλικά δεδομένα.

3.2.10 Cox's test

Το Cox's test ελέγχει το αν η κατανομή του πληθυσμού από τον οποίο προέκυψε το τυχαίο δείγμα διαφέρει σημαντικά από τη θεωρητική κατανομή von Mises. Η απόκλιση από την κατανομή Von Mises μπορεί να παρουσιαστεί υπό τη μορφή της διπλής κορυφής, της λοξότητας ή της κυρτότητας.

Με δεδομένο ένα τυχαίο δείγμα n γωνιών $\phi_1, \phi_2, \dots, \phi_n$, οι οποίες αναπαριστούν κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου, προσαρμόζουμε την κατανομή von Mises με τη μέθοδο των εκτιμητών μέγιστης πιθανοφάνειας για τις παραμέτρους της μέσης κατεύθυνσης ($\hat{\theta}_1$) και συγκέντρωσης (\hat{k}). Η βασική ιδέα του τεστ είναι ο έλεγχος των δεύτερων τριγωνομετρικών ροπών

$$1/n \sum_{i=1}^n \cos z(\phi_i - \hat{\theta}_1)$$

και

$$1/n \sum_{i=1}^n \sin z(\phi_i - \hat{\theta}_1)$$

3.2. Έλεγχοι τυχειότητας και καλής προσαρμογής

Πίνακας 3.8: Πίνακας αποτελεσμάτων της ανάλυσης των μεταβλητών WINDIR_D10 και WINDIR_D24 κατά το μήνα μελέτης. Η εφαρμογή των Kuiper's test και Watson's U^2 test τόσο για την ομοιομορφία, όσο και για την κανονικότητα των κατανομών τους έδωσε για το σύνολο, σχεδόν, των μεταβλητών WINDIR_Dxx πολύ σημαντικά αποτελέσματα. Ως εξαιρέσεις, επιλέξαμε τις WINDIR_D10 και WINDIR_D24, οι οποίες έδωσαν υψηλότερα P -value. Πάντως, πλην των εξαιρέσεων αυτών, φαίνεται ότι, γενικά, η κατεύθυνση του ανέμου, κατά τη διάρκεια της κάθε ημέρας, δεν κατανέμεται ούτε τυχαία ούτε κανονικά.

<i>Variable</i>	<i>WINDIR_D10</i>	<i>WINDIR_D24</i>
Data Type	Angles	Angles
Number of Observations	22	22
Mean Vector (μ)	125.129°	91.033°
Length of Mean Vector (r)	0,333	0,241
Median	135°	78.75°
Concentration	0,706	0,496
Circular Variance	0,667	0,759
Circular Standard Deviation	85.009°	96.706°
Standard Error of Mean	25.215°	35.36°
95% Confidence Interval (-/+ for μ)	75.697°	21.713°
	174.561°	160.353°
99% Confidence Interval (-/+ for μ)	60.169°	359.939°
	190.089°	182.127°
Rayleigh Test (Z)	2,434	1,274
Rayleigh Test (p)	0,087	0,283
Rao's Spacing Test (U)	196,364	180
Rao's Spacing Test (p)	< 0.01	< 0.01
Watson's U^2 Test (Uniform, U^2)	0,181	0,173
Watson's U^2 Test (p)	0.1 > p > 0.05	0.1 > p > 0.05
Kuiper's Test (Uniform, V)	1,92	1,697
Kuiper's Test (p)	< 0.025	0.10 > p > 0.05
Watson's U^2 Test (von Mises, U^2)	0,062	0,122
Watson's U^2 Test (p)	0.15 > p > 0.1	< 0.005
Kuiper's Test (von Mises, V)	1,068	1,417
Kuiper's Test (p)	> 0.15	> 0.15

\underline{x} = αναξιόπιστη τιμή λόγω χαμηλής συγκέντρωσης
(ομοιόμορφη κατανομή)

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.9: Εφαρμογή των ελέγχων καλής προσαρμογής στις μεταβλητές δεύτερης τάξης WINDIR_HOUR και WINDIR_DAY. Σε πλήρη συμφωνία, όλοι οι έλεγχοι ομοιομορφίας απορρίπτουν την τυχαιότητά τους, δίνοντας πολύ σημαντικά αποτελέσματα. Οι έλεγχοι καλής προσαρμογής της κατανομής von Mises, Watson's U^2 test και Kuiper's test, απορρίπτουν την κανονικότητα της WINDIR_DAY, δίνοντας πολύ σημαντικά αποτελέσματα, αλλά δεν απορρίπτουν την κανονικότητα της WINDIR_HOUR, δίνοντας, μάλιστα, υψηλά P -value. Φαίνεται, λοιπόν, ότι η μέση ημερήσια κατεύθυνση του ανέμου (WINDIR_DAY) δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises, ενώ η μέση ωριαία κατεύθυνση του ανέμου (WINDIR_HOUR) κατανέμεται κανονικά (von Mises). Δηλαδή, φαίνεται ότι υπάρχει μια κανονικότητα στις τιμές της κατεύθυνσης του ανέμου, κατά το μέσο όρο κάθε ώρας του εικοσιτετραώρου.

<i>Variable</i>	<i>WINDIR_HOUR</i>	<i>WINDIR_DAY</i>
Data Type	Angles	Angles
Number of Observations	24	31
Mean Vector (μ)	207,643°	194,672°
Length of Mean Vector (r)	0,961	0,49
Median	207,286°	211,579°
Concentration	13,236	1,121
Circular Variance	0,039	0,51
Circular Standard Deviation	16,066°	68,447°
Standard Error of Mean	3,279°	13,887°
95% Confidence Interval (-/+ for μ	201,215°	167,447°
	214,07°	221,896°
99% Confidence Interval (-/+ for μ	199,196°	158,896°
	216,089°	230,448°
Rayleigh Test (Z)	22,185	7,44
Rayleigh Test (p)	1,13E-09	4,03E-04
V Test (V ; expected mean)	0,961	0,49
V Test (u)	6,661	3,857
V Test (p)	7,57E-11	3,70E-05
Rao's Spacing Test (U)	283,32	237,807
Rao's Spacing Test (p)	< 0.01	< 0.01
Watson's U^2 Test (Uniform, U^2)	1,489	0,769
Watson's U^2 Test (p)	< 0.005	< 0.005
Kuiper's Test (Uniform, V)	4,229	3,49
Kuiper's Test (p)	< 0.01	< 0.01
Watson's U^2 Test (von Mises, U^2)	0,047	0,595
Watson's U^2 Test (p)	> 0.5	< 0.005
Kuiper's Test (von Mises, V)	1,073	3,101
Kuiper's Test (p)	> 0.15	< 0.01

Υπό την υπόθεση ενός von Mises πληθυσμού, για τα δεδομένα $\hat{\theta}_1$ και \hat{k} , οι παραπάνω ποσότητες είναι προσεγγιστικά ανεξάρτητες και κανονικά κατανοημένες για μεγάλο n , με μέσους $I_2(\hat{k})/I_0(\hat{k})$ και 0, αντίστοιχα. Εδώ, τα I_m αποτελούν τροποποιημένες *Bessel* συναρτήσεις, για $m = 0, 2$. Απορρίπτουμε την υπόθεση της von Mises κατανομής, όταν οι παραπάνω ποσότητες είναι πολύ μακριά από τις μέσες τιμές τους.

3.2.11 Σύνοψη

Στην ανάλυση των κατευθύνσεων ή των ρυθμών με γνωστή περίοδο, τα συχνότερα απαιτούμενα τεστ είναι οι έλεγχοι τυχειότητας με χρήση ενός δείγματος. Τα πιο απλά στην εφαρμογή τους τεστ είναι το Hodges-Ajne test και το Range test, καθώς και το Rao's spacing test. Για το Rayleigh test, το chi-squared test και το Watson's U^2 test απαιτείται υπολογιστική μηχανή. Εξαιρέση αποτελεί το Kuiper's test το οποίο είναι γραφικό τεστ.

Οι ερευνητές φαίνεται να προτιμούν το Rayleigh test καθώς έχει το πλεονέκτημα να χρησιμοποιεί μέτρα συγκέντρωσης, διασποράς και μέσης κατεύθυνσης. Αν δοθεί μια υποθετική κατεύθυνση για έλεγχο, τότε βελτιώνεται η ισχύς των ελέγχων με τη χρησιμοποίηση του V test ή του Modified Hodges-Ajne test. Στην περίπτωση των πολυκόρυφων κατανομών καλούμαστε να διακρίνουμε μεταξύ δύο περιπτώσεων: είτε οι πολλαπλές κορυφές μπορούν να αποκλειστούν (αν πρόκειται για δεδομένα διεύθυνσης, με το διπλασιασμό των γωνιών), είτε δεν μπορούν να αποκλειστούν. Στην πρώτη υποπερίπτωση όλα τα τεστ ελέγχου της τυχειότητας είναι εφαρμοστέα. Στη δεύτερη υποπερίπτωση συνιστώνται το Rao's spacing test, το chi-squared test, το Kuiper's test και το Watson's U^2 test.

Υπάρχουν περιπτώσεις κατά τις οποίες τίθενται γενικότερα προβλήματα καθώς η μηδενική υπόθεση δεν αναφέρεται στην ομοιόμορφη κατανομή, αλλά σε κάποια άλλη κατανομή. Για τις περιπτώσεις αυτές συνιστώνται το Chi-squared test, το Kuiper's test και το Watson's U^2 test. Ένα συγκεκριμένο τεστ για την καλή προσαρμογή της von Mises κατανομής αποτελεί το Cox's test.

Πίνακας 3.10: Έλεγχοι κατά της προσομοίωσης.

Έλεγχος	Είδος ελέγχου	Παραδοχές	Στατιστικό	Κριτήριο απόρριψης	Προσομοίωση	Παρατηρήσεις
<i>chi-squared test</i>	Μη παραμετρικός	Οι αναμενόμενες συχνότητες σε κάθε ομάδα ≥ 4	$\chi^2 = \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i}$	$\chi^2 > \chi_{\alpha, k-1}^2$	Ομοτιμότητα/ von Mises	
<i>Kuiper's test</i>	Μη παραμετρικός	Μη ομοδοποιημένα δεδομένα ή, τουλάχιστον, μέγεθος κλάσης $\leq 5^\circ$	$K = n^{1/2} V_n, V_n = D^+ + D^-, D^+ / D^-$ η μεγαλύτερη/μικρότερη κάθετη απόσταση της δειγματικής κατανομής από τη συνεχή θεωρητική καμπύλη.	$P < \alpha$	Ομοτιμότητα/ von Mises	Πιο ισχυρό από το χ^2 . Κατάλληλο για μονοκόρυφες και πολυκόρυφες κατανομές.
<i>Watson's U^2 test</i>	Μη παραμετρικός	Μη ομοδοποιημένα δεδομένα. Διάταξη σε αύξουσα σειρά.	$U_n^2 = \sum_{i=1}^n \frac{e_i u_i}{n} + n [1/3 - (\bar{u} - 1/2)^2], u_i = F(\phi_i)$ οι τιμές της θεωρητικής σ.κ. $F(\cdot)$	$P < \alpha$	Ομοτιμότητα/ von Mises	Πολύ ισχυρό για μικρά δείγματα. Κατάλληλο για μονοκόρυφες και πολυκόρυφες κατανομές.
<i>Cox's test</i>	Παραμετρικός	-	$1/n \sum_{i=1}^n \cos z(\phi_i - \hat{\theta}_1)$ $N(I_2(k)/I_0(k), \sigma^2),$ $1/n \sum_{i=1}^n \sin z(\phi_i - \hat{\theta}_1)$ $N(0, \sigma^2)$	\sim \sim	von Mises	

3.3 Έλεγχοι δύο δειγμάτων και πολλαπλών δειγμάτων

Στις πρακτικές στατιστικές εφαρμογές, οι έλεγχοι δύο δειγμάτων αποτελούν ίσως τα πιο σημαντικά εργαλεία. Παρακάτω παρατίθενται διάφοροι τέτοιου τύπου έλεγχοι, παραμετρικοί και απαραμετρικοί.

Οι βασικές παραδοχές για όλα αυτά τα τεστ είναι οι εξής:

- Κάθε δείγμα πρέπει να είναι *τυχαίο*, δηλαδή, κάθε απλό στοιχείο (παρατήρηση) του δείγματος πρέπει να προέρχεται *ανεξάρτητα* από τον ίδιο πληθυσμό.
- Τα δύο ή περισσότερα δείγματα που συγκρίνονται πρέπει να είναι *μεταξύ τους ανεξάρτητα*.

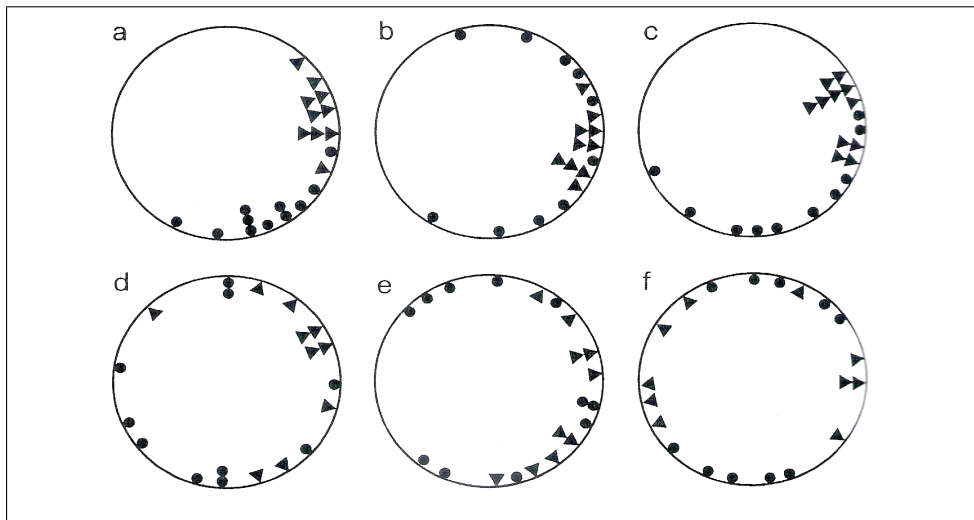
Όλοι οι παρακάτω έλεγχοι, εκτός από το chi-squared test, απαιτούν οι κατανομές των πληθυσμών να είναι συνεχείς. Πάντως, σε αρκετές περιπτώσεις οι φυσικοί και τεχνικοί περιορισμοί στην ακρίβεια των μετρήσεων επιβάλλουν την ομαδοποίηση των δεδομένων. Όσο το μέγεθος της κλάσης ομαδοποίησης είναι σχετικά μικρό (μικρότερο από 5°), τα παρακάτω τεστ δεν χάνουν την εγκυρότητά τους. Παρ' όλ' αυτά, οι ισοβαθμίες που προκύπτουν σε τέτοιες περιπτώσεις αποτελούν τροχοπέδη για κάποια από αυτά τα τεστ.

Οι περισσότεροι από τους παρακάτω ελέγχους έχουν σχεδιαστεί έτσι, ώστε να εντοπίζουν μια απόκλιση στη μέση κατεύθυνση. Πάντως, η ύπαρξη σημαντικών αποτελεσμάτων δεν συνεπάγεται το ότι οι μέσες κατευθύνσεις διαφέρουν μεταξύ τους. Κάποιες φορές, ο λόγος ύπαρξης σημαντικών αποτελεσμάτων σε έναν τέτοιο έλεγχο μπορεί να είναι κάτι διαφορετικό, όπως για παράδειγμα διαφορετικές διασπορές ή η ύπαρξη πολλών κορυφών. Κατά συνέπεια, αν το αποτέλεσμα κάποιου από αυτούς τους ελέγχους υποδείξει την απόρριψη της μηδενικής υπόθεσης, στον ερευνητή παραμένουν ακόμα αμφιβολίες σχετικά με το αίτιο. Τότε, μόνο με προσεκτική διερεύνηση των δειγμάτων μπορεί να δοθεί επαρκής ερμηνεία (Batschelet, 1981, p. 93).

Εν περιλήψει, ως πρώτο βήμα οφείλουμε να εξασφαλίσουμε τη σημαντικότητα των αποτελεσμάτων και, κατόπιν, ως δεύτερο βήμα, να διερμηνεύσουμε προσεκτικά το αίτιο της σημαντικότητας.

3.3.1 Watson-Williams test

Το Watson-Williams test είναι ένας παραμετρικός έλεγχος που χρησιμοποιείται όταν, δεδομένων δύο ανεξάρτητων τυχαίων δειγμάτων, θέλουμε να



Σχήμα 3.11: Δύο κυκλικά δείγματα μπορεί να διαφέρουν: (a) ως προς τις μέσες γωνίες, (b) ως προς την κυκλική διακύμανση, (c) ως προς τις μέσες γωνίες και την κυκλική διακύμανση, (d, e, f) κατά τρόπο που δεν μπορεί εύκολα να προσδιοριστεί.

ελέγξουμε το αν οι μέσες κατευθύνσεις των πληθυσμών από τους οποίους προέρχονται τα δείγματα αυτά είναι ίσες, δηλαδή αν τα δείγματα προέρχονται από τον ίδιο πληθυσμό. Μπορεί να εφαρμοστεί μόνο σε περιπτώσεις όπου (Batschelet, 1981, p. 96):

1. Τα δύο δείγματα επιλέγονται από πληθυσμούς με κατανομή von Mises
2. Η παράμετρος συγκέντρωσης είναι η ίδια και στους δύο πληθυσμούς, δηλαδή είναι:

$$\kappa_1 = \kappa_2 = \kappa$$

3. Η παράμετρος κ είναι επαρκώς μεγάλη, συγκεκριμένα είναι:

$$\kappa > 2$$

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n_1 και n_2 :

$$\phi_1, \phi_2, \dots, \phi_{n_1} \quad \psi_1, \psi_2, \dots, \psi_{n_2}$$

τα οποία προέρχονται από πληθυσμούς που ακολουθούν την κατανομή von Mises, με ίσες διασπορές. Αν θ_1 και θ_2 είναι οι (άγνωστες) μέσες κατευθύνσεις

των πληθυσμών, θέλουμε να ελέγξουμε τη μηδενική υπόθεση

$$H_0 : \theta_1 = \theta_2$$

Για κάθε ένα δείγμα υπολογίζουμε τις συνιστώσες των συνισταμένων διανυσμάτων \vec{R}_1 και \vec{R}_2 :

$$\begin{aligned} C_1 &= \sum^i \cos \phi_i, & S_1 &= \sum^i \sin \phi_i \\ C_2 &= \sum^i \cos \psi_i, & S_2 &= \sum^i \sin \psi_i \end{aligned} \quad (3.15)$$

Τότε, τα συνισταμένα διανύσματα έχουν τα ακόλουθα μήκη:

$$R_1 = \sqrt{C_1^2 + S_1^2}, \quad R_2 = \sqrt{C_2^2 + S_2^2} \quad (3.16)$$

Οι κατευθύνσεις των συνισταμένων διανυσμάτων δίνονται από τις μέσες γωνίες των δύο δειγμάτων, $\bar{\phi}$ και $\bar{\psi}$, αντίστοιχα.

Θεωρώντας τα δύο δείγματα ως ένα, παίρνουμε ένα νέο, από κοινού δείγμα (combined sample) μεγέθους

$$n = n_1 + n_2$$

Για το νέο αυτό δείγμα, οι συνιστώσες της συνισταμένης του \vec{R} είναι οι:

$$C = C_1 + C_2, \quad S = S_1 + S_2$$

και, συνεπώς, το μήκος της είναι:

$$R = \sqrt{C^2 + S^2}.$$

Στην ειδική περίπτωση, όπου είναι

$$\bar{\phi} = \bar{\psi}$$

και οι τρεις συνισταμένες έχουν την ίδια κατεύθυνση και είναι

$$R_1 + R_2 = R$$

Πάντως, αν είναι

$$\bar{\phi} \neq \bar{\psi},$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

τότε παίρνουμε

$$R_1 + R_2 > R,$$

δηλαδή, η ποσότητα

$$R_1 + R_2 - R,$$

είναι θετική. Όσο περισσότερο αποκλίνουν μεταξύ τους τα $\bar{\phi}$ και $\bar{\psi}$, τόσο αυξάνεται το $R_1 + R_2 - R$. Επομένως, το τεστ για τη διαφορά $\bar{\phi} - \bar{\psi}$ βασίζεται στο $R_1 + R_2 - R$.

Για τον έλεγχο της ισότητας των μέσων κατευθύνσεων των πληθυσμών χρησιμοποιούμε το στατιστικό

$$F = g \cdot (n - 2) \frac{R_1 + R_2 - R}{n - (R_1 + R_2)} \quad (3.17)$$

Στην παραπάνω εξίσωση ο παράγοντας g διορθώνει τη μεροληψία του στατιστικού για την F -κατανομή και υπολογίζεται από την εξίσωση

$$g = 1 + \frac{3}{8\tilde{\kappa}} \quad (3.18)$$

όπου η παράμετρος $\tilde{\kappa}$ υπολογίζεται από τους πίνακες τιμών της παραμέτρου συγκέντρωσης κ της κατανομής von Mises, για τις τιμές των

$$\bar{r} = (R_1 + R_2)/n,$$

ως τη μέση τιμή των μηκών των μέσων διανυσμάτων και

$$\bar{n} = n/2 = (n_1 + n_2)/2,$$

ως το μέσο μέγεθος δείγματος. Η τρίτη ειδική παραδοχή, πρέπει να ισχύει και για το $\tilde{\kappa}$, δηλαδή να είναι $\tilde{\kappa} > 2$.

Όταν το στατιστικό (3.17) παίρνει τιμές μεγαλύτερες από τις κρίσιμες τιμές των πινάκων, δηλαδή $F > F_{1,n-2}$, απορρίπτουμε τη μηδενική υπόθεση της ισότητας των μέσων κατευθύνσεων των πληθυσμών και υιοθετούμε το συμπέρασμα ότι οι μέσες κατευθύνσεις των πληθυσμών διαφέρουν σημαντικά μεταξύ τους.

Πρέπει, ωστόσο, να σημειωθεί ότι το στατιστικό (3.17), υπό την H_0 , ακολουθεί μόνο κατά προσέγγιση τη γνωστή F -κατανομή. Επίσης, λίγα είναι γνωστά για το σφάλμα προσέγγισης (Batschelet, 1981, p. 97).

Πίνακας 3.11: Έλεγχος των συνθηκών τής σελ. 68 για την εφαρμογή του Watson-Williams test στις μεταβλητές δεύτερης τάξης WINDIR_DAY[1-17] και WINDIR_DAY[18-31]. Τα πολύ σημαντικά αποτελέσματα των ελέγχων καλής προσαρμογής της κατανομής von Mises, υποδεικνύουν ότι οι μεταβλητές WINDIR_DAY[1-17] και WINDIR_DAY[18-31] αποκλίνουν από την κανονικότητα, δηλαδή, οι μέσες ημερήσιες κατεύθυνσεις του ανέμου δεν κατανομούνται σύμφωνα με την κατανομή von Mises, ούτε με φυσιολογικές τιμές θερμοκρασίας ούτε με καύσωνα. Επίσης, οι τιμές της παραμέτρου συγκέντρωσης κ (Concentration) είναι μικρότερες από 2. Πάντως, ο έλεγχος της υποενοότητας 3.1.2, σελ. 38, για την ισότητα των παραμέτρων συγκέντρωσης έδωσε τιμή του στατιστικού $F = 1.11 < F_{0.05(2),16,13} = 3.03$, δηλαδή η μηδενική υπόθεση της ισότητας των παραμέτρων συγκέντρωσης ($\kappa_1 = \kappa_2$) δεν απορρίπτεται. Συμπερασματικά, μπορούμε να πούμε ότι οι συγκεκριμένες μεταβλητές δεν πληρούν τις συνθήκες για την εφαρμογή του ελέγχου Watson-Williams.

Variable	WINDIR_DAY	WINDIR_DAY- [1-17]	WINDIR_DAY- [18-31]
Data Type	Angles	Angles	Angles
Number of Observations	31	17	14
Mean Vector (μ)	194.672°	192.444°	197.051°
Length of Mean Vector (r)	0,49	0,462	0,525
Median	211.579°	211.897°	211.29°
Concentration	1,121	1,039	1,112
Circular Variance	0,51	0,538	0,475
Circular Standard Deviation	68.447°	71.238°	65.037°
Standard Error of Mean	13.887°	20.063°	<u>20.042°</u>
Watson's U ² Test (von Mises, U ²)	0,595	0,283	0,341
Watson's U ² Test (p)	< 0.005	< 0.005	< 0.005
Kuiper's Test (von Mises, V)	3,101	2,225	2,399
Kuiper's Test (p)	< 0.01	< 0.01	< 0.01

\underline{x} = αναξιόπιστη τιμή λόγω χαμηλής συγκέντρωσης
(ομοιόμορφη κατανομή)

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.12: Εφαρμογή του Watson-Williams test για τις μεταβλητές WINDIR_DAY[1-17] και WINDIR_DAY[18-31]. Όπως προκύπτει από τα στοιχεία του πίνακα 3.11, δεν πληρούνται όλες οι συνθήκες για την εφαρμογή αυτού του παραμετρικού ελέγχου, οπότε το αποτέλεσμα του ελέγχου είναι αναξιόπιστο.

<i>Variables (& observations)</i>	<i>F</i>	<i>p</i>	<i>df</i>	<i>df2</i>	<i>Est. Mean</i>
WINDIR_DAY[1-17] WINDIR_DAY[18-31] (17 & 14)	& 0,03	0,864	1	29	194,672

Μεταβάλλεται η μέση ημερήσια κατεύθυνση του ανέμου μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας και των ημερών του καύσωνα; (Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών δεύτερης τάξης WINDIR_DAY[1-17] και WINDIR_DAY[18-31] με το Watson-Williams test.)

Από τα στοιχεία του πίνακα (3.12), οι μεταβλητές WINDIR_DAY[1-17] και WINDIR_DAY[18-31] φαίνεται να πληρούν το κριτήριο της ισότητας των παραμέτρων συγκέντρωσης, όμως δεν πληρούν τα άλλα δύο κριτήρια της σελ. 68. Συγκεκριμένα, για τις παραμέτρους συγκέντρωσης είναι $Concentration < 2$ και, ακόμα, οι έλεγχοι Watson's U^2 test και Kuiper's test απορρίπτουν την κανονικότητα των δύο μεταβλητών δίνοντας πολύ σημαντικά αποτελέσματα. Έτσι, η εφαρμογή του παραμετρικού ελέγχου Watson-Williams test στην περίπτωση αυτή δίνει αποτελέσματα αμφίβολης αξιοπιστίας.

Η εφαρμογή του συγκεκριμένου ελέγχου έδωσε υψηλή τιμή του P -value ($P = 0.864$), κάτι το οποίο θα μπορούσε να υποδεικνύει ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν μεταβάλλεται κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας και τις ημέρες με καύσωνα. Το αποτέλεσμα αυτό, όμως, δεν είναι αξιόπιστο και ο έλεγχος για την ισονομία των πληθυσμών, δηλαδή το αν η μέση ημερήσια κατεύθυνση μεταβάλλεται ή όχι μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας και των ημερών του καύσωνα πρέπει να γίνει με κάποιο κατάλληλο παραμετρικό τεστ, όπως το Mardia-Watson-Wheeler test της επόμενης ενότητας (3.3.2) (βλ. πίνακα 3.15, σελ. 82).

Ισοδύναμα, μπορεί να χρησιμοποιηθεί το στατιστικό (Zar, 1999, p. 627)

$$t = \sqrt{g \cdot (n - 2) \frac{R_1 + R_2 - R}{n - (R_1 + R_2)}} \quad (3.19)$$

το οποίο ακολουθεί την κατανομή t -Student. Τότε το t συγκρίνεται με τις κρίσιμες τιμές $t_{\alpha(2),n-2}$. Αντίστοιχα με το παραπάνω στατιστικό F , απορρίπτουμε τη μηδενική υπόθεση της ισότητας των μέσων κατευθύνσεων των πληθυσμών, όταν είναι $t > t_{\alpha(2),n-2}$, ενώ αν είναι $t < t_{\alpha(2),n-2}$, τότε δεν μπορούμε να απορρίψουμε την ομοιογένεια.

Η λογική του παραπάνω ελέγχου μπορεί να παρουσιαστεί και με ένα διαφορετικό τρόπο (Batschelet, 1981, p. 98). Έστω s_1^2 και s_2^2 οι γωνιακές διακυμάνσεις των δύο δειγμάτων και s^2 η από κοινού διακύμανση. Αν οι μέσες κατευθύνσεις των δύο δειγμάτων δεν αποκλίνουν η μία από την άλλη, τότε η s^2 δεν είναι πολύ διαφορετική από την από κοινού διακύμανση

$$\bar{s}^2 = (n_1 s_1^2 + n_2 s_2^2) / n$$

Ωστόσο, αν τα δύο δείγματα διαφέρουν μεταξύ τους στις μέσες κατευθύνσεις τους, τότε είναι προφανές ότι θα ισχύει:

$$s^2 > \bar{s}^2,$$

και, συνεπώς, θα είναι

$$s^2 / \bar{s}^2 > 1 \quad \text{ή} \quad s^2 / \bar{s}^2 - 1 > 0$$

Η τελευταία αυτή διαφορά μπορεί να χρησιμοποιηθεί ως στατιστικό του παραπάνω ελέγχου, οπότε η (3.17) θα πάρει τη μορφή:

$$F = g \cdot (n - 2) \left(\frac{s^2}{\bar{s}^2} - 1 \right) \quad (3.20)$$

Πολυδειγματικός έλεγχος Αν θέλουμε να συγκρίνουμε περισσότερα από δύο δείγματα, έστω m , ως προς τις μέσες κατευθύνσεις τους, ακολουθούμε παρόμοια διαδικασία. Χρησιμοποιώντας τις (3.15) και (3.16) υπολογίζουμε τα

$$R_1, R_2, \dots, R_m$$

και στη συνέχεια τα

$$\begin{aligned} n &= n_1 + n_2 + \dots + n_m \\ C &= C_1 + C_2 + \dots + C_m \\ S &= S_1 + S_2 + \dots + S_m \end{aligned} \quad (3.21)$$

Το στατιστικό του ελέγχου είναι τώρα το

$$F = g \frac{(n - m) \left(\sum^i R_i - R \right)}{(m - 1) \left(n - \sum^i R_i \right)} \quad (3.22)$$

Για τον προσδιορισμό του παράγοντα g χρησιμοποιούμε την τιμή του $\bar{\kappa}$, όπως αυτή προκύπτει από τους πίνακες τιμών του κ , βάσει των παραμέτρων

$$\bar{n} = n/m$$

και

$$\bar{r} = \sum^i R_i/n,$$

ή, παίρνουμε την τιμή του g απευθείας από τους πίνακες [π.χ. Zar (1999, App. 193-194)].² Η κρίσιμη τιμή του στατιστικού ελέγχου είναι η $F_{m-1, n-m}$, με $m - 1$ και $n - m$ βαθμούς ελευθερίας.

Το Watson-Williams test μπορεί να ερμηνευθεί ως ένας έλεγχος ομοιογένειας. Στην περίπτωση των σημαντικών αποτελεσμάτων, μπορούμε μόνο να ισχυριστούμε ότι οι m μέσες κατευθύνσεις δεν είναι όλες ίσες. Το τεστ δεν επιτρέπει το διαχωρισμό των αποτελεσμάτων μεταξύ των επιμέρους (απλών) μέσων κατευθύνσεων.

Το Watson-Williams test είναι ένας πολύ ισχυρός έλεγχος. Αν χρειαστεί, τα δεδομένα μπορεί να ομαδοποιηθούν σε σχετικά μικρά διαστήματα κλάσεων ($\leq 10^\circ$). Δεν προκύπτουν προβλήματα ισοβαθμιών, όπως στα απαραμετρικά

²Ο Zar (1999) δίνει πίνακα τιμών του παράγοντα g , συμβολιζόμενου ως $K = 1 + \frac{3}{8}\kappa$.

τεστ. Οι υπολογισμοί μπορούν να γίνονται εύκολα με τη βοήθεια μιας υπολογιστικής μηχανής. Για τους λόγους αυτούς, ο έλεγχος αυτός είναι ευρέως χρησιμοποιούμενος. Από την άλλη πλευρά, το Watson-Williams test θέτει κάποιες, αρκετά περιοριστικές, παραδοχές. Παρ' όλ' αυτά, ο Zar (1999, p. 629) αναφέρει ότι κάποιες μικρές αποκλίσεις από την κανονικότητα (von Mises κατανομή) και την απαίτηση για ισότητα των θεωρητικών διασπορών μπορεί να είναι ανεκτές, καθώς το συγκεκριμένο τεστ είναι ανθεκτικό σ' αυτές. Αν, πάντως, οι παραδοχές αυτές παραβιάζονται σοβαρά (π.χ. οι κατανομές δεν είναι μονοκόρυφες), τότε ο έλεγχος χάνει την αξιοπιστία του. Πάντως, η απαίτηση $\tilde{\kappa} > 2$ πρέπει να ικανοποιείται οπωσδήποτε. Επιπλέον, όπως προαναφέρθηκε, υπό την H_0 , το χρησιμοποιούμενο στατιστικό F ακολουθεί μόνο κατά προσέγγιση την F -κατανομή.

Το Watson-Williams test προτάθηκε από τους Watson and Williams (1956), αρχικά για $g = 1$. Η εισαγωγή του παράγοντα g προτάθηκε από τον Stephens (1972), με σκοπό να μειωθεί το σφάλμα προσέγγισης. Για μεγάλα δείγματα, οι Rao (1966) και Rao and Sengupta (1972) έχουν προτείνει έναν εναλλακτικό, απαραμετρικό, πολυδειγματικό έλεγχο της ομοιογένειας.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.13: Έλεγχος των συνθηκών για την εφαρμογή του ελέγχου Watson-Williams test στις μεταβλητές WINDIR_D5, -21, -24, -30. Οι συγκεκριμένες μεταβλητές επιλέχθηκαν με κριτήριο τη σχετική εγγύτητα των μέσων γωνιών τους (από 72.319° έως 84.690°). Η μεταβλητή WINDIR_D30 δεν φαίνεται να κατανέμεται κανονικά. Οι μεταβλητές WINDIR_D21 και WINDIR_D24 δεν πληρούν τη συνθήκη $\kappa > 2$. Μπορούμε να πούμε ότι οι συγκεκριμένες μεταβλητές δεν πληρούν τις συνθήκες για την εφαρμογή του ελέγχου.

Variable	WINDIR_D5	WINDIR_D21	WINDIR_D24	WINDIR_D30
Data Type	Angles	Angles	Angles	Angles
Number of Observations	21	22	22	22
Mean Vector ($\bar{\mu}$)	84.69°	82.701°	91.033°	72.319°
Length of Mean Vector (r)	0,758	0,351	0,241	0,347
Concentration	2,428	0,749	0,496	0,74
Circular Variance	0,242	0,649	0,759	0,653
Circular Standard Deviation	42,678°	82,923°	96,706°	83,358°
Standard Error of Mean	9,218°	<u>23,822°</u>	<u>35,36°</u>	<u>24,104°</u>
95% Confidence Interval (-/+) for $\bar{\mu}$	66,62°	36,001°	21,713°	25,066°
99% Confidence Interval (-/+) for $\bar{\mu}$	102,761°	129,402°	160,353°	119,573°
	60,944°	21,332°	359,939°	10,223°
	108,437°	144,071°	182,127°	134,416°
Watson's U ² Test (von Mises, U ²)	0,067	0,117	0,122	0,281
Watson's U ² Test (p)	0.25 > p > 0.15	< 0,005	< 0,005	< 0,005
Kuiper's Test (von Mises, V)	1,369	1,558	1,417	2,226
Kuiper's Test (p)	> 0.15	0.15 > p > 0.10	> 0.15	< 0,01

\bar{x} = αναξιόπιστη τιμή λόγω χαμηλής συγκέντρωσης
(ομοιόμορφη κατανομή)

Πίνακας 3.14: Αποτελέσματα του ελέγχου Watson-Williams test στις μεταβλητές WINDIR_D5, -21, -24, -30. Όπως προκύπτει από τα στοιχεία του πίνακα 3.13, οι συγκεκριμένες μεταβλητές δεν πληρούν τις συνθήκες για την εφαρμογή αυτού του παραμετρικού ελέγχου. Ο έλεγχος δίνει υψηλά P -value, κάτι ίσως αναμενόμενο, καθώς υπάρχει σχετική εγγύτητα των μέσων κατευθύνσεων ($72.319^\circ - 84.690^\circ$), όμως τα περισσότερα αποτελέσματα χαρακτηρίζονται από το πρόγραμμα ως αναξιόπιστα, λόγω χαμηλής συγκέντρωσης, δηλαδή οι μεταβλητές δεν κατανέμονται κανονικά, αλλά ομοιόμορφα.

<i>Variables (& observations)</i>	<i>F</i>	<i>p</i>	<i>df</i>	<i>df2</i>	<i>Est. Mean</i>
<i>Από κοινού έλεγχος</i>					
WINDIR_D5 (21)					
WINDIR_D21 (22)					
WINDIR_D24 (22)					
WINDIR_D30 (22)	<u>0,147</u>	0,931	3	83	82,613
<i>Ανά δύο έλεγχος</i>					
WINDIR_D5 & WINDIR_D21 (21 & 22)	0,009	0,927	1	41	84,041
WINDIR_D5 & WINDIR_D24 (21 & 22)	0,061	0,806	1	41	86,273
WINDIR_D5 & WINDIR_D30 (21 & 22)	0,326	0,571	1	41	80,687
WINDIR_D21 & WINDIR_D24 (22 & 22)	<u>0,072</u>	0,789	1	42	86,09
WINDIR_D21 & WINDIR_D30 (22 & 22)	<u>0,139</u>	0,711	1	42	77,539
WINDIR_D24 & WINDIR_D30 (22 & 22)	<u>0,362</u>	0,551	1	42	79,968
<i>x = αναξιόπιστο αποτέλεσμα, λόγω χαμηλής συγκέντρωσης (ομοιόμορφη κατανομή)</i>					

Έλεγχος των μεταβλητών WINDIR_D5, 21, 24, 30 με το Watson-Williams test για την ισονομία των αντίστοιχων πληθυσμών

Από το σύνολο των μεταβλητών WINDIR_Dxx, επιλέχθηκαν οι WINDIR_D5, 21, 24, 30, με κριτήριο τη σχετική εγγύτητα των μέσων γωνιών τους (από 72.319° έως 84.690°), και εφαρμόστηκε ο από κοινού έλεγχος και ο ανά δύο έλεγχος, δηλαδή ο έλεγχος για όλα τα δυνατά ζεύγη, Watson-Williams test (βλ. πίνακα (3.14), σελ. 77).

Η μεταβλητή WINDIR_D30 δεν φαίνεται να κατανέμεται κανονικά. Οι μεταβλητές WINDIR_D21 και WINDIR_D24 δεν πληρούν τη συνθήκη $\kappa > 2$. Μόνο η μεταβλητή WINDIR_D5 με τιμή συγκέντρωσης *Concentration* = 2.428 φαίνεται να πληροί τις συνθήκες της σελ. 68 για την εφαρμογή του ελέγχου. Μπορούμε, λοιπόν, να πούμε ότι οι συγκεκριμένες μεταβλητές δεν πληρούν τις συνθήκες για την εφαρμογή του ελέγχου, συνεπώς θα πρέπει να εφαρμοστεί κάποιος κατάλληλος παραμετρικός έλεγχος.

Προκειμένου να αναδειχθεί το πρόβλημα της μη πλήρωσης των συνθηκών της εφαρμογής του συγκεκριμένου παραμετρικού ελέγχου, στις μεταβλητές WINDIR_D5, -21, -24, -30 έγινε ο έλεγχος Watson-Williams test για την ισονομία των πληθυσμών τους. Όπως προκύπτει από τα στοιχεία του πίνακα 3.13, ο έλεγχος δίνει υψηλά *P*-value, κάτι ίσως αναμενόμενο, καθώς δεν υπάρχει μεγάλη γωνιακή απόσταση μεταξύ των μέσων κατευθύνσεων (72.319°– 84.690°).

Σε ό,τι αφορά στην ομοιομορφία των μεταβλητών WINDIR_D5, -21, -24, -30, τα αποτελέσματα των σχετικών ελέγχων (κυρίως του Rayleigh test) υποδεικνύουν γι' αυτές την ομοιόμορφη κατανομή. Ειδικά για τη μεταβλητή WINDIR_D24, η τυχαιότητα φαίνεται να μην μπορεί να απορριφθεί ούτε από τα Watson's U^2 test και Kuiper's test. Γι' αυτό, τα περισσότερα αποτελέσματα χαρακτηρίζονται από το πρόγραμμα *Oriana* ως αναξιόπιστα, λόγω χαμηλής συγκέντρωσης, δηλαδή οι μεταβλητές δεν κατανέμονται κανονικά, αλλά ομοιόμορφα.

3.3.2 Mardia-Watson-Wheeler test

Το Mardia-Watson-Wheeler test είναι ένα απαραμετρικό τεστ για τον έλεγχο της ισονομίας των κατανομών των πληθυσμών από τους οποίους προέρχονται τα δειγματικά δεδομένα. Η μέθοδος βασίζεται στην (αύξουσα) ταξινομήση των τιμών της κυκλικής μεταβλητής και την απόδοση βαθμών στις αντίστοιχες παρατηρήσεις.

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n_1 και n_2 :

$$\phi_1, \phi_2, \dots, \phi_{n_1} \quad \psi_1, \psi_2, \dots, \psi_{n_2}$$

Θέλουμε να ελέγξουμε αν τα δύο αυτά δείγματα ανήκουν σε ταυτόσημους πληθυσμούς ή διαφέρουν σημαντικά μεταξύ τους, ως προς τη μέση κατεύθυνση, τη γωνιακή διακύμανση ή και τα δύο. Δεχόμαστε ότι οι δύο πληθυσμοί έχουν συνεχείς κυκλικές κατανομές και ότι δεν προκύπτουν ισοβαθμίες στα δείγματα. Παρατηρούμε την τάξη με την οποία διατάσσονται οι παρατηρήσεις ϕ_i και ψ_i . Έτσι, ταξινομούμε τις παρατηρήσεις και των δύο δειγμάτων κατ' αύξουσα σειρά και αναδιαμορφώνουμε τα διαστήματα (τόξα) μεταξύ διαδοχικών παρατηρήσεων και των δύο δειγμάτων πάνω στο μοναδιαίο κύκλο έτσι, ώστε αυτά να αθροίζουν το ίδιο μέγεθος, δηλαδή

$$\delta = 360^\circ/n,$$

όπου

$$n = n_1 + n_2$$

Έχοντας ταξινομήσει και επανατοποθετήσει ισομερώς πάνω στον κύκλο τα δειγματικά σημεία, βαθμονομούμε τις αρχικές παρατηρήσεις (γωνίες ϕ και ψ) με τα βάρη

$$r_j, \quad j = 1, 2, \dots, n$$

και παίρνουμε τις νέες (τροποποιημένες) παρατηρήσεις,

$$d_j = r_j \delta \quad j = 1, 2, \dots, n \quad (3.23)$$

Τότε, το συνισταμένο διάνυσμα \vec{R} των n τροποποιημένων γωνιών έχει συ-

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

νιστώσες

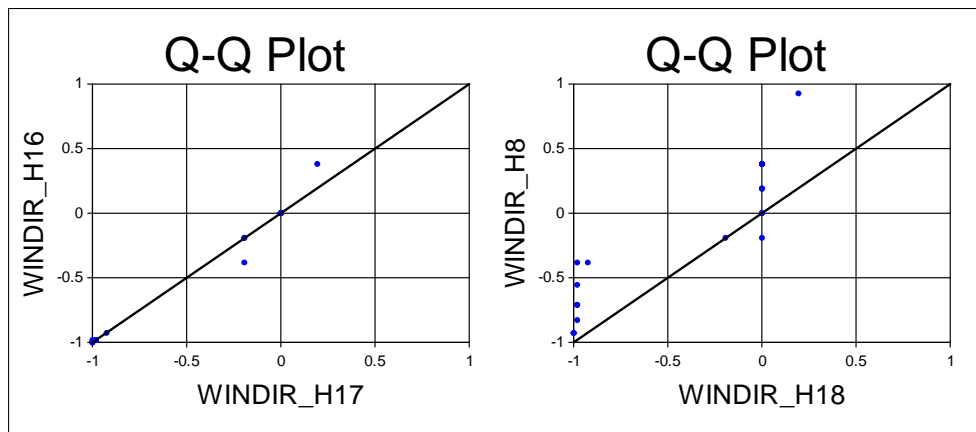
$$C_i = \sum_{j=1}^{n_i} \cos d_j \quad (3.24)$$

$$S_i = \sum_{j=1}^{n_i} \sin d_j$$

όπου το $i = 1, 2$ αναπαριστά το κάθε δείγμα. Το στατιστικό του ελέγχου είναι

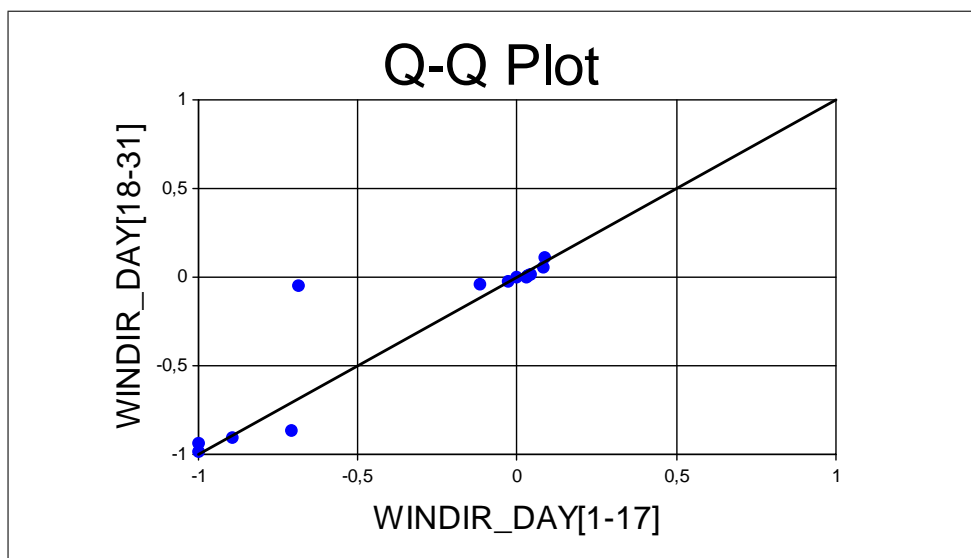
$$W = 2 \left(\frac{C_1^2 + S_1^2}{n_1} + \frac{C_2^2 + S_2^2}{n_2} \right) \quad (3.25)$$

το οποίο ακολουθεί τη χ^2 κατανομή με 2 βαθμούς ελευθερίας (Mardia, 1972). Ένα παράδειγμα παρουσιάζεται στο διάγραμμα Q-Q plot³ του σχήματος (3.12).



Σχήμα 3.12: Εποπτικός έλεγχος ισονομίας πληθυσμών. Στο αριστερό διάγραμμα Q-Q plot, οι μεταβλητές WINDIR_H16 και WINDIR_H17 φαίνεται να προσαρμόζουν καλά στην ευθεία αναφοράς των 45° , όπως απαιτείται, ώστε να ισχυριστούμε ότι οι κατανομές τους συμπίπτουν. Στο δεξί διάγραμμα, οι μεταβλητές WINDIR_H8 και WINDIR_H18 φαίνεται να αποκλίνουν αισθητά από την ευθεία αναφοράς των 45° , ώστε να μπορούμε να ισχυριστούμε ότι οι κατανομές τους διαφέρουν. Το Mardia-Watson-Wheeler test επιβεβαιώνει αυτές τις εντυπώσεις, δίνοντας πιθανότητες P -value 0.906 και 0.013, αντίστοιχα.

³Το διάγραμμα Q-Q plot συγκρίνει τις κατανομές δύο δειγμάτων, κατά τον τρόπο με τον οποίο ένα διάγραμμα καλής προσαρμογής (distribution plot) συγκρίνει μια δειγματική κατανομή ως προς μια θεωρητική. Τα δείγματα στοιχίζονται ως προς τις διαμέσους τους, κατόπιν ταξινομούνται από κοινού σε αύξουσα σειρά και προβάλλονται το ένα ως προς το άλλο. Αν υπάρχουν διαφορές στο μέγεθος των δειγμάτων, επιλέγεται ο μικρότερος αριθμός παρατηρήσεων. Αν οι δύο κατανομές συμπίπτουν, όλα τα σημεία θα βρίσκονται κατά μήκος της ευθείας αναφοράς των 45° . Στην αντίθετη περίπτωση, όπου οι κατανομές διαφέρουν σημαντικά, θα υπάρχουν αισθητές αποκλίσεις των δειγματικών σημείων από την ευθεία αναφοράς.



Σχήμα 3.13: Εποπτικός έλεγχος της ισονομίας των πληθυσμών των μεταβλητών δεύτερης τάξης WINDIR_DAY [1-17] και WINDIR_DAY [18-31] με χρήση διαγράμματος Q-Q plot. Τα σημεία των δύο δειγμάτων φαίνεται να προσαρμόζουν καλά στην ευθεία αναφοράς των 45°, όπως απαιτείται, ώστε να ισχυριστούμε ότι οι κατανομές τους συμπίπτουν. Το Mardia-Watson-Wheeler test επιβεβαιώνει αυτή την εντύπωση, δίνοντας κρίσιμο επίπεδο $P\text{-value}=0,471$.

Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών WINDIR_DAY[1-17] και WINDIR_DAY[18-31] με το Mardia-Watson-Wheeler test

Τα στοιχεία του πίνακα (3.15), σελ. 82, με την υψηλή τιμή του $P = 0.471$, υποδεικνύουν ότι, για τις μεταβλητές WINDIR_DAY[1-17] και WINDIR_DAY[18-31], οι αντίστοιχοι πληθυσμοί δεν διαφοροποιούνται, δηλαδή η κατεύθυνση του ανέμου δεν μεταβάλλεται μεταξύ των φυσιολογικών περιόδων και των περιόδων καύσωνα.⁴

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.15: Αποτελέσματα της εφαρμογής του ελέγχου Mardia-Watson-Wheeler test στις μεταβλητές δεύτερης τάξης WINDIR_DAY [1-17] και WINDIR_DAY [18-31].

<i>Variables (& observations)</i>	<i>W</i>	<i>p</i>
<i>Από κοινού έλεγχος</i>		
WINDIR_DAY (31)		
WINDIR_DAY[1-17] (17)		
WINDIR_DAY[18-31] (14)	1,435	0,838
<i>Ανά δύο έλεγχος</i>		
WINDIR_DAY & WINDIR_DAY[1-17] (31 & 17)	0,396	0,82
WINDIR_DAY & WINDIR_DAY[18-31] (31 & 14)	0,454	0,797
WINDIR_DAY[1-17] WINDIR_DAY[18-31] (17 & 14)	& 1,506	0,471

Προσέγγιση Batschelet Ο Batschelet (1981, pp. 101-105), σε μια κάπως διαφορετική προσέγγιση της διαδικασίας, σημειώνει ότι, αν κατά την προετοιμασία του ελέγχου (ταξινόμηση και επανατοποθέτηση των παρατηρήσεων), τα σημεία του ενός δείγματος είναι εμφανώς διαχωρισμένα από τα αντίστοιχα σημεία του άλλου δείγματος, τότε αναμένουμε σημαντική διαφορά μεταξύ των δύο δειγμάτων. Ένας τέτοιος διαχωρισμός υποδεικνύεται από το μέγεθος των δύο αντίθετων (αντίρροπων και ίσων, ως προς το μήκος) συνισταμένων διάνυσμάτων. Το τελευταίο προκύπτει από το γεγονός ότι, αν υπολογίσουμε το συνισταμένο διάνυσμα του συνδυασμένου δείγματος, αυτό θα ισούται με το μηδενικό διάνυσμα, λόγω της προηγούμενης τεχνητής ομοιόμορφης κατανομής όλων των δειγματικών σημείων. Δηλαδή, θα είναι

$$R_1 = R_2$$

και, κατά συνέπεια, όσο μεγαλύτερο είναι το R_1 (ή το R_2), τόσο περισσότερο διαχωρισμένα είναι τα δύο δείγματα. Τότε, αν είναι

$$r_1, r_2, \dots, r_{n_1}$$

οι βαθμοί για τα σημεία, έστω, του πρώτου δείγματος, οι αντίστοιχες τροποποιημένες γωνίες είναι

$$d_i = \delta r_i \quad (i = 1, 2, \dots, n_1),$$

και το συνισταμένο διάνυσμα (\vec{R}_1) για το πρώτο δείγμα έχει συνιστώσες

$$C_1 = \sum_i \cos i, \quad S_1 = \sum_i \sin i$$

με το μήκος του να είναι

$$R_1 = \sqrt{C^2 + S^2}$$

Ως στατιστικό ελέγχου χρησιμοποιούμε το

$$B = R_1^2 \tag{3.26}$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Ο Batschelet (1981, p. 344) δίνει πίνακα τιμών του στατιστικού $B(\alpha)$, για διάφορα επίπεδα σημαντικότητας α . Για δειγματικά μέγεθη $n > 17$, η ποσότητα

$$X^2 = 2(n - 1) \frac{R_1^2}{n_1 n_2} \quad (3.27)$$

κατανέμεται προσεγγιστικά ως χ^2 κατανομή με δύο βαθμούς ελευθερίας, υπό την H_0 . Στην περίπτωση αυτή μπορεί να χρησιμοποιούνται οι πιο συνήθεις πίνακες κρίσιμων τιμών της χ^2 κατανομής. Όταν είναι $B > B(\alpha)$ απορρίπτουμε τη μηδενική υπόθεση της ομοιογένειας και συμπεραίνουμε ότι τα δύο δείγματα διαφέρουν σημαντικά μεταξύ τους. Όταν είναι $B < B(\alpha)$, δεν μπορούμε να αποκλείσουμε τη (μηδενική) υπόθεση του να ανήκουν τα δείγματα στον ίδιο πληθυσμό.

Το τεστ είναι αρκετά ισχυρό. Επίσης, η εφαρμογή του είναι εύκολα υλοποιήσιμη. Όπως προαναφέρθηκε, σημαντικά αποτελέσματα μπορεί να προκύψουν λόγω αποκλίσεων είτε των μέσων κατευθύνσεων, είτε των γωνιακών διακυμάνσεων. Αυτή η αμφιβολία απαιτεί διερεύνηση των αρχικών δειγμάτων πριν την εφαρμογή της διαδικασίας της ομοιόμορφης ανακατανομής. Επίσης, δεν επιτρέπονται οι ισοβαθμίες, δηλαδή οι συμπτώσεις γωνιών από το ένα και το άλλο δείγμα. Αν προκύψει κάτι τέτοιο, απαιτείται η διόρθωσή του με τυχαία ανακατανομή των σημείων. Κατά συνέπεια, ο έλεγχος δεν είναι εφαρμόσιμος σε μεγάλα δείγματα (με υψηλή πιθανότητα πολλών ισοβαθμιών) ή σε δείγματα με ομαδοποιημένα δεδομένα (Batschelet, 1981, p. 103).

Το τεστ είναι εφαρμόσιμο, όταν τα δειγματικά μεγέθη n_1 και n_2 είναι τουλάχιστον ίσα με 10. Αν τα δειγματικά μεγέθη είναι μικρότερα του 10 ή υπάρχουν ισοβαθμίες, τότε είναι προτιμότερο το Watson's U^2 test της ενότητας 3.3.5, σελ. 91 (Zar, 1999, p. 633).

Το Mardia-Watson-Wheeler test προτάθηκε από τους Watson and Wheeler (1964) και ανακαλύφθηκε, ανεξάρτητα, από τον Mardia (1967), ο οποίος το έκανε εφαρμόσιμο παρέχοντας έναν πίνακα κρίσιμων τιμών. Το τεστ είναι επίσης γνωστό και ως *uniform-score test* (Mardia, 1972).

Πολυδειγματικός έλεγχος Ο Mardia (1972) πρότεινε τη γενίκευση του Mardia-Watson-Wheeler test για k δείγματα. Έστω n_1, n_2, \dots, n_k τα δειγματικά μεγέθη και

$$n = n_1 + n_2 + \dots + n_k$$

το μέγεθος του συνδυαστικού δείγματος. Υποθέτουμε ότι δεν υπάρχουν ισοβαθμίες ή, υπάρχουν λίγες που μπορεί να εξαλειφθούν, και εφαρμόζοντας την παραπάνω διαδικασία παίρνουμε τους βαθμούς

$$r_{ij} \quad (j = 1, 2, \dots, n_i)$$

που δίνουν τις τροποποιημένες γωνίες

$$d_{ij} = r_{ij} \delta$$

του i -οστού δείγματος. Κατόπιν, υπολογίζουμε τις συνιστώσες του συνισταμένου διανύσματος κάθε δείγματος

$$C_i = \sum_{j=1}^{n_i} \cos \beta_{ij}, \quad S_i = \sum_{j=1}^{n_i} \sin \beta_{ij} \quad (i = 1, 2, \dots, k)$$

Οι ισότητες

$$\sum_{i=1}^k C_i = 0 \quad \sum_{i=1}^k S_i = 0$$

επαληθεύουν τη διαδικασία. Τα τετραγωνικά μήκη των μέσων διανυσμάτων είναι

$$R_i^2 = C_i^2 + S_i^2 \quad (i = 1, 2, \dots, k)$$

Ως στατιστικό έλεγχου χρησιμοποιούμε το

$$W = 2 \sum_{i=1}^k R_i^2 / n_i \quad (3.28)$$

το οποίο ακολουθεί τη χ^2 κατανομή με $2(k-1)$ βαθμούς ελευθερίας. Η διαδικασία είναι εφαρμόσιμη, μόνο αν κάθε $n_i \geq 10$ (Fisher, 1993) και μόνο αν δεν υπάρχουν ισοβαθμίες (Batschelet, 1981, p. 102). Οι Mardia and Spurr (1973) δίνουν πίνακα κρίσιμων τιμών του W , για $k = 3$ και μικρά δειγματικά μεγέθη.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.16: Αποτελέσματα της εφαρμογής του απαραμετρικού ελέγχου Mardia-Watson-Wheeler test στις μεταβλητές WINDIR_D5, -21, -24, -30. Παρατηρούμε ότι ο έλεγχος για την ισονομία των πληθυσμών της ομάδας των μεταβλητών WINDIR_D5, -21, -24, -30 με το Mardia-Watson-Wheeler test, σε αντίθεση με τον έλεγχο Watson-Williams test της προηγούμενης ενότητας (βλ. πίνακα 3.14, σελ. 77), δίνει πολύ σημαντικά αποτελέσματα κατά τον έλεγχο της μεταβλητής WINDIR_D5 με τις υπόλοιπες (ανά δύο έλεγχος) και απορρίπτει τη μηδενική υπόθεση της ισονομίας των πληθυσμών (από κοινού έλεγχος).

<i>Variables (& observations)</i>	<i>W</i>	<i>p</i>
<i>Από κοινού έλεγχος</i>		
WINDIR_D5 (21)		
WINDIR_D21 (22)		
WINDIR_D24 (22)		
WINDIR_D30 (22)	17,281	0,008
<i>Ανά δύο έλεγχος</i>		
WINDIR_D5 & WINDIR_D21 (21 & 22)	9,691	0,008
WINDIR_D5 & WINDIR_D24 (21 & 22)	10,249	0,006
WINDIR_D5 & WINDIR_D30 (21 & 22)	12,248	0,002
WINDIR_D21 & WINDIR_D24 (22 & 22)	0,625	0,732
WINDIR_D21 & WINDIR_D30 (22 & 22)	5,199	0,074
WINDIR_D24 & WINDIR_D30 (22 & 22)	3,104	0,212

Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών WINDIR_D5, -21, -24, -30 με το Mardia-Watson-Wheeler test

Από τα στοιχεία του πίνακα (3.16), σελ. 86, παρατηρούμε ότι ο έλεγχος για την ισονομία των πληθυσμών της ομάδας των μεταβλητών WINDIR_D5, -21, -24, -30 με το Mardia-Watson-Wheeler test, σε αντίθεση με τον έλεγχο Watson-Williams test της προηγούμενης ενότητας, δίνει πολύ σημαντικά αποτελέσματα κατά τον έλεγχο της μεταβλητής WINDIR_D5 με τις υπόλοιπες και απορρίπτει τη μηδενική υπόθεση της ισονομίας των πληθυσμών.

Η εφαρμογή του παραμετρικού ελέγχου Watson-Williams test για τις μεταβλητές αυτές έδωσε δεν έδωσε σημαντικά αποτελέσματα, τα οποία υποδείκνυαν ότι οι μέσες κατευθύνσεις των συγκεκριμένων μεταβλητών δεν διαφέρουν, από κοινού ή ανά δύο. Όμως, ο έλεγχος είναι αμφίβολης αξιοπιστίας, καθώς δεν πληρούνταν οι συνθήκες για την εφαρμογή του. Από την άλλη, το Mardia-Watson-Wheeler test δίνει σημαντικά αποτελέσματα, τόσο στον από κοινού, όσο και στον ανά δύο έλεγχο, κάτι που σημαίνει ότι οι μέσες κατευθύνσεις των αντίστοιχων πληθυσμών, τελικά, διαφοροποιούνται.

Ακόμα, ο απαραμετρικός έλεγχος φαίνεται να μην απορρίπτει την ισονομία των πληθυσμών για τις υπόλοιπες, πλην της WINDIR_D5, μεταβλητές. Κατά την εφαρμογή του στις μεταβλητές WINDIR_D21, -24, -30, δεν δίνει σημαντικό αποτέλεσμα με υψηλό, μάλιστα, $P\text{-value}=0.583$, δηλαδή, υποδεικνύει την ισονομία τους. Η παρουσία, πάντως, πολλών ισοβαθμιών μεταξύ των τιμών των μεταβλητών αυτών πιθανόν επηρεάζει την αξιοπιστία του ελέγχου.

3.3.3 chi-squared test

Το ευρύτατα χρησιμοποιούμενο στατιστικό ελέγχου chi-squared test προτάθηκε από τον Pearson, περίπου το 1900, ενώ η θεωρία και η εφαρμογή του αναπτύχθηκαν περαιτέρω από τον ίδιο και τον Fisher. Χρησιμοποιείται για τον έλεγχο κυρίως κατηγορικών δεδομένων.

Έστω Δ_1 και Δ_2 δύο ανεξάρτητα τυχαία δείγματα ομαδοποιημένων κυκλικών παρατηρήσεων. Παριστάνουμε με

$$\nu_{ij} \quad (j = 1, 2, \dots, c)$$

την j -οστή συχνότητα του δείγματος Δ_i . Θέλουμε να ελέγξουμε το αν τα δύο αυτά δείγματα προέρχονται από πληθυσμούς, οι οποίοι διαφέρουν σημαντικά μεταξύ τους. Δεν εξειδικεύεται το αίτιο της τυχόν απόκλισης.

Πίνακας 3.17: Πίνακας συχνοτήτων και αθροιστικών συχνοτήτων για το Chi-squared test.

					ΣΥΝΟΛΟ
Δ_1	n_{11}	n_{12}	\dots	n_{1c}	M_1
Δ_2	n_{21}	n_{22}	\dots	n_{2c}	M_2
ΣΥΝΟΛΟ	N_1	N_2	\dots	N_c	N

Οι συχνότητες των παρατηρήσεων σχηματίζουν τον παραπάνω $2 \times c$ πίνακα (3.17). Βάσει των αθροιστικών συχνοτήτων των δύο δειγμάτων M_i , N_j και της ολικής συχνότητας N , υπολογίζουμε, κατά τον ίδιο τρόπο όπως και στα μη κυκλικά δεδομένα, τις αναμενόμενες συχνότητες

$$e_{ij} = \frac{M_i N_j}{N} \quad (i = 1, 2; \quad j = 1, 2, \dots, c) \quad (3.29)$$

Κατά παραδοχή, οι αναμενόμενες συχνότητες πρέπει να είναι τουλάχιστον

$$e_{ij} \geq 5$$

Τότε, η μηδενική υπόθεση ότι τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό ελέγχεται με το στατιστικό

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (3.30)$$

Το παραπάνω στατιστικό, υπό τη μηδενική υπόθεση, ακολουθεί τη χ^2 κατανομή με $(r - 1)(c - 1)$ βαθμούς ελευθερίας για πίνακα συχνοτήτων $r \times c$. Στην υποθετική περίπτωση που εξετάζουμε ο πίνακας συχνοτήτων είναι μεγέθους $2 \times c$ και οι βαθμοί ελευθερίας που αντιστοιχούν από τους πίνακες κρίσιμων τιμών της χ^2 κατανομής είναι $c - 1$.

Αν είναι $\chi^2 > \chi^2(\alpha)$, τότε μπορούμε να απορρίψουμε, σε επίπεδο σημαντικότητας α , τη μηδενική υπόθεση της ισονομίας των κατανομών των πληθυσμών και να συμπεράνουμε ότι τα δύο δείγματα προέρχονται από διαφορετικούς πληθυσμούς. Για $\chi^2 < \chi^2(\alpha)$ δεν μπορούμε να απορρίψουμε την ομοιογένεια. Μπορούμε, επίσης, να καταλήξουμε σε συμπέρασμα σχετικά με την απόρριψη ή όχι της H_0 βασιζόμενοι στο κρίσιμο επίπεδο P -value. Αν είναι $P < \alpha$, τότε υπάρχουν σημαντικά αποτελέσματα. Πάντως, η ύπαρξη σημαντικών αποτελεσμάτων δεν συνεπάγεται διαφορά στη μέση κατεύθυνση. Για την εξακρίβωση του αιτίου της απόκλισης, απαιτείται προσεκτική διερεύνηση των δειγμάτων (Batschelet, 1981, p. 110).

Το chi-squared test είναι ιδιαίτερα χρήσιμο στην περίπτωση όπου η περιορισμένη ακρίβεια των μετρήσεων αναγκάζει τον ερευνητή να χρησιμοποιήσει σχετικά μεγάλα διαστήματα κλάσεων (μεγαλύτερα των 10°). Από την άλλη, το τεστ απαιτεί μεγάλο αριθμό παρατηρήσεων, προκειμένου να εξασφαλίζεται η συνθήκη $e_{ij} \geq 5$. Αν η συνθήκη αυτή δεν ικανοποιείται, τότε πρέπει να προβούμε σε αναδιαμόρφωση των ομάδων, μειώνοντας τους βαθμούς ελευθερίας.

3.3.4 Kuiper's test

Το Kuiper's test βασίζεται στη σύγκριση των δειγματικών συναρτήσεων κατανομής, όπου η παρατήρηση σχετικά μεγάλων αποκλίσεων οδηγεί στο συμπέρασμα ότι οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα είναι διαφορετικοί.

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n και m :

$$\phi_1, \phi_2, \dots, \phi_n \quad \psi_1, \psi_2, \dots, \psi_m$$

Θέλουμε να ελέγξουμε αν τα δύο αυτά δείγματα διαφέρουν σημαντικά μεταξύ τους. Και εδώ, το αίτιο της απόκλισης δεν εξειδικεύεται: οι κατανομές των πληθυσμών μπορεί να διαφέρουν ως προς τη μέση κατεύθυνση, τη γωνιακή διακύμανση ή κάποια άλλη ιδιότητα.

Η βασική ιδέα αυτού του ελέγχου είναι η σύγκριση των δειγματικών συ-

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

ναρτήσεων κατανομών: όσο μεγαλύτερη απόκλιση παρατηρείται μεταξύ των κατανομών αυτών, τόσο πιθανότερο είναι να υπάρχουν σημαντικές διαφορές στους πληθυσμούς. Παραδεχόμαστε ότι τα δείγματα προέρχονται από πληθυσμούς με συνεχείς κατανομές και ότι στην περίπτωση ομαδοποιημένων δεδομένων, το διάστημα της κάθε κλάσης είναι σχετικά μικρό (όχι μεγαλύτερο από 5°).

Ξεκινάμε τη διαδικασία αυτού του τεστ, σχεδιάζοντας τις δύο δειγματικές συναρτήσεις κατανομής, έστω $F_1(\phi)$ και $F_2(\psi)$, οι οποίες είναι κλιμακωτές συναρτήσεις. Ορίζουμε ως D^+ τη μέγιστη (απόλυτη) παρατηρούμενη διαφορά

$$F_2(\psi) - F_1(\phi)$$

και ως D^- την αντίθετη μέγιστη διαφορά

$$F_1(\phi) - F_2(\psi)$$

Αν θέσουμε

$$V_{n,m} = D^+ - D^-$$

τότε το στατιστικό του ελέγχου είναι το

$$k = n m V_{n,m} \tag{3.31}$$

Ο Batschelet (1981, pp.346-347) δίνει πίνακα κρίσιμων επιπέδων P -value για μικρές τιμές των n και m . Χάριν απλούστευσης, υποθέτουμε ότι είναι $n \leq m$. Για τιμές των n και m εκτός του παραπάνω πίνακα, προτείνει, για $n \leq m$, την ποσότητα

$$L(\alpha) = \sqrt{nm(n+m)}K(\alpha) \tag{3.32}$$

όπου το $K(\alpha)$ είναι η κρίσιμη τιμή για το αντίστοιχο Kuiper's test για ένα δείγμα, μεγέθους n .

Η μηδενική υπόθεση της ισονομίας των πληθυσμών απορρίπτεται σε επίπεδα σημαντικότητας α μεγαλύτερα από το κρίσιμο επίπεδο P -value, $P < \alpha$. Αν ο παραπάνω πίνακας δεν μπορεί να χρησιμοποιηθεί, καταφεύγουμε σε μια προσεγγιστική διαδικασία, κατά την οποία, για δεδομένο επίπεδο σημαντικότητας α , προκύπτουν σημαντικά αποτελέσματα, όταν η τιμή του στατιστικού ελέγχου k υπερβαίνει την κρίσιμη τιμή $L(\alpha)$, όπως αυτή δίνεται από την (3.32).

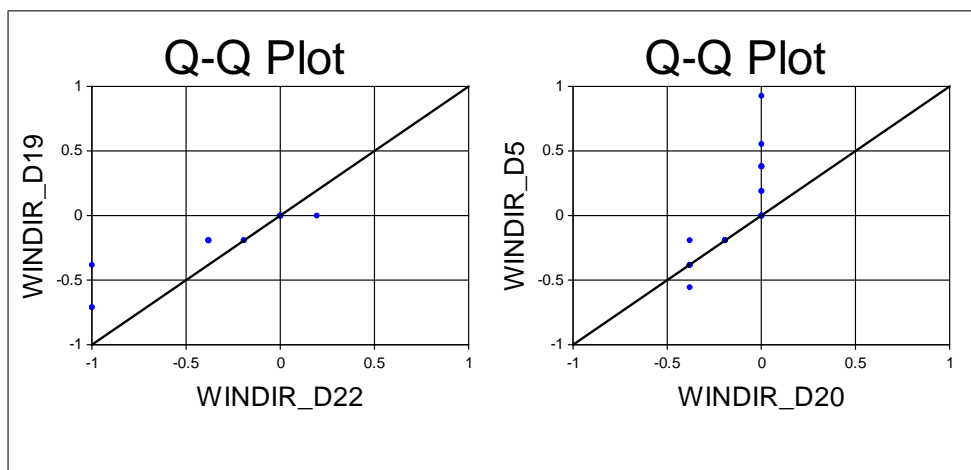
Το Kuiper's test έχει μεγάλη πρακτική αξία, καθώς είναι πολύ ισχυρό στον

εντοπισμό των σημαντικών διαφορών, οι οποίες μπορεί να οφείλονται σε κάθε είδος απόκλισης. Μπορεί εύκολα να εφαρμοστεί με τη χρήση ενός διαγράμματος. Η εφαρμογή του είναι επίσης εφικτή με τον υπολογισμό του στατιστικού K του μονοδειγματικού ελέγχου.

Το τεστ αυτό έχει προταθεί από τους Kolmogorov και Smirnov για μη κυκλικά δεδομένα. Ο Kuiper (1960) προσαρμοσε τον έλεγχο για την περίπτωση των κυκλικών δεδομένων, κάνοντάς το στατιστικά ανεξάρτητο από την (όποια) αρχική κατεύθυνση.

3.3.5 Watson's U^2 test

Το Watson's U^2 test μοιάζει στη βασική του ιδέα, της σύγκρισης των δειγματικών συναρτήσεων κατανομής, με το Kuiper's test, μόνο που αντί της μέτρησης των μέγιστων αποκλίσεων μεταξύ των δύο γραφημάτων, χρησιμοποιεί τη (διορθωμένη) μέση τετραγωνική απόκλιση.



Σχήμα 3.14: Εποπτικός έλεγχος της ισονομίας πληθυσμών. Στο αριστερό διάγραμμα Q-Q plot, οι μεταβλητές WINDIR_D19 και WINDIR_D22 φαίνεται να προσαρμόζουν σχετικά καλά στην ευθεία αναφοράς των 45° , όπως απαιτείται, ώστε να ισχυριστούμε ότι οι κατανομές τους συμπίπτουν. Στο δεξί διάγραμμα, οι μεταβλητές WINDIR_D5 και WINDIR_D20 φαίνεται να αποκλίνουν αισθητά από την ευθεία αναφοράς των 45° , ώστε να μπορούμε να ισχυριστούμε ότι οι κατανομές τους διαφέρουν. Το Watson's U^2 test επιβεβαιώνει αυτές τις εντυπώσεις, δίνοντας πιθανότητες P -value 0.486 και 0, αντίστοιχα.

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n και m :

$$\phi_1, \phi_2, \dots, \phi_n \quad \psi_1, \psi_2, \dots, \psi_m$$

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Θέλουμε να ελέγξουμε αν τα δύο αυτά δείγματα διαφέρουν σημαντικά μεταξύ τους. Και σ' αυτό τον έλεγχο, το αίτιο της απόκλισης δεν εξειδικεύεται: οι κατανομές των πληθυσμών μπορεί να διαφέρουν ως προς τη μέση κατεύθυνση, τη γωνιακή διακύμανση ή κάποια άλλη ιδιότητα. Κάνουμε τις παραδοχές ότι οι πληθυσμοί απ' όπου προέρχονται τα δείγματα έχουν συνεχείς κατανομές και ότι στην περίπτωση ομαδοποιημένων δεδομένων το πλάτος των κλάσεων δεν ξεπερνά τις 5°.

Αν θέσουμε ως $n + m = N$ το σύνολο των δειγματικών σημείων και ως d_1, d_2, \dots, d_N τις διαφορές μεταξύ των συναρτήσεων κατανομών σε κάθε σημείο και από τα δύο δείγματα, τότε, συμβολίζοντας με \bar{d} τον αριθμητικό μέσο των διαφορών αυτών, θεωρούμε την ποσότητα

$$U^2 = \frac{nm}{N^2} \sum_{k=1}^N (d_k - \bar{d})^2. \quad (3.33)$$

Όσο μεγαλύτερο είναι το U^2 , τόσο πιθανότερο είναι τα δύο δείγματα να ανήκουν σε διαφορετικούς πληθυσμούς.

Η ποσότητα της (3.33) αποτελεί το στατιστικό του ελέγχου, το οποίο, πάντως, μπορεί να αναδιατυπωθεί σε μια μορφή καταλληλότερη για αριθμητικούς υπολογισμούς (Zar, 1999, p. 630)

$$U^2 = \frac{nm}{N^2} \left[\sum_{k=1}^N d_k^2 - \frac{1}{N} \left(\sum_{k=1}^N d_k \right)^2 \right] \quad (3.34)$$

Δίνονται πίνακες με κρίσιμες τιμές του $U^2(\alpha)$ για $n \geq 5$ και $m \geq 5$ [Batschelet (1981, p. 348), Zar (1999, App. 195-197), με διάταξη $n \leq m$, Mardia and Jupp (2000, p. 377), για $N \geq 9$ και $n \geq 4$]. Για μεγάλες τιμές των n και m , οι κρίσιμες τιμές είναι σχεδόν ανεξάρτητες από τα n και m . Η μηδενική υπόθεση της ισονομίας των πληθυσμών απορρίπτεται, σε επίπεδο σημαντικότητας α , όταν είναι $U^2 > U^2(\alpha)$, οπότε παρατηρούμε ότι τα δύο δείγματα διαφέρουν σημαντικά. Όταν είναι $U^2 < U^2(\alpha)$, δεν μπορούμε να απορρίψουμε την ομοιότητα.

Υπό τη μηδενική υπόθεση της ισότητας των μέσων κατευθύνσεων των πλη-

θυσμών, είναι (Mardia and Jupp, 2000, p. 151)

$$E(U^2) = \frac{N + 1}{12N} \quad (3.35)$$

$$Var(U^2) = \frac{(n - 1)(m - 1)(N + 1)}{360Nnm} \quad (3.36)$$

Το Watson's U^2 test είναι χρήσιμο ειδικά για κυκλικά δεδομένα, διότι το αρχικό σημείο για τον προσδιορισμό των αθροιστικών συχνοτήτων δεν έχει σημασία, καθώς δεν συμπεριλαμβάνεται στους υπολογισμούς. Επίσης, τυπικά, μπορεί να χρησιμοποιηθεί σε κάθε περίπτωση γραμμικών δεδομένων, τα οποία πληρούν τις συνθήκες για την εφαρμογή του Mann-Witney test, αλλά, γενικά, δεν ενδείκνυται ως υποκατάστατο του Mann-Witney test, καθώς ο δεύτερος έλεγχος είναι ευκολότερος στην υλοποίησή του και διαθέτει πιο εκτεταμένους πίνακες κρίσιμων τιμών και, ακόμα, ο πρώτος έλεγχος μπορεί να υποδείξει σημαντικότητα, λόγω διαφορών στις διασπορές των ομάδων των παρατηρήσεων (Zar, 1999, p. 630).

Το Watson's U^2 test είναι αρκετά ισχυρό και μπορεί να εντοπίσει κάθε είδους απόκλιση μεταξύ των συγκρινόμενων δειγμάτων. Ο Beran (1969b) έχει δείξει ότι το τεστ είναι συνεπές έναντι κάθε εναλλακτικής. Τυχόν ισοβαθμίες πρέπει να εξαλειφθούν, για παράδειγμα με τυχαία ανακατανομή. Κατά συνέπεια, το τεστ αυτό δεν μπορεί να εφαρμοστεί σε μεγάλα δείγματα με ομαδοποιημένα δεδομένα (Batschelet, 1981, p. 115).

Η ιδέα της χρησιμοποίησης του αθροίσματος των τετραγωνικών αποκλίσεων μεταξύ συναρτήσεων κατανομών εντοπίζεται στους Cramér και Kolmogorov. Ο Watson (1962) τροποποίησε τον έλεγχο για την περίπτωση των κυκλικών δεδομένων. Το αποτέλεσμα του ελέγχου είναι ανεξάρτητο της επιλογής του αρχικού σημείου, δηλαδή το στατιστικό είναι αμετάβλητο υπό περιστροφή (Mardia and Jupp, 2000, p. 151).

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.18: Αποτελέσματα της εφαρμογής του ελέγχου Watson's U^2 test στις μεταβλητές δεύτερης τάξης WINDIR_DAY [1-17] και WINDIR_DAY [18-31].

<i>Variables (& observations)</i>	U^2	p	df
WINDIR_DAY[1-17] & WINDIR_DAY[18-31] (17 & 14)	0,068	> 0.5	14

Μεταβάλλεται η μέση ημερήσια κατεύθυνση του ανέμου μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας και των ημερών του καύσωνα; (Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών δεύτερης τάξης WINDIR_DAY[1-17] και WINDIR_DAY[18-31] με το Watson's U^2 test.)

Από τα στοιχεία του πίνακα (3.18), σελ. 94, παρατηρούμε ότι ο έλεγχος για ισότητα των μέσων κατευθύνσεων των πληθυσμών των μεταβλητών WINDIR_DAY[1-17] και WINDIR_DAY[18-31] με το Watson's U^2 test δεν δίνει σημαντικά αποτελέσματα. Δηλαδή, όπως έδειξε και ο παραμετρικός έλεγχος Mardia-Watson-Wheeler test της ενότητας (3.3.2) (βλ. πίνακα 3.15, σελ. 82], φαίνεται ότι οι μέσες ημερήσιες κατευθύνσεις του ανέμου για το διάστημα των πρώτων 17 ημερών (φυσιολογικές τιμές θερμοκρασίας) και των τελευταίων 14 ημερών (περίοδος του καύσωνα) του μήνα δεν διαφοροποιούνται.

Πίνακας 3.19: Αποτελέσματα της εφαρμογής του Watson's U^2 test για τον έλεγχο της ισονομίας των μεταβλητών WINDIR_D5, -21, -24, -30.

<i>Variables (& observations)</i>	U^2	p	df	$df2$
WINDIR_D5 & WINDIR_D21 (21 & 22)	0,259	< 0.02	21	22
WINDIR_D5 & WINDIR_D24 (21 & 22)	0,311	< 0.005	21	22
WINDIR_D5 & WINDIR_D30 (21 & 22)	0,469	< 0.001	21	22
WINDIR_D21 & WINDIR_D24 (22 & 22)	0,016	> 0.5	22	22
WINDIR_D21 & WINDIR_D30 (22 & 22)	0,055	> 0.5	22	22
WINDIR_D24 & WINDIR_D30 (22 & 22)	0,047	> 0.5	22	22

Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών WINDIR_D5, -21, -24, -30 με το Watson's U^2 test

Από τα στοιχεία του πίνακα 3.19, σελ. 95, παρατηρούμε ότι ο έλεγχος για την ισονομία των πληθυσμών για τα όλα τα ζεύγη της ομάδας των μεταβλητών WINDIR_D5, WINDIR_D21, WINDIR_D24 και WINDIR_D30 με το Watson's U^2 test, σε αντίθεση με τον έλεγχο Watson-Williams test της ενότητας (3.3.1) σελ. 67, δίνει πολύ σημαντικά αποτελέσματα κατά τον έλεγχο της μεταβλητής WINDIR_D5 με κάθε μία από τις υπόλοιπες και, κατά συνέπεια, η μηδενική υπόθεση της ισονομίας των πληθυσμών τους απορρίπτεται.

Ακόμα, ο έλεγχος φαίνεται να μην απορρίπτει την ισονομία των πληθυσμών για τις υπόλοιπες, πλην της WINDIR_D5, μεταβλητές (WINDIR_D21, WINDIR_D24 και WINDIR_D30).

Σε κάθε περίπτωση, επισημαίνουμε ότι η μορφή των δεδομένων, με τα σχετικά μεγάλα διαστήματα (22.5°) ανάμεσα σε κάθε κλάση τιμών, παραβιάζει τη συνθήκη του ελέγχου για μικρό πλάτος κλάσης.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.20: Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών WINDIR_DPREV [1-17] και WINDIR_DPREV [18-31] με το Watson's U^2 test.

<i>Variables (& observations)</i>	U^2	p	df	df^2
WINDIR_DPREV[1-17] WINDIR_DPREV[18-31] (17 & 14)	& 0,025	> 0.5	14	17

Υπάρχει διαφορά στην ημερήσια επικρατούσα κατεύθυνση μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας (1-17) και των ημερών της περιόδου του καύσωνα (18-31); (Έλεγχος για την ισονομία των πληθυσμών των δεύτερης τάξης μεταβλητών WINDIR_DPREV[1-17] και WINDIR_DPREV[18-31] με το Watson's U^2 test.)

Για τον έλεγχο της ισονομίας των αντίστοιχων πληθυσμών, δεν μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος Watson-Williams test της ενότητας (3.3.1), καθώς δεν φαίνεται να πληρείται καμία από τις συνθήκες τής σελ. 68.

Από τα αποτελέσματα του ελέγχου Watson's U^2 test, όπως παρουσιάζονται στον πίνακα 3.20, σελ. 96, φαίνεται ότι οι επικρατούσες κατευθύνσεις του ανέμου δεν διαφέρουν μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας και των ημερών της περιόδου του καύσωνα.

Πίνακας 3.21: Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών WINDIR_DAY και WINDIR_DPREV, για τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, για τις ημέρες του καύσωνα και για το σύνολο των ημερών του μήνα μελέτης με το Watson's U^2 test.

<i>Variables (& observations)</i>	U^2	p	df	$df2$
WINDIR_DAY & WINDIR_DPREV (31 & 31)	0,429	< 0.001	31	31
WINDIR_DAY[1-17] WINDIR_DPREV[1-17] (17 & 17)	& 0,214	< 0.05	17	17
WINDIR_DAY[18-31] WINDIR_DPREV[18-31] (14 & 14)	& 0,307	< 0.005	14	14

Υπάρχει διαφορά ανάμεσα στη μέση ημερήσια κατεύθυνση του ανέμου και στην ημερήσια επικρατούσα κατεύθυνση, για τις ημέρες με φυσιολογικές τιμές θερμοκρασίας (1–17), τις ημέρες του καύσωνα (18–31) και το σύνολο των ημερών του μήνα μελέτης; (Έλεγχος για την ισονομία των πληθυσμών των μεταβλητών των δεύτερης τάξης δειγμάτων των μεταβλητών WINDIR_DAY και WINDIR_DPREV, για κάθε ένα από τα παραπάνω διαστήματα του μήνα μελέτης με το Watson's U^2 test)

Στο πλαίσιο αυτό, γίνεται έλεγχος της ισονομίας των πληθυσμών μεταξύ μιας μεταβλητής πρώτης τάξης (ημερήσια επικρατούσα κατεύθυνση του ανέμου — WINDIR_DPREV) και μιας μεταβλητής δεύτερης τάξης (μέση ημερήσια κατεύθυνση του ανέμου — WINDIR_DAY). Η WINDIR_DPREV θεωρείται μεταβλητή πρώτης τάξης, καθώς από τη φύση του προβλήματος δεν μπορεί να έχει μέσο μέτρο μικρότερο της μονάδας.

Για τον έλεγχο της ισότητας των μέσων κατευθύνσεων των αντίστοιχων πληθυσμών, δεν μπορεί να εφαρμοστεί το Watson-Williams test της ενότητας (3.3.1), καθώς δεν φαίνεται να πληρείται καμία από τις συνθήκες της σελ. 68.

Από τα αποτελέσματα του ελέγχου Watson's U^2 test, όπως αυτά παρουσιάζονται στον πίνακα 3.21, σελ. 97, φαίνεται ότι οι επικρατούσες κατευθύνσεις του ανέμου διαφέρουν από τις αντίστοιχες μέσες ημερήσιες κατευθύνσεις (σημαντικά έως πολύ σημαντικά αποτελέσματα) σε κάθε περίπτωση, δηλαδή, είτε σε φυσιολογικές περιόδους, είτε σε περιόδους καύσωνα, είτε σε μικτές περιόδους.

3.3.6 rank-sum test

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n και m :

$$\phi_1, \phi_2, \dots, \phi_n \quad \psi_1, \psi_2, \dots, \psi_m$$

Θέλουμε να ελέγξουμε αν τα δύο αυτά δείγματα διαφέρουν σημαντικά μεταξύ τους. Κατά παραδοχή, τα δείγματα προέρχονται από πληθυσμούς με συνεχείς κατανομές και, ακόμα, στην περίπτωση ομαδοποιημένων δεδομένων, το πλάτος των κλάσεων είναι μικρό έτσι, ώστε να αποφεύγονται, κατά το δυνατό, οι ισοβαθμίες μεταξύ των δειγμάτων.

Όπως και στο τεστ των Wilcoxon, Mann και Whitney, συνδυάζουμε τα δύο δείγματα και βαθμονομούμε τις $N = n + m$ παρατηρήσεις. Χωρίς βλάβη της γενικότητας, υποθέτουμε ότι είναι $m \leq n$. Προκειμένου να κάνουμε το στατιστικό του ελέγχου ανεξάρτητο από την αυθαίρετη επιλογή της μηδενικής κατεύθυνσης, υιοθετούμε έναν ειδικό κανόνα: βαθμονομούμε τις N παρατηρήσεις με τέτοιο τρόπο, ώστε το άθροισμα των βαθμών των ψ_i (δηλαδή, των παρατηρήσεων του μικρότερου δείγματος) να είναι ελάχιστο. Με μια απλή διερεύνηση βρίσκεται η σωστή φορά της περιστροφής.

Αν S είναι το ελάχιστο άθροισμα των βαθμών των ψ_i , όπως περιγράφηκε παραπάνω, τότε το στατιστικό του ελέγχου είναι το

$$U = S - \frac{1}{2}m(m + 1) \quad (3.37)$$

Ο Batschelet (1981, p. 349) δίνει πίνακα με κρίσιμα επίπεδα, P -value, για διάφορες τιμές του στατιστικού (3.37), με περιορισμό, πάντως, σε μικρά δείγματα. Αν το κρίσιμο επίπεδο, $P(\alpha)$, του ελέγχου είναι μικρότερο από το προκαθορισμένο επίπεδο σημαντικότητας α , τότε προκύπτει σημαντικότητα των αποτελεσμάτων και η μηδενική υπόθεση της ισονομίας των πληθυσμών μπορεί να απορριφθεί. Αν είναι $P(\alpha) > \alpha$, δεν μπορούμε να απορρίψουμε την ομοιογένεια.

Το rank-sum test αποτελεί μια τροποποιημένη εκδοχή του Wilcoxon-Mann-Whitney test, ειδικά σχεδιασμένη για τον εντοπισμό διαφορών στη μέση κατεύθυνση, για μικρά δείγματα. Το κύριο πλεονέκτημά του είναι η ταχεία υλοποίηση. Πριν την εφαρμογή του τεστ, τυχόν ισοβαθμίες μεταξύ των δειγμάτων πρέπει να εξαλειφθούν, π.χ. με τυχαία ανακατανομή.

Ο συγκεκριμένος έλεγχος προτάθηκε από τον Batschelet (1965).

3.3.7 run test

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n και m :

$$\phi_1, \phi_2, \dots, \phi_n \quad \psi_1, \psi_2, \dots, \psi_m$$

Θέλουμε να ελέγξουμε αν τα δύο αυτά δείγματα διαφέρουν σημαντικά μεταξύ τους, κυρίως ως προς τη μέση κατεύθυνση. Κατά παραδοχή, τα δείγματα προέρχονται από πληθυσμούς με συνεχείς κατανομές και, ακόμα, στην περίπτωση ομαδοποιημένων δεδομένων, το πλάτος των κλάσεων είναι μικρό έτσι, ώστε να αποφεύγονται, κατά το δυνατό, οι ισοβαθμίες μεταξύ των δειγμάτων.

Ως *ροή* (run) νοείται μια ακολουθία διαδοχικών παρατηρήσεων (σημείων στον κύκλο). Έτσι, μια απλή ροή μπορεί να αποτελείται από ένα ή περισσότερα σημεία του ίδιου δείγματος. Στην περίπτωση δύο δειγμάτων πάνω στον κύκλο, ο αριθμός των ροών είναι πάντοτε περιττός. Αν ο αριθμός των ροών είναι μικρός, τα δείγματα είναι καλώς διαχωρισμένα και η σημαντικότητα των αποτελεσμάτων είναι πολύ πιθανή.

Ως εκ τούτου, το στατιστικό του ελέγχου είναι το

$$h = \text{ο συνολικός αριθμός των ροών} \quad (3.38)$$

Ο Batschelet (1981, p. 350) δίνει πίνακα με κρίσιμα επίπεδα, P -value, σχετικά με τις διάφορες τιμές του στατιστικού (3.38). Αν το κρίσιμο επίπεδο, $P(\alpha)$, του ελέγχου είναι μικρότερο από το προκαθορισμένο επίπεδο σημαντικότητας α , τότε προκύπτει σημαντικότητα των αποτελεσμάτων και η μηδενική υπόθεση της ισονομίας των πληθυσμών μπορεί να απορριφθεί. Αν είναι $P(\alpha) > \alpha$, δεν μπορούμε να απορρίψουμε την ομοιογένεια.

Το run test είναι ο ταχύτερος από τους ελέγχους δύο δειγμάτων. Σε περιπτώσεις ύπαρξης ισοβαθμιών, αυτές θα πρέπει να εξαλειφθούν, π.χ. με τυχαία ανακατανομή. Το συγκεκριμένο τεστ δεν ενδείκνυται για μεγάλα δείγματα, καθώς η ισχύς του είναι μικρή (Batschelet, 1981, p. 121).

Το run test των Wald και Wolfowitz για δείγματα γραμμικών κατανομών τροποποιήθηκε για τα κυκλικά δεδομένα από τους Barton and David (1958) και David and Barton (1962).

3.3.8 Ένα παραμετρικό τεστ για την παράμετρο συγκέντρωσης

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n και m :

$$\phi_1, \phi_2, \dots, \phi_n \quad \psi_1, \psi_2, \dots, \psi_m$$

Υπό την παραδοχή ότι τα δύο δείγματα προέρχονται από πληθυσμούς με κατανομή von Mises, θέλουμε να ελέγξουμε αν οι παράμετροι συγκέντρωσης κ_1 και κ_2 των δύο πληθυσμών διαφέρουν μεταξύ τους.

Η παράμετρος συγκέντρωσης κ ενός πληθυσμού με κατανομή von Mises αυξάνεται, καθ' όσον η γωνιακή διακύμανση μειώνεται. Γι' αυτό, για τον παραπάνω έλεγχο, μπορούμε να συγκρίνουμε τις γωνιακές διακυμάνσεις των δύο δειγμάτων.

Η γωνιακή διακύμανση ενός δείγματος μεγέθους n και μήκους μέσου διανύσματος r ορίζεται ως

$$s^2 = 2(1 - r)$$

ή, ισοδύναμα,

$$s^2 = 2 \frac{n - R}{n},$$

όπου $R = nr$ είναι το μήκος της συνισταμένης. Τότε, αν s_1^2 και s_2^2 είναι οι δύο δειγματικές διακυμάνσεις, για τον παραπάνω έλεγχο μπορεί να χρησιμοποιηθεί η ποσότητα s_1^2/s_2^2 ή οποιοσδήποτε μετασχηματισμός της.

Την παραπάνω παραδοχή της "κανονικότητας" των πληθυσμών (κατανομή von Mises) συμπληρώνει η απαίτηση

$$\bar{r} = \frac{R_1 + R_2}{n_1 + n_2} > 0.70,$$

όπου R_1 και R_2 τα μήκη των συνισταμένων των δύο δειγμάτων.

Το στατιστικό που χρησιμοποιείται είναι το

$$F = \frac{(n_2 - 1)(n_1 - R_1)}{(n_1 - 1)(n_2 - R_2)} \quad (3.39)$$

Υπό την ισχύ της μηδενικής υπόθεσης ότι οι παράμετροι συγκέντρωσης των δύο πληθυσμών είναι ίσες, δηλαδή $\kappa_1 = \kappa_2$, το στατιστικό (3.39) ακολουθεί προσεγγιστικά την F_{n_1-1, n_2-1} κατανομή. Αν α είναι το προεπιλεγέν επίπεδο σημαντικότητας του ελέγχου, τότε η H_0 απορρίπτεται, όταν η τιμή του στατιστικού υπερβεί την κρίσιμη τιμή $F_{n_1-1, n_2-1}(\alpha/2)$, η οποία δίνεται από τους

πίνακες των κρίσιμων F -τιμών. Τότε, μπορούμε να συμπεράνουμε ότι οι παράμετροι συγκέντρωσης των πληθυσμών διαφέρουν.

Το συγκεκριμένο τεστ είναι τυπικά παραμετρικό και μπορεί να εφαρμοστεί εύκολα με τη βοήθεια μιας υπολογιστικής μηχανής. Τυχόν ισοβαθμίες δεν δημιουργούν προβλήματα και, επίσης, είναι επιτρεπτή η ομαδοποίηση δεδομένων σε κλάσεις πλάτους έως 15° .

Ο έλεγχος αυτός προτάθηκε από τους Watson and Williams (1956) και τροποποιήθηκε από τον Mardia (1972).

3.3.9 Έλεγχος μεταβλητότητας γωνιακών αποκλίσεων

Ως γωνιακή απόκλιση ή απόσταση (angular distance) μεταξύ δύο σημείων στον κύκλο ορίζεται η μικρότερη γωνία που σχηματίζουν τα σημεία αυτά με το κέντρο του. Το μέτρο αυτό είναι χρήσιμο για το σχηματισμό συμπερασμάτων σχετικά με παρατηρούμενες αποκλίσεις από συγκεκριμένες κατευθύνσεις, π.χ. την πορεία μιας ομάδας ζώων ως προς μια προκαθορισμένη κατεύθυνση (πιθανώς, την κατεύθυνση επαναπατρισμού ή οσμής της λείας) ή την ώρα της ημέρας που λαμβάνει χώρα μια φυσιολογική ή συμπεριφορική δραστηριότητα μιας ομάδας υποκειμένων ως προς την ώρα ενός συγκεκριμένου ερεθίσματος. Οι παραπάνω περιπτώσεις (με μία ομάδα ζώων ή υποκειμένων) εξετάζονται με τους κατάλληλους ελέγχους της προηγούμενης ενότητας.

Στις περιπτώσεις όπου θέλουμε να εξετάσουμε το αν συμπίπτουν ή όχι οι μεταβλητότητες των γωνιακών αποστάσεων από αντίστοιχες προκαθορισμένες κατευθύνσεις (όχι απαραίτητα ίσες μεταξύ τους) παρατηρήσεων από δύο ή περισσότερα δείγματα (ομάδες), μπορεί να εφαρμοστούν οι αντίστοιχοι για κυκλικά δεδομένα έλεγχοι Mann-Whitney test (δύο δείγματα) και Kruskal-Wallis test (τρία ή περισσότερα δείγματα).

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n_1 και n_2 :

$$\phi_1, \phi_2, \dots, \phi_{n_1} \quad \psi_1, \psi_2, \dots, \psi_{n_2}$$

Θέλουμε να ελέγξουμε αν οι δύο πληθυσμοί, από τους οποίους προέρχονται τα δείγματα, παρουσιάζουν την ίδια μεταβλητότητα γωνιακών αποκλίσεων

$$d_{\phi_i - \theta_1}, i = 1, 2, \dots, n_1 \quad \text{και} \quad d_{\psi_j - \theta_2}, j = 1, 2, \dots, n_2,$$

ως προς δύο αντίστοιχες προκαθορισμένες κατευθύνσεις θ_1 και θ_2 .

Ακολουθώντας τη μεθοδολογία του ελέγχου Mann-Whitney test, με το σύνολο των παρατηρήσεων από τα δύο δείγματα, σχηματίζουμε το συνδυασμένο δείγμα και ταξινομούμε τις παρατηρήσεις ως προς αύξουσα ή φθίνουσα σειρά των γωνιακών τους αποκλίσεων από τις αντίστοιχες κατευθύνσεις και, στη συνέχεια, εφαρμόζουμε τον αμφίπλευρο⁵ (two-tailed) έλεγχο Mann-Whitney test.

Στην περίπτωση όπου έχουμε περισσότερα από δύο δείγματα, μετά το σχηματισμό του συνδυασμένου δείγματος, εφαρμόζουμε τον έλεγχο Kruskal-Wallis test.

Ο συγκεκριμένος έλεγχος της μεταβλητότητας των γωνιακών αποκλίσεων με τη χρήση του Mann-Whitney test προτάθηκε από τον Wallraff (1979).

3.3.10 Ένας απαραμετρικός έλεγχος της γωνιακής διασποράς

Η διαδικασία της ενότητας (3.3.9) που αφορά στην ανάλυση των γωνιακών αποκλίσεων δύο ή περισσότερων πληθυσμών από αντίστοιχες προκαθορισμένες κατευθύνσεις μπορεί να εφαρμοστεί και για τον έλεγχο της ισότητας των γωνιακών τους διασπορών, δηλαδή των γωνιακών τους αποκλίσεων από τις μέσες κατευθύνσεις τους.

Έστω δύο ανεξάρτητα τυχαία δείγματα κυκλικών παρατηρήσεων, μεγέθους n_1 και n_2 :

$$\phi_1, \phi_2, \dots, \phi_{n_1} \quad \psi_1, \psi_2, \dots, \psi_{n_2}$$

Θέλουμε να ελέγξουμε αν οι γωνιακές διασπορές των πληθυσμών από τους οποίους προέρχονται τα δείγματα διαφέρουν μεταξύ τους.

Οι δειγματικές γωνιακές διασπορές προσδιορίζονται από τις γωνιακές αποκλίσεις ως προς τις αντίστοιχες μέσες γωνίες, δηλαδή τις

$$d_{\phi_i - \bar{\phi}}, i = 1, 2, \dots, n_1 \quad \text{και} \quad d_{\psi_j - \bar{\psi}}, j = 1, \dots, n_2.$$

Οι γωνιακές αυτές αποκλίσεις συνδυαζόμενες σχηματίζουν το από κοινού δείγμα, το οποίο ταξινομείται ως προς το μέγεθός τους και, κατόπιν, εφαρμόζεται ο αμφίπλευρος⁶ έλεγχος Mann-Whitney test.

Οι ισοβαθμίες μεταξύ των δειγμάτων γενικά προκαλούν προβλήματα στην

⁵Η το μονόπλευρο (one-tailed) έλεγχο, όταν ενδιαφερόμαστε να διαπιστώσουμε το αν η γωνιακή απόκλιση του ενός πληθυσμού από μια προκαθορισμένη κατεύθυνση είναι μεγαλύτερη από τη γωνιακή απόκλιση του άλλου πληθυσμού από μια, αντίστοιχη, κατεύθυνση.

⁶Η ο μονόπλευρος έλεγχος, όταν ενδιαφερόμαστε να διαπιστώσουμε το αν η γωνιακή διασπορά του ενός πληθυσμού είναι μεγαλύτερη από τη γωνιακή διασπορά του άλλου.

εφαρμογή αυτού του τεστ, ειδικά στην περίπτωση των ομαδοποιημένων δεδομένων. Οι ισοβαθμίες πρέπει να εξαλείφονται, π.χ. με τυχαία ανακατανομή.

Αν θέλουμε να συγκρίνουμε τις γωνιακές διασπορές περισσότερων από δύο δειγμάτων, μπορούμε να εφαρμόσουμε την επέκταση του Mann-Whitney test για τρία ή περισσότερα δείγματα, δηλαδή τον έλεγχο Kruskal-Wallis test.

Παράδειγμα Έστω μια ομάδα δεκαπέντε πουλιών ενός συγκεκριμένου είδους, τα οποία μεταφέρθηκαν σε μια περιοχή νότια της φωλιάς τους και, κατόπιν, αφέθηκαν ελεύθερα, τα $n_1 = 8$ με ηλιοφάνεια και τα υπόλοιπα $n_2 = 7$ με συννεφιά. Στον πίνακα 3.23 δίνεται για κάθε πουλί η κατεύθυνση επαναπαυρισμού, σε μοίρες από το Βορρά και κατά τη φορά των δεικτών του ωρολογίου. Θέλουμε να εξετάσουμε το αν η μεταβλητότητα της κατεύθυνσης επαναπαυρισμού είναι ίδια και στις δύο περιπτώσεις, καθώς και το αν τα πουλιά προσανατολίζονται καλύτερα προς τη φωλιά τους (βόρεια, 0°) με ηλιοφάνεια ή με συννεφιά.

Πίνακας 3.23: Κατευθύνσεις σε μοίρες από το Βορρά και κατά τη φορά των δεικτών του ωρολογίου ομάδας πουλιών ενός συγκεκριμένου είδους.

Κατεύθυνση πορείας με ηλιοφάνεια	Κατεύθυνση πορείας με συννεφιά
350	340
340	305
315	255
10	270
20	305
355	320
345	335
360	

Οι μέσες γωνίες των δύο δειγμάτων είναι $\bar{\theta}_1 = 352^\circ$ και $\bar{\theta}_2 = 305^\circ$. Υπολογίζουμε τις γωνιακές αποστάσεις των παρατηρήσεων κάθε δείγματος από την αντίστοιχη δειγματική μέση κατεύθυνση και εφαρμόζουμε αμφίπλευρο έλεγχο Mann-Whitney test για τη μηδενική υπόθεση

H_0 : Η μεταβλητότητα της κατεύθυνσης της πορείας των πουλιών είναι ίδια με ηλιοφάνεια και με συννεφιά

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

έναντι της εναλλακτικής υπόθεσης

H_1 : Η μεταβλητότητα της κατεύθυνσης της πορείας των πουλιών δεν είναι ίδια με ηλιοφάνεια και με συννεφιά

Πίνακας 3.24: Κατευθύνσεις σε μοίρες, γωνιακές αποστάσεις (από τις αντίστοιχες μέσες γωνίες, σε μοίρες) και βαθμοί.

Με ηλιοφάνεια			Με συννεφιά		
Κατεύθυνση	Γωνιακή απόσταση	Τάξη	Κατεύθυνση	Γωνιακή απόσταση	Τάξη
θ_{1i}	$d_{352^\circ - \theta_{1i}}$		θ_{2j}	$d_{305^\circ - \theta_{2j}}$	
350	2	3	340	35	12,5
340	12	7	305	0	1,5
315	37	14	255	50	15
10	18	9	270	35	12,5
20	28	10	305	0	1,5
355	3	4	320	15	8
345	7	5	335	30	11
360	8	6			
		58			62

Στον πίνακα 3.24 αναγράφεται για κάθε πουλί η κατεύθυνση σε μοίρες, η γωνιακή απόσταση από την αντίστοιχη μέση γωνία σε μοίρες και ο βαθμός (τάξη) της. Επίσης, υπολογίζεται το σύνολο των βαθμών κατά περίπτωση. Τότε, είναι

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1 = 34$$

$$U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - S_2 = 22$$

και

$$U_{0.05(2),8,7} = U_{0.05(2),7,8} = 46 > U = 34, U' = 22$$

οπότε, η H_0 σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται.

Ακόμα, κάνουμε τον έλεγχο της μηδενικής υπόθεσης

H_0 : Τα πουλιά προσανατολίζονται προς τη φωλιά τους το ίδιο με ηλιοφάνεια και με συννεφιά

ως προς την εναλλακτική υπόθεση

H_1 : Τα πουλιά προσανατολίζονται προς τη φωλιά τους καλύτερα με ηλιοφάνεια απ' ό,τι με συννεφιά

Πίνακας 3.25: Κατευθύνσεις σε μοίρες, γωνιακές αποστάσεις (από τη βόρεια κατεύθυνση, σε μοίρες) και βαθμοί.

Με ηλιοφάνεια			Με συννεφιά		
Κατεύθυνση	Γωνιακή απόσταση	Τάξη	Κατεύθυνση	Γωνιακή απόσταση	Τάξη
θ_{1i}	$d_{360^\circ - \theta_{1i}}$		θ_{2j}	$d_{360^\circ - \theta_{2j}}$	
350	10	3.5	340	20	7
340	20	7	305	55	12,5
315	45	11	255	105	15
10	10	3.5	270	90	14
20	20	7	305	55	12,5
355	5	2	320	40	10
345	15	5	335	25	9
360	0	1			
40			80		

Στον πίνακα 3.25 αναγράφεται για κάθε πουλί η κατεύθυνση σε μοίρες, η γωνιακή απόσταση από τη βόρεια κατεύθυνση σε μοίρες, και ο βαθμός της. Εφαρμόζοντας το μονόπλευρο έλεγχο Mann-Whitney test παίρνουμε

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1 = 52 > U_{0.05(1),8,7} = U_{0.05(1),7,8} = 43$$

οπότε, η H_0 σε επίπεδο σημαντικότητας 5% απορρίπτεται.

ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Πίνακας 3.22: Έλεγχοι δύο δειγμάτων για την ισονομία των πληθυσμών.

Έλεγχος	Είδος ελέγχου	Παραδοχές	Στατιστικό	Κανόνας απόρριψης	Παρατηρήσεις
Watson-Williams test	Παραμετρικός	Πληθυσμοί με κατανομή von Mises. $k_1 = k_2 = k, k > 2$	$F = g(n - 2) \frac{R_1 + R_2 - R}{n - (R_1 + R_2)},$ R_1, R_2, R συνισταμένες	$F > F_{\alpha, 1, n-2}$	Έλεγχος ομοιογένειας (ισότητα όλων των μέσων καταυθύνσεων. Πολύ ισχυρός έλεγχος.
Mardia-Watson-Wheeler test	Μη παραμετρικός	Πληθυσμοί με συνεχείς κυκλικές κατανομές. Όχι ισοτιμίες στα δείγματα.	$W = 2 \left(\frac{C_1^2 + S_1^2}{n_1} + \frac{C_2^2 + S_2^2}{n_2} \right),$ C_2^2, S_2 συνιστώσες του συνισταμένου διανύσματος	$\chi^2 > \chi^2(\alpha)$	$n_1, n_2 \geq 10$
Chi-squared test	Μη παραμετρικός	Οι αναμενόμενες συχνότητες $e_{ij} \geq 5$	$\chi^2 = \sum^i \sum^j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$	$P < \alpha$	Αρκετά ισχυρός.
Kuiper's test	Μη παραμετρικός	Πληθυσμοί με συνεχείς κατανομές. Σε ομαδοποιημένα δεδομένα, το διάστημα της κλάσης δεν είναι μεγαλύτερο από 5° .	$V_{n,m} = D^+ - D^-, k = \frac{nm}{mn} V_{n,m}, D^+ / D^-$ η μέγιστη /ελάχιστη παρατηρούμενη διαφορά $F_2(\psi) - F_1(\phi)$	$P < \alpha$	Πολύ ισχυρός στον εντοπισμό διαφορών που οφείλονται σε κάθε είδος απόκλισης.
Watson's U^2 test	Μη παραμετρικός	Πληθυσμοί με συνεχείς κατανομές. Σε ομαδοποιημένα δεδομένα το πλάτος των κλάσεων δεν ξεπερνά τις 5° .	$U^2 = \frac{nm}{N^2} \times \left[\sum_{k=1}^N d_k^2 - \frac{1}{N} \left(\sum_{k=1}^N d_k \right)^2 \right]$	$P < \alpha$	Αρκετά ισχυρός για τον εντοπισμό κάθε είδους απόκλισης.
Rank-sum test	Μη παραμετρικός	Πληθυσμοί με συνεχείς κατανομές. Σε ομαδοποιημένα δεδομένα, μικρό πλάτος των κλάσεων (όχι ισοβαθμίες)	$S = \text{το ελάχιστο άθροισμα βαθμών}, U = S - \frac{1}{2}m(m+1)$	$P < \alpha$	-
Run test	Μη παραμετρικός	Πληθυσμοί με συνεχείς κατανομές. Σε ομαδοποιημένα δεδομένα, μικρό πλάτος των κλάσεων (όχι ισοβαθμίες)	$h = \text{ο συνολικός αριθμός ροών}$	$P < \alpha$	-

Πίνακας 3.26: Άλλοι έλεγχοι δύο ή περισσότερων δειγμάτων.

Έλεγχος	Είδος ελέγχου	Μηδενική υπόθεση	Παραδοχές	Στατιστικό	Κανόνας απόρριψης	Παρατηρήσεις
Watson-Williams (Mardia) test	Παραμετρικός	$H_0 : \kappa_1 = \kappa_2$	Πληθυσμοί με κανονική von Mises. $\bar{r} = \frac{R_1 + R_2}{n_1 + n_2} > 0.70$	$F = \frac{(n_2 - 1)(n_1 - R_1)}{(n_1 - 1)(n_2 - R_2)}$	F $F_{n_1 - 1, n_2 - 1, (\alpha/2)}$	> -
Mann-Whitney test (1)	Μη παραμετρικός	$H_0 : -$ $Var(d_{\phi_i - \theta_1}) = Var(d_{\psi_j - \theta_2})$		$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1 U'$ $S_1 U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - S_2$ S_2, S_1, S_2 τα αδροίσματα των βαδμών	U, U' $U_{\alpha(2), n_1, n_2}$	> Για τρία ή περισσότερα δείγματα, εφαρμόζεται το Kruskal-Wallis test
Mann-Whitney test (2)	Μη παραμετρικός	$H_0 : -$ $Var(d_{\phi_i - \bar{\phi}}) = Var(d_{\psi_j - \bar{\psi}})$		$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1 U'$ $S_1 U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - S_2$ S_2, S_1, S_2 τα αδροίσματα των βαδμών	U, U' $U_{\alpha(2), n_1, n_2}$	> Για τρία ή περισσότερα δείγματα, εφαρμόζεται το Kruskal-Wallis test

*ΚΕΦΑΛΑΙΟ 3. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΙ
ΥΠΟΘΕΣΕΩΝ*

Κεφάλαιο 4

Ανάλυση δεύτερης τάξης

4.1 Εισαγωγή

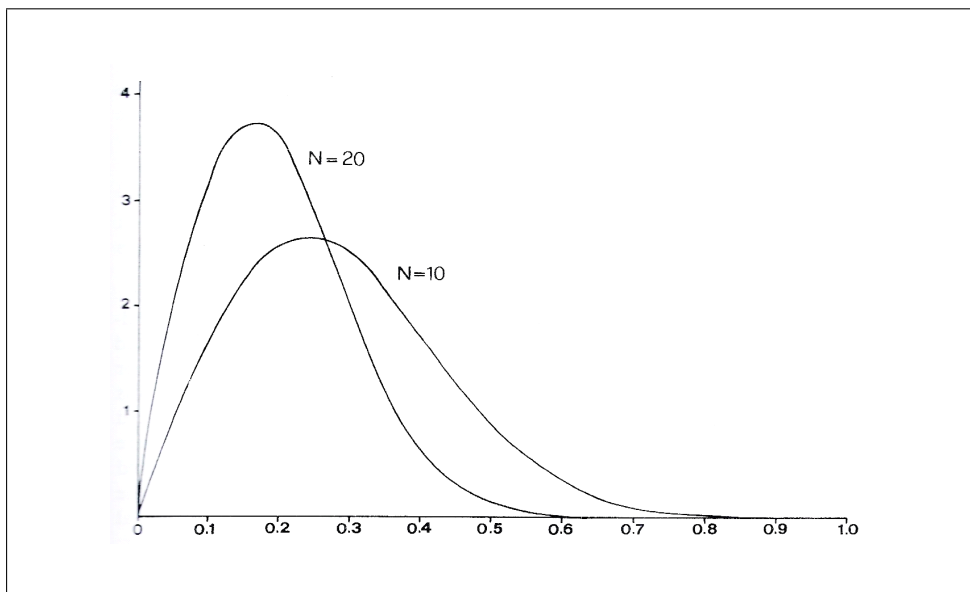
Είναι κοινή στη στατιστική ανάλυση η κατ' επανάληψη παρατήρηση και μέτρηση των υπό μελέτη χαρακτηριστικών του ίδιου υποκειμένου. Στην κατηγορία των δεδομένων κατεύθυνσης, μπορεί, για παράδειγμα να πάρουμε δεκάδες μετρήσεις για τον ίδιο φυσικό οργανισμό. Τότε, από τα δεδομένα αυτά θα προσδιοριστεί ένα μέσο διάνυσμα, το οποίο θα περιέχει πληροφορία για τη μέση γωνία και τη μέση γωνιακή απόκλιση. Αφού οι μετρήσεις γίνονται πάνω στο ίδιο υποκείμενο και *σειριακά*, θεωρούνται ως εξαρτημένες μεταξύ τους. Ως εκ τούτου, δεν μπορεί να εφαρμοστεί κάποιο τεστ σημαντικότητας (ως προς τη συγκέντρωση ή τη μέση κατεύθυνση). Γενικά, κατά τις περιπτώσεις όπου έχουμε επαναλαμβανόμενες μετρήσεις πάνω στο ίδιο υποκείμενο, από τα παρατηρούμενα μέτρα στο υποκείμενο αυτό δεν μπορούμε να σχηματίζουμε στατιστικές επαγωγές για τη συμπεριφορά των υπολοίπων υποκειμένων του πληθυσμού. Η στατιστική ανάλυση τέτοιων παρατηρήσεων αφορά στην ελάττωση των δεδομένων και υπηρετεί μόνο περιγραφικούς σκοπούς (Batschelet, 1981, p. 197).

Αν θέλουμε, σ' αυτές τις περιπτώσεις, να εξαγάγουμε κάποια στατιστικά συμπεράσματα για τον υπό μελέτη πληθυσμό, είμαστε υποχρεωμένοι να εξετάσουμε μια *ομάδα* υποκειμένων. Τότε, η στατιστική ανάλυση¹ πραγματοποιείται σε δύο στάδια:

¹ Στη βιβλιογραφία, αυτό το είδος ανάλυσης αναφέρεται ως *πρώτου* και *δεύτερου σταδίου ανάλυση*. Για την περίπτωση των κυκλικών δεδομένων, ο Wallraff (1970) πρότεινε τους όρους *πρώτης* και *δεύτερης τάξης ανάλυση* (first order/second order analysis).

ΚΕΦΑΛΑΙΟ 4. ΑΝΑΛΥΣΗ ΔΕΥΤΕΡΗΣ ΤΑΞΗΣ

1. Για κάθε υποκείμενο της ομάδας, ελαττώνουμε τον όγκο των δεδομένων, υπολογίζοντας τα κατάλληλα στατιστικά μέτρα.
2. Χρησιμοποιούμε τα στατιστικά του σταδίου (1) και ελέγχουμε τη σημαντικότητά τους. Μόνον τότε μπορούμε να σχηματίσουμε στατιστικές επαγωγές για τον πληθυσμό στον οποίο ανήκει η ομάδα των υποκειμένων.



Σχήμα 4.1: Κατανομές συχνοτήτων του μέσου μέτρου r για δείγματα μεγέθους $N = 10$ και $N = 20$, υπό την παραδοχή της ομοιόμορφης κατανομής των κατευθύνσεων. Παρ' όλο που το θεωρητικό μήκος του μέσου διανύσματος είναι μηδέν, το δειγματικό μήκος μπορεί να παίρνει τιμές στο εύρος από 0 έως 1. Όπως φαίνεται στο σχήμα, για μέγεθος δείγματος $N = 10$, το r συνήθως παίρνει τιμές από 0.1 έως 0.4, με κορυφή την τιμή 0.23, ενώ για $N = 20$, το r είναι γενικά μικρότερο (κορυφή, η τιμή 0.17). Επομένως, αυξανόμενου του μεγέθους του δείγματος, το r γενικά μειώνεται.

Αναλυτικότερα, αν για μια ομάδα k υποκειμένων, προσδιορίσουμε, με ανάλυση πρώτης τάξης, για το σύνολο των παρατηρήσεων n_j σε κάθε υποκείμενο j , $j = 1, 2, \dots, k$ το μήκος του μέσου διανύσματος r_j και τη μέση γωνία $\bar{\phi}_j$, τότε καλούμε το σύνολο των διατεταγμένων ζευγών

$$(r_j, \bar{\phi}_j), \quad j = 1, 2, \dots, k$$

ως *δεύτερης τάξης δείγμα* (second-order sample). Υποθέτουμε ότι τα ζεύγη $(r_j, \bar{\phi}_j)$ είναι *αμοιβαία ανεξάρτητα*.

Έτσι, ένα δεύτερης τάξης δείγμα $(r_j, \bar{\phi}_j)$ θεωρείται ως ένα σύνολο αμοιβαία ανεξάρτητων μέσων διανυσμάτων \mathbf{m}_j , $j = 1, 2, \dots, k$. Τα διανύσματα

αυτά μπορεί να αναπαρίστανται στο επίπεδο είτε δια των πολικών τους συντεταγμένων r_j και $\bar{\phi}_j$, είτε δια των καρτεσιανών τους συντεταγμένων \bar{x}_j και \bar{y}_j , οι οποίες συνδέονται μεταξύ τους μέσω των σχέσεων

$$\begin{aligned}\bar{x}_j &= r_j \cos \bar{\phi}_j = \frac{\sum_{i=1}^{n_j} \cos \phi_i}{n_j} \\ \bar{y}_j &= r_j \sin \bar{\phi}_j = \frac{\sum_{i=1}^{n_j} \sin \phi_i}{n_j}\end{aligned}\quad (4.1)$$

Προκειμένου να υλοποιήσουμε μια ανάλυση δεύτερης τάξης, ζητούμε να πληρείται η συνθήκη τού να έχουν τα ζεύγη $(r_j, \bar{\phi}_j)$ το ίδιο βάρος, δηλαδή κάθε υποκείμενο της ομάδας να φέρει τον ίδιο αριθμό παρατηρήσεων $n_1 = n_2 = \dots = n_k = n$. Ο κύριος λόγος για την απαίτηση αυτή είναι ότι τα μέσα μέτρα r_j , όχι μόνο υπόκεινται στην τυχαία διακύμανση από υποκείμενο σε υποκείμενο, αλλά και εξαρτώνται από το μέγεθος του δείγματος. Το τελευταίο φαίνεται διαγραμματικά στο σχήμα 4.1.

Μικρές αποκλίσεις από την απαίτηση της ισότητας του πλήθους των παρατηρήσεων για κάθε υποκείμενο είναι ανεκτές. Μεγάλες, όμως, αποκλίσεις από την απαίτηση αυτή περιπλέκουν σημαντικά τη στατιστική ανάλυση (Batschelet, 1981, p. 198).

4.2 Γενικός μέσος k μέσων γωνιών

Έστω ότι έχουμε k ομάδες κυκλικών παρατηρήσεων, με n_j , $j = 1, 2, \dots, k$ παρατηρήσεις για κάθε ομάδα (η οποία μπορεί να λογίζεται ως δείγμα για την ανάλυση αυτή) και προσδιορίζουμε ένα μέσο για κάθε ομάδα, οπότε προκύπτει ένα σύνολο k μέσων γωνιών. Αν θέλουμε να υπολογίσουμε το γενικό μέσο αυτών των επιμέρους μέσων γωνιών, τότε οι μέθοδοι που περιγράφηκαν στα πρώτα κεφάλαια δεν είναι κατάλληλες. Δηλαδή, δεν είναι σωστό το να θεωρήσουμε κάθε έναν από τους δειγματικούς μέσους ως μια γωνία, καθώς κάτι τέτοιο θα σήμαινε ότι υποθέτουμε μέσο μέτρο r ίσο με τη μονάδα, για κάθε έναν από αυτούς τους n μέσους. Κατά συνέπεια, υποθέτουμε μηδενική γωνιακή απόκλιση s για κάθε ένα από τα k δείγματα, πράγμα εξαιρετικά απίθανο να συμβεί (Zar, 1999, pp. 608–609).

Η γενική μέση γωνία $\bar{\phi}$ αυτών των k δειγμάτων, η οποία καλείται *μέση γωνία δεύτερης τάξης*, μπορεί να υπολογιστεί κατάλληλα σύμφωνα με την παρακάτω διαδικασία, την οποία υπέδειξε ο Batschelet (1981, ch. 7, 10). Ο γενικός

αυτός μέσος $\bar{\phi}$ έχει καρτεσιανές συντεταγμένες

$$\begin{aligned}\bar{x} &= \frac{\sum_{j=1}^k \bar{x}_j}{k} \\ \bar{y} &= \frac{\sum_{j=1}^k \bar{y}_j}{k}\end{aligned}\quad (4.2)$$

όπου οι ποσότητες \bar{x}_j και \bar{y}_j είναι οι καρτεσιανές συντεταγμένες των μέσων γωνιών του j δείγματος, $j = 1, 2, \dots, k$, από τις εξ. (4.1), σελ. 111.

Χρησιμοποιώντας τις πολικές, αντί των καρτεσιανών, συντεταγμένες κάθε μέσης γωνίας, παίρνουμε μια ισοδύναμη της (4.2) έκφραση των καρτεσιανών συντεταγμένων του γενικού μέσου $\bar{\phi}$

$$\begin{aligned}\bar{x} &= \frac{\sum_{j=1}^k r_j \cos \bar{\phi}_j}{k} \\ \bar{y} &= \frac{\sum_{j=1}^k r_j \sin \bar{\phi}_j}{k}\end{aligned}\quad (4.3)$$

όπου με $\bar{\phi}_j$ συμβολίζεται η μέση γωνία της j ομάδας και με r_j το αντίστοιχο μέσο μέτρο. Επομένως, η κατανομή του γενικού μέσου ενός δεύτερης τάξης δείγματος κυκλικών δεδομένων μπορεί να θεωρηθεί ως μια διμεταβλητή κατανομή των καρτεσιανών του συντεταγμένων \bar{x} και \bar{y} . Για την παρακάτω ανάλυση, υποθέτουμε ότι τα \bar{x} και \bar{y} κατανέμονται σύμφωνα με τη διμεταβλητή κανονική κατανομή και, ακόμα, ότι όλα τα δειγματικά μεγέθη n_j (πλήθος των παρατηρήσεων της κάθε ομάδας) είναι ίσα. Η συνθήκη της διμεταβλητής κανονικότητας είναι πολύ βασική,² ενώ μικρές αποκλίσεις από την παραδοχή της ισότητας των δειγματικών μεγεθών δεν επηρεάζουν σοβαρά τα αποτελέσματα της ανάλυσης (Zar, 1999, p. 610).

Από τον πίνακα 4.1, παρατηρούμε ότι, ενώ κατά την ανάλυση πρώτης τάξης οι καρτεσιανές συντεταγμένες των μέσων γωνιών υπολογίζονται *τριγωνομετρικά* (από τα μέσα αθροίσματα των τριγωνομετρικών αριθμών των γωνιών), κατά την ανάλυση δεύτερης τάξης οι αντίστοιχες συντεταγμένες του γενικού μέσου προκύπτουν ως *οι αριθμητικοί μέσοι* των συντεταγμένων των μέσων της πρώτης τάξης.

²Η παραδοχή της διμεταβλητής κανονικότητας, εκτός από βασικό περιορισμό για το συγκεκριμένο είδος ανάλυσης, αποτελεί και επαγωγικό εργαλείο, καθώς η μη συσχέτιση των μεταβλητών μιας διμεταβλητής κανονικής κατανομής *ισοδυναμεί* με την ανεξαρτησία τους.

Ανάλυση 1ης τάξης	Ανάλυση 2ης τάξης
$\bar{x} = \frac{\sum^i \cos \phi_i}{n}$	$\bar{\bar{x}} = \frac{\sum^j \bar{x}_j}{k}$
$\bar{y} = \frac{\sum^i \sin \phi_i}{n}$	$\bar{\bar{y}} = \frac{\sum^j \bar{y}_j}{k}$

Πίνακας 4.1: Σύγκριση του υπολογισμού των μέσων κατά την ανάλυση πρώτης και δεύτερης τάξης.

Υπολογισμός της μέσης γωνίας της μεταβλητής δεύτερης τάξης WINDIR_DAY (μέση ημερήσια κατεύθυνση του ανέμου)

Η μεταβλητή WINDIR_DAY εκφράζει την κατεύθυνση του ανέμου κατά το μέσο όρο ημέρας. Δηλαδή, η WINDIR_DAY παίρνει ως τιμές τις μέσες κατευθύνσεις των 31 ημερών του μήνα μελέτης, οπότε περιγράφει ένα χαρακτηριστικό ενός δεύτερης τάξης δείγματος. Συνεπώς, για τον υπολογισμό της μέσης γωνίας της μεταβλητής WINDIR_DAY πρέπει να εφαρμοστεί η αλγεβρική διαδικασία που περιγράφηκε παραπάνω.

Τα αποτελέσματα της επεξεργασίας των μεταβλητών WINDIR_Dxx με το στατιστικό πρόγραμμα *Oriana* δίνουν τις τιμές των μέσων γωνιών και των μέσων μέτρων που χαρακτηρίζουν το δεύτερης τάξης δείγμα των (WINDIR_DAY). Από τις τιμές αυτών των δύο χαρακτηριστικών μπορούμε να υπολογίσουμε τις ποσότητες των σχέσεων 4.1, σελ. 111, για $j = 1, 2, \dots, 31$. Ακόμα, από το άθροισμα των \bar{x}_j, \bar{y}_j μπορούμε να υπολογίσουμε τις ποσότητες των σχέσεων 4.3, σελ. 112. Τότε, είναι

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{31} r_j \cos \bar{\phi}_j}{31} = \frac{-12.69}{31} = -0.4093$$

$$\bar{\bar{y}} = \frac{\sum_{j=1}^{31} r_j \sin \bar{\phi}_j}{31} = \frac{-5.76}{31} = -0.1856$$

Αν συμβολίσουμε με ϕ το γενικό μέσο και με r το μέσο μέτρο του, είναι

$$r = \sqrt{\bar{x}^2 + \bar{y}^2} = 0.4495$$

οπότε είναι

$$\cos \phi = \frac{\bar{x}}{r} = -0.9106$$

$$\sin \phi = \frac{\bar{y}}{r} = -0.4131$$

Τότε, η εφαπτομένη της γωνίας ϕ είναι

$$\tan \phi = 0.4536$$

και η ϕ ισούται με

$$\arctan \phi = 0.4258 \text{ ακτίνια ή } 24.4000 \text{ μοίρες}$$

Επομένως, ο ζητούμενος γενικός μέσος ισούται με

$$24.4000 + 180 = 204.40000 \text{ μοίρες}$$

Το αποτέλεσμα αυτό διαφέρει από το αντίστοιχο της στατιστικής επεξεργασίας με το πρόγραμμα *Oriana* (192.6720 μοίρες) κατά 9.7280 μοίρες, δίνοντας μια δυτικότερη κατεύθυνση. Η διαφορά αυτή εκφράζει απώλεια πληροφορίας, καθώς το πρόγραμμα *Oriana* χειρίζεται το δείγμα των μέσων ημερήσιων κατευθύνσεων WINDIR_DAY ως πρώτης τάξης, δηλαδή αποτελούμενο από κατευθύνσεις (με μέσο μέτρο ίσο με τη μονάδα) και όχι διανύσματα (μέσου μέτρου μικρότερου ή ίσου της μονάδας), όπως είναι στην πραγματικότητα.

4.3 Όρια εμπιστοσύνης για τη μέση γωνία και το μέσο μέτρο δεύτερης τάξης

Η μέση γωνία που υπολογίζεται με τη διαδικασία της προηγούμενης ενότητας (4.2) αποτελεί μια εκτίμηση για την αντίστοιχη παράμετρο του πληθυ-

4.3. Όρια εμπιστοσύνης για τη μέση γωνία και το μέσο μέτρο δεύτερης τάξης

σμού, δηλαδή τη μέση κατεύθυνση, έστω θ_1 . Κατά την ανάλυση πρώτης τάξης, τα όρια εμπιστοσύνης $\bar{\phi} \pm d$ (όπου $\bar{\phi}$ η δειγματική μέση γωνία) για τη μέση κατεύθυνση θ_1 δίνονται με τον υπολογισμό της ποσότητας d από τις εξ. (3.1) και (3.2), σελ. 36, για καθορισμένο επίπεδο εμπιστοσύνης α . Για την ανάλυση δεύτερης τάξης, μπορούμε να υπολογίσουμε τα όρια εμπιστοσύνης για τη μέση κατεύθυνση του πληθυσμού, αφού πρώτα συμπεράνουμε, κάνοντας κάποιο σχετικό έλεγχο, ότι υπάρχει στατιστικά σημαντική κατεύθυνση (προσανατολισμός) στα δεδομένα.

Ο Batschelet (1981, p. 262–265) περιγράφει, γεωμετρικά και αναλυτικά, έναν τρόπο με τον οποίο προκύπτουν τα όρια εμπιστοσύνης για την ανάλυση δεύτερης τάξης.

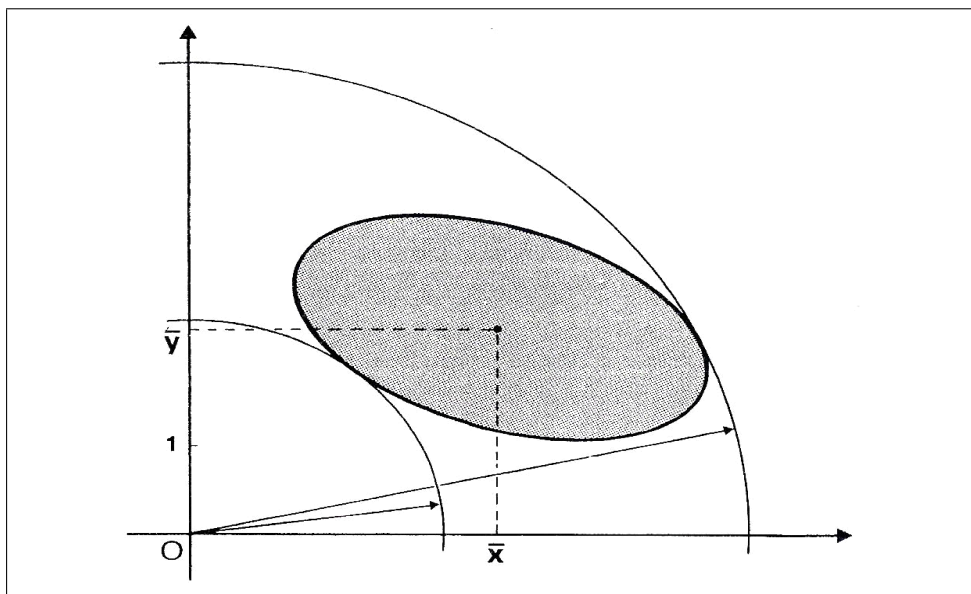
Γεωμετρική περιγραφή Έστω ένα τυχαίο δείγμα δεύτερης τάξης για το οποίο υποθέτουμε ότι ο πληθυσμός από τον οποίο προέρχεται είναι κανονικός, με μέσο διάνυσμα μ_1 . Αν είναι

$$\rho_1 = |\mu_1| > 0$$

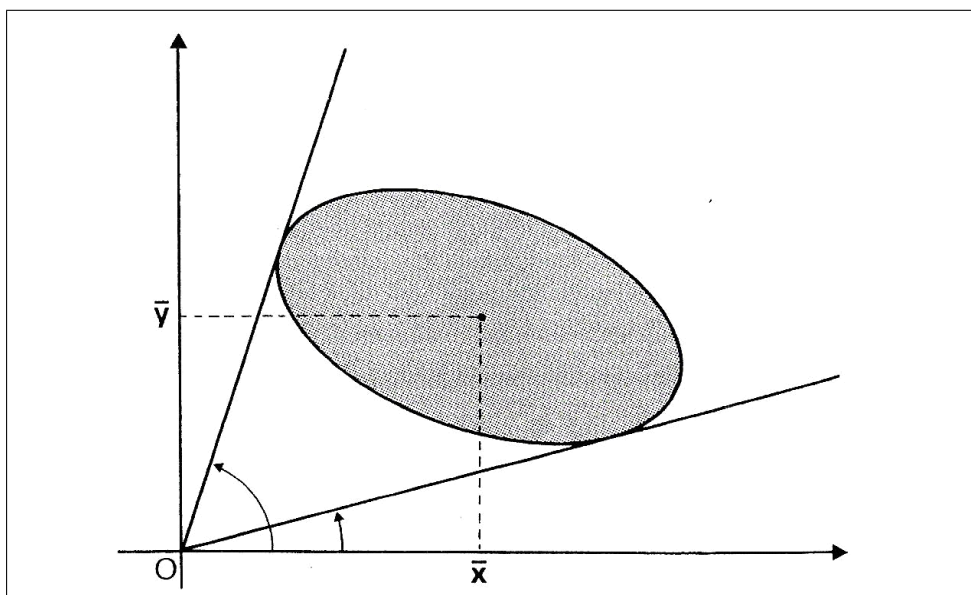
μπορούμε να εισάγουμε την πολική γωνία θ_1 του μ_1 . Για τον υπολογισμό ορίων εμπιστοσύνης για τα ρ_1 και θ_1 , προσδιορίζουμε, σε ένα ορθοκανονικό σύστημα συντεταγμένων xOy την *ελλειπτική περιοχή εμπιστοσύνης*³ για τον προκαθορισμένο συντελεστή εμπιστοσύνης $Q = 1 - \alpha$. Λαμβάνουμε υπ' όψη μόνο την περίπτωση κατά την οποία η έλλειψη δεν περιλαμβάνει το κέντρο O των αξόνων (σημαντικά αποτελέσματα).

Για τον προσδιορισμό των ορίων εμπιστοσύνης για το μήκος του μέσου διανύσματος ρ_1 του πληθυσμού, διαγράφουμε τους εφαπτόμενους στη έλλειψη κύκλους με κέντρο το O . Η άγνωστη παράμετρος ρ_1 βρίσκεται μέσα στο διάστημα μεταξύ των δύο ακτίνων (βλ. σχήμα 4.2).

³Η ελλειπτική περιοχή εμπιστοσύνης περιγράφηκε από τον Hotelling (1931) και, ως μέθοδος, εφαρμόζεται γενικά για την εύρεση μιας περιοχής στο xy -επίπεδο, η οποία περιλαμβάνει, με μια προκαθορισμένη πιθανότητα, το άγνωστο κέντρο ενός διμεταβλητού κανονικού πληθυσμού. Είναι το ανάλογο του διαστήματος εμπιστοσύνης $\bar{x} \pm t \frac{s}{\sqrt{n}}$ στα μονομεταβλητά γραμμικά στατιστικά. Υπό τη βασική παραδοχή της διμεταβλητής κανονικότητας, η περιοχή αυτή έχει το σχήμα μιας έλλειψης, της οποίας το μέγεθος μειώνεται καθώς το μέγεθος του δείγματος αυξάνεται. Δηλαδή, η ελλειπτική περιοχή εμπιστοσύνης δεν υπόκειται μόνο στην τυχαία διακύμανση, αλλά εξαρτάται από το μέγεθος του δείγματος, n και δεν περιγράφει τη μεταβλητότητα των δειγματικών σημείων. Στην περίπτωση των κυκλικών δεδομένων, το κέντρο της ελλειπτικής περιοχής εμπιστοσύνης είναι το σημείο (\bar{x}, \bar{y}) , με τις συντεταγμένες του να ορίζονται από την εξ. (4.3), σελ. 112.



Σχήμα 4.2: Εφαπτόμενοι κύκλοι σε μια ελλειπτική περιοχή εμπιστοσύνης. Οι ακτίνες τους προσδιορίζουν τα όρια εμπιστοσύνης για την απόσταση του κέντρου του πληθυσμού από την αρχή O των αξόνων, δηλαδή το μήκος ρ_1 του μέσου διανύσματος μ_1 του πληθυσμού.



Σχήμα 4.3: Εφαπτόμενες ευθείες από την αρχή O των αξόνων σε μια ελλειπτική περιοχή εμπιστοσύνης. Οι γωνίες που ορίζονται από τις ευθείες αυτές ως προς τον x -άξονα καθορίζουν τα όρια εμπιστοσύνης για την άγνωστη παράμετρο θ_1 του πληθυσμού.

4.3. Όρια εμπιστοσύνης για τη μέση γωνία και το μέσο μέτρο δεύτερης τάξης

Παρόμοια, αν διαγράψουμε τις δύο εφαπτόμενες προς την έλλειψη ευθείες από το κέντρο O των αξόνων, η άγνωστη παράμετρος θ_1 θα βρίσκεται μεταξύ των δύο γωνιών που ορίζονται από τις ευθείες αυτές ως προς τον x -άξονα (βλ. σχήμα 4.3).

Αλγεβρική περιγραφή Έστω το δεύτερης τάξης δείγμα $(r_j, \bar{\phi}_j)$, $j = 1, 2, \dots, k$, μεγέθους k , με τις καρτεσιανές συντεταγμένες των $\bar{\phi}_j$ να προκύπτουν από την εξ. (4.2). Υποθέτουμε ότι το δείγμα αυτό προέρχεται από διμεταβλητή κανονική κατανομή (δηλαδή, από έναν πληθυσμό όπου τα x_j και y_j κατανέμονται κανονικά). Αν θέσουμε

$$D = \frac{2(k-1) \left[1 - \frac{(\sum xy)^2}{\sum x^2 \sum y^2} \right] F_{\alpha(1), 2, k-2}}{k(k-2)}$$

όπου είναι

$$\begin{aligned} \sum x^2 &= \sum x_j^2 - \frac{(\sum x_j)^2}{k} \\ \sum y^2 &= \sum y_j^2 - \frac{(\sum y_j)^2}{k} \\ \sum xy &= \sum x_j y_j - \frac{\sum x_j \sum y_j}{k} \end{aligned} \quad (4.4)$$

με το \sum να αναφέρεται σε άθροιση επί όλων των k μέσων (δηλαδή, $\sum = \sum_{j=1}^k$), τότε, για

$$\begin{aligned} A &= \frac{k-1}{\sum x^2} \\ B &= -\frac{(k-1) \sum xy}{\sum x^2 \sum y^2} \\ C &= \frac{k-1}{\sum y^2} \end{aligned} \quad (4.5)$$

η τετραγωνική εξίσωση

$$A(x - \bar{x}^2) + 2B(x - \bar{x})(y - \bar{y}) + Cy - \bar{y}^2 = D \quad (4.6)$$

προσδιορίζει την αλγεβρική μορφή της ελλειπτικής περιοχής εμπιστοσύνης για τα δεδομένα. Τότε, οι κρίσιμες ποσότητες b_1 και b_2 για τον προσδιορισμό των ορίων εμπιστοσύνης για τη μέση κατεύθυνση του πληθυσμού, υπολογίζονται

με βάση τις παρακάτω βοηθητικές ποσότητες

$$\begin{aligned} H &= AC - B^2 & (4.7) \\ G &= A\bar{x}^2 + 2B\bar{x}\bar{y} + C\bar{y}^2 - D \\ U &= H\bar{x}^2 - CD \\ V &= \sqrt{DGH} \\ W &= H\bar{x}\bar{y} + BD \end{aligned}$$

Τα b_1 και b_2 προκύπτουν από τις εξισώσεις

$$\begin{aligned} b_1 &= \frac{W + V}{U} & (4.8) \\ b_2 &= \frac{W - V}{U} \end{aligned}$$

και εξετάζονται χωριστά, προκειμένου να παραχθεί το κάθε όριο. Έτσι, αν είναι

$$M = \sqrt{1 + b_i^2}, \quad i = 1, 2,$$

τότε κάθε όριο προσδιορίζεται από το τόξο με τριγωνομετρικούς αριθμούς

$$\begin{aligned} \text{sine} &= \frac{b_i}{M} & (4.9) \\ \text{cosine} &= \frac{1}{M} \end{aligned}$$

Τα όρια εμπιστοσύνης που προκύπτουν κατ' αυτό τον τρόπο προσδιορίζονται περαιτέρω από το τόξο που προκύπτει από τις παραπάνω εξισώσεις (4.9) ή αυτό το τόξο $+ 180^\circ$, ανάλογα με το ποιο από τα δύο είναι πιο κοντά στη μέση γωνία $\bar{\phi}$ (αν το τόξο $+ 180^\circ$ είναι μεγαλύτερο των 360° , τότε αφαιρούμε από αυτό τις 360°) (Zar, 1999, p. 612).

4.4 Ένας έλεγχος για την παράμετρο συγκέντρωσης κ

Έστω το δεύτερης τάξης δείγμα $(r_j, \bar{\phi}_j)$, $j = 1, 2, \dots, k$. Μπορούμε να αναλύσουμε τις μέσες γωνίες $\bar{\phi}_j$, μη λαμβάνοντας υπόψη τα δειγματικά μήκη r_j , εφαρμόζοντας κάποια από τις κατάλληλες για κυκλικά δεδομένα μεθόδους. Ανάλογα, μπορούμε να αναλύσουμε τα μήκη r_j , χωρίς να λάβουμε υπόψη τις μέσες γωνίες $\bar{\phi}_j$. Στη δεύτερη περίπτωση, ο έλεγχος αφορά στην τάση συγκέ-

ντρωσης του δεύτερης τάξης δείγματος των r_1, r_2, \dots, r_k . Αν οι τιμές αυτές είναι σχετικά υψηλές, τότε η μηδενική υπόθεση της τυχαιότητας μπορεί να απορριφθεί. Σημειωτέον ότι, υπό τη μηδενική υπόθεση της ομοιομορφίας, παραδεχόμαστε την ανεξαρτησία των επί μέρους παρατηρήσεων (γωνιών) για κάθε στοιχείο του δεύτερης τάξης δείγματος. Αυτό αποτελεί μια πολύ αυστηρή παραδοχή πάνω σε ένα πρώτη τάξης δείγμα, που συνήθως δεν συναντάται στην ανάλυση των πρώτης τάξης δειγμάτων (Batschelet, 1981, p. 199).

Χάριν απλούστευσης, υποθέτουμε ότι οι τιμές r_i είναι διατεταγμένες σε αύξουσα σειρά, δηλαδή είναι

$$0 \leq r_1 \leq r_2 \leq \dots \leq r_k \leq 1$$

Συγκρίνουμε τη δειγματική κατανομή των r_j με την αντίστοιχη θεωρητική κατανομή των r , δεδομένου του k , υπό τη μηδενική υπόθεση της ομοιομορφίας, όπως δίνεται από τους πίνακες του Rayleigh test [π.χ. Batschelet (1981, pp. 334–335)]. Η συνάρτηση κατανομής $F(r)$ μπορεί να παραχθεί μέσω της σχέσης

$$F(r) = 1 - P \quad (4.10)$$

Για τον έλεγχο της μηδενικής υπόθεσης της ανεξαρτησίας, ένα κατάλληλο τεστ είναι το Kolmogorov's one-sample test.

4.5 Ένας παραμετρικός μονοδειγματικός έλεγχος για τη μέση κατεύθυνση — Hotelling's one-sample test

Έστω το δεύτερης τάξης δείγμα $(r_j, \bar{\phi}_j)$, $j = 1, 2, \dots, k$. Για το δείγμα αυτό, υπολογίζουμε το γενικό μέσο, με τις συντεταγμένες του να προκύπτουν από την εξ. (4.2). Υποθέτουμε ότι το δείγμα αυτό προέρχεται από διμεταβλητή κανονική κατανομή (δηλαδή, από έναν πληθυσμό όπου τα x_j και y_j κατανέμονται κανονικά). Θέλουμε να ελέγξουμε τη στατιστική σημαντικότητα του γενικού μέσου.⁴

Ο Hotelling (1931) πρότεινε, για την περίπτωση αυτή, μια διαδικασία ελέγχου που χρησιμοποιεί τη στατιστική ποσότητα

$$F = \frac{k(k-2)}{2} \left[\frac{\bar{x}^2 \sum y^2 - 2\bar{x}\bar{y} \sum xy + \bar{y}^2 \sum x^2}{\sum x^2 \sum y^2 - (\sum xy)^2} \right] \quad (4.11)$$

⁴Δηλαδή την $H_0 : \boldsymbol{\mu} = 0$, όπου $\boldsymbol{\mu}$ το μέσο διάνυσμα του πληθυσμού (Batschelet, 1981, p. 146).

όπου τα $\sum x^2$, $\sum y^2$ και $\sum xy$ υπολογίζονται από τις εξ. (4.4).

Υπό τη μηδενική υπόθεση της ομοιομορφίας του πληθυσμού, το στατιστικό της (4.11) ακολουθεί την F -κατανομή με 2 και $k-2$ βαθμούς ελευθερίας. Επομένως, απορρίπτουμε τη μηδενική υπόθεση ($\rho = 0$), όταν είναι $F > F_{\alpha(1),2,k-2}$.

Ο έλεγχος υποθέτει ότι τα δεδομένα δεν είναι ομαδοποιημένα. Η παραδοχή της διμεταβλητής κανονικότητας είναι κρίσιμη.⁵ Παρ' όλο που το τεστ εμφανίζεται ανθεκτικό σε αποκλίσεις λόγω κύρτωσης, μπορεί να επηρεάζεται σοβαρά από την παρουσία μεγάλης λοξότητας, απορρίπτοντας μια αληθινή μηδενική υπόθεση πολύ πιο συχνά απ' ό,τι υποδεικνύει το επίπεδο σημαντικότητας (Zar, 1999, p. 639).

4.6 Ένας απαραμετρικός μονοδειγματικός έλεγχος για τη μέση κατεύθυνση — Moore's test

Έστω το δεύτερης τάξης δείγμα $(r_j, \bar{\phi}_j)$, $j = 1, 2, \dots, k$. Θέλουμε να ελέγξουμε αν οι μέσοι $\bar{\phi}_j$ συγκεντρώνονται προς μια κατεύθυνση ή είναι κατανομημένοι ομοιόμορφα.

Μια προσέγγιση για τον παραπάνω έλεγχο αποτελεί η χρησιμοποίηση ενός από τους κατάλληλους ελέγχους για τον προσανατολισμό των δεδομένων (π.χ. το Rayleigh test), αψηφώντας τα δειγματικά μήκη r_j . Όμως, η προσέγγιση αυτή παραβλέπει την πληροφορία που περιέχεται στα μέσα συνισταμένα μήκη r_j . Αν το r_j είναι σχετικά μεγάλο, η γωνία $\bar{\phi}_j$ του j στοιχείου του δεύτερης τάξης δείγματος προσδιορίζεται πιο αξιόπιστα, απ' ό,τι αν το r_j είναι σχετικά μικρό.

Μια τροποποίηση του Rayleigh test, η οποία χρησιμοποιεί την πληροφορία από τα δειγματικά μέσα μήκη r_j , προτάθηκε από τον Moore (1980). Ο έλεγχος είναι απαραμετρικός, καθώς χρησιμοποιεί μόνο τους βαθμούς (ranks) των r_j . Έτσι, τα μήκη των μέσων διανυσμάτων r_j κατατάσσονται σε αυξουσα σειρά

⁵To Hotelling's one-sample test δεν μπορεί να εφαρμοστεί σε κυκλικά δεδομένα πρώτης τάξης (Batschelet, 1981, p. 146).

και τους αποδίδονται οι αντίστοιχοι βαθμοί, έστω t_j . Αν είναι

$$\begin{aligned} C &= \sum_{j=1}^n t_j \cos \bar{\phi}_j \\ S &= \sum_{j=1}^n t_j \sin \bar{\phi}_j \\ D &= \sqrt{C^2 + S^2}, \end{aligned} \quad (4.12)$$

τότε η μηδενική υπόθεση της ανεξαρτησίας των $\bar{\phi}_j$ μεταξύ τους, της ομοιόμορφης κατανομής των $\bar{\phi}_j$ στον κύκλο και της ανεξαρτησίας των μέσων συνισταμένων μηκών r_j από τις μέσες γωνίες $\bar{\phi}_j$, ελέγχεται με το στατιστικό⁶

$$D^* = \frac{D}{k^{3/2}} \quad (4.13)$$

Ο Batschelet (1981, p. 343) δίνει πίνακα με τις κρίσιμες τιμές $D^*(\alpha)$ του στατιστικού της (4.13) για διάφορα επίπεδα σημαντικότητας α . Απορρίπτουμε τη μηδενική υπόθεση, υπέρ της εναλλακτικής υπόθεσης του προσανατολισμού των $\bar{\phi}_j$, όταν το D^* είναι μεγάλο, δηλαδή όταν είναι

$$D^* \geq D^*(\alpha).$$

Η συγκεκριμένη τροποποίηση του Rayleigh test μπορεί να εφαρμοστεί και σε περιπτώσεις όπου έχουμε δεύτερης τάξης δείγματα σταθμισμένων γωνιών. Τότε, μπορούμε να ταξινομήσουμε τις γωνίες βάσει των βαρών τους, αντί των μηκών των μέσων διανυσμάτων r_j . Για παράδειγμα, μπορεί σε ένα πείραμα επαναλαμβανόμενων απελευθερώσεων, από ένα συγκεκριμένο σημείο, μιας ομάδας πτηνών να καταγράφεται κάθε φορά η γωνία πτήσης, ως προς μια αρχική κατεύθυνση, από κοινού με την απόσταση που διανύθηκε μέσα σε κάποια όρια. Τότε, κάθε γωνία αποκτά ένα βάρος (τη διανυθείσα απόσταση) και η εφαρμογή του ελέγχου του Moore μπορεί να βασιστεί στην ταξινόμηση αυτών των βαρών (Zar, 1999, p. 641).

⁶Ο Zar (1999, p.640) αναφέρει το στατιστικό του Moore ως $R' = \sqrt{\frac{C^2+S^2}{k}}$, δίνοντας τον ίδιο πίνακα κρίσιμων τιμών.

4.7 Ένας παραμετρικός έλεγχος δύο δειγμάτων δεύτερης τάξης για τη μέση κατεύθυνση — Hotelling's two-sample test

Έστω τα δεύτερης τάξης δείγματα

$$(r_{j_1}, \bar{\phi}_{j_1}), j_1 = 1, 2, \dots, k_1$$

και

$$(r_{j_2}, \bar{\phi}_{j_2}), j_2 = 1, 2, \dots, k_2,$$

τα οποία προέρχονται από πληθυσμούς που ακολουθούν τη διμεταβλητή κανονική κατανομή και έχουν ίσες διακυμάνσεις και συνδιακυμάνσεις. Υποθέτουμε, ακόμα, ότι τα δύο δείγματα είναι ανεξάρτητα μεταξύ τους και τα δεδομένα δεν είναι ομαδοποιημένα. Θέλουμε να ελέγξουμε αν οι μέσες κατευθύνσεις $\theta_1^{(1)}$ και $\theta_1^{(2)}$ των δύο πληθυσμών ταυτίζονται.⁷

Αν θέσουμε ως

$$(\bar{x}_1, \bar{y}_1) \quad (\bar{x}_2, \bar{y}_2)$$

τις καρτεσιανές συντεταγμένες των δύο δειγματικών γενικών μέσων $\bar{\phi}_1$ και $\bar{\phi}_2$, τότε με βάση τις ποσότητες

$$\begin{aligned} (\sum x^2)_c &= (\sum x^2)_1 + (\sum x^2)_2 \\ (\sum y^2)_c &= (\sum y^2)_1 + (\sum y^2)_2 \\ (\sum xy)_c &= (\sum xy)_1 + (\sum xy)_2, \end{aligned} \quad (4.14)$$

με τα $\sum x^2$, $\sum y^2$ και $\sum xy$ να υπολογίζονται από τις εξ. (4.4), για κάθε το δείγμα (1) και (2) αντίστοιχα, υπολογίζουμε το στατιστικό

$$\begin{aligned} F &= \frac{k_1 + k_2 - 3}{2 \left(\frac{1}{k_1} + \frac{1}{k_2} \right)} \\ &\times \left[\frac{(\bar{x}_1 - \bar{x}_2)^2 (\sum y^2)_c - 2(\bar{x}_1 - \bar{x}_2)(\bar{y}_1 - \bar{y}_2)(\sum xy)_c + (\bar{y}_1 - \bar{y}_2)^2 (\sum x^2)_c}{(\sum x^2)_c (\sum y^2)_c - (\sum xy)_c^2} \right] \end{aligned} \quad (4.15)$$

⁷ Δεδομένης της ισχύος των συνθηκών, αυτό συνεπάγεται ταύτιση των πληθυσμών. Επί της ουσίας, ελέγχεται η $H_0 : \mu_1 = \mu_2$, όπου μ_1 και μ_2 είναι τα μέσα διανύσματα των πληθυσμών (Batschelet, 1981, p. 151).

Το στατιστικό της (4.15) κατανέμεται σύμφωνα με τη μονοδιάστατη F -κατανομή με 2 και $k_1 + k_2 - 3$ βαθμούς ελευθερίας. Η μηδενική υπόθεση της ισότητας των μέσων κατευθύνσεων των πληθυσμών απορρίπτεται για τιμές του στατιστικού $F > F_{\alpha(1),2,k_1+k_2-3}$. Ο έλεγχος είναι περισσότερο ανθεκτικός σε αποκλίσεις από την κανονικότητα απ' όσο είναι ο αντίστοιχος μονοδειγματικός έλεγχος της ενότητας (4.5). Πάντως, η μη κανονικότητα των πληθυσμών οδηγεί σε πιο συντηρητικούς ελέγχους, δηλαδή το τεστ τείνει να απορρίπτει μια ψευδή H_0 πιο συχνά απ' όσο υποδεικνύει το επίπεδο σημαντικότητας α . Επίσης, μεγάλες διαφορές στα μεγέθη των δειγμάτων δεν φαίνεται να προκαλούν σοβαρές επιπτώσεις στην εγκυρότητα του ελέγχου (Zar, 1999, p. 641).

4.8 Ένας απαραμετρικός έλεγχος δύο δειγμάτων δεύτερης τάξης για τη μέση κατεύθυνση — Mardia's two-sample test

Ο έλεγχος της προηγούμενης ενότητας (4.7) θέτει κάποιες ιδιαίτερα αυστηρές παραδοχές, οι οποίες δύσκολα ικανοποιούνται στην πράξη. Ως έναν εναλλακτικό έλεγχο της ταύτισης των μέσων κατευθύνσεων δύο πληθυσμών, ο Mardia (1976) έχει προτείνει το παρακάτω απαραμετρικό τεστ. Υποθέτουμε ότι οι δύο πληθυσμοί έχουν *συνεχείς* κατανομές πυκνότητας, τα δύο δείγματα είναι ανεξάρτητα μεταξύ τους και τα δεδομένα δεν είναι ομαδοποιημένα.

Έστω τα δεύτερης τάξης δείγματα

$$(r_{j_1}, \bar{\phi}_{j_1}), j_1 = 1, 2, \dots, k_1$$

και

$$(r_{j_2}, \bar{\phi}_{j_2}), j_2 = 1, 2, \dots, k_2.$$

Συνδυάζοντας όλα τα δειγματικά σημεία, υπολογίζουμε το γενικό μέσο διάνυσμα του από κοινού δείγματος. Στη συνέχεια, οι καρτεσιανές συντεταγμένες \bar{x} και \bar{y} του γενικού μέσου αφαιρούνται από τις αντίστοιχες συντεταγμένες x και y των σημείων κάθε δείγματος.⁸ Τα σημεία που προκύπτουν από αυτήν τη διαδικασία σχηματίζουν δύο νέα δείγματα, τα οποία μπορεί να ελεγχθούν με κάποιο κατάλληλο απαραμετρικό τεστ, όπως το Watson's U^2 test ή το Mardia-Watson-Wheeler test (Zar, 1999, p. 643).

⁸Ο Batschelet (1981, p. 154) το αναφέρει ως αλλαγή της αρχής των αξόνων του συστήματος.

Μεταβάλλεται η μέση ημερήσια κατεύθυνση του ανέμου μεταξύ των ημερών με φυσιολογικές τιμές θερμοκρασίας και των ημερών της περιόδου του καύσωνα; (Έλεγχος της ισότητας των μέσων κατευθύνσεων των μεταβλητών δεύτερης τάξης $WINDIR_DAY[1-17]$ και $WINDIR_DAY[18-31]$ με το *Mardia's two-sample test*)

Οι μεταβλητές $WINDIR_DAY[1-17]$ και $WINDIR_DAY[18-31]$ εκφράζουν την κατεύθυνση του ανέμου κατά το μέσο όρο ημέρας για τα αντίστοιχα διαστήματα του μήνα μελέτης. Δηλαδή, παίρνουν ως τιμές τις μέσες κατευθύνσεις των 17 πρώτων και 14 τελευταίων ημερών του συγκεκριμένου μήνα, οπότε περιγράφουν ένα χαρακτηριστικό δύο δεύτερης τάξης δειγμάτων.

Τα δύο δείγματα μπορεί να θεωρηθεί ότι είναι ανεξάρτητα μεταξύ τους και έχουν συνεχείς κατανομές πυκνότητας, δηλαδή πληρούν τις παραδοχές για την εφαρμογή του *Mardia's two-sample test*.

Συνδυάζοντας τα δειγματικά σημεία, παίρνουμε το από κοινού δείγμα που είναι, βεβαίως, το δείγμα των $WINDIR_DAY$. Οι συνιστώσες του γενικού μέσου $\phi = 204.4000^\circ$ του από κοινού δείγματος, όπως υπολογίστηκαν στο προηγούμενο πλαίσιο, σύμφωνα με τις μεθόδους ανάλυσης δεύτερης τάξης, είναι

$$\bar{x} = -0.4093$$

$$\bar{y} = -0.1856$$

Αφαιρώντας τα \bar{x} και \bar{y} από τα αντίστοιχα μεγέθη των επιμέρους μέσων γωνιών των δύο δειγμάτων $WINDIR_DAY[1-17]$ και $WINDIR_DAY[18-31]$, παίρνουμε τα νέα δείγματα των διαφορών $WDDIFF_DAY[1-17]$ και $WDDIFF_DAY[18-31]$, αντίστοιχα.

Η εφαρμογή των ελέγχων *Watson's U^2 test* και *Mardia-Watson-Wheeler test*, με τη χρησιμοποίηση του στατιστικού προγράμματος *Oriana* έδειξε ότι οι μέσες κατευθύνσεις των νέων δειγμάτων δεν διαφέρουν σημαντικά (βλ. πίνακα ??). Έτσι, δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση της ισότητας των μέσων κατευθύνσεων, δηλαδή φαίνεται ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν διαφοροποιείται κατά τις ημέρες με καύσωνα. Πάντως, η *P-value* και στους δύο παραπάνω ελέγχους είναι σχετικά κοντά στην περιοχή του 0.10.

Πίνακας 4.2: Αποτελέσματα των απαραμετρικών ελέγχων Watson's U^2 και Mardia-Watson-Wheeler για τον έλεγχο της ισότητας των μέσων γωνιών των βοηθητικών δειγμάτων που περιγράφουν οι μεταβλητές WDDIFF_DAY[1-17] και WDDIFF_DAY[18-31]. Οι συγκεκριμένοι έλεγχοι έγιναν κατά την εφαρμογή του ελέγχου Mardia's two-sample για την ισότητα της μέσης κατεύθυνσης του ανέμου κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας (μέσος όρος των 17 μέσων ημερήσιων κατευθύνσεων WINDIR_DAY[1-17]) με τη μέση κατεύθυνση του ανέμου κατά τις ημέρες με καύσωνα (μέσος όρος των 14 μέσων ημερήσιων κατευθύνσεων WINDIR_DAY[18-31]). Φαίνεται, τελικά, ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν διαφοροποιείται κατά τις ημέρες με καύσωνα.

<i>[MARDIA-WATSON-WHEELER]</i>					
<i>Variables (& observations)</i>		<i>W</i>	<i>p</i>		
WDDIFF_DAY[1-17]	&	4,595	0,101		
WDDIFF_DAY[18-31] (17 & 14)					
<i>[WATSON'S U^2]</i>					
<i>Variables (& observations)</i>		<i>U²</i>	<i>p</i>	<i>df</i>	<i>df2</i>
WDDIFF_DAY[1-17]	&	0,144	0.2 > p > 0.1	14	17
WDDIFF_DAY[18-31] (17 & 14)					

4.9 Ένας παραμετρικός έλεγχος δύο δεύτερης τάξης δειγμάτων ζευγαρωτών παρατηρήσεων για τη μέση κατεύθυνση

Όπως συμβαίνει στην περίπτωση των γραμμικών μεταβλητών, έτσι και για τις κυκλικές μεταβλητές, δύο δείγματα κυκλικών δεδομένων μπορεί να είναι ζευγαρωτά⁹ (paired). Υπολογίζοντας τις διαφορές μεταξύ των ζευγαρωτών παρατηρήσεων, προκειμένου να ελέγξουμε την ταύτιση των μέσων κατευθύνσεων των πληθυσμών από τους οποίους προέρχονται τα δύο δείγματα, μπορούμε να εφαρμόσουμε στο νέο δείγμα που προκύπτει το μονοδειγματικό Hotelling test της ενότητας (4.5).

Έστω τα δεύτερης τάξης δείγματα ζευγαρωτών παρατηρήσεων¹⁰ $(r_{lj}, \bar{\phi}_{lj})$, όπου το $l = 1, 2$ εκφράζει το πρώτο ή το δεύτερο δείγμα (π.χ. το πριν και το μετά), και το $j = 1, 2, \dots, k$ εκφράζει το j ζεύγος των παρατηρήσεων (μέσων γωνιών και μηκών μέσων διανυσμάτων). Δηλαδή, η $\bar{\phi}_{lj}$ είναι η j μέση γωνία και το r_{lj} είναι το μήκος του μέσου διανύσματος της j ομάδας παρατηρήσεων (δείγμα πρώτης τάξης) για το δείγμα 1 (τα αντίστοιχα ισχύουν για το δείγμα 2). Τότε μπορούμε να σχηματίσουμε ένα νέο δείγμα καρτεσιανών συντεταγμένων x_j και y_j από τις διαφορές

$$\begin{aligned} x_j &= r_{2j} \cos \bar{\phi}_{2j} - r_{1j} \cos \bar{\phi}_{1j} \\ y_j &= r_{2j} \sin \bar{\phi}_{2j} - r_{1j} \sin \bar{\phi}_{1j} \end{aligned} \quad (4.16)$$

Ο έλεγχος της μέσης κατεύθυνσης για το νέο αυτό δείγμα¹¹ μπορεί να γίνει με τη διαδικασία του Hotelling's one-sample test της ενότητας (4.5), σελ. 119.

⁹ Δύο δείγματα είναι ζευγαρωτά, όταν κάθε ένα από αυτά συντίθεται από το ίδιο σύνολο πειραματικών μονάδων, οι οποίες ελέγχονται δις η κάθε μία.

¹⁰ Η ίδια διαδικασία μπορεί να εφαρμοστεί και σε πρώτης τάξης δείγματα κυκλικών δεδομένων, δηλαδή για ζευγαρωτές παρατηρήσεις απλών γωνιών ϕ_{lj} , $l = 1, 2$, $j = 1, 2, \dots, k$. Τότε, ο υπολογισμός των καρτεσιανών συντεταγμένων (x_j, y_j) γίνεται από τις (4.16), με τις ϕ_{lj} αντί των $\bar{\phi}_{lj}$ και, βέβαια, με τα $r_{lj} = 1$.

¹¹ Για το δείγμα των διαφορών ελέγχουμε την $H_0 : \rho = 0$, δηλαδή, επί της ουσίας, την ταύτιση των μέσων κατευθύνσεων των δύο αρχικών δειγμάτων των ζευγαρωτών παρατηρήσεων (π.χ. την ταύτιση του πριν με το μετά).

4.10 Ένας απαραμετρικός έλεγχος δύο δεύτερης τάξης δειγμάτων ζευγαρωτών παρατηρήσεων για τη μέση κατεύθυνση

Δεδομένων δύο δεύτερης τάξης δειγμάτων ζευγαρωτών παρατηρήσεων, ο έλεγχος για την ταύτιση των μέσων κατευθύνσεων των πληθυσμών από τους οποίους προέρχονται τα δείγματα, μπορεί να γίνει και απαραμετρικά, υπολογίζοντας το νέο δείγμα των διαφορών των ζευγαρωτών παρατηρήσεων και εφαρμόζοντας σ' αυτό το Moore's test της ενότητας (4.6).

Έστω τα δεύτερης τάξης δείγματα ζευγαρωτών παρατηρήσεων¹² $(r_{lj}, \bar{\phi}_{lj})$, όπου το $l = 1, 2$ εκφράζει το πρώτο ή το δεύτερο δείγμα (π.χ. το πριν και το μετά), και το $j = 1, 2, \dots, k$ εκφράζει το j ζεύγος των παρατηρήσεων (μέσων γωνιών και μηκών μέσων διανυσμάτων). Δηλαδή, η $\bar{\phi}_{lj}$ είναι η j μέση γωνία και το r_{lj} είναι το μήκος του μέσου διανύσματος της j ομάδας παρατηρήσεων (δείγμα πρώτης τάξης) για το δείγμα 1 (τα αντίστοιχα ισχύουν για το δείγμα 2). Τότε μπορούμε να σχηματίσουμε ένα νέο δείγμα καρτεσιανών συντεταγμένων x_j και y_j από τις διαφορές (4.16). Για κάθε ζευγαρωτή διαφορά j , ($j = 1, 2, \dots, k$), υπολογίζουμε τα

$$\begin{aligned} r_j &= \sqrt{x_j^2 + y_j^2} \\ \cos \bar{\phi}_j &= \frac{x_j}{r_j} \\ \sin \bar{\phi}_j &= \frac{y_j}{r_j} \end{aligned} \quad (4.17)$$

Στη συνέχεια, ταξινομούμε κατ' αύξουσα σειρά τις τιμές των r_j , αποδίδοντας τους αντίστοιχους βαθμούς, έστω $t_j = 1, 2, \dots, k$. Ο έλεγχος της ταύτισης των μέσων κατευθύνσεων των πληθυσμών γίνεται με το απαραμετρικό Moore's test της ενότητας (4.6), χρησιμοποιώντας τις εξ. (4.12) και υπολογίζοντας την τιμή του στατιστικού της εξ. (4.13).

¹²Η ίδια διαδικασία μπορεί να εφαρμοστεί και σε πρώτης τάξης δείγματα κυκλικών δεδομένων, δηλαδή για ζευγαρωτές παρατηρήσεις απλών γωνιών ϕ_{lj} , $l = 1, 2$, $j = 1, 2, \dots, k$. Τότε, ο υπολογισμός των καρτεσιανών συντεταγμένων (x_j, y_j) γίνεται από τις (4.16), με τις ϕ_{lj} αντί των $\bar{\phi}_{lj}$ και τα $r_{lj} = 1$.

Πίνακας 4.3: Συνοπτική επισκόπηση των ελέγχων για την ανάλυση δεύτερης τάξης.

Δεδομένα	Έλεγχος	Είδος ελέγχου	Μηδενική υπόθεση	Παράδοχές	Στατιστικό / Κριτήριο προτίχης
Ένα δείγμα	Έλεγχος για t_0	Παραμετρικός	Τυχασιότητα	Ανεξαρτησία των επί μέρους παρατηρήσεων	Έλεγχος της κανονικότητας των μέσων μετρών r με το Rayleigh Z
	Hotelling's one-sample test	Παραμετρικός	Τυχασιότητα	Διμεταβλητή κανονικότητα, μη ομαδοποιημένα δεδομένα	$F > F_{\alpha(1), 2, k-2}$
	Moore's test	Μη παραμετρικός	Τυχασιότητα	-	$D^* \geq D^*(\alpha)$
Δύο δείγματα	Hotelling's two-sample test	Παραμετρικός	$H_0 : \mu_1 = \mu_2$	Διμεταβλητή κανονικότητα, ίσες διακυμάνσεις—συνδιακυμάνσεις, ανεξαρτησία δειγμάτων, μη ομαδοποιημένα δεδομένα	$F > F_{\alpha(1), 2, k_1+k_2-3}$
	Mardia's two-sample test	Μη παραμετρικός	$H_0 : \mu_1 = \mu_2$	Συνεχείς κατανομές πυκνότητας, ανεξαρτησία δειγμάτων, μη ομαδοποιημένα δεδομένα	Έλεγχος του από κοινού δείγματος με το Watson's U^2 ή το Mardia-Watson-Wheeler test
Ζευγαρωτές παρατηρήσεις	Έλεγχος των μέσων κατευθύνσεων	Παραμετρικός	$H_0 : \rho = 0$ (για το δείγμα των διαφορών)	-	Έλεγχος με το Hotelling's one-sample test
	Έλεγχος των μέσων κατευθύνσεων	Μη παραμετρικός	$H_0 : \rho = 0$ (για το δείγμα των διαφορών)	-	Έλεγχος με το Moore's test
	Έλεγχος των μέσων κατευθύνσεων	Παραμετρικός	$H_0 : \rho = 0$ (για το δείγμα των διαφορών)	-	Έλεγχος με το Moore's test

Κεφάλαιο 5

Κυκλική συσχέτιση και παλινδρόμηση

Η συσχέτιση κυκλικών μεταβλητών αποτελεί τη σχετικά νεώτερη θεματική τής στατιστικής κυκλικών δεδομένων. Θεωρητικές μελέτες πάνω στο αντικείμενο αυτό έχουν γραφτεί από τη δεκαετία του 1970 και μετά. Αρχικά, λήφθηκε υπ' όψη μόνο η περίπτωση της συσχέτισης μεταξύ δύο μεταβλητών. Τα τελευταία χρόνια, μελετάται η ανάλυση συσχέτισης σε υψηλότερη διάσταση, με την εισαγωγή μερικών συντελεστών συσχέτισης για τα διάφορα προτεινόμενα υποδείγματα.

5.1 Συσχέτιση κυκλικής μεταβλητής με κυκλική μεταβλητή (C-C correlation)

Ελέγχουμε την περίπτωση κατά την οποία δύο κυκλικές μεταβλητές παρατηρούνται πάνω στο ίδιο υποκείμενο. Τέτοιες περιπτώσεις αποτελούν οι κατευθύνσεις αναχώρησης και άφιξης αποδημητικών ζώων, η κατεύθυνση πτήσης των πουλιών σε σχέση με την κατεύθυνση του ανέμου, τα ακρότατα ενός βιολογικού ρυθμού πάνω στο ίδιο ανθρώπινο υποκείμενο πριν και μετά τη θεραπεία κ.ά. Με επαναλαμβανόμενες παρατηρήσεις σε n άτομα παίρνουμε ένα δείγμα ζευγαρωτών παρατηρήσεων.

Έστω (ϕ_i, ψ_i) το ζεύγος των κυκλικών μετρήσεων που παρατηρούνται πάνω στο υποκείμενο i , $i = 1, 2, \dots, n$. Υποθέτουμε ότι οι ϕ_i και ψ_i είναι μεταβλητές ίδιας φύσης, για παράδειγμα και οι δύο είναι κατευθύνσεις ή και οι δύο είναι ακρότατα ενός βιολογικού ρυθμού. Επειδή αποτελούν μετρήσεις πάνω στο

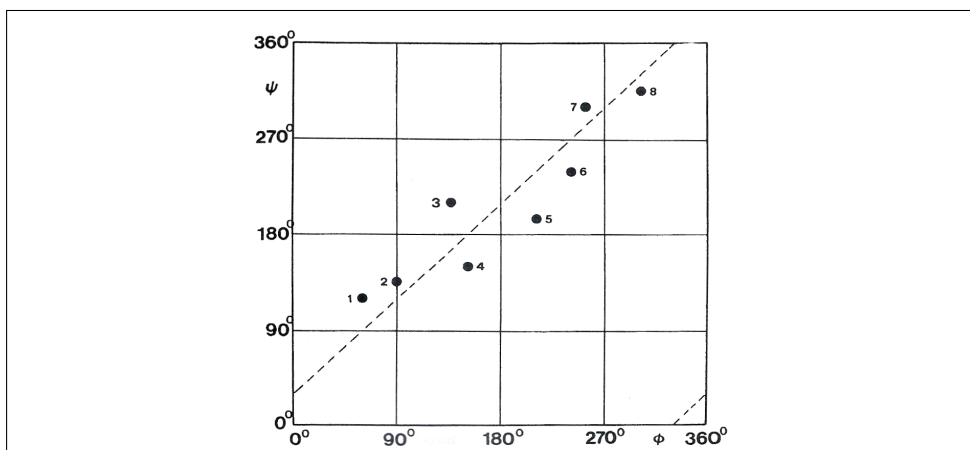
ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

ίδιο υποκείμενο, οι ϕ_i και ψ_i συσχετίζονται, διαφορετικά όμως ζεύγη παρατηρήσεων (ϕ_i, ψ_i) και (ϕ_j, ψ_j) , $i \neq j$ λογίζονται ως ανεξάρτητα μεταξύ τους.

Υπό την παραδοχή ότι οι ϕ και ψ κατανέμονται ομοιόμορφα (τυχαία) στον κύκλο, η διαδικασία ελέγχου της συσχέτισης των δύο κυκλικών μεταβλητών έχει ως εξής:

Θετική συσχέτιση Στην περίπτωση μιας αναμενόμενης θετικής συσχέτισης, είναι δόκιμο το να βασίσουμε την ανάλυση στις διαφορές

$$\delta_i = \psi_i - \phi_i$$



Σχήμα 5.1: Γραμμικό διάγραμμα τυχαίου δείγματος ζευγαρωτών παρατηρήσεων δύο κυκλικών μεταβλητών ϕ και ψ . Η διακεκομμένη γραμμή παλινδρόμησης υποδεικνύει θετική συσχέτιση των ϕ και ψ .

Οι διαφορές δ_i αποτελούν μια επίσης κυκλική μεταβλητή, καθώς προσδιορίζονται μόνο modulo 360° . Στην περίπτωση της θετικής συσχέτισης, αναμένουμε τις γωνίες δ_i να είναι κάπως ομαδοποιημένες γύρω από μια μέση γωνία. Τότε, μπορούμε να προσδιορίσουμε το μέγεθος της ομαδοποίησης με κάποιο μέτρο συγκέντρωσης, όπως είναι το μήκος του μέσου διανύσματος r , για τις γωνίες δ_i (βλ. και σχήμα 5.1)

$$r = \frac{1}{n} \left[\left(\sum_{i=1}^n \cos \delta_i \right)^2 + \left(\sum_{i=1}^n \sin \delta_i \right)^2 \right]^{1/2} \quad (5.1)$$

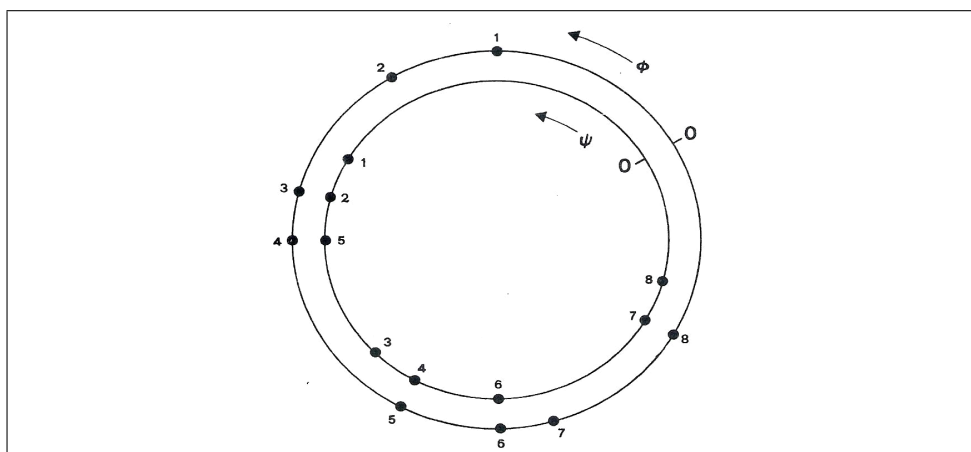
Μεγάλες τιμές του r υποδεικνύουν ισχυρή θετική συσχέτιση, ενώ μικρές

τιμές σημαίνουν ασθενή ή και καθόλου συσχέτιση των ϕ και ψ . Από τη στιγμή που το r κυμαίνεται μεταξύ 0 και 1, μπορεί να χρησιμοποιηθεί ως συντελεστής κυκλικής συσχέτισης. Προκειμένου να υποδείξουμε ότι το r αφορά σε θετική συσχέτιση, χρησιμοποιούμε το συμβολισμό r_+ . Ο συντελεστής συσχέτισης r_+ πλησιάζει τη μέγιστη τιμή του, όταν οι διαφορές $\psi_i - \phi_i$ παραμένουν σταθερές (modulo 360°) για κάθε $i = 1, 2, \dots, n$.

Το αν το r_+ διαφέρει σημαντικά από το 0 είναι ένα ερώτημα ισοδύναμο με το αν τα δ_i κατανέμονται μη ομοιόμορφα στον κύκλο. Οι διαφορές δ_i είναι συνεχείς κυκλικές μεταβλητές, οπότε για τον έλεγχο της ομοιομορφίας τους μπορεί να εφαρμοστεί το Rayleigh test. Για μεγάλες τιμές του r_+ , δηλαδή υψηλό βαθμό συγκέντρωσης, η μηδενική υπόθεση της τυχαιότητας μπορεί να απορριφθεί και τότε η συσχέτιση θεωρείται σημαντική.

Τυχόν ισοψηφίες μεταξύ των μετρήσεων δεν προκαλούν προβλήματα. Η ομαδοποίηση των δεδομένων συνεπάγεται το ότι οι διαφορές δ_i ομαδοποιούνται επίσης. Ως εκ τούτου, ο συντελεστής διόρθωσης της εξ. (1.27), σελ. 20, μπορεί να εφαρμοστεί στο r_+ , τουλάχιστον για μικρό αριθμό ομάδων.

Επίσης, σημειώνουμε ότι το r_+ είναι ανεξάρτητο από την όποια επιλογή αρχικού σημείου στον κύκλο (μηδέν γωνία), είτε για τη ϕ , είτε για την ψ . Κατά συνέπεια, σε ένα κυκλικό διάγραμμα, όπως αυτό του σχήματος 5.2, οι δύο κύκλοι μπορεί να περιστρέφονται ο ένας ως προς τον άλλο, χωρίς να μεταβάλλεται ο βαθμός της συσχέτισης των δειγματικών τους σημείων. Δηλαδή, ο συντελεστής συσχέτισης r_+ είναι αμετάβλητος υπό περιστροφή.



Σχήμα 5.2: Κυκλικό διάγραμμα του ίδιου με το σχήμα 5.1 δείγματος ζευγαρωτών παρατηρήσεων των κυκλικών μεταβλητών ϕ και ψ . Ο συντελεστής συσχέτισης των ϕ και ψ είναι αμετάβλητος υπό περιστροφή.

Μπορούμε, επίσης, να εφαρμόσουμε την εξ. (1.14), σελ. 12, στις γωνίες δ_i . Τότε, αν υπάρχει μια μέση γωνία $\bar{\delta}$, ο συντελεστής συσχέτισης r_+ δίνεται από την

$$r_+ = \frac{1}{n} \sum^i \cos(\delta_i - \bar{\delta}) \quad (5.2)$$

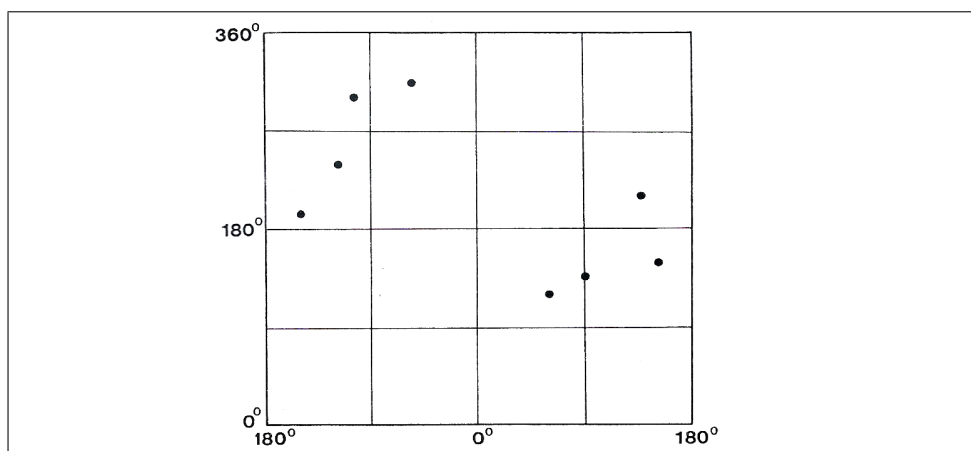
Αρνητική συσχέτιση Μπορούμε να χειριστούμε την αρνητική συσχέτιση με παρόμοιο τρόπο. Εδώ, υποθέτουμε ότι, κατά μέσο όρο, το ϕ μειώνεται, όσο το ψ αυξάνεται. Αυτό συνεπάγεται ότι τα $-\phi$ και ψ , κατά μέσο όρο, αυξάνονται ταυτόχρονα ή, αλλιώς, είναι θετικά συσχετισμένα. Ως εκ τούτου, μπορούμε να εφαρμόσουμε την προηγούμενη διαδικασία αντικαθιστώντας το ϕ με $-\phi$ και σημειώνοντας το νέο συντελεστή συσχέτιση ως r_- .

Στην περίπτωση των κυκλικών μεταβλητών, η θετική και η αρνητική συσχέτιση διαχειρίζονται ξεχωριστά: προσδιορίζονται δύο συντελεστές συσχέτισης, οι r_+ και r_- . Στο ερώτημα αν δύο κυκλικές μεταβλητές μπορούν να είναι συγχρόνως θετικά και αρνητικά συσχετισμένες, ο Batschelet (1981, p. 182) απαντά ότι αυτό είναι δυνατό, σε αντίθεση με την περίπτωση των γραμμικών μεταβλητών, όπου οι δύο τύποι συσχέτισης διαχωρίζονται καθαρά από το πρόσημο του συντελεστή γραμμικής συσχέτισης r . Το επιχείρημα αυτό παρουσιάζεται διαγραμματικά στο σχήμα 5.3, σελ. 133, το οποίο περιέχει ακριβώς τα ίδια δειγματικά σημεία με το προηγούμενο σχήμα 5.1, σελ. 130 (που χρησιμοποιήθηκε για την αναπαράσταση της θετικής συσχέτισης), με τη διαφορά ότι το αρχικό σημείο στον οριζόντιο άξονα έχει μετατοπιστεί από 0° σε 180° . Η τάση φαίνεται τώρα να είναι φθίνουσα, υποδηλώνοντας αρνητική συσχέτιση.

Στην περίπτωση αυτή ερχόμαστε αντιμέτωποι με το παράδοξο γεγονός τού να υπάρχουν ταυτόχρονα και θετική και αρνητική συσχέτιση σε δείγμα κυκλικών μεταβλητών. Ένας τρόπος για να αποφύγουμε το δίλημμα είναι μια ιδέα που πρότεινε ο Mardia (1975): επιλέγουμε το *μεγαλύτερο* από τους δύο συντελεστές συσχέτισης, δηλαδή

$$r = \max(r_+, r_-) \quad (5.3)$$

Επισημαίνουμε ότι οι παραμετρικές μέθοδοι υπολογισμού του συντελεστή συσχέτισης είναι εφαρμόσιμες μόνο υπό την αυστηρή συνθήκη της ομοιόμορφης κατανομής των κυκλικών μεταβλητών ϕ και ψ .



Σχήμα 5.3: Γραμμικό διάγραμμα, ανάλογο με αυτό του σχήματος 5.1, σελ. 130, για το ίδιο δείγμα ζευγαρωτών παρατηρήσεων των κυκλικών μεταβλητών ϕ και ψ , με διαφορετικό, όμως, αρχικό σημείο στον οριζόντιο άξονα (180°). Η διάταξη αυτή των παρατηρήσεων φαίνεται να υποδηλώνει αρνητική συσχέτιση. Ο συντελεστής συσχέτισης των ϕ και ψ είναι θετικός ή αρνητικός; Μια απάντηση στο ερώτημα αυτό δίνει ο Mardia (1975) με το συντελεστή συσχέτισης της εξ. (5.3).

5.2 Συσχέτιση κυκλικής μεταβλητής με γραμμική μεταβλητή (C-L correlation)

Είναι πολλές οι περιπτώσεις όπου μια κυκλική μεταβλητή ϕ σχετίζεται με μία γραμμική μεταβλητή y .¹ Τυπικό παράδειγμα αποτελεί η κατεύθυνση του ανέμου μετρούμενη σε μοίρες (κυκλική μεταβλητή — γωνία) που είναι συχνά συσχετισμένη με γραμμικές ποσότητες, όπως υγρασία, ύψος ατμοσφαιρικών κατακρημνισμάτων, συγκεντρώσεις ρυπαντών, ένταση οσμών κ.ά. Ακόμα, στις περιπτώσεις προσανατολισμού των ζώων, η κατεύθυνση του ανέμου μπορεί να συσχετίζεται με τον αριθμό των μετακινούμενων ζώων προς τον προορισμό

¹ Στη βιβλιογραφία, η συσχέτιση κυκλικής με γραμμική μεταβλητή αναφέρεται ως linear-circular (L-C) correlation [όπως στον Mardia (1976)], αλλά και ως circular-linear (C-L) correlation [όπως στον Batschelet (1981)]. Η διάκριση δεν είναι προφανής. Πιθανώς, η σειρά αναφοράς σημειώνει τη μελετούμενη σχέση εξάρτησης — π.χ. αν θεωρούμε ότι η γραμμική μεταβλητή πρέπει να εξαρτάται (γραμμικά) από την κυκλική, τότε αναφέρουμε L-C συσχέτιση, δηλαδή μια πιθανή σχέση L(C). Μάλιστα, οι Fisher and Lee (1992, p. 665) αναφέρουν ότι η συσχέτιση μεταξύ μιας κυκλικής ϕ και μιας γραμμικής y μεταβλητής προσδιορίζεται από τη φύση του μοντέλου παλινδρόμησης (δηλαδή, αν η ϕ παλινδρομείται επί της y ή η y επί της ϕ), με αρκετά διαφορετικούς τύπους συντελεστών συσχέτισης να είναι κατάλληλοι κατά περίπτωση. Σε ό,τι αφορά στην παραμετρική μέθοδο υπολογισμού του συντελεστή συσχέτισης που πρότεινε ο Mardia (1976), η σειρά αναφοράς φαίνεται να σημειώνει τη μεταβλητή που κατατάσσεται και τη μεταβλητή στην οποία αποδίδονται οι βαθμοί. Έτσι, οι Mardia and Jupp (2000, pp. 246–248) στη μέθοδο που αναφέρουν ως L-C correlation κατατάσσουν τη γραμμική μεταβλητή και αποδίδουν βαθμούς στην κυκλική, ενώ ο Batschelet (1981, pp. 191–195) στη μέθοδο που αναφέρει ως C-L correlation εφαρμόζει το αντίστροφο.

τους ή με τη διανυθείσα απόσταση. Υπάρχουν επίσης πολλά παραδείγματα στο χώρο των βιολογικών ρυθμών, όπως για παράδειγμα, ο ρυθμός των γεννήσεων που μπορεί να σχετίζεται με το χρόνο σε καθημερινή (24ωρη) ή σε ετήσια (με περίοδο 365 ημέρες) βάση. Αντίστροφα, το ακρότατο ενός βιολογικού ρυθμού μπορεί να εξαρτάται από την ηλικία ή τη συγκέντρωση στο αίμα κάποιου παράγοντα.

Αν (ϕ, y) είναι μια διμεταβλητή κατανομή, όπου ο δειγματικός χώρος της τυχαίας μεταβλητής ϕ είναι ο κύκλος και ο δειγματικός χώρος της τυχαίας μεταβλητής y είναι η πραγματική ευθεία, τότε δεν μπορούν να εφαρμοστούν οι συνήθεις τεχνικές της ανάλυσης συσχέτισης, καθώς οι τεχνικές αυτές προϋποθέτουν ως δειγματικό χώρο το επίπεδο. Στην περίπτωση όμως μιας τυχαίας κυκλικής-γραμμικής διμεταβλητής κατανομής, ο δειγματικός χώρος είναι ο *κύλινδρος* (Mardia, 1976, p. 403) (Solow et al., 1988, p. 220).

Ένα βασικό υπόδειγμα αποτελεί η παλινδρόμηση μιας κυκλικής μεταβλητής ϕ επί μιας μόνο ερμηνευτικής γραμμικής μεταβλητής x . Τότε, η συνάρτηση παλινδρόμησης

$$\phi(x) = E(\phi|x)$$

δηλαδή, η υπό συνθήκη αναμενόμενη τιμή της ϕ με δεδομένη την τιμή x , μπορεί να εννοηθεί ως μια καμπύλη πάνω στην επιφάνεια ενός άπειρου κυλίνδρου, η οποία εκτελεί μία περιστροφή γύρω από τον κύλινδρο, καθώς το x εκτείνεται από $-\infty$ έως $+\infty$. Άλλοι πιθανοί τύποι συνάρτησης παλινδρόμησης αυτού του είδους είναι συναρτήσεις όπου η καμπύλη παλινδρόμησης περιστρέφεται περισσότερες από μία φορές γύρω από την επιφάνεια του κυλίνδρου (Fisher and Lee, 1992, p. 667).

5.2.1 Παραμετρική μέθοδος

Έστω τυχαίο δείγμα ζευγαρωτών παρατηρήσεων

$$(\phi_1, y_1), (\phi_2, y_2), \dots, (\phi_n, y_n)$$

n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν τις τιμές μιας κυκλικής μεταβλητής ϕ (κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου) και των αντίστοιχων ποσοτήτων y_1, y_2, \dots, y_n μιας γραμμικής μεταβλητής y , η οποία, κατά κάποιον τρόπο, εξαρτάται από τη ϕ . Υποθέτουμε ότι η γραμμική μεταβλητή y κατανέμεται κανονικά με σταθερή διακύμανση σ^2 και

μέση τιμή $E(y_i)$ που δίνεται από την παρακάτω εξ. (5.4) και ότι τα n ζεύγη των παρατηρήσεων είναι ανεξάρτητα μεταξύ τους. Θέλουμε να ελέγξουμε την τυχόν ύπαρξη κάποιας σχέσης μεταξύ των δύο μεταβλητών, υπολογίζοντας έναν κατάλληλο συντελεστή συσχέτισης. Οι Mardia (1976) και Johnson and Wehrly (1977) πρότειναν ένα παραμετρικό τεστ για τον έλεγχο αυτό, το οποίο μπορεί να βασιστεί στο υπόδειγμα

$$E(y_i) = a_0 + a_1 \cos(\phi - \phi_0) \quad (5.4)$$

όπου a_0 είναι το μέσο επίπεδο (mean level), a_1 το πλάτος² (amplitude) και ϕ_0 η ακρότατη γωνία³ (acrophase angle). Αν και η παραπάνω εξίσωση είναι μη γραμμική ως προς τις παραμέτρους της, μπορεί να μετασχηματιστεί στη γραμμική εξίσωση

$$E(y_i) = a_0 + b_1 \cos \phi_i + b_2 \sin \phi_i \quad (5.5)$$

όπου είναι

$$a_1 = \sqrt{b_1^2 + b_2^2} \quad (5.6)$$

$$\phi_0 = \begin{cases} \arctan(b_2/b_1), & (b_1 > 0) \\ \pi + \arctan(b_2/b_1) & (b_1 < 0) \end{cases} \quad (5.7)$$

Με βάση τα παραπάνω δεδομένα, μπορούμε να ισχυριστούμε ότι η γραμμή παλινδρόμησης θα είναι προσεγγιστικά ημιτονοειδής (Batschelet, 1981, p. 191).

Τότε, ο συντελεστής συσχέτισης του πληθυσμού $\rho_{y\phi}^2$, ισούται με

$$\rho_{y\phi}^2 = \frac{4(V_C^2 E_{YS}^2 + V_S^2 E_{YC}^2 - \beta_2 E_{YC} E_{YS})}{\sigma^2(4V_C^2 V_S^2 - \beta_2^2)} \quad (5.8)$$

όπου είναι

$$E(y) = 0, E(\sin \phi) = 0, E(y \sin \phi) = E_{yS}, E(y \cos \phi) = E_{yC}, \beta_2 = E(\sin 2\phi)$$

και

$$Var(y) = \sigma^2, Var(\sin \phi) = V_S^2, Var(\cos \phi) = V_C^2$$

(Mardia, 1976, p. 403) (Liddell and Ord, 1978, p. 448).

Κατά την προσέγγιση αυτή, την οποία οι Mardia and Jupp (2000, p. 245) χα-

²Το πλάτος αναπαριστά την ένταση της επίδρασης της κατεύθυνσης.

³Η ακρότατη γωνία σημειώνει την κατεύθυνση της μέγιστης επίδρασης.

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

ρακτηρίζουν ως *ενσωματωμένη*,⁴ η κυκλική μεταβλητή ϕ μπορεί να εκληφθεί ως το τυχαίο διάνυσμα

$$\phi = (\cos \phi, \sin \phi)^T$$

στο επίπεδο. Ένα κατάλληλο μέτρο της (γραμμικής) εξάρτησης μεταξύ των ϕ και y , βασισμένο στο τυχαίο δείγμα

$$(\phi_1, y_1), (\phi_2, y_2), \dots, (\phi_n, y_n)$$

είναι ο δειγματικός συντελεστής συσχέτισης $r_{y\phi}$ μεταξύ των ϕ και y , δηλαδή της *μέγιστης* δειγματικής συσχέτισης μεταξύ της y και γραμμικών συναρτήσεων $\mathbf{a}^T \phi$ του ϕ . Αυτό σημαίνει ότι, προκειμένου να εφαρμοστεί στην (5.5) η συνήθης ανάλυση γραμμικής συσχέτισης, θα πρέπει να προσδιοριστεί κατάλληλα η ακρότατη γωνία ϕ_0 , κάτι που μπορεί να γίνει με την εφαρμογή μιας συνάρτησης συνημίτονο στην εξ. (5.4). Η γωνία ϕ_0 προσδιορίζεται έτσι, ώστε να μεγιστοποιείται ο συντελεστής συσχέτισης $r_{y\phi}^2$ (Batschelet, 1981, pp. 192–193).

Αν θέσουμε

$$\begin{aligned} r_{yC} &= \text{corr}(y, \cos \phi) \\ r_{yS} &= \text{corr}(y, \sin \phi) \\ r_{CS} &= \text{corr}(\cos \phi, \sin \phi), \end{aligned}$$

ως τους μερικούς συντελεστές συσχέτισης στην εξ. (5.5), τότε ο συντελεστής συσχέτισης μεταξύ των μεταβλητών y και ϕ ορίζεται ως (Mardia, 1976, p. 404)

$$r_{y\phi}^2 = \frac{r_{yC}^2 + r_{yS}^2 - 2r_{yC}r_{yS}r_{CS}}{1 - r_{CS}^2} \quad (5.9)$$

Υπό τις παραπάνω παραδοχές, δηλαδή της ανεξαρτησίας μεταξύ των ζευγών (ϕ, y_i) και της κανονικότητας και της σταθερής διακύμανσης των y_i , η ποσότητα

$$\frac{(n-3)r_{y\phi}^2}{1 - r_{y\phi}^2} \sim F_{2, n-3} \quad (5.10)$$

κατανέμεται σύμφωνα με την F -κατανομή με 2 και $n-3$ βαθμούς ελευθερίας (Mardia and Jupp, 2000, p. 246).

Οι Liddell and Ord (1978, pp.448-450) έδωσαν την ακριβή κατανομή αυτού

⁴Embedding approach — η συγκεκριμένη μεθοδολογία λαμβάνει τις κατευθύνσεις ως *μιγαδικούς αριθμούς* στο επίπεδο.

του συντελεστή συσχέτισης. Αν θεωρήσουμε την υπό συνθήκη κατανομή της y , δεδομένης της ϕ , τότε, αφού είναι $\rho_{y\phi} = 1$, όταν είναι

$$y = a + b_1 \cos \phi + b_2 \sin \phi \quad \text{με πιθανότητα 1,}$$

μια φυσική παραμετρική εξειδίκευση είναι να θέσουμε

$$y_i | \phi_i \sim N(a + b_1 \cos \phi_i + b_2 \sin \phi_i, \sigma^2) \quad (i = 1, 2, \dots, n).$$

Αν στο παραπάνω σχήμα εφαρμόσουμε γραμμική παλινδρόμηση με τη συνήθη μέθοδο των ελαχίστων τετραγώνων και με e_i συμβολίσουμε τα κατάλοιπα

$$e_i = y - \hat{y} = y - \hat{a} + \hat{b}_1 \cos \phi + \hat{b}_2 \sin \phi,$$

τότε, με $\phi^* = (\phi_1, \phi_2, \dots, \phi_n)$, είναι (Mardia and Sutton, 1978)

$$\frac{\sum_{i=1}^n e_i^2}{\sigma^2} | \phi^* \sim \chi_{n-3}^2$$

και

$$\frac{\sum_{i=1}^n (y_i - \hat{a})^2 - \sum_{i=1}^n e_i^2}{\sigma^2} | \phi^* \sim \chi_2^2(\Delta)$$

όπου το $\chi^2(\Delta)$ δηλώνει τη μη κεντρική χ^2 κατανομή, με παράμετρο

$$\Delta = \frac{\sum_{i=1}^n (b_1 \cos \phi_i + b_2 \sin \phi_i)^2}{\sigma^2}$$

Όταν είναι $\Delta = 0$, το ερμηνευόμενο άθροισμα στην παραπάνω παλινδρόμηση έχει κατανομή, η οποία είναι ανεξάρτητη από το ϕ^* και δίνει το τυπικό αποτέλεσμα

$$U = \frac{(n-3)r_{y\phi}^2}{2(1-r_{y\phi}^2)} \sim F_{2, n-3},$$

το οποίο ισχύει για κάθε δειγματική κατανομή ϕ^* . Το αποτέλεσμα αυτό εξειδικεύει το στατιστικό της εξ. (5.10).

Υπό τη μηδενική υπόθεση της μη συσχέτισης ($\rho_{y\phi} = 0$) και για μεγάλο n , μπορεί να αποδειχθεί ότι είναι (Batschelet, 1981, p. 193)

$$n r_{y\phi}^2 \simeq \chi_2^2 \quad (5.11)$$

Ακόμα, αν υποθέσουμε ότι οι μεταβλητές y και ϕ κατανέμονται ακολου-

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

θώντας τη διμεταβλητή κανονική κατανομή, τότε μπορούμε να σχηματίσουμε όρια εμπιστοσύνης για τον άγνωστο συντελεστή συσχέτισης ρ του πληθυσμού. Αυτό αναμφισβήτητα αποτελεί πλεονέκτημα της παραπάνω μεθόδου.

Πρέπει να επισημανθεί ότι στην περίπτωση της C-L συσχέτισης δεν γίνεται διάκριση μεταξύ θετικής και αρνητικής συσχέτισης, όπως γίνεται στην περίπτωση δύο κυκλικών μεταβλητών.

Πίνακας 5.1: Έλεγχος συσχέτισης των μεταβλητών πρώτης τάξης WINDIR_Hxx και TEMPER_Hxx με το πρόγραμμα *Oriana*. Υπολογίζεται ο συντελεστής συσχέτισης $r_{y\phi}^2$ της 5.9, του οποίου η σημαντικότητα ελέγχεται από τη σχέση 5.10. Γενικά, τα αποτελέσματα υποδεικνύουν τη μη συσχέτιση των μεταβλητών.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_H1 & TEMPER_H1 (26)	0,456	0,008
WINDIR_H2 & TEMPER_H2 (30)	0,122	0,671
WINDIR_H3 & TEMPER_H3 (26)	0,18	0,475
WINDIR_H4 & TEMPER_H4 (28)	0,227	0,276
WINDIR_H5 & TEMPER_H5 (26)	0,176	0,492
WINDIR_H6 & TEMPER_H6 (25)	0,192	0,444
WINDIR_H7 & TEMPER_H7 (29)	0,159	0,518
WINDIR_H8 & TEMPER_H8 (0)	***	***
WINDIR_H9 & TEMPER_H9 (30)	0,321	0,061
WINDIR_H10 & TEMPER_H10 (31)	0,352	0,03
WINDIR_H11 & TEMPER_H11 (31)	0,387	0,015
WINDIR_H12 & TEMPER_H12 (31)	0,107	0,727
WINDIR_H13 & TEMPER_H13 (31)	0,159	0,492
WINDIR_H14 & TEMPER_H14 (31)	0,117	0,68
WINDIR_H15 & TEMPER_H15 (31)	0,255	0,162
WINDIR_H16 & TEMPER_H16 (31)	0,139	0,583
WINDIR_H17 & TEMPER_H17 (31)	0,198	3,33E-01
WINDIR_H18 & TEMPER_H18 (31)	0,251	0,171
WINDIR_H19 & TEMPER_H19 (31)	0,168	0,456
WINDIR_H20 & TEMPER_H20 (30)	0,239	0,213
WINDIR_H21 & TEMPER_H21 (30)	0,551	0,000208
WINDIR_H22 & TEMPER_H22 (30)	0,312	0,072
WINDIR_H23 & TEMPER_H23 (30)	0,128	0,642
WINDIR_H24 & TEMPER_H24 (25)	0,202	0,408

*** = δεν μπορεί να γίνει υπολογισμός
(διαφορετικός αριθμός παρατηρήσεων)

Συσχετίζεται η ωριαία κατεύθυνση του ανέμου με την, αντίστοιχη, ωριαία θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών $WINDIR_Hxx$ και $TEMPER_Hxx$ με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Η εφαρμογή των ελέγχων κανονικότητας για τις γραμμικές μεταβλητές $TEMPER_Hxx$ δεν έδωσε σημαντικά αποτελέσματα, παρά μόνο σε μια, οριακή περίπτωση ($TEMPER_H5$). Αυτό σημαίνει ότι οι πρώτης τάξης μεταβλητές $TEMPER_Hxx$ δεν αποκλίνουν από την κανονικότητα. Τόσο τα τεστ κανονικότητας^{α'}, όσο και τα τεστ καλής προσαρμογής (της κανονικής κατανομής),^{β'} υποδεικνύουν ισχυρές κανονικότητες.

Συνεπώς, μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος κυκλικής-γραμμικής συσχέτισης του *Mardia*.

Στα δεδομένα του πρώτου συνόλου δειγμάτων $\{(WINDIR_Hxx, TEMPER_Hxx)\}$ εφαρμόστηκε το παρακάτω μοντέλο παλινδρόμησης με δύο ερμηνευτικές μεταβλητές

$$TEMPER_Hxx = a_0 + b_1 \cos(WINDIR_Hxx) + b_2 \sin(WINDIR_Hxx) \quad (5.12)$$

Οι παλινδρομήσεις (5.12), στο σύνολό τους, δεν δίνουν σημαντικούς συντελεστές b_1 και b_2 , παρά μόνον σε τέσσερις περιπτώσεις, όπου δίνουν τουλάχιστον ένα σημαντικό συντελεστή, από τις οποίες μόνον η παλινδρόμηση $WINDIR_H21/TEMPER_H21$ δίνει και τους δύο συντελεστές κλίσης σημαντικούς. Τα R^2 , γενικά, είναι χαμηλά έως πολύ χαμηλά.

^{α'} Chi-squared, Shapiro-Wilks W , Z score for skewness και Z score for kurtosis.

^{β'} Chi-squared, Kolmogorov-Smirnov D και Anderson-Darling A^2 .

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Με βάση τα παραπάνω αποτελέσματα και λόγω της συγραμμικότητας που παρουσιάζεται στις (5.12), αφού για τις δύο ερμηνευτικές μεταβλητές είναι, βεβαίως,

$$\sin(WINDIR_Hxx) = \sqrt{1 - \cos^2(WINDIR_Hxx)}$$

οι παλινδρομήσεις (5.12) εγκαταλήφθηκαν και εφαρμόστηκαν παλινδρομήσεις με μόνη ερμηνευτική την $\cos(WINDIR_Hxx)$

$$TEMPER_Hxx = a_0 + a_1 \cos(WINDIR_Hxx) \quad (5.13)$$

Οι παλινδρομήσεις (5.13), στο σύνολό τους, δεν δίνουν σημαντικούς συντελεστές a_1 . Τα R^2 , γενικά, είναι πολύ χαμηλά.

Σε συμφωνία με τα παραπάνω, το πρόγραμμα *Oriana* δίνει μόνο μόνον τέσσερα σημαντικά αποτελέσματα (βλ. πίνακα 5.1, σελ. 138).

Μια σύνοψη των παραπάνω παρουσιάζεται στους πίνακες 5.20 και 5.21, σσ. 174–175

Πίνακας 5.2: Αποτελέσματα των ελέγχων κανονικότητας και καλής προσαρμογής για τις μεταβλητές $TEMPER_Dxx$. Οι έλεγχοι έγιναν με το στατιστικό πρόγραμμα *STATGRAPHICS* και αφορούν σε τεστ κανονικότητας (Chi-squared, Shapiro-Wilks W , Z score for skewness και Z score for kurtosis και τεστ καλής προσαρμογής (της κανονικής κατανομής) (Chi-squared, Kolmogorov-Smirnov D και Anderson-Darling A^2). Από τα στοιχεία του πίνακα φαίνεται ότι οι μεταβλητές $TEMPER_Dxx$ γενικά αποκλίνουν από την κανονικότητα. Οι 28 από τις 31 εμφανίζουν σημαντικά αποτελέσματα τουλάχιστον σε 1 τεστ. Οι 25 από τις 31 εμφανίζουν σημαντικά αποτελέσματα σε 3 ή περισσότερα τεστ. Μόνο 3 μεταβλητές (οι $TEMPER_D6$, -12 , -26) δεν δίνουν σημαντικότητες σε κανένα τεστ.

Αριθμός ελέγχων με σημαντικά αποτελέσματα	Μεταβλητές ($WINDIR_$)	Συχνότητα
0	D6, D12, D26	3
1	D4, D15	2
2	D19	1
3	D2, D3, D5, D8, D9, D11, D14, D16, D17	9
4	D1, D18, D22, D23, D24, D28	6
5	D7, D20, D25, D29, D31	5
6	D10, D13, D21, D27, D30	5
7	-	0

Συσχετίζεται η ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών $WINDIR_Dxx$ και $TEMPER_Dxx$ με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Όπως φαίνεται από τον πίνακα (5.2), σελ. 140, οι μεταβλητές $TEMPER_Dxx$ γενικά αποκλίνουν από την κανονικότητα. Αυτό σημαίνει ότι η θερμοκρασία δεν κατανέμεται κανονικά κατά τη διάρκεια της ημέρας.

Συνεπώς, για τα δεδομένα του συνόλου των δειγμάτων $\{(WINDIR_Dxx, TEMPER_Dxx)\}$ δεν μπορεί να γίνει εφαρμογή του παραμετρικού ελέγχου κυκλικής-γραμμικής συσχέτισης του *Mardia*.

Παρά το ότι παραβιάζεται η συνθήκη της κανονικότητας της γραμμικής μεταβλητής, έγινε εφαρμογή των παλινδρομήσεων $WINDIR_Dxx / TEMPER_Dxx$, σύμφωνα με το υπόδειγμα (5.4), σελ. 135. Όπως φαίνεται από τον πίνακα 5.4, σελ. 143, για την παλινδρόμηση με δύο ερμηνευτικές $[\cos(TEMPER_Dxx), \sin(TEMPER_Dxx)]$ έχουμε 20 περιπτώσεις με τουλάχιστον 1 σημαντικό συντελεστή και 11 περιπτώσεις με κανένα σημαντικό συντελεστή. Μόνο 7 όμως είχαν 2 συντελεστές σημαντικούς. Για την παλινδρόμηση με ερμηνευτική μεταβλητή μόνο την $\cos(TEMPER_Dxx)$ έχουμε 10 σημαντικά αποτελέσματα, ενώ για την παλινδρόμηση με ερμηνευτική μεταβλητή την $\sin(TEMPER_Dxx)$ έχουμε 6 σημαντικά αποτελέσματα. 16 μεταβλητές δεν έδωσαν σημαντικά αποτελέσματα σε καμμία από τις παραπάνω δύο απλές παλινδρομήσεις. Γενικά, φαίνεται ότι η παλινδρόμηση με ερμηνευτική μεταβλητή την $\cos(TEMPER_Dxx)$ δίνει ελαφρώς περισσότερα σημαντικά αποτελέσματα από τις άλλες παλινδρομήσεις.

Επισημαίνουμε ότι τα παραπάνω αποτελέσματα είναι αναξιόπιστα και για τον έλεγχο της συσχέτισης των συγκεκριμένων μεταβλητών θα πρέπει να εφαρμοστεί κάποιο κατάλληλο απαραμετρικό τεστ.

Μια σύνοψη των παραπάνω παρουσιάζεται στους πίνακες 5.17, 5.18 και 5.19, σσ. 171–173

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πίνακας 5.3: Παραμετρικός έλεγχος συσχέτισης των μεταβλητών πρώτης τάξης WINDIR_Dxx και TEMPER_Dxx με το πρόγραμμα *Oriana*. Υπολογίζεται ο συντελεστής συσχέτισης $r_{y\phi}^2$ της 5.9, του οποίου η σημαντικότητα ελέγχεται από τη σχέση 5.10. Τα αποτελέσματα υποδεικνύουν τη συσχέτιση των μεταβλητών, κάτι που φαίνεται να επαληθεύεται και από τον απαραμετρικό έλεγχο, της επόμενης υποενοότητας.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_D1 & TEMPER_D1 (24)	0,443	0,015
WINDIR_D2 & TEMPER_D2 (24)	0,559	0,001
WINDIR_D3 & TEMPER_D3 (0)	***	***
WINDIR_D4 & TEMPER_D4 (18)	0,619	0,002
WINDIR_D5 & TEMPER_D5 (21)	0,53	0,006
WINDIR_D6 & TEMPER_D6 (0)	***	***
WINDIR_D7 & TEMPER_D7 (24)	0,366	0,059
WINDIR_D8 & TEMPER_D8 (21)	0,414	0,044
WINDIR_D9 & TEMPER_D9 (24)	0,572	8,06E-04
WINDIR_D10 & TEMPER_D10 (0)	***	***
WINDIR_D11 & TEMPER_D11 (23)	0,516	0,004
WINDIR_D12 & TEMPER_D12 (24)	0,622	1,92E-04
WINDIR_D13 & TEMPER_D13 (24)	0,752	1,42E-06
WINDIR_D14 & TEMPER_D14 (24)	0,371	0,054
WINDIR_D15 & TEMPER_D15 (24)	*****	*****
WINDIR_D16 & TEMPER_D16 (24)	*****	*****
WINDIR_D17 & TEMPER_D17 (24)	*****	*****
WINDIR_D18 & TEMPER_D18 (24)	*****	*****
WINDIR_D19 & TEMPER_D19 (22)	0,658	1,52E-04
WINDIR_D20 & TEMPER_D20 (23)	0,804	2,10E-07
WINDIR_D21 & TEMPER_D21 (22)	0,622	4,33E-04
WINDIR_D22 & TEMPER_D22 (0)	***	***
WINDIR_D23 & TEMPER_D23 (20)	0,62	0,001
WINDIR_D24 & TEMPER_D24 (0)	***	***
WINDIR_D25 & TEMPER_D25 (23)	0,781	7,04E-07
WINDIR_D26 & TEMPER_D26 (23)	0,346	0,09
WINDIR_D27 & TEMPER_D27 (23)	0,444	0,018
WINDIR_D28 & TEMPER_D28 (23)	0,715	1,26E-05
WINDIR_D29 & TEMPER_D29 (24)	0,64	1,08E-04
WINDIR_D30 & TEMPER_D30 (0)	***	***
WINDIR_D31 & TEMPER_D31 (23)	0,428	0,025

*** = δεν μπορεί να γίνει υπολογισμός
(διαφορετικός αριθμός παρατηρήσεων)

***** = δεν μπορεί να γίνει υπολογισμός

Πίνακας 5.4: Αποτελέσματα των παλινδρομήσεων $WINDIR_{Dxx} / TEMPER_{Dxx}$. Τα αποτελέσματα αυτά είναι αναξιόπιστα, καθώς δεν πληρείται η συνθήκη της κανονικότητας των γραμμικών μεταβλητών $TEMPER_{Dxx}$ για την εφαρμογή του υποδείγματος (5.4), σελ. 135.

Αριθμός σημαντικών συντελεστών (Παλινδρομητές)	Συχνότητα (Παλινδρομητές)			
	$\cos(TEMPER_{-})$	$\cos(TEMPER_{-})$	$\sin(TEMPER_{-})$	$\sin(TEMPER_{-})$
2	7 (D10, D11, D19, D21, D22, D23, D28)	-	-	-
1 [$\cos(TEMPER_{-})$]	7 (D4, D12, D13, D20, D24, D25, D31)	10 (D3, D4, D9, D12, D13, D17, D19, D20, D24, D25)	-	-
1 [$\sin(TEMPER_{-})$]	6 (D1, D2, D5, D6, D27, D29)	-	-	6 (D2, D5, D10, D19, D28, D29)

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πίνακας 5.5: Παραμετρικός έλεγχος συσχέτισης των μεταβλητών δεύτερης τάξης WINDIR_HOUR και TEMPER_HOUR με το πρόγραμμα *Oriana*. Υπολογίζεται ο συντελεστής συσχέτισης $r_{y\phi}^2$ της 5.9, του οποίου η σημαντικότητα ελέγχεται από τη σχέση 5.10. Το πάρα πολύ σημαντικό αποτέλεσμα που προκύπτει υποδεικνύει ισχυρή συσχέτιση της μέσης ωριαίας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέση ωριαία θερμοκρασία.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_HOUR & TEMPER_HOUR (24)	0,841	9,27E-09

Συσχετίζεται η μέση ωριαία κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ωριαία θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_HOUR και TEMPER_HOUR με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Από τις συνθήκες για την εφαρμογή αυτού του παραμετρικού ελέγχου, όπως αυτές αναφέρθηκαν παραπάνω, θεωρούμε ότι η ανεξαρτησία των μεταβλητών της κατεύθυνσης του ανέμου και της θερμοκρασίας πληρείται σε κάθε περίπτωση.

Σε ό,τι αφορά στη δεύτερη συνθήκη της κανονικότητας της γραμμικής μεταβλητής της θερμοκρασίας, τα τεστ κανονικότητας φαίνεται να απορρίπτουν την κανονικότητα της TEMPER_HOUR (πλην του ελέγχου της λοξότητας, οι υπόλοιποι έλεγχοι δίνουν πολύ σημαντικά αποτελέσματα). Από τα τεστ καλής προσαρμογής, το Anderson-Darling απορρίπτει την κανονικότητα της TEMPER_HOUR, όμως τα άλλα τη δέχονται.

Με την επιφύλαξη των παραπάνω, εφαρμόστηκε στα δεδομένα το μοντέλο παλινδρόμησης με δύο ερμηνευτικές μεταβλητές

$$TEMPER_HOUR = 55.116 + 27.3243 \cos(WINDIR_HOUR) + 2.8982 \sin(WINDIR_HOUR) \quad (5.14)$$

Η παλινδρόμηση της (5.14) δίνει πολύ σημαντικό το συντελεστή της $\cos(WINDIR_HOUR)$ ($P = 0.0013$) και μη σημαντικό το συντελεστή της $\sin(WINDIR_HOUR)$ ($P = 0.4510$). Το $R^2 = 70.68\%$ ^{α'} είναι πολύ ικανοποιητικό.

^{α'} Δηλαδή, το $r_{y\phi}^2$ της (5.10).

5.2. Κυκλική-γραμμική συσχέτιση

Λόγω της συγραμμικότητας των ερμηνευτικών μεταβλητών της (5.14), η συγκεκριμένη παλινδρόμηση εγκαταλήφθηκε και ακολούθησε παλινδρόμηση με μόνη ερμηνευτική την $\cos(WINDIR_HOUR)$

$$TEMPER_HOUR = 49.4408 + 22.1788 \cos(WINDIR_HOUR) \quad (5.15)$$

Η παλινδρόμηση της (5.15) δίνει πάρα πολύ σημαντικό το συντελεστή της $\cos(WINDIR_HOUR)$ ($P = 0$). Το $R^2 = 69.85\%$ είναι, επίσης, πολύ ικανοποιητικό. Τότε, για μέγεθος του δείγματος $n = 24$ και $r_{y\phi}^2 = 0.6985$, η τιμή του στατιστικού της (5.10) είναι $48.66 > F_{0.005,2,21} \simeq 6.99$, επομένως η μηδενική υπόθεση $H_0 : \rho = 0$ απορρίπτεται σε επίπεδο σημαντικότητας $\alpha = 0.005$, συνεπώς και σε επίπεδο σημαντικότητας $\alpha = 0.01$ ή $\alpha = 0.05$.

Σε συμφωνία με τα παραπάνω, ο έλεγχος με το στατιστικό πρόγραμμα *Oriana* δίνει τον πάρα πολύ σημαντικό συντελεστή συσχέτισης $r = 0.841$ ($P = 9.27E - 09$) (βλ. πίνακα 5.5).

Δηλαδή, μπορούμε να ισχυριστούμε ότι η μέση ωριαία κατεύθυνση του ανέμου συσχετίζεται με την, αντίστοιχη, μέση ωριαία θερμοκρασία, με τον τρόπο που υποδεικνύει το μοντέλο της (5.15).

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πίνακας 5.6: Παραμετρικός έλεγχος συσχέτισης των μεταβλητών δεύτερης τάξης WINDIR_DAY και TEMPER_DAY, για το σύνολο των παρατηρήσεων του μήνα, τις παρατηρήσεις των πρώτων 17 ημερών (φυσιολογικές τιμές θερμοκρασίας) και τις παρατηρήσεις των υπολοίπων 14 ημερών (περίοδος καύσωνα), με το πρόγραμμα *Oriana*. Υπολογίζεται ο συντελεστής συσχέτισης $r_{y\phi}^2$ της 5.9, του οποίου η σημαντικότητα ελέγχεται από τη σχέση 5.10. Το σημαντικό αποτέλεσμα που προκύπτει για το διάστημα των πρώτων 17 ημερών του μήνα, υποδεικνύει συσχέτιση της μέσης ημερήσιας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, υπό φυσιολογικές τιμές της θερμοκρασίας.

Variables (& observations)	r	P (Prob.)
WINDIR_DAY & TEMPER_DAY (31)	0,135	0,601
WINDIR_DAY[1-17] & TEMPER_DAY[1-17] (17)	0,463	0,048
WINDIR_DAY[18-31] & TEMPER_DAY[18-31] (14)	0,364	0,232

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το σύνολο των παρατηρήσεων του μήνα μελέτης;
(Έλεγχος συσχέτισης των μεταβλητών WINDIR_DAY και TEMPER_DAY με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Θεωρούμε ότι η συνθήκη της ανεξαρτησίας των μεταβλητών της κατεύθυνσης του ανέμου και της θερμοκρασίας πληρείται και σ' αυτή την περίπτωση.

Σε ό,τι αφορά στη συνθήκη της κανονικότητας της γραμμικής μεταβλητής της θερμοκρασίας, όλα τα τεστ κανονικότητας και όλα τα τεστ καλής προσαρμογής φαίνεται να δέχονται την κανονικότητα της TEMPER_DAY, δίνοντας μεγάλα P -value.

Στα δεδομένα εφαρμόστηκε το μοντέλο παλινδρόμησης με δύο ερμηνευτικές μεταβλητές

$$TEMPER_DAY = 30.6234 + 0.026189 \cos(WINDIR_DAY) + 0.485111 \sin(WINDIR_DAY) \quad (5.16)$$

Η παλινδρόμηση της (5.16) δεν δίνει σημαντικούς τους συντελεστές των $\cos(WINDIR_DAY)$ και $\sin(WINDIR_DAY)$ (μάλιστα, με πολύ υψηλά κρίσιμα επίπεδα, $P = 0.9813$ και $P = 0.6495$, αντίστοιχα). Το $R^2 = 1.81\%$ είναι πολύ χαμηλό.

5.2. Κυκλική-γραμμική συσχέτιση

Η συγκεκριμένη παλινδρόμηση εγκαταλήφθηκε (λόγω και της συγραμμικότητας των ερμηνευτικών) και ακολούθησε παλινδρόμηση με μόνη ερμηνευτική την $\cos(WINDIR_DAY)$

$$TEMPER_DAY = 30.7442 + 0.40804 \cos(WINDIR_DAY) \quad (5.17)$$

Η παλινδρόμηση της (5.17) δεν δίνει σημαντικό το συντελεστή της $\cos(WINDIR_HOUR)$ ($P = 0.5783$). Το $R^2 = 1.07\%$ είναι, επίσης, πολύ χαμηλό.

Μπορούμε να πούμε ότι ούτε η πρώτη, ούτε η δεύτερη παλινδρόμηση είναι στο ελάχιστο ικανοποιητικές.

Σε συμφωνία με τα παραπάνω, ο έλεγχος με το στατιστικό πρόγραμμα *Oriana* δεν δίνει σημαντικό συντελεστή συσχέτισης $r = 0.135$ ($P = 0.601$) (βλ. πίνακα 5.6).

Δηλαδή, μπορούμε να ισχυριστούμε ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το διάστημα των πρώτων 17 ημερών του μήνα, δηλαδή των ημερών με φυσιολογικές τιμές θερμοκρασίας; (Έλεγχος συσχέτισης των μεταβλητών $WINDIR_DAY[1-17]$ και $TEMPER_DAY[1-17]$ με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Θεωρούμε ότι η ανεξαρτησία των μεταβλητών της κατεύθυνσης του ανέμου και της θερμοκρασίας ισχύει και σ' αυτή την περίπτωση. Ακόμα, η γραμμική μεταβλητή $TEMPER_DAY[1-17]$ κατανέμεται κανονικά.

Στα δεδομένα εφαρμόστηκε το μοντέλο παλινδρόμησης με δύο ερμηνευτικές μεταβλητές

$$TEMPER_DAY[1-17] = 28.782 + 0.745242 \cos(WINDIR_DAY[1-17]) + 1.41352 \sin(WINDIR_DAY[1-17]) \quad (5.18)$$

Η παλινδρόμηση της (5.18) δεν δίνει σημαντικό το συντελεστή της $\cos(WINDIR_DAY[1-17])$ ($P = 0.3501$). Μπορούμε, όμως, να πούμε ότι ο συντελεστής της $\sin(WINDIR_DAY[1-17])$ είναι σημαντικός, έστω σε επίπεδο $\alpha = 0.10$ ($P = 0.0818$). Το $R^2 = 21.4082\%$ είναι χαμηλό.

Ακολούθησαν οι παλινδρομήσεις με μία ερμηνευτική, την $\cos(WINDIR_DAY[1-17])$ και την $\sin(WINDIR_DAY[1-17])$. Η πρώτη παλινδρόμηση δεν έδωσε σημαντικό συντελεστή ($P = 0.6215$), ενώ η δεύτερη παλινδρόμηση μπορούμε να πούμε ότι έδωσε έναν οριακό συντελεστή, σε επίπεδο σημαντικότητας $\alpha = 0.10$, όπως φαίνεται από την παρακάτω εξ. (5.19)

$$TEMPER_DAY[1-17] = 29.0121 + 0.890512 \sin(WINDIR_DAY[1-17]) \quad (5.19)$$

Το $R^2 = 16.1618\%$ της (5.19) είναι, επίσης, χαμηλό.

Πάντως, ο έλεγχος με το στατιστικό πρόγραμμα *Oriana* δίνει, έστω οριακά, σημαντικό συντελεστή συσχέτισης $r = 0.463$ ($P = 0.048$) (βλ. πίνακα 5.6).

Δηλαδή, μπορούμε να ισχυριστούμε ότι η μέση ημερήσια κατεύθυνση του ανέμου συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, υπό φυσιολογικές τιμές θερμοκρασίας.

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το διάστημα των τελευταίων 14 ημερών του μήνα, δηλαδή της περιόδου του καύσωνα; (Έλεγχος συσχέτισης των μεταβλητών $WINDIR_DAY[18-31]$ και $TEMPER_DAY[18-31]$ με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Θεωρούμε, επίσης, ότι πληρείται η συνθήκη της ανεξαρτησίας των μεταβλητών της κατεύθυνσης του ανέμου και της θερμοκρασίας. Ακόμα, η γραμμική μεταβλητή $TEMPER_DAY[18-31]$ κατανέμεται κανονικά.

Στα δεδομένα εφαρμόστηκε το μοντέλο παλινδρόμησης με δύο ερμηνευτικές μεταβλητές

$$TEMPER_DAY[1-17] = 28.782 + 0.745242 \cos(WINDIR_DAY[1-17]) + 1.41352 \sin(WINDIR_DAY[1-17]) \quad (5.20)$$

Η παλινδρόμηση της (5.20) δεν δίνει σημαντικούς τους συντελεστές των $\sin(WINDIR_DAY[18-31])$ και $\cos(WINDIR_DAY[18-31])$ ($P = 0.4814$ και $P = 0.2437$, αντίστοιχα).

Οι παλινδρομήσεις με μία ερμηνευτική, την $\cos(WINDIR_DAY[1-17])$ και την $\sin(WINDIR_DAY[1-17])$. Καμμία από τις δύο παλινδρομήσεις δεν έδωσε σημαντικό συντελεστή ($P = 0.2966$ και $P = 0.7041$, αντίστοιχα).

Σε συμφωνία με τα παραπάνω, ο έλεγχος με το στατιστικό πρόγραμμα *Oriana* δεν δίνει σημαντικό συντελεστή συσχέτισης r (βλ. πίνακα 5.6).

Δηλαδή, μπορούμε να ισχυριστούμε ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, υπό συνθήκες καύσωνα.

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πίνακας 5.7: Αποτελέσματα του παραμετρικού ελέγχου συσχέτισης των μεταβλητών WINDIR_DAY και TEMPER_DMAX. Φαίνεται ότι κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας υπάρχει συσχέτιση της μέσης ημερήσιας κατεύθυνσης του ανέμου με τη μέγιστη ημερήσια θερμοκρασία, κάτι που δεν φαίνεται να ισχύει για τις περιόδους με καύσωνα.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_DAY & TEMPER_DMAX (31)	0,27	0,129
WINDIR_DAY[1-17] & TEMPER_DMAX[1-17] (17)	0,65	0,002
WINDIR_DAY[18-31] & TEMPER_DMAX[18-31] (14)	0,296	0,381

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DAY και TEMPER_DMAX με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Ο έλεγχος κανονικότητας της μεταβλητής TEMPER_DMAX με το Kolmogorov-Smirnov test έδειξε ότι η μέγιστη ημερήσια θερμοκρασία κατανέμεται κανονικά ($P = 0.3574$).

Συνεπώς, για τις μεταβλητές WINDIR_DAY και TEMPER_DMAX μπορεί να γίνει εφαρμογή του παραμετρικού ελέγχου κυκλικής-γραμμικής συσχέτισης του *Mardia*. Από τα αποτελέσματα του πίνακα 5.7, σελ. 150, φαίνεται ότι για τη χρονική περίοδο του μήνα μελέτης, η μέση ημερήσια κατεύθυνση του ανέμου (WINDIR_DAY) δεν φαίνεται να συσχετίζεται με τη μέγιστη ημερήσια θερμοκρασία (TEMPER_DMAX). Το ίδιο φαίνεται να ισχύει και για την περίοδο του καύσωνα (ημέρες 18–31). Όμως, για το διάστημα των ημερών 1–17, τα πολύ σημαντικά αποτελέσματα που προκύπτουν ($P = 0.002$) υποδεικνύουν τη συσχέτιση των μεταβλητών. Φαίνεται, δηλαδή, ότι κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας υπάρχει συσχέτιση της μέσης ημερήσιας κατεύθυνσης του ανέμου με τη μέγιστη ημερήσια θερμοκρασία, κάτι που δεν φαίνεται να ισχύει για τις περιόδους με καύσωνα.

Πίνακας 5.8: Αποτελέσματα του παραμετρικού ελέγχου συσχέτισης των μεταβλητών WINDIR_DPREV και TEMPER_DAY. Φαίνεται ότι κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας υπάρχει συσχέτιση της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου με τη μέση ημερήσια θερμοκρασία, κάτι που δεν φαίνεται να ισχύει για τις περιόδους με καύσωνα.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_DPREV & TEMPER_DAY (31)	0,176	0,42
WINDIR_DPREV[1-17] & TEMPER_DAY[1-17] (17)	0,472	0,042
WINDIR_DPREV[18-31] & TEMPER_DAY[18-31] (14)	0,253	0,494

Συσχετίζεται η επικρατούσα ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DPREV και TEMPER_DAY με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Έχειδειχθεί ότι η μέση ημερήσια θερμοκρασία (TEMPER_DAY) κατανέμεται κανονικά.

Συνεπώς, για τις μεταβλητές WINDIR_DPREV και TEMPER_DAY μπορεί να γίνει εφαρμογή του παραμετρικού ελέγχου κυκλικής-γραμμικής συσχέτισης του *Mardia*. Σύμφωνα με τα στοιχεία του πίνακα (5.8), για το διάστημα του μήνα (31 ημέρες) δεν φαίνεται να υπάρχει συσχέτιση της επικρατούσας ημερήσιας κατεύθυνσης και της μέσης ημερήσιας θερμοκρασίας και το ίδιο φαίνεται να ισχύει για την περίοδο του κύσωνα. Όμως, τα σημαντικά αποτελέσματα του ελέγχου για το διάστημα των πρώτων 17 ημερών του μήνα ($P = 0.042$) δείχνουν ότι, υπό φυσιολογικές τιμές θερμοκρασίας, ίσως υπάρχει κάποια σχέση μεταξύ των δύο μεταβλητών.

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πίνακας 5.9: Αποτελέσματα του παραμετρικού ελέγχου συσχέτισης των μεταβλητών WINDIR_DPREV και TEMPER_DMAX. Φαίνεται ότι κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας υπάρχει συσχέτιση της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου με τη μέγιστη ημερήσια θερμοκρασία, κάτι που δεν φαίνεται να ισχύει για τις περιόδους με καύσωνα.

<i>Variables (& observations)</i>	<i>r</i>	<i>P (Prob.)</i>
WINDIR_DPREV & TEMPER_DMAX (31)	0,292	0,091
WINDIR_DPREV[1-17] & TEMPER_DMAX[1-17] (17)	0,66	0,001
WINDIR_DPREV[18-31] & TEMPER_DMAX[18-31] (14)	0,269	0,45

Συσχετίζεται η επικρατούσα ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DPREV και TEMPER_DMAX με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Από τα στοιχεία του πίνακα 5.9, φαίνεται να υπάρχει ισχυρή συσχέτιση μεταξύ της επικρατούσας ημερήσιας κατεύθυνσης (WINDIR_DPREV) και της μέγιστης ημερήσιας θερμοκρασίας (TEMPER_DMAX), όπως υποδεικνύουν τα πάρα πολύ σημαντικά αποτελέσματα του ελέγχου, για το διάστημα των 17 πρώτων ημερών του μήνα, δηλαδή των ημερών με φυσιολογικές τιμές θερμοκρασίας.

Ακόμα, μπορεί να θεωρηθεί ότι υπάρχει συσχέτιση των δύο μεταβλητών για το σύνολο των ημερών του μήνα, έστω σε επίπεδο σημαντικότητας $\alpha = 0.10$. Όμως, για το διάστημα των ημερών του καύσωνα, το υψηλό $P = 0.45$ υποδεικνύει ότι η επικρατούσα ημερήσια κατεύθυνση, δεν επηρεάζει την τιμή της μέγιστης ημερήσιας θερμοκρασίας.

5.2.2 Απαραμετρική μέθοδος

Οι συνθήκες για την εφαρμογή της παραπάνω παραμετρικής μεθόδου είναι αυστηρές. Μία εναλλακτική μέθοδος για την ανάλυση της L-C συσχέτισης, όπου δεν απαιτούνται ισχυρές προϋποθέσεις, αποτελεί η *συσχέτιση των βαθμών* (rank correlation) (Mardia, 1976).

Έστω τυχαίο δείγμα ζευγαρωτών παρατηρήσεων

$$(\phi_1, y_1), (\phi_2, y_2), \dots, (\phi_n, y_n)$$

n γωνιών $\phi_1, \phi_2, \dots, \phi_n$ οι οποίες αναπαριστούν τις τιμές μιας κυκλικής μεταβλητής ϕ (κατευθύνσεις στο επίπεδο ή χρονικές στιγμές ενός κυκλικού φαινομένου) και των αντίστοιχων ποσοτήτων y_1, y_2, \dots, y_n μιας γραμμικής μεταβλητής y , η οποία, κατά κάποιον τρόπο, εξαρτάται από τη ϕ . Δεν κάνουμε κάποια υπόθεση σχετικά με το πρότυπο της κατανομής των δειγματικών σημείων.⁵ Θεωρούμε μόνο ότι τα n ζεύγη των παρατηρήσεων είναι μεταξύ τους ανεξάρτητα. Θέλουμε να εξετάσουμε τη συσχέτιση μεταξύ των δύο μεταβλητών, υπολογίζοντας έναν κατάλληλο συντελεστή συσχέτισης.

Δεδομένου του παραπάνω δείγματος, ταξινομούμε τις παρατηρήσεις της γραμμικής μεταβλητής έτσι, ώστε να είναι

$$y_1 \leq y_2 \leq \dots \leq y_n$$

Αν θέσουμε ως r_1, r_2, \dots, r_n τους αντίστοιχους *κυκλικούς βαθμούς* (circular ranks) των παρατηρήσεων $\phi_1, \phi_2, \dots, \phi_n$ της κυκλικής μεταβλητής (όπως αυτές έχουν αναδιαταχθεί) και ορίσουμε τα *ομοιόμορφα βάρη* (uniform scores) ή απλώς *βάρη* $\beta_1, \beta_2, \dots, \beta_n$ ως

$$\beta_i = \frac{2\pi r_i}{n} \quad (5.21)$$

τότε παίρνουμε τα νέα ζεύγη παρατηρήσεων

$$(1, \beta_1), (2, \beta_2), \dots, (n, \beta_n)$$

Ως συντελεστής συσχέτισης μπορεί να χρησιμοποιηθεί η ποσότητα D_n (Mardia,

⁵ Πάντως, οι Solow et al. (1988, p. 221) αναφέρουν ότι το συγκεκριμένο απαραμετρικό τεστ είναι ευαίσθητο μόνο σε περιπτώσεις όπου η κατανομή της γραμμικής μεταβλητής εμφανίζει μία μόνο κορυφή και ένα μόνο κοίλο.

1976, p. 404), όπου

$$D_n = \alpha_n(C^2 + S^2) \quad (5.22)$$

με

$$C = \sum_{i=1}^n i \cos \beta_i, \quad S = \sum_{i=1}^n i \sin \beta_i$$

και

$$\alpha_n = \begin{cases} [1 + 5\cotan^2(\pi/n) + 4\cotan^4(\pi/n)]^{-1} & (n \text{ άρτιος}) \\ 2 \sin^4(\pi/n)/[1 + \cos(\pi/n)]^3 & (n \text{ περιττός}) \end{cases} \quad (5.23)$$

Αποδεικνύεται ότι είναι $0 \leq D_n \leq 1$. Ακόμα, επειδή το D_n βασίζεται σε βαθμούς και ομοιόμορφα βάρη, δεν επηρεάζεται από τυχόν αλλαγή της αρχής του άξονα της γραμμικής μεταβλητής y ή περιστροφή των αξόνων της κυκλικής μεταβλητής ϕ . Κατά συνέπεια, μπορούμε να βαθμονομήσουμε τις τιμές της κυκλικής μεταβλητής ϕ ξεκινώντας από οποιοδήποτε αρχικό σημείο και κατά την οποιαδήποτε φορά (των δεικτών του ρολογιού ή την αντίστροφη) (Mardia, 1976, p. 404) (Mardia and Jupp, 2000, p. 247).

Απορρίπτουμε τη μηδενική υπόθεση της μη συσχέτισης των y και ϕ για μεγάλες τιμές του D_n (Mardia, 1976, p. 404).

Για μεγάλο n , αποδεικνύεται ότι

$$U_n = \frac{24(C^2 + S^2)}{n^2(n+1)} \sim \chi_2^2 \quad (5.24)$$

(Mardia, 1976, p. 404).

Υπολογιστικά, είναι προτιμότερο να βασίσουμε τον έλεγχο στο U_n παρά στο D_n . Πάντως, με τον υπολογισμό του “κανονικοποιημένου” συντελεστή συσχέτισης D_n εκτιμούμε το βαθμό της συσχέτισης, καθώς το στατιστικό κείται στο διάστημα $[0,1]$. Βεβαίως, είναι

$$D_n = \frac{U_n}{c_n}$$

όπου

$$c_n = \frac{24}{\alpha_n n^2 (n+1)}$$

(Mardia, 1976, p. 404).

Ο συντελεστής συσχέτισης U_n έχει την ακόλουθη γεωμετρική ερμηνεία:

αφού είναι

$$\frac{C^2 + S^2}{n^2} = \left| \sum_{j=1}^n \frac{j}{n} e^{i\beta_j} \right|^2$$

το U_n είναι ανάλογο με το τετράγωνο του σταθμισμένου αθροίσματος των μοναδιαίων μιγαδικών αριθμών που αναπαριστώνται από τα ομοιόμορφα βάρη (Mardia and Jupp, 2000, p. 247).

Ο Mardia (1976) έδωσε πίνακα με τις κρίσιμες τιμές του U_n μαζί με τις βοηθητικές ποσότητες a_n και c_n .

Για τον υπολογισμό των στατιστικών D_n και U_n του απαραμετρικού ελέγχου του Mardia, κατά τις περιπτώσεις που δεν καλύπτονται από τους πίνακες που υπάρχουν στη βιβλιογραφία, δημιουργήθηκε πρόγραμμα προσομοίωσης, σε γλώσσα προγραμματισμού Visual Basic, το οποίο περιγράφεται στο Παράρτημα, σελ. α'. Με βάση το συγκεκριμένο πρόγραμμα έγιναν οι σχετικοί υπολογισμοί αυτού του κεφαλαίου.

Συσχετίζεται η ωριαία κατεύθυνση του ανέμου με την, αντίστοιχη, ωριαία θερμοκρασία; (Απαραμετρικός έλεγχος συσχέτισης των πρώτης τάξης μεταβλητών WINDIR_Hxx και TEMPER_Hxx.)

Με τη χρησιμοποίηση του προγράμματος προσομοίωσης που αναπτύχθηκε στο πλαίσιο αυτής της μελέτης, εφαρμόστηκε ο απαραμετρικός έλεγχος συσχέτισης του Mardia στα δεδομένα των πρώτης τάξης μεταβλητών WINDIR_Hxx και TEMPER_Hxx. Πλην μιας οριακής περίπτωσης, ο έλεγχος δεν δίνει σημαντικά αποτελέσματα. Επισημαίνουμε, ότι η εφαρμογή του παραμετρικού ελέγχου της προηγούμενης υποενότητας, έδωσε, εν πολλοίς, τα ίδια αποτελέσματα (βλ. πίνακα 5.1, σελ. 138).

Επομένως, μπορούμε να πούμε ότι για το σύνολο των μεταβλητών φαίνεται ότι δεν υπάρχει συσχέτιση της ωριαίας κατεύθυνση του ανέμου με την, αντίστοιχη, ωριαία θερμοκρασία. Ο απαραμετρικός έλεγχος επιβεβαίωσε τα αποτελέσματα του αντίστοιχου παραμετρικού ελέγχου.

Πίνακας 5.10: Αποτελέσματα του παραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών WINDIR_Hxx και TEMPER_Hxx. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχο του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Η μηδενική υπόθεση της μη συσχέτισης μπορεί να απορριφθεί μόνο σε μια περίπτωση και μάλιστα οριακά. Από τα δεδομένα του μήνα μελέτης, φαίνεται ότι, για κάθε ώρα του εικοσιτετραώρου, η θερμοκρασία δεν συσχετίζεται με την κατεύθυνση του ανέμου.

<i>Variables (& observations)</i>	<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_H1 & TEMPER_H1 (26)	0,2156	6,3759	0,06 (*)
WINDIR_H2 & TEMPER_H2 (30)	0,2512	7,4288	0,0463
WINDIR_H3 & TEMPER_H3 (26)	0,0166	0,4915	0,8
WINDIR_H4 & TEMPER_H4 (28)	0,0759	2,2468	0,3395
WINDIR_H5 & TEMPER_H5 (26)	0,0007	2,1457	0,9892
WINDIR_H6 & TEMPER_H6 (25)	0,0157	0,4671	0,8
WINDIR_H7 & TEMPER_H7 (29)	0,1336	3,9534	0,1557
WINDIR_H8 & TEMPER_H8 (0)	0,0221	0,6556	0,7299
WINDIR_H9 & TEMPER_H9 (30)	0,0309	0,914	0,6944
WINDIR_H10 & TEMPER_H10 (31)	0,4679	13,8365	0,1078 (*)
WINDIR_H11 & TEMPER_H11 (31)	0,2491	7,9682	0,4777 (*)
WINDIR_H12 & TEMPER_H12 (31)	0,4798	14,1906	0,8386
WINDIR_H13 & TEMPER_H13 (31)	0,0555	1,6432	0,7495
WINDIR_H14 & TEMPER_H14 (31)	0,0451	1,3352	0,912
WINDIR_H15 & TEMPER_H15 (31)	0,0136	0,404	0,8976
WINDIR_H16 & TEMPER_H16 (31)	0,0624	1,8463	0,7766
WINDIR_H17 & TEMPER_H17 (31)	0,0239	0,7075	0,9163
WINDIR_H18 & TEMPER_H18 (31)	0,243	7,187	0,6556
WINDIR_H19 & TEMPER_H19 (31)	0,3197	9,4569	0,3922
WINDIR_H20 & TEMPER_H20 (30)	0,184	5,4435	0,3851
WINDIR_H21 & TEMPER_H21 (30)	0,1409	4,1679	0,3773 (*)
WINDIR_H22 & TEMPER_H22 (30)	0,0049	0,147	0,9502
WINDIR_H23 & TEMPER_H23 (30)	0,0022	0,0659	0,9756
WINDIR_H24 & TEMPER_H24 (25)	0,0199	0,591	0,7676

* = ο αντίστοιχος παραμετρικός έλεγχος
δίνει σημαντικό αποτέλεσμα

Συσχετίζεται η ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, ημερήσια θερμοκρασία; (Απαραμετρικός έλεγχος συσχέτισης των πρώτης τάξης μεταβλητών WINDIR_Dxx και TEMPER_Dxx.)

Με τη χρησιμοποίηση του προγράμματος προσομοίωσης, εφαρμόστηκε ο απαραμετρικός έλεγχος συσχέτισης του Mardia στα δεδομένα των πρώτης τάξης μεταβλητών WINDIR_Dxx και TEMPER_Dxx. Όπως φαίνεται από τα στοιχεία του πίνακα 5.11, ο έλεγχος δίνει 21 σημαντικά αποτελέσματα, δηλαδή η μηδενική υπόθεση της μη συσχέτισης μπορεί να απορριφθεί στις περισσότερες περιπτώσεις. Φαίνεται ότι η ημερήσια θερμοκρασία συσχετίζεται με την ημερήσια κατεύθυνση του ανέμου.

Ο αντίστοιχος παραμετρικός έλεγχος της προηγούμενης υποεπότητας δίνει σημαντικά αποτελέσματα (βλ. πίνακα 5.3, σελ. 142), τα οποία, όμως, είναι αναξιόπιστα, καθώς οι γραμμικές μεταβλητές δεν κατανέμονται κανονικά. Ο απαραμετρικός έλεγχος της αυτής της υποεπότητας δείχνει ότι, γενικά, η ημερήσια θερμοκρασία συσχετίζεται με την ημερήσια κατεύθυνση του ανέμου.

Πίνακας 5.11: Αποτελέσματα του παραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών WINDIR_Dxx και TEMPER_Dxx. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχο του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Η μηδενική υπόθεση της μη συσχέτισης μπορεί να απορριφθεί στις περισσότερες περιπτώσεις, καθώς προκύπτουν 21 σημαντικά αποτελέσματα. Ο έλεγχος δείχνει ότι η ημερήσια θερμοκρασία συσχετίζεται με την ημερήσια κατεύθυνση του ανέμου.

<i>Variables (& observations)</i>	<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_D1 & TEMPER_D1 (24)	0,8151	18,4837	0,3943
WINDIR_D2 & TEMPER_D2 (24)	1,9705	44,6797	0,0025
WINDIR_D3 & TEMPER_D3 (0)	0,7935	17,9931	0,0013
WINDIR_D4 & TEMPER_D4 (18)	1,2319	27,9336	0,0002
WINDIR_D5 & TEMPER_D5 (21)	0,6699	15,1913	0
WINDIR_D6 & TEMPER_D6 (0)	0,535	12,1322	0,0042
WINDIR_D7 & TEMPER_D7 (24)	0,0464	1,0538	0,6056
WINDIR_D8 & TEMPER_D8 (21)	1,985	45,0099	0,0155
WINDIR_D9 & TEMPER_D9 (24)	1,0825	24,5454	0,0045
WINDIR_D10 & TEMPER_D10 (0)	0,2624	5,9506	0,0466
WINDIR_D11 & TEMPER_D11 (23)	2,5836	58,5815	0,0011
WINDIR_D12 & TEMPER_D12 (24)	0,189	4,2871	0,1739
WINDIR_D13 & TEMPER_D13 (24)	2,4756	56,1328	0
WINDIR_D14 & TEMPER_D14 (24)	1,1323	25,6754	0,1016
WINDIR_D15 & TEMPER_D15 (24)	3,2415	73,5	0,3567
WINDIR_D16 & TEMPER_D16 (24)	0,1296	2,94	0,8386
WINDIR_D17 & TEMPER_D17 (24)	2,7096	61,44	0,0014
WINDIR_D18 & TEMPER_D18 (24)	0,2305	5,2266	0,4374
WINDIR_D19 & TEMPER_D19 (22)	1,0591	24,0157	0,0011
WINDIR_D20 & TEMPER_D20 (23)	2,8107	63,7311	0
WINDIR_D21 & TEMPER_D21 (22)	0,1374	3,1157	0,2537
WINDIR_D22 & TEMPER_D22 (0)	0,3233	7,3318	0,0382
WINDIR_D23 & TEMPER_D23 (20)	0,7038	15,9594	0,0078
WINDIR_D24 & TEMPER_D24 (0)	0,914	20,7259	0
WINDIR_D25 & TEMPER_D25 (23)	0,888	20,1358	0
WINDIR_D26 & TEMPER_D26 (23)	2,2489	50,9942	0,0252
WINDIR_D27 & TEMPER_D27 (23)	0,867	19,6591	0,0326
WINDIR_D28 & TEMPER_D28 (23)	4,463	101,1958	0
WINDIR_D29 & TEMPER_D29 (24)	2,1295	48,2851	0,006
WINDIR_D30 & TEMPER_D30 (0)	0,3712	8,4187	0,098
WINDIR_D31 & TEMPER_D31 (23)	0,1204	2,7308	0,3081

Πίνακας 5.12: Απαραμετρικός έλεγχος συσχέτισης των μεταβλητών δεύτερης τάξης WINDIR_HOUR και TEMPER_HOUR. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχό του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Το πάρα πολύ σημαντικό αποτέλεσμα που προκύπτει υποδεικνύει ισχυρή συσχέτιση της μέσης ωριαίας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέση ωριαία θερμοκρασία.

<i>Variables (& observations)</i>	<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_HOUR TEMPER_HOUR (24)	& 0,6808	15,4366	0

Συσχετίζεται η μέση ωριαία κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ωριαία θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_HOUR και TEMPER_HOUR.)

Το πάρα πολύ σημαντικό αποτέλεσμα της εφαρμογής του απαραμετρικού ελέγχου της συσχέτισης αυτών των μεταβλητών, όπως παρουσιάζεται στον πίνακα 5.12, υποδεικνύει ότι μέση ωριαία κατεύθυνση του ανέμου συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

Σε συμφωνία με τα παραπάνω, ο αντίστοιχος παραμετρικός έλεγχος υποδεικνύει, δίνοντας, επίσης, πάρα πολύ σημαντικό αποτέλεσμα, ότι οι μεταβλητές WINDIR_HOUR και TEMPER_HOUR συσχετίζονται (βλ. πίνακα 5.5, σελ. 144).

Πίνακας 5.13: Αποτελέσματα του απαραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών δεύτερης τάξης WINDIR_DAY και TEMPER_DAY. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχό του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Ο απαραμετρικός έλεγχος υποδεικνύει ότι, είτε κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, είτε κατά τις περιόδους καύσωνα, είτε κατά τις μικτές περιόδους, η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

<i>Variables (& observations)</i>	<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_DAY & TEMPER_DAY (31)	0,1654	4,893	0,0895
WINDIR_DAY[1-17] TEMPER_DAY[1-17] (17)	& 0,2502	3,9489	0,14
WINDIR_DAY[18-31] TEMPER_DAY[18-31] (14)	& 0,0269	0,3459	0,85

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το σύνολο των παρατηρήσεων του μήνα, τις παρατηρήσεις των πρώτων 17 ημερών (φυσιολογικές τιμές θερμοκρασίας) και τις παρατηρήσεις των υπολοίπων 14 ημερών (περίοδος καύσωνα); (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DAY και TEMPER_DAY, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, την περίοδο του καύσωνα και το μικτό διάστημα (σύνολο των ημερών του μήνα), με τον απαραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Τα αποτελέσματα της εφαρμογής του απαραμετρικού ελέγχου της συσχέτισης αυτών των μεταβλητών, όπως παρουσιάζονται στον πίνακα 5.13, υποδεικνύουν ότι, είτε κατά το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, είτε κατά την περίοδο του καύσωνα, είτε κατά τη μικτή περίοδο, η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

Πάντως, σε επίπεδο σημαντικότητας $\alpha = 0.10$, μπορούμε να πούμε ότι, για το σύνολο των ημερών του μήνα, τα δύο αυτά μεγέθη συσχετίζονται ($P = 0.0895$). Όπως έχει δειχθεί (βλ. πίνακα 5.6, σελ. 146), ο αντίστοιχος παραμετρικός έλεγχος υποδεικνύει ότι οι μεταβλητές WINDIR_DAY και TEMPER_DAY δεν συσχετίζονται κατά την περίοδο του καύσωνα και τη μικτή περίοδο, αλλά συσχετίζονται κατά το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας.

Πίνακας 5.14: Αποτελέσματα του απαραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών δεύτερης τάξης WINDIR_DAY και TEMPER_DMAX. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχό του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Ο απαραμετρικός έλεγχος υποδεικνύει ότι, σε περιόδους καύσωνα, η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία. Αντίθετα, τόσο κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, όσο και κατά τις μικτές περιόδους (σύνολο των ημερών του μήνα), τα δύο αυτά μεγέθη φαίνεται να συσχετίζονται.

<i>Variables (& observations)</i>	<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_DAY & TEMPER_DMAX (31)	0,2127	6,2915	0,03
WINDIR_DAY[1-17] TEMPER_DMAX[1-17] (17)	& 0,3482	5,494	0,05
WINDIR_DAY[18-31] TEMPER_DMAX[18-31] (14)	& 0,1503	1,9277	0,4

Συσχετίζεται η μέση ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DAY και TEMPER_DMAX με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Με τη χρησιμοποίηση του προγράμματος προσομοίωσης, εφαρμόστηκε ο παραμετρικός έλεγχος συσχέτισης του *Mardia* στα δεδομένα των δεύτερης τάξης μεταβλητών WINDIR_DAY και TEMPER_DMAX. Τα αποτελέσματα αυτού του ελέγχου, όπως παρουσιάζονται στον πίνακα 5.14, υποδεικνύουν ότι, σε περιόδους καύσωνα, η μέση ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία ($P = 0.40$). Αντίθετα, τόσο κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας ($P = 0.05$), όσο και κατά τις μικτές περιόδους (σύνολο των ημερών του μήνα) ($P = 0.03$), τα δύο αυτά μεγέθη φαίνεται να συσχετίζονται.

Όπως έχει δειχθεί, ο αντίστοιχος παραμετρικός έλεγχος υποδεικνύει, επίσης, τη συσχέτιση της μέσης ημερήσιας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, δίνοντας συντελεστή συσχέτισης $r = 0.65$ ($P = 0.002$).

Όμως, σε ό,τι αφορά στη συσχέτιση των μεταβλητών WINDIR_DAY και TEMPER_DMAX κατά τις μικτές περιόδους (σύνολο των ημερών του μήνα), ο αντίστοιχος παραμετρικός έλεγχος υποδεικνύει ότι οι μεταβλητές αυτές δεν συσχετίζονται, δίνοντας συντελεστή συσχέτισης $r = 0.27$ ($P = 0.129$) (βλ. πίνακα 5.7, σελ. 150).

Πίνακας 5.15: Αποτελέσματα του απαραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών δεύτερης τάξης WINDIR_DPREV και TEMPER_DAY. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχό του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Ο απαραμετρικός έλεγχος υποδεικνύει ότι, είτε κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, είτε κατά τις περιόδους καύσωνα, είτε κατά τις μικτές περιόδους, η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

<i>Variables (& observations)</i>		<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_DPREV TEMPER_DAY (31)	&	0,1554	4,5962	0,8
WINDIR_DPREV[1-17] TEMPER_DAY[1-17] (17)	&	0,006	0,0951	0,9887
WINDIR_DPREV[18-31] TEMPER_DAY[18-31] (14)	&	0,3591	4,6049	0,5323

Συσχετίζεται η επικρατούσα ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών *WINDIR_DPREV* και *TEMPER_DAY* με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Με τη χρησιμοποίηση του προγράμματος προσομοίωσης, εφαρμόστηκε ο παραμετρικός έλεγχος συσχέτισης του *Mardia* στα δεδομένα των δεύτερης τάξης μεταβλητών *WINDIR_DPREV* και *TEMPER_DAY*. Τα αποτελέσματα αυτού του ελέγχου, όπως παρουσιάζονται στον πίνακα 5.15, υποδεικνύουν ότι, είτε κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, είτε κατά τις περιόδους καύσωνα, είτε κατά τις μικτές περιόδους, η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

Σε ό,τι αφορά στη συσχέτιση των μεταβλητών *WINDIR_DPREV* και *TEMPER_DAY* κατά τις μικτές περιόδους και κατά τις περιόδους του καύσωνα, ο αντίστοιχος παραμετρικός έλεγχος δίνει σύμφωνα αποτελέσματα, υποδεικνύοντας ότι οι μεταβλητές αυτές δεν συσχετίζονται.

Όμως, σε αντίθεση με τα παραπάνω, ο παραμετρικός έλεγχος υποδεικνύει, τη συσχέτιση της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, δίνοντας συντελεστή συσχέτισης $r = 0.472$ ($P = 0.042$) (βλ. πίνακα 5.8, σελ. 151).

Πίνακας 5.16: Αποτελέσματα του απαραμετρικού ελέγχου του Mardia για τη συσχέτιση των μεταβλητών δεύτερης τάξης WINDIR_DPREV και TEMPER_DMAX. Υπολογίζεται το στατιστικό U_n της εξ. (5.24) και το προσαρμοσμένο αντίστοιχό του D_n της εξ. (5.22). Τα P -value προκύπτουν με τη χρησιμοποίηση του προγράμματος προσομοίωσης και μπορεί να θεωρηθούν μόνον κατά προσέγγιση. Ο απαραμετρικός έλεγχος υποδεικνύει ότι, είτε κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, είτε κατά τις περιόδους καύσωνα, είτε κατά τις μικτές περιόδους, η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία.

<i>Variables (& observations)</i>		<i>D</i>	<i>U</i>	<i>P (Prob.) -approx.</i>
WINDIR_DPREV TEMPER_DMAX (31)	&	0,0693	2,0499	0,9323
WINDIR_DPREV[1-17] TEMPER_DMAX[1-17] (17)	&	0,0411	0,6494	0,9215
WINDIR_DPREV[18-31] TEMPER_DMAX[18-31] (14)	&	0,2448	3,1402	0,6376

Συσχετίζεται η επικρατούσα ημερήσια κατεύθυνση του ανέμου με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία; (Έλεγχος συσχέτισης των μεταβλητών WINDIR_DPREV και TEMPER_DMAX με τον παραμετρικό έλεγχο κυκλικής-γραμμικής συσχέτισης του *Mardia*.)

Με τη χρησιμοποίηση του προγράμματος προσομοίωσης, εφαρμόστηκε ο παραμετρικός έλεγχος συσχέτισης του *Mardia* στα δεδομένα των δεύτερης τάξης μεταβλητών WINDIR_DPREV και TEMPER_DMAX. Τα αποτελέσματα του ελέγχου, όπως παρουσιάζονται στον πίνακα 5.16, υποδεικνύουν ότι, κατά τις περιόδους καύσωνα και κατά τις μικτές περιόδους, η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν συσχετίζεται με την, αντίστοιχη, μέση ημερήσια θερμοκρασία.

Όμως, κατά τις ημέρες με φυσιολογικές τιμές θερμοκρασίας, τα δύο αυτά μεγέθη συσχετίζονται.

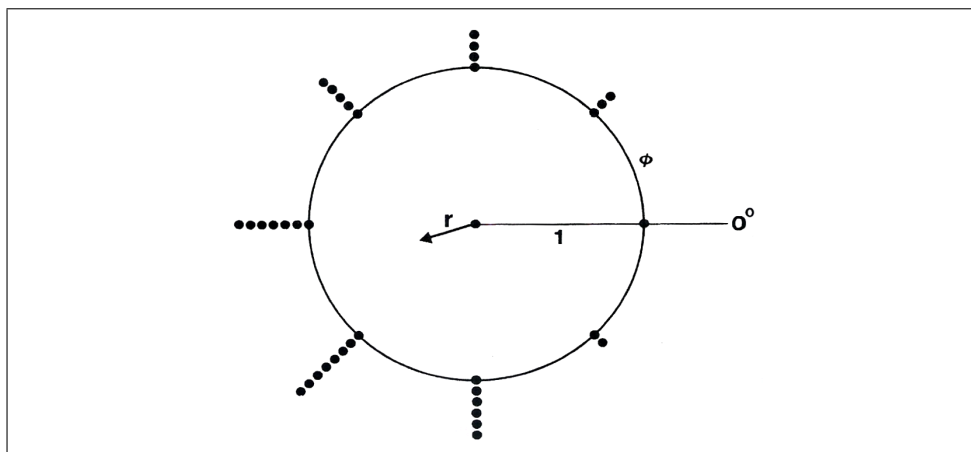
Σε ό,τι αφορά στη συσχέτιση των μεταβλητών WINDIR_DPREV και TEMPER_DMAX κατά τις μικτές περιόδους και κατά τις περιόδους του καύσωνα, ο αντίστοιχος παραμετρικός έλεγχος δίνει σύμφωνα αποτελέσματα, υποδεικνύοντας ότι οι μεταβλητές αυτές δεν συσχετίζονται.

Σε συμφωνία με τα παραπάνω, ο παραμετρικός έλεγχος υποδεικνύει, επίσης, τη συσχέτιση της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου με την, αντίστοιχη, μέση ημερήσια θερμοκρασία, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, δίνοντας συντελεστή συσχέτισης $r = 0.66$ ($P = 0.001$) (βλ. πίνακα 5.9, σελ. 152).

Προσέγγιση Batschelet Ο Batschelet (1981, pp. 193-195) παρουσιάζει την ίδια ιδέα με μια ελαφρώς διαφορετική προσέγγιση. Αυτός κατατάσσει σε αύξουσα σειρά την κυκλική⁶ και όχι τη γραμμική μεταβλητή έτσι, ώστε οι δείκτες (indices) i της κυκλικής μεταβλητής (ϕ) να αποτελούν τα βάρη για τον υπολογισμό των β_i (εδώ, ϕ'). Οι αντίστοιχοι βαθμοί (ranks) της γραμμικής μεταβλητής y_i σημειώνονται ως r_i . Μετατρέπουμε τις γωνίες ϕ στις αντίστοιχες γωνίες ϕ'_i έτσι, ώστε να είναι

$$\phi'_i = i \epsilon$$

όπου το $\epsilon = 360^\circ/n$. Προκειμένου να δοθεί έμφαση στο συλλογισμό αυτό, στο



Σχήμα 5.4: Διαισθητική ερμηνεία της κατασκευής του συντελεστή κυκλικής-γραμμικής συσχέτισης. Οι βαθμοί της γραμμικής μεταβλητής αναπαρίστανται ως στοίβες μοναδιαίων μαζών.

σχήμα 5.4 οι βαθμοί r_i αναπαρίστανται ως απλές σημειακές μάζες στο μοναδιαίο κύκλο. Κατ' αυτό τον τρόπο, ο βαθμός $r_1 = 3$ αναπαρίσταται στο σχήμα με τρεις τελείες, ο βαθμός $r_2 = 4$ με τέσσερις κ.ο.κ. Σχετική συσσώρευση των σημειακών μαζών υποδεικνύει συσχέτιση των y και ϕ . Ως μέτρο αυτής της συσσώρευσης επιλέγουμε το μέτρο του διανύσματος της μέσης κατεύθυνσης ή μήκος του μέσου διανύσματος r . Για να υπολογίσουμε το r , εισάγουμε τις δύο συνιστώσες του συνισταμένου διανύσματος \vec{R}

$$C = \sum_{i=1}^n r_i \cos \phi'_i, \quad S = \sum_{i=1}^n r_i \sin \phi'_i \quad (5.25)$$

⁶Με αυθαίρετο αρχικό σημείο, ασφαλώς.

εδώ, με τους βαθμούς r_i στη θέση των δεικτών i σε σχέση με τη διατύπωση του Mardia — οι δύο προσεγγίσεις *αντιστρέφουν* τους ρόλους των δεικτών i και των βαθμών r_i . Το μήκος του συνισταμένου διανύσματος r δίνεται βεβαίως από τον τύπο

$$r = (C^2 + S^2)^{1/2}$$

Είναι και εδώ προφανές ότι όσο μεγαλύτερο είναι το r , τόσο πιο ισχυρή είναι η συσχέτιση. Αφού υπάρχει τουλάχιστον μία σημειακή μάζα που σχετίζεται με κάθε ϕ'_i , το μήκος του μέσου διανύσματος r δεν μπορεί να πλησιάσει τη μέγιστη τιμή 1. Ως εκ τούτου, προκειμένου το r να αποτελέσει συντελεστή συσχέτισης, θα πρέπει να προσαρμοστεί βάσει κάποιου συγκεκριμένου παράγοντα. Ένας τέτοιος παράγοντας δεν είναι εύκολο να παραχθεί. Ο Mardia πάντως, όπως αναφέρθηκε παραπάνω, κατάφερε να βρει τον κατάλληλο παράγοντα για την προσαρμογή του r , προτείνοντας την ποσότητα D_n της (5.22) ως συντελεστή συσχέτισης κυκλικής με γραμμική μεταβλητή (circular-linear correlation coefficient), όπου ο συντελεστής α_n δίνεται από την (5.23).

Ένας άλλος συντελεστής συσχέτισης βάσει κατάταξης (rank correlation coefficient) έχει προταθεί από τους Batschelet et al. (1973). Δόθηκε πίνακας των κρίσιμων τιμών για $n = 5, 6, \dots, 15$, όμως η εφαρμογή του τεστ είναι ιδιαίτερα επίπονη.

5.2.3 Σύγκριση παραμετρικής και απαραμετρικής μεθόδου

Το παραμετρικό τεστ απαιτεί την εκπλήρωση κάποιων ιδιαίτερα αυστηρών συνθηκών. Σε αντίθεση, ο απαραμετρικός έλεγχος δεν απαιτεί κάποια παραδοχή για τη μορφή της σχέσης των δύο μεταβλητών, υπό την εναλλακτική υπόθεση. Από την άλλη, το απαραμετρικό τεστ δεν παρέχει εκτιμήσεις του πλάτους α_1 ή της κατεύθυνσης της μέγιστης επίδρασης ϕ_0 [βλ. εξ. (5.5, σελ. 135)] (Solow et al., 1988, p. 221).

Ο (Batschelet, 1981, p. 195) αναφέρει ότι μια περίπτωση όπου θα πρέπει να περιμένουμε υπεροχή της απαραμετρικής μεθόδου της συσχέτισης των βαθμών έναντι του παραμετρικού ελέγχου, προφανώς είναι όταν, στο σχήμα της γραμμής παλινδρόμησης, η γραμμή φαίνεται να είναι έντονα λοξή, ξεφεύγοντας από την ημιτονοειδή μορφή που απαιτεί το παραμετρικό τεστ.

Πίνακας 5.17: Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx [1-10] (συνερίζεται στον πίνακα 5.18)

Παραμετρικοί έλεγχοι		STATGRAPHICS		Απαραμ. έλεγχος (Mardia)		
Ορίσiana	Συντ. συ-σχέτισης r	P (Prob.)	Παραμετρικό μοντέλο	P (Prob.)	R^2/R_{adj}^2 (%)	P (Prob.) approx.
Μεταβλητές (Παρατηρήσεις)						
WINDIR_D1	& 0,443	0,015	TEMPER_D1 = 19.1914 - Pr(cos) = 0.0538,		19.61 /	0,3821
TEMPER_D1 (24)			7.98574 cos(WINDIR_D1) - Pr(sin) = 0.0478		11.95	
WINDIR_D2	& 0,559	0,001	4.60809 sin(WINDIR_D1)			
TEMPER_D2 (24)			TEMPER_D2 = 22.2938 - Pr(cos) = 0.0637,		31.25 /	0,0030
			5.14506 cos(WINDIR_D2) - Pr(sin) = 0.0060		24.70	
WINDIR_D3	& ***		4.91517 sin(WINDIR_D2)			
TEMPER_D3 (0)			TEMPER_D3 = 27.0923 + Pr(cos) = 0.0006		43.39 /	0,0015
WINDIR_D4	& 0,619	0,002	3.18105 cos(WINDIR_D3)		40.69	
TEMPER_D4 (18)			TEMPER_D4 = 27.7367 + Pr(cos) = 0.0071		37.29 /	0,0001
WINDIR_D5	& 0,53	0,006	2.56616 cos(WINDIR_D4)		33.37	
TEMPER_D5 (21)			TEMPER_D5 = 27.1922 + Pr(sin) = 0.0200		25.35 /	0,0002
WINDIR_D6	& ***		5.36883 sin(WINDIR_D5)		21.43	
TEMPER_D6 (0)			TEMPER_D6 = 25.2711 - Pr(cos) = 0.1421,		24.98 /	0,0052
			3.75507 cos(WINDIR_D6) - Pr(sin) = 0.0249		17.08	
WINDIR_D7	& 0,366	0,059	1.91316 sin(WINDIR_D6)			0,6162
TEMPER_D7 (24)			Μη συσχέτιση			
WINDIR_D8	& 0,414	0,044	Μη συσχέτιση			0,0153
TEMPER_D8 (21)						
WINDIR_D9	& 0,572	8,06E-04	TEMPER_D9 = 43.8 + Pr(cos) = 0.0039		32.14 /	0,0027
TEMPER_D9 (24)			15.7826 cos(WINDIR_D9)		29.05	
WINDIR_D10	& ***		TEMPER_D10 = 33.5806 + Pr(cos) = 0.0049,		49.75 /	0,0446
TEMPER_D10 (0)			3.34864 cos(WINDIR_D10) - Pr(sin) = 0.0010		44.46	
			4.874 sin(WINDIR_D10)			

*** = Δεν μπορεί να γίνει υπολογισμός - διαφορετικά δειγματικά μεγέθη

Πίνακας 5.18: (συνέχεια του πίνακα 5.17) Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx [11-21] (συνεχίζεται στον πίνακα 5.19)

Μεταβλητές (Παρατηρήσεις)	Συντ. συσχέτισης r	P (Prob.)	Παραμετρικό μοντέλο	P (Prob.)	R^2/R^2_{adj} (%)	P (Prob.)- approx.	Παραμετρικοί έλεγχοι	Ασταθμ. έλεγχος (Mardia)
							STATGRAPHICS	
WINDIR_D11 TEMPER_D11 (23)	& 0,516	0,004	$TEMPER_{D11} = 25.161 + 4.39872 \cos(WINDIR_{D11}) + 5.18704 \sin(WINDIR_{D11})$	$P^*(\cos) = 0.0181,$ $P^*(\sin) = 0.0384$	26.62 / 19.28	0,0014		
WINDIR_D12 TEMPER_D12 (24)	& 0,622	1,92E-04	$TEMPER_{D12} = 30.5063 + 3.55885 \cos(WINDIR_{D12}) + 3.55885 \cos(WINDIR_{D12})$	$P^*(\cos)=0.0016$	37.12 / 34.26	0,1741		
WINDIR_D13 TEMPER_D13 (24)	& 0,752	1,42E-06	$TEMPER_{D13} = 37.8476 + 13.7436 \cos(WINDIR_{D13})$	$P^*(\cos) = 0.0001$	52.19 / 50.02	0,0000		
WINDIR_D14 TEMPER_D14 (24)	& 0,371	0,054	Μη συσχέτιση			0,1062		
WINDIR_D15 TEMPER_D15 (24)	& *****	0	Μη συσχέτιση			0,3590		
WINDIR_D16 TEMPER_D16 (24)	& *****	0	Μη συσχέτιση			0,8344		
WINDIR_D17 TEMPER_D17 (24)	& *****	0	$TEMPER_{D17} = 18.8023 - 14.1565 \cos(WINDIR_{D17})$	$P^*(\cos) = 0.0091$	33.46 / 30.43	0,0018		
WINDIR_D18 TEMPER_D18 (24)	& *****	0	Μη συσχέτιση			0,4357		
WINDIR_D19 TEMPER_D19 (22)	& 0,658	1,52E-04	$TEMPER_{D19} = 43.2551 + 14.1141 \cos(WINDIR_{D19}) - 3.656 \sin(WINDIR_{D19})$	$P^*(\cos) = 0.0355,$ $P^*(\sin) = 0.0323$	43.23 / 37.26	0,0008		
WINDIR_D20 TEMPER_D20 (23)	& 0,804	2,10E-07	$TEMPER_{D20} = 50.0082 + 21.7418 \cos(WINDIR_{D20})$	$P^*(\cos) = 0.0000$	61.87 / 60.06	0,0000		
WINDIR_D21 TEMPER_D21 (22)	& 0,622	4,33E-04	$TEMPER_{D21} = 32.6611 + 3.19295 \cos(WINDIR_{D21}) - 3.56232 \sin(WINDIR_{D21})$	$P^*(\cos) = 0.0089,$ $P^*(\sin) = 0.0087$	38.64 / 32.18	0,2463		

***** = Υποδεικνύει ένα αποτέλεσμα που δεν μπορεί να υπολογιστεί

Πίνακας 5.19: (συνέχεια του πίνακα 5.18) Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Dxx και TEMPER_Dxx [22-31]

Παραμετρικοί έλεγχοι		STATGRAPHICS			Απαραμ. έλεγχος (Mardia)	
Ορίσιμα	Συντ. συ-σχέτισης r	P (Prob.)	Παραμετρικό μοντέλο	P (Prob.)	R^2/R_{adj}^2 (%)	P (Prob.) approx.
Μεταβλητές (Παρατηρήσεις)						
WINDIR_D22	& ***		$TEMPER_D22 = 35.6628 +$	$Pr(\cos) = 0.0062,$	40.75 /	0,0344
TEMPER_D22 (0)			$6.36307 \cos(WINDIR_D22) -$	$Pr(\sin) = 0.0274$	32.85	
			$5.30372 \sin(WINDIR_D22)$			
WINDIR_D23	& 0,62	0,001	$TEMPER_D23 = 11.1623 -$	$Pr(\cos) = 0.0075,$	38.44 /	0,0094
TEMPER_D23 (20)			$24.4769 \cos(WINDIR_D23) -$	$Pr(\sin) = 0.0066$	31.19	
			$11.4107 \sin(WINDIR_D23)$			
WINDIR_D24	& ***		$TEMPER_D24 = 35.8286 +$	$Pr(\cos) = 0.0000$	77.99 /	0,0000
TEMPER_D24 (0)			$4.5036 \cos(WINDIR_D24)$		76.89	
WINDIR_D25	& 0,781	7,04E-07	$TEMPER_D25 = 33.6551 +$	$Pr(\cos) = 0.0000$	59.48 /	0,0000
TEMPER_D25 (23)			$4.56618 \cos(WINDIR_D25)$		57.55	
WINDIR_D26	& 0,346	0,09	Μη συσχέτιση			0,0235
TEMPER_D26 (23)						
WINDIR_D27	& 0,444	0,018	$TEMPER_D27 = 31.3769 +$	$Pr(\cos) = 0.4097,$	19.73 /	0,0307
TEMPER_D27 (23)			$1.81292 \cos(WINDIR_D27) -$	$Pr(\sin) = 0.0404$	11.70	
			$3.02062 \sin(WINDIR_D27)$			
WINDIR_D28	& 0,715	1,26E-05	$TEMPER_D28 = 27.9929 -$	$Pr(\cos) = 0.0219,$	51.08 /	0,0000
TEMPER_D28 (23)			$5.35838 \sin(WINDIR_D28)$	$Pr(\sin) = 0.0008$	46.19	
WINDIR_D29	& 0,64	1,08E-04	$TEMPER_D29 = 28.5294 -$	$Pr(\cos) = 0.0008$	40.61 /	0,0011
TEMPER_D29 (24)			$5.10063 \sin(WINDIR_D29)$		37.91	
WINDIR_D30	& ***		Μη συσχέτιση			0,0947
TEMPER_D30 (0)						
WINDIR_D31	& 0,428	0,025	$TEMPER_D31 = 33.0052 +$	$Pr(\cos) = 0.0468,$	18.33 /	0,3161
TEMPER_D31 (23)			$3.11334 \cos(WINDIR_D31) -$	$Pr(\sin) = 0.5362$	10.16	
			$1.06412 \sin(WINDIR_D31)$			

*** = Δεν μπορεί να γίνει υπολογισμός - διαφορετικά δειγματικά μεγέθη

Πίνακας 5.20: Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Hxx και TEMPER_Hxx [1-11] (συνεχίζεται στον πίνακα 5.21)

Μεταβλητές (Παρατηρήσεις)	Συντ. συσχέτισης r	P (Prob.)	Παραμετρικό μοντέλο	P (Prob.)	R^2/R^2_{act} (%)	P (Prob.)-approx.	Παραμετρικοί	Αταραχ. έλεγχοι
							Όρια	STATGRAPHICS
WINDIR_H1 TEMPER_H1 (26)	& 0,456	0,008	$TEMPER_{H1} = 28.0236 + 1.53775 \cos(WINDIR_{H1}) - 1.69727 \sin(WINDIR_{H1})$	$Pr(\cos) = 0.10,$ $Pr(\sin) = 0.04$	20,82	Μη συσχέτιση ($P = 0.06$)		
WINDIR_H2 TEMPER_H2 (30)	& 0,122	0,671	$TEMPER_{H2} = 26.776 - 0.269609 \cos(WINDIR_{H2}) - 0.427679 \sin(WINDIR_{H2})$	$Pr(\cos) = 0.78,$ $Pr(\sin) = 0.59$	1,47	0,05		
WINDIR_H3 TEMPER_H3 (26)	& 0,18	0,475	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H4 TEMPER_H4 (28)	& 0,227	0,276	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H5 TEMPER_H5 (26)	& 0,176	0,492	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H6 TEMPER_H6 (25)	& 0,192	0,444	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H7 TEMPER_H7 (29)	& 0,159	0,518	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H8 TEMPER_H8 (0)	& ***	***	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H9 TEMPER_H9 (30)	& 0,321	0,061	Μη συσχέτιση			Μη συσχέτιση		
WINDIR_H10 TEMPER_H10 (31)	& 0,352	0,03	$TEMPER_{H10} = 33.3979 + 2.05315 \cos(WINDIR_{H10}) - 1.48482 \sin(WINDIR_{H10})$	$Pr(\cos) = 0.06,$ $Pr(\sin) = 0.15$	12,41	Μη συσχέτιση ($P = 0.11$)		
WINDIR_H11 TEMPER_H11 (31)	& 0,387	0,015	$TEMPER_{H11} = 34.1502 + 2.33818 \cos(WINDIR_{H11}) - 2.39494 \sin(WINDIR_{H11})$	$Pr(\cos) = 0.04,$ $Pr(\sin) = 0.06$	14,95	Μη συσχέτιση ($P = 0.48$)		

*** = Δεν μπορεί να γίνει υπολογισμός - διαφορετικά δειγματικά μεγέθη

5.2. Κυκλική-γραμμική συσχέτιση

Πίνακας 5.21: (συνέχεια του πίνακα 5.20) Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών WINDIR_Hxx και TEMPER_Hxx [1.2-24]

Παραμετρικοί έλεγχοι		STATGRAPHICS			Απαραμ. έλεγχος (Mardia)
Μεταβλητές (Παρατηρήσεις)	Ορίσται	Συντ. συσχέτισης r	P (Prob.)	Παραμετρικό μοντέλο	P (Prob.)-approx.
					R^2/R_{adj}^2 (%)
WINDIR_H12 TEMPER_H12 (31)	& 1,07E-01		7,27E-01	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H13 TEMPER_H13 (31)	& 0,159		0,492	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H14 TEMPER_H14 (31)	& 0,117		0,68	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H15 TEMPER_H15 (31)	& 0,255		0,162	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H16 TEMPER_H16 (31)	& 0,139		0,583	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H17 TEMPER_H17 (31)	& 0,198		0,333	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H18 TEMPER_H18 (31)	& 0,251		0,171	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H19 TEMPER_H19 (31)	& 0,168		0,456	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H20 TEMPER_H20 (30)	& 0,239		0,213	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H21 TEMPER_H21 (30)	& 0,551		2,08E-04	$TEMPER_{H21} = 28,5479 - 2,44406 \cos(WINDIR_{H21}) + 2,20737 \sin(WINDIR_{H21})$	Μη συσχέτιση (P=0,38) 30,40
WINDIR_H22 TEMPER_H22 (30)	& 0,312		0,072	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H23 TEMPER_H23 (30)	& 0,128		0,642	Μη συσχέτιση	Μη συσχέτιση
WINDIR_H24 TEMPER_H24 (25)	& 0,202		0,408	Μη συσχέτιση	Μη συσχέτιση

ΚΕΦΑΛΑΙΟ 5. ΚΥΚΛΙΚΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Κεφάλαιο 6

Συμπεράσματα

Στο παρόν κεφάλαιο παρουσιάζουμε τα συμπεράσματα της στατιστικής επεξεργασίας των μετεωρολογικών δεδομένων της μελέτης, όπως καταγράφηκε στα πλαίσια των προηγούμενων κεφαλαίων.

6.1 Αποτελέσματα των ελέγχων τυχειότητας και κανονικότητας της μεταβλητής της κατεύθυνσης του ανέμου

Ο έλεγχος ομοιομορφίας των μεταβλητών της κατεύθυνσης του ανέμου κατά περίπτωση (ωριαία — WINDIR_Hxx, ημερήσια — WINDIR_Dxx, μέση ωριαία — WINDIR_HOUR, μέση ημερήσια — WINDIR_DAY και επικρατούσα ημερήσια — WINDIR_DPREV) έγινε με το πρόγραμμα *Oriana*, σύμφωνα με τα Rayleigh test, Rao's Spacing test, V test, Kuiper's test και Watson's U^2 test. Ο έλεγχος της κανονικότητας (δηλαδή, της καλής προσαρμογής της κατανομής von Mises) των μεταβλητών της κατεύθυνσης του ανέμου κατά τις ως άνω περιπτώσεις έγινε, με το ίδιο στατιστικό πρόγραμμα, σύμφωνα με τα Kuiper's test και Watson's U^2 test.

6.1.1 Έλεγχοι τυχειότητας και κανονικότητας των πρώτης τάξης δειγμάτων

Ωριαία κατεύθυνση του ανέμου (WINDIR_Hxx)

Όλα τα τεστ απορρίπτουν την τυχειότητα όλων των WINDIR_Hxx. Και τα δύο τεστ απορρίπτουν την κανονικότητα όλων των WINDIR_Hxx (πλην των

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

WINDIR_H3, -4, -5, -7, -8, οι οποίες αποτυγχάνουν σε ένα τεστ). *Φαίνεται, λοιπόν, ότι η κατεύθυνση του ανέμου σε συγκεκριμένες ώρες της κάθε ημέρας δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises.*

Ημερήσια κατεύθυνση του ανέμου (WINDIR_Dxx)

Όλα τα τεστ απορρίπτουν την τυχαιότητα όλων (πλην δύο οριακών περιπτώσεων, των WINDIR_D10 και WINDIR_D24) των WINDIR_Dxx. Και οι δύο έλεγχοι καλής προσαρμογής της κατανομής von Mises απορρίπτουν την κανονικότητα όλων των WINDIR_Dxx (πλην πέντε περιπτώσεων — των WINDIR_D5, -7, -10, -12, -21, οι οποίες αποτυγχάνουν σε ένα από τα δύο τεστ). *Φαίνεται, λοιπόν, ότι, επίσης, η ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises.*

6.1.2 Έλεγχοι τυχαιότητας και κανονικότητας των δεύτερης τάξης δειγμάτων

Μέση ωριαία κατεύθυνση του ανέμου (WINDIR_HOUR)

Όλα τα τεστ απορρίπτουν την τυχαιότητα της WINDIR_HOUR. Από την άλλη, και τα δύο τεστ δέχονται την κανονικότητα της WINDIR_HOUR. *Φαίνεται, λοιπόν, ότι η μέση ωριαία κατεύθυνση του ανέμου (δηλαδή, η κατεύθυνση του ανέμου κατά το μέσο όρο κάθε ώρας του εικοσιτετραώρου) κατανέμεται κανονικά (δηλαδή, σύμφωνα με την κατανομή von Mises).*

Μέση ημερήσια κατεύθυνση του ανέμου (WINDIR_DAY)

Όλα τα τεστ απορρίπτουν την τυχαιότητα της WINDIR_DAY, για κάθε διάστημα του μήνα μελέτης, δηλαδή, τις πρώτες 17 ημέρες (φυσιολογικές τιμές θερμοκρασίας) (WINDIR_DAY[1-17]), τις τελευταίες 14 ημέρες (περίοδος καύσωνα) (WINDIR_DAY[1-17]) και το σύνολο των ημερών του μήνα. Επίσης, και τα δύο τεστ απορρίπτουν την κανονικότητα της WINDIR_DAY, επίσης για κάθε ένα από τα παραπάνω διαστήματα. *Φαίνεται, λοιπόν, ότι η μέση ημερήσια κατεύθυνση του ανέμου (δηλαδή, η κατεύθυνση του ανέμου κατά το μέσο όρο της κάθε ημέρας) δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises.*

6.1. Αποτελέσματα των ελέγχων τυχειότητας και κανονικότητας της μεταβλητής της κατεύθυνσης του ανέμου

Επικρατούσα ημερήσια κατεύθυνση του ανέμου (WINDIR_DPREV)

Όλα τα τεστ απορρίπτουν την τυχειότητα των WINDIR_DPREV, για κάθε διάστημα του μήνα μελέτης, δηλαδή, τις πρώτες 17 ημέρες, με φυσιολογικές τιμές θερμοκρασίας (WINDIR_DPREV[1-17]), τις υπόλοιπες 14 ημέρες της περιόδου του καύσωνα (WINDIR_DPREV[18-31]) και το σύνολο των ημερών του μήνα (πλην μιας αποτυχίας της WINDIR_DPREV[18-31] σε ένα τεστ). Ακόμα, και τα δύο τεστ απορρίπτουν την κανονικότητα των WINDIR_DPREV. Φαίνεται, λοιπόν, ότι η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises, για κάθε διάστημα του μήνα μελέτης.

6.1.3 Έλεγχοι ισονομίας των κατανομών των δεύτερης τάξης δειγμάτων

Έλεγχος ισότητας των μέσων κατευθύνσεων των κατανομών της μέσης ημερήσιας κατεύθυνσης του ανέμου για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας (WINDIR_DAY[1-17]) και για την περίοδο των ημερών του καύσωνα (WINDIR_DAY[18-31])

Η ισότητα των μέσων κατευθύνσεων των κατανομών της μέσης ημερήσιας κατεύθυνσης του ανέμου για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας και για την περίοδο των ημερών του καύσωνα ελέγχθηκε, κατ' αρχήν, με το Watson-Williams test, χωρίς, όμως, τα δύο δείγματα να πληρούν όλες τις συνθήκες της σελ. 68 για την εφαρμογή του. Έτσι, το αποτέλεσμα του ελέγχου ($P = 0.864$, βλ. πίνακα 3.12, σελ. 72) κρίνεται αναξιόπιστο. Ο έλεγχος με το παραμετρικό Mardia-Watson-Wheeler test έδωσε υψηλό P -value ($P = 0.471$, βλ. πίνακα 3.15, σελ. 82 και σχήμα ??, σελ. ??). Σε συμφωνία, το επίσης παραμετρικό Watson's U^2 test έδωσε $P > 0.5$ (βλ. πίνακα 3.18, σελ. 94). Φαίνεται, λοιπόν, ότι η μέση ημερήσια κατεύθυνση του ανέμου δεν παρουσιάζει διαφοροποίηση ανάμεσα στα διαστήματα με φυσιολογικές τιμές θερμοκρασίας και σε αυτά με καύσωνα.

Έλεγχος ισότητας των μέσων κατευθύνσεων των κατανομών της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας (WINDIR_DPREV[1-17]) και για την περίοδο των ημερών του καύσωνα (WINDIR_DPREV[18-31])

Η ισότητα των μέσων κατευθύνσεων των κατανομών της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας και για την περίοδο των ημερών του καύσωνα ελέγχθηκε με το Watson's U^2 , δίνοντας υψηλό P -value ($P > 0.5$, βλ. πίνακα 3.20, σελ. 96). Φαίνεται, λοιπόν, ότι η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν παρουσιάζει διαφοροποίηση ανάμεσα στα διαστήματα με φυσιολογικές τιμές θερμοκρασίας και σε αυτά με καύσωνα.

Έλεγχος ισότητας των μέσων κατευθύνσεων των κατανομών της μέσης ημερήσιας κατεύθυνσης του ανέμου (WINDIR_DAY) και της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου (WINDIR_DPREV), για κάθε διάστημα του μήνα μελέτης.

Η ισότητα των μέσων κατευθύνσεων των κατανομών της μέσης ημερήσιας και της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου ελέγχθηκε με το Watson's U^2 , δίνοντας σημαντικά έως πολύ σημαντικά αποτελέσματα (βλ. πίνακα 3.21, σελ. 97). Φαίνεται, λοιπόν, ότι η μέση ημερήσια και η επικρατούσα ημερήσια κατεύθυνση του ανέμου διαφοροποιούνται για κάθε διάστημα, δηλαδή, για τα διαστήματα με φυσιολογικές τιμές θερμοκρασίας, γι' αυτά με καύσωνα και για το σύνολο των ημερών του μήνα μελέτης.

6.2 Συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου (C–L correlation)

Η συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου, κατά περίπτωση, δηλαδή ως προς τα ημερήσια και τα ανά ώρα μεγέθη, ελέγχθηκε με δύο τρόπους, παραμετρικά και μη παραμετρικά.

- Ο παραμετρικός έλεγχος έγινε με το πρόγραμμα *Oriana*, όπου υπολογίστηκε ο συντελεστής συσχέτισης $r_{y\phi}^2$ τής 5.9, σελ. 136, του οποίου η σημαντικότητα ελέγχεται από τη σχέση 5.10, σελ. 136. Οι συνθήκες που πρέπει να πληρούνται για το συγκεκριμένο έλεγχο είναι η ανεξαρτησία των μεταβλητών και η κανονικότητα της γραμμικής μεταβλητής (Mardia and Jupp, 2000, pp. 245-246). Τα αποτελέσματα των ελέγχων κανονικότητας των γραμμικών μεταβλητών της μελέτης παρατίθενται στον πίνακα 6.1, σελ. 187. Επίσης, με τη χρήση του στατιστικού προγράμματος *STATGRAPHICS*, μελετήθηκε η παλινδρόμηση των γραμμικών μεταβλητών `TEMPER_` επί των κυκλικών `WINDIR_`, σύμφωνα με το μοντέλο του Mardia (Mardia, 1976, p. 403) (Mardia and Jupp, 2000, p. 257):

$$X|\theta \sim N(a_0 + b_1 \cos \theta + b_2 \sin \theta, \sigma^2)$$

- Για τις περιπτώσεις όπου δεν πληρούνταν οι προϋποθέσεις για την εφαρμογή του παραμετρικού ελέγχου εφαρμόστηκε ο αντίστοιχος μη παραμετρικός έλεγχος που έχει προτείνει ο Mardia (Mardia, 1976, p. 404). Ο απαραμετρικός αυτός έλεγχος έγινε με τη χρήση του στατιστικού

$$U_n = \frac{24(C^2 + S^2)}{n^2(n+1)}$$

(Mardia and Jupp, 2000, pp. 246-248) (Solow et al., 1988, p. 221) και τη βοήθεια του προγράμματος προσομοίωσης της κατανομής του που δημιουργήσαμε στο πλαίσιο αυτής της μελέτης.

6.2.1 Έλεγχος συσχέτισης των πρώτης τάξης δειγμάτων

Έλεγχος συσχέτισης της ωριαίας κατεύθυνσης του ανέμου (WINDIR_Hxx) με την, αντίστοιχη, ωριαία θερμοκρασία (TEMPER_Hxx)

Η θερμοκρασία σε συγκεκριμένη ώρα του εικοσιτετραώρου φαίνεται να κατανέμεται κανονικά και, ως εκ τούτου, πληρείται η συνθήκη για τον υπολογισμό του $r_{y\phi}^2$ και την εφαρμογή του παραμετρικού μοντέλου. Από τους υπολογισμούς με το *Oriana* προκύπτουν μόνο 4 σημαντικοί συντελεστές, με το συντελεστή της TEMPER_H21 να είναι πάρα πολύ σημαντικός (βλ. πίνακα 5.1, σελ. 138). Οι παλινδρομήσεις που εφαρμόστηκαν είτε με δύο ερμηνευτικές μεταβλητές, τις $\sin(WINDIR_Hxx)$ και $\cos(WINDIR_Hxx)$, είτε με μία ερμηνευτική, τη $\sin(WINDIR_Hxx)$ ή τη $\cos(WINDIR_Hxx)$, δεν είναι ικανοποιητικές.

Σε συμφωνία, το μη παραμετρικό μοντέλο δίνει μόνο ένα οριακά σημαντικό αποτέλεσμα (βλ. πίνακα 5.10, σελ. 157). *Φαίνεται, λοιπόν, ότι η ανά ώρα θερμοκρασία, δηλαδή η θερμοκρασία σε συγκεκριμένη ώρα του εικοσιτετραώρου, δεν συσχετίζεται με την αντίστοιχη κατεύθυνση του ανέμου.*

Έλεγχος συσχέτισης της ημερήσιας κατεύθυνσης του ανέμου (WINDIR_Dxx) με την, αντίστοιχη, ημερήσια θερμοκρασία (TEMPER_Dxx)

Οι μεταβλητές TEMPER_Dxx γενικά αποκλίνουν από την κανονικότητα (βλ. πίνακα 5.2, σελ. 140), συνεπώς, η εφαρμογή του παραμετρικού ελέγχου μπορεί να δώσει αναξιόπιστα αποτελέσματα.

Στην περίπτωση αυτή εφαρμόστηκε ο μη παραμετρικός έλεγχος που έχει προτείνει ο Mardia, όπου, για την εκτίμηση της κατανομής του κρίσιμου επιπέδου, *P-value*, του ελέγχου χρησιμοποιήθηκε το πρόγραμμα προσομοίωσης που αναπτύξαμε (βλ. Παράρτημα, σελ. α').

Έτσι, ο μη παραμετρικός έλεγχος έδωσε 21 σημαντικά αποτελέσματα σε συνολικά 31 περιπτώσεις (βλ. πίνακα 5.11, σελ. 159). Ειδικότερα, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας παρατηρήθηκαν 11 σημαντικά αποτελέσματα από 17 το σύνολο, ενώ για την περίοδο των ημερών του καύσωνα παρατηρήθηκαν 10 σημαντικά αποτελέσματα από 14 το σύνολο. *Φαίνεται, λοιπόν, ότι η ημερήσια θερμοκρασία, δηλαδή η θερμοκρασία κατά τη διάρκεια της ημέρας μπορεί να συσχετίζεται με την αντίστοιχη κατεύθυνση του ανέμου.*

6.2.2 Έλεγχος συσχέτισης των δεύτερης τάξης δειγμάτων

Έλεγχος συσχέτισης της μέσης ωριαίας κατεύθυνσης του ανέμου ($WINDIR_HOUR$) με την, αντίστοιχη, μέση ωριαία θερμοκρασία ($TEMPER_HOUR$)

Η κανονικότητα της γραμμικής μεταβλητής $TEMPER_HOUR$ είναι αμφίβολη. Τα τεστ κανονικότητας φαίνεται να απορρίπτουν την κανονικότητά της. Από τα τεστ καλής προσαρμογής, το Anderson-Darling test απορρίπτει την κανονικότητά της, όμως τα άλλα τη δέχονται.

Από το πρόγραμμα *Oriana* υπολογίστηκε ένας πάρα πολύ σημαντικός συντελεστής (βλ. πίνακα 5.5, σελ. 144). Η παλινδρόμηση με δύο ερμηνευτικές μεταβλητές, τις $\sin(WINDIR_HOUR)$ και $\cos(WINDIR_HOUR)$, δίνει σημαντικό το συντελεστή του παλινδρομητή $\cos(WINDIR_HOUR)$ και μη σημαντικό (με πολύ μεγάλο $P = 0.45$) το συντελεστή του $\sin(WINDIR_HOUR)$. Το $R^2 = 70.68\%$ είναι πολύ υψηλό.

Η παλινδρόμηση με μόνη ερμηνευτική, την $\cos(WINDIR_HOUR)$, φαίνεται πολύ ικανοποιητική. Το $R^2 = 69.85\%$ είναι πολύ υψηλό. Το μοντέλο που φαίνεται να ερμηνεύει καλύτερα τη σχέση της μέσης ωριαίας θερμοκρασίας ($TEMPER_HOUR$) και της, αντίστοιχης, μέσης ωριαίας κατεύθυνσης του ανέμου ($WINDIR_HOUR$) είναι το

$$TEMPER_HOUR = 49.4408 + 22.1788 \cos(WINDIR_HOUR)$$

Ο παραμετρικός έλεγχος συσχέτισης έδωσε το πολύ σημαντικό αποτέλεσμα $F = 48.66 > F_{0.005,2,21} = 6.99$. Σε συμφωνία με τα παραπάνω, ο μη παραμετρικός έλεγχος έδωσε το πάρα πολύ σημαντικό αποτέλεσμα $P = 0$ (βλ. πίνακα 5.12, σελ. 160). Φαίνεται, λοιπόν, ότι η μέση ωριαία θερμοκρασία συσχετίζεται με την, αντίστοιχη, μέση ωριαία κατεύθυνση του ανέμου.

Έλεγχος συσχέτισης της μέσης ημερήσιας κατεύθυνσης του ανέμου ($WINDIR_DAY$) με την, αντίστοιχη, μέση ημερήσια θερμοκρασία ($TEMPER_DAY$)

Όλα τα τεστ κανονικότητας και όλα τα τεστ καλής προσαρμογής φαίνεται να δέχονται την κανονικότητα της $TEMPER_DAY$. Κατά συνέπεια, μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος με αξιόπιστα αποτελέσματα.

Από το πρόγραμμα *Oriana* υπολογίστηκαν, για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα, μη σημαντικοί συντελεστές συσχέτισης ($P = 0.601$ και $P = 0.232$, αντίστοιχα). Όμως, για το διάστημα των

ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει ένας, έστω οριακά, σημαντικός συντελεστής συσχέτισης $P = 0.048$ (βλ. πίνακα 5.6, σελ. 146).

Οι παλινδρομήσεις που εφαρμόστηκαν είτε με δύο ερμηνευτικές μεταβλητές, τις $\sin(WINDIR_DAY)$ και $\cos(WINDIR_DAY)$, είτε με μία μόνο ερμηνευτική μεταβλητή, τη $\sin(WINDIR_DAY)$ ή τη $\cos(WINDIR_DAY)$, δεν είναι ούτε στο ελάχιστο ικανοποιητικές. Ο παραμετρικός έλεγχος συσχέτισης δίνει μη σημαντικό αποτέλεσμα $F = 0.31 < F_{0.005,2,28} = 6.45$.

Ο μη παραμετρικός έλεγχος, για κάθε ένα από τα παραπάνω διαστήματα, δεν δίνει σημαντικά αποτελέσματα (βλ. πίνακα 5.13, σελ. 161). *Φαίνεται, λοιπόν, ότι η μέση ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη μέση κατεύθυνση του ανέμου.*

Έλεγχος συσχέτισης της μέσης ημερήσιας κατεύθυνσης του ανέμου (WINDIR_DAY) με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία (TEMPER_DMAX)

Η γραμμική μεταβλητή TEMPER_DMAX κατανέμεται κανονικά, οπότε, μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος συσχέτισης με αξιόπιστα αποτελέσματα.

Από το πρόγραμμα *Oriana* υπολογίστηκαν, για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα, μη σημαντικοί συντελεστές συσχέτισης ($P = 0.129$ και $P = 0.381$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει ένας πολύ σημαντικός συντελεστής συσχέτισης $P = 0.002$ (βλ. πίνακα 5.7, σελ. 150).

Σε αντιστοιχία με τα παραπάνω, οι παλινδρομήσεις που εφαρμόστηκαν, για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα, είτε με δύο ερμηνευτικές μεταβλητές, είτε με μία μόνο ερμηνευτική μεταβλητή, δεν είναι ικανοποιητικές. Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, η παλινδρόμηση

$$\begin{aligned} \text{TEMPER_DMAX}[1 - 17] = & 33.7197 \\ & + 3.04624 \sin(\text{WINDIR_DPREV}[1 - 17]) \\ & - 1.58218 \cos(\text{WINDIR_DPREV}[1 - 17]) \end{aligned}$$

έδωσε υψηλό $R^2 = 43.5244\%$.

Ο μη παραμετρικός έλεγχος, δίνει σημαντικά αποτελέσματα για το σύνολο των ημερών του μήνα και για το διάστημα των ημερών με φυσιολογικές τιμές

6.2. Συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου

θερμοκρασίας ($P = 0.03$ και $P = 0.05$, αντίστοιχα). Από την άλλη, ο παραμετρικός έλεγχος δεν δίνει σημαντικά αποτελέσματα ($P = 0.4$) για τις ημέρες της περιόδου του καύσωνα (βλ. πίνακα 5.14, σελ. 163). Φαίνεται, λοιπόν, ότι η μέγιστη ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη μέση κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.

Έλεγχος συσχέτισης της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου (WINDIR_DPREV) με την, αντίστοιχη, μέση ημερήσια θερμοκρασία (TEMPER_DAY)

Η γραμμική μεταβλητή TEMPER_DAY κατανέμεται κανονικά, οπότε, μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος συσχέτισης με αξιόπιστα αποτελέσματα.

Από το πρόγραμμα *Oriana* υπολογίστηκαν, για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα, μη σημαντικοί συντελεστές συσχέτισης ($P = 0.42$ και $P = 0.494$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει σημαντικός συντελεστής συσχέτισης $P = 0.042$ (βλ. πίνακα 5.8, σελ. 151).

Ο μη παραμετρικός έλεγχος, δεν δίνει σημαντικά αποτελέσματα για κανένα από τα παραπάνω διαστήματα (βλ. πίνακα 5.15, σελ. 165). Φαίνεται, λοιπόν, ότι η μέση ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη επικρατούσα κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.

Έλεγχος συσχέτισης της επικρατούσας ημερήσιας κατεύθυνσης του ανέμου (WINDIR_DPREV) με την, αντίστοιχη, μέγιστη ημερήσια θερμοκρασία (TEMPER_DMAX)

Η γραμμική μεταβλητή TEMPER_DMAX κατανέμεται κανονικά, οπότε, μπορεί να εφαρμοστεί ο παραμετρικός έλεγχος συσχέτισης με αξιόπιστα αποτελέσματα.

Από το πρόγραμμα *Oriana* υπολογίστηκαν, για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα, μη σημαντικοί συντελεστές συσχέτισης ($P = 0.091$ και $P = 0.45$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει πολύ σημαντικός συντελεστής συσχέτισης $P = 0.001$ (βλ. πίνακα 5.9, σελ. 152).

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο μη παραμετρικός έλεγχος, δεν δίνει σημαντικά αποτελέσματα για κανένα από τα παραπάνω διαστήματα (βλ. πίνακα 5.16, σελ. 167). Φαίνεται, λοιπόν, ότι η μέγιστη ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη επικρατούσα κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.

6.2. Συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου

Πίνακας 6.1: Αποτελέσματα των ελέγχων για την ομοιομορφία και την κανονικότητα των κυκλικών μεταβλητών και την κανονικότητα των γραμμικών μεταβλητών της μελέτης.

Μεταβλητές	Έλεγχος ομοιομορφίας*	Έλεγχος κανονικότητας**	Συμπέρασμα
<i>1ης τάξης δείγματα</i>			
WINDIR_Hxx	Σημαντικά αποτελέσματα σε όλα τα τεστ	Σημαντικά αποτελέσματα σε όλα τα τεστ (πλην 5 μεταβλητών που αποτυγχάνουν σε 1 τεστ)	Φαίνεται ότι η κατεύθυνση του ανέμου σε συγκεκριμένες ώρες της κάθε ημέρας δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises.
WINDIR_Dxx	Σημαντικά αποτελέσματα σε όλα τα τεστ	Σημαντικά αποτελέσματα σε όλα τα τεστ	Φαίνεται ότι η ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises.
<i>2ης τάξης δείγματα</i>			
WINDIR_HOUR	Σημαντικά αποτελέσματα σε όλα τα τεστ	Μη σημαντικά αποτελέσματα σε όλα τα τεστ	Φαίνεται ότι η μέση ωριαία κατεύθυνση του ανέμου (δηλαδή, η κατεύθυνση του ανέμου κατά το μέσο όρο κάθε ώρας του εικοσιτετραώρου) κατανέμεται κανονικά (δηλαδή, σύμφωνα με την κατανομή von Mises).
WINDIR_DAY	Σημαντικά αποτελέσματα σε όλα τα τεστ, για κάθε διάστημα του μήνα μελέτης.	Σημαντικά αποτελέσματα σε όλα τα τεστ, για κάθε διάστημα του μήνα μελέτης.	Φαίνεται ότι η μέση ημερήσια κατεύθυνση του ανέμου (δηλαδή, η κατεύθυνση του ανέμου κατά το μέσο όρο της κάθε ημέρας) δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises, για κάθε διάστημα του μήνα μελέτης.
WINDIR_DPREV	Σημαντικά αποτελέσματα σε όλα τα τεστ, για κάθε διάστημα του μήνα μελέτης.	Σημαντικά αποτελέσματα σε όλα τα τεστ, για κάθε διάστημα του μήνα μελέτης.	Φαίνεται ότι η επικρατούσα ημερήσια κατεύθυνση του ανέμου δεν κατανέμεται ούτε ομοιόμορφα ούτε von Mises, για κάθε διάστημα του μήνα μελέτης.

(*) Τεστ ομοιομορφίας: Rayleigh's Uniformity Test, Rao's Spacing Test, Watson's U^2 Test και Kuiper's Test

(**) Τεστ κανονικότητας: Watson's U^2 Test και Kuiper's Test

Πίνακας 6.2: Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών $TEMPER_ / WINDIR_ (1)$.

Μεταβλητές	Κανονικότητα γραμμικής μεταβλητής	Παραμετρικός έλεγχος	Μη παραμετρικός έλεγχος <i>Mardia (1976)</i>	Συμπέρασμα
<i>1ης τάξης δείγματα</i>				
WINDIR_Hxx TEMPER_Hxx	/ Ισχυρή κανονικότητα	Μη σημαντικά αποτελέσματα (μόνο 4 σημαντικοί συντελεστές)	Μη σημαντικοί συντελεστές συσχέτισης (1 οριακά σημαντικοί αποτελέσματα)	Φαίνεται ότι η ανά ώρα θερμοκρασία δεν συσχετίζεται με την αντίστοιχη κατεύθυνση του ανέμου.
WINDIR_Dxx TEMPER_Dxx	/ Μη κανονική κατανομή	Αναξιόπιστα αποτελέσματα	Αμφίβολα αποτελέσματα (21 σημαντικά αποτελέσματα από συνολικά 31 — για το διάστημα των ημερών με φουσιολογικές τιμές θερμοκρασίας παρατηρήθηκαν 11 σημαντικά αποτελέσματα από 17 το σύνολο, ενώ για την περίοδο των ημερών του καύσωνα παρατηρήθηκαν 10 σημαντικά αποτελέσματα από 14 το σύνολο.)	Η ημερήσια θερμοκρασία μπορεί να συσχετίζεται με την αντίστοιχη κατεύθυνση του ανέμου.
<i>2ης τάξης δείγματα</i>				
WINDIR_HOUR TEMPER_HOUR	/ Αμφίβολη κανονικότητα	Πολύ σημαντικό αποτέλεσμα ($P = 9,27E - 09$). Πολύ ικανοποιητική παλινδρόμηση $TEMPER_HOUR = 49,4408 + 22,1788 \cos(WINDIR_HOUR)$ ($R^2 = 69,86\%$)	Πάρα πολύ σημαντικό αποτέλεσμα ($P = 0$)	Φαίνεται ότι η μέση ωριαία θερμοκρασία συσχετίζεται ισχυρά με την, αντίστοιχη, μέση ωριαία κατεύθυνση του ανέμου.
WINDIR_DAY TEMPER_DAY	/ Κανονική κατανομή	Μη σημαντικοί συντελεστές συσχέτισης για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα ($P = 0,601$ και $P = 0,232$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φουσιολογικές τιμές θερμοκρασίας, προκύπτει οριακά σημαντικός συντελεστής συσχέτισης ($P = 0,048$). Καθόλου ικανοποιητικές παλινδρόμησης	Μη σημαντικοί συντελεστές συσχέτισης	Φαίνεται ότι η μέση ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη μέση κατεύθυνση του ανέμου.

(*) Για τον παραμετρικό έλεγχο, οι παλινδρόμησης $TEMPER_ / \sin(WINDIR_)$, $\cos(WINDIR_)$ έγιναν σύμφωνα με το μοντέλο του *Mardia (Mardia, 1976, p. 403): $X|\theta \sim N(a_0 + b_1 \cos \theta + b_2 \sin \theta, \sigma^2)$.*

Πίνακας 6.3: Αποτελέσματα των ελέγχων συσχέτισης των μεταβλητών TEMPER_ / WINDIR_ (II)

Μεταβλητές	Κανονικότητα γραμμικής μεταβλητής	Παραμετρικός έλεγχος	Μη παραμετρικός έλεγχος Mardia (1976)	Συμπέρασμα
<i>2ης τάξης δείγματα</i>				
WINDIR_DAY TEMPER_DMAX	/ Κανονική κατανομή	Μη σημαντικοί συντελεστές συσχέτισης, εκτός από το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας ($P = 0.002$). Μη ικανοποιητικές παλινδρομήσεις.	Σημαντικά αποτελέσματα για το σύνολο των ημερών του μήνα και για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας ($P = 0.03$ και $P = 0.05$, αντίστοιχα). Μη σημαντικά αποτελέσματα ($P = 0.4$) για τις ημέρες της περιόδου του καύσωνα	Φαίνεται ότι η μέγιστη ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη μέση κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.
WINDIR_DPREV TEMPER_DAY	/ Κανονική κατανομή	Μη σημαντικοί συντελεστές συσχέτισης για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα ($P = 0.42$ και $P = 0.494$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει σημαντικός συντελεστής συσχέτισης ($P = 0.042$).	Μη σημαντικά αποτελέσματα	Φαίνεται ότι η μέση ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη επικρατούσα κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.
WINDIR_DPREV TEMPER_DMAX	/ Κανονική κατανομή	Μη σημαντικοί συντελεστές συσχέτισης για το σύνολο των ημερών του μήνα και για τις ημέρες της περιόδου του καύσωνα ($P = 0.091$ και $P = 0.45$, αντίστοιχα). Όμως, για το διάστημα των ημερών με φυσιολογικές τιμές θερμοκρασίας, προκύπτει σημαντικός συντελεστής συσχέτισης ($P = 0.001$).	Μη σημαντικά αποτελέσματα	Φαίνεται ότι η μέγιστη ημερήσια θερμοκρασία δεν συσχετίζεται με την αντίστοιχη επικρατούσα κατεύθυνση του ανέμου, υπό συνθήκες καύσωνα, ενώ, υπό φυσιολογικές τιμές θερμοκρασίας, υπάρχει συσχέτιση των μεταβλητών.

6.2. Συσχέτιση των μεταβλητών της θερμοκρασίας και της κατεύθυνσης του ανέμου

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Παράρτημα

Για την εφαρμογή του απαραμετρικού ελέγχου του Mardia για τη συσχέτιση κυκλικής με γραμμική μεταβλητή, σε περιπτώσεις που δεν καλύπτονται από τους πίνακες τιμών που υπάρχουν στη βιβλιογραφία, δημιουργήθηκε πρόγραμμα, γραμμένο σε γλώσσα *Visual Basic 6.0*, το οποίο υπολογίζει, για κάθε ζεύγος γραμμικής—κυκλικής μεταβλητής, τις τιμές του στατιστικού U_n , του τυποποιημένου στατιστικού D_n και της P -value με προσομοίωση.

Σύμφωνα με τη μεθοδολογία του Mardia, η γραμμική μεταβλητή ταξινομείται κατ' αύξουσα σειρά (όπου αποδίδονται οι δείκτες i — μεταβλητή *index*), ενώ η κυκλική μεταβλητή βαθμονομείται κατά τάξη μεγέθους (όπου αποδίδονται οι βαθμοί r_i — μεταβλητή *ranks*). Για τον υπολογισμό των παραμέτρων C και S του στατιστικού U_n χρησιμοποιούνται οι γωνίες *theta_star* (που ισοδυναμούν με τα ομοιόμορφα βάρη β_i), οι οποίες δίνονται από την εξ. 5.21. Οι τιμές που προκύπτουν είναι οι παρατηρούμενες τιμές του στατιστικού U_n (U_{obs}) για το δείγμα. Παραθέτουμε τις κύριες ρουτίνες του προγράμματος:

Παραγωγή τυχαίων μεταθέσεων Ο προσδιορισμός του κρίσιμου επιπέδου P -value του στατιστικού U_n έγινε με βάση τον κανόνα $U_n > U_{obs}$, όπου οι τιμές του U_n υπολογίστηκαν με προσομοίωση. Συγκεκριμένα, στο δείγμα εφαρμόστηκαν 10.000 τυχαίες (μεταθέσεις των ζευγών των τιμών των δύο μεταβλητών (y_i, ϕ_i)). Για την παραγωγή των τυχαίων αυτών μεταθέσεων μελετήθηκαν τρεις αλγόριθμοι, από το βιβλίο "Discrete Algorithmic Mathematics" των Stephen B. Maurer and Anthony Ralston (2004), οι οποίοι αφορούν στην παραγωγή μιας τυχαίας μετάθεσης n στοιχείων και παρουσιάζονται παρακάτω σε μορφή ψευδοκώδικα.

1. Σύμφωνα με τον πρώτο αλγόριθμο (βλ. πίνακα 6.4), χρησιμοποιούμε σε έναν εσωτερικό βρόχο μια συνάρτηση παραγωγής τυχαίων αριθμών (τη συνάρτηση *RAND* για τους αριθμούς από 1 έως n , δηλαδή τη *RAND*[1, n]). Χρησιμοποιείται η βοηθητική δίτιμη συνάρτηση

$Flag$, η οποία σημειώνει τις επαναλήψεις. Αν ο αριθμός που παράγει η $RAND$ εμφανίζεται για πρώτη φορά ($Flag = 0$), αποδίδει τιμή στη μεταβλητή $Perm$, ενώ κάθε φορά που η $RAND$ επαναλαμβάνει κάποιον ήδη δοθέντα αριθμό ($Flag = 1$), αυτός απορρίπτεται. Ο βρόχος τερματίζεται όταν δοθεί ο τελευταίος αριθμός. Αποδεικνύεται ότι ο αλγόριθμος παράγει με τυχαίο τρόπο τις μεταθέσεις, αλλά είναι τάξης $O(n \log n)$.

Πίνακας 6.4: Αλγόριθμος απόδοσης τυχαίων ακέραιων τιμών PERMUTE-1 (Maurer and Ralston, 2004, pp. 387–389). Δεν είναι αποτελεσματικός, καθώς είναι τάξης $O(n \log n)$.

Input n	[Length of permutation]
Output Perm	[A random permutation given as a sequence]
Algorithm PERMUTE-1	
Flag $\leftarrow 0$	[Set <i>all</i> components of Flag to 0]
for $i = 1$ to n	[i th pass will determine Perm(i)]
repeat	
$r \leftarrow RAND[1, n]$	[Pick a random number
endrepeat when Flag(r)=0	until an unused one found]
Perm(i) $\leftarrow r$	
Flag(r) $\leftarrow 1$	[r has been used]
endfor	

Πίνακας 6.5: Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-1 για $n = 5$. Με υπογράμμιση σημειώνονται οι αριθμοί που επαναλαμβάνονται και, κατά συνέπεια, απορρίπτονται.

3	1	4	<u>1</u>	5	<u>4</u>	<u>3</u>	2
---	---	---	----------	---	----------	----------	---

2. Ο δεύτερος αλγόριθμος (βλ. πίνακα 6.6) είναι αποδοτικότερος του πρώτου στο ότι καλεί τη συνάρτηση $RAND$ ακριβώς n φορές, δηλαδή, μετά από την τυχαία επιλογή k αριθμών από 1 έως n , καλείται η $RAND[1, n - k]$. Τότε, ο αριθμός που παράγεται προσδιορίζει τη θέση τής $k + 1$ επιλογής από τους εναπομείναντες αριθμούς (διατεταγμένους σε αύξουσα σειρά). Όμως, η βοηθητική συνάρτηση $Flag$ καλείται περισσότερες φορές απ' ό,τι στον πρώτο αλγόριθμο, καθώς σε κάθε κλήση τής $RAND$ πρέπει να γίνεται ο έλεγχος των θέσεων προς επιλογή ($Flag = 0$). Αποδεικνύεται ότι ο αλγόριθμος

αυτός παράγει με τυχαίο τρόπο τις μεταθέσεις, αλλά είναι τάξης $O(n^2) > O(n \log n)$.

Πίνακας 6.6: Αλγόριθμος απόδοσης τυχαίων ακέραιων τιμών PERMUTE-2 (Maurer and Ralston, 2004, pp. 389–390). Ο λιγότερο αποτελεσματικός, τάξης $O(n^2)$.

Input n
Output Perm
Algorithm PERMUTE-2
 Flag $\leftarrow 0$
for $i = 1$ to n [Pick i th entry in Perm]
 $r \leftarrow \text{RAND}[1, n - i + 1]$
 $c \leftarrow 0$ [Number of unchosen values passed over]
 $j \leftarrow 0$ [Initializes number considered for Perm(i)]
 repeat while $c < r$
 $j \leftarrow j + 1$
 if Flag(j)=0 **then** $c \leftarrow c + 1$ [Another unchosen number found]
 endrepeat
 Perm(i) $\leftarrow j$
 Flag(j) $\leftarrow 1$
endfor

Πίνακας 6.7: Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-2 για $n = 5$. Εντός των τετραγώνων σημειώνονται οι αριθμοί που επιλέγονται τυχαία, οι οποίοι δεν συμμετέχουν στις επόμενες επαναλήψεις (υπογράμμιση).

RAND[1,5]=3	1	2	<u>3</u>	4	5
RAND[1,4]=4	1	2	<u>3</u>	4	<u>5</u>
RAND[1,3]=1	<u>1</u>	2	<u>3</u>	4	<u>5</u>
RAND[1,2]=2	<u>1</u>	2	<u>3</u>	<u>4</u>	<u>5</u>
RAND[1,1]=1	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>

3. Ο τρίτος αλγόριθμος (βλ. πίνακα 6.8) υπερέχει του δεύτερου στο ότι σε κάθε επιλογή (κλήση της $RAND$), αντιμετωπίζει τον επιλεγμένο αριθμό με τον πρώτο μη επιλεγμένο. Ως αποτέλεσμα, αν k είναι οι αριθμοί που έχουν επιλεγεί, αυτοί συγκεντρώνονται και καλείται η $RAND[k + 1, n]$ για τον προσδιορισμό της θέσης της $k + 1$ επιλογής από τους εναπομείναντες αριθμούς. Στην περίπτωση αυτή δεν απαιτείται η χρησιμοποίηση της βοηθητικής συνάρτησης $Flag$, αφού οι ήδη επιλεγμένοι αριθμοί είναι διαχωρισμένοι από

τους εναπομείναντες. Αποδεικνύεται ότι ο αλγόριθμος αυτός παράγει με τυχαίο τρόπο τις μεταθέσεις και είναι τάξης $O(n)$. Επομένως, ο συγκεκριμένος αλγόριθμος είναι μικρότερης τάξης σε σχέση με τους δύο προηγούμενους και γι' αυτόν το λόγο τον επιλέξαμε για την παραγωγή προκαθορισμένου αριθμού (10.000) τυχαίων μεταθέσεων των δεικτών i .

Πίνακας 6.8: Αλγόριθμος απόδοσης τυχαίων ακέραιων τιμών PERMUTE-3 (Maurer and Ralston, 2004, pp. 390–392). Ο πιο αποτελεσματικός απ' όλους, τάξης $O(n)$.

Input n
Output Perm
Algorithm PERMUTE-3
for $i = 1$ to n
 Perm(i) $\leftarrow i$ [Initialize Perm]
endfor
for $i = 1$ to n
 $r \leftarrow \text{Rand}[i, n]$
 Perm(i) \leftrightarrow Perm(r) [Exchange values]
endfor

Πίνακας 6.9: Τυπικό αποτέλεσμα του αλγόριθμου PERMUTE-3 για $n = 5$. Με αστερίσκο σημειώνονται οι αριθμοί που αντιμετατίθενται και με διπλό αστερίσκο σημειώνονται οι αριθμοί που αντιμετατίθενται με τον εαυτό τους (δηλαδή δεν αλλάζουν θέση). Σε κάθε τυχαία επιλογή, ο επιλεγμένος αριθμός μετατίθεται στη θέση του πρώτου στη σειρά μη επιλεγμένου αριθμού και σημειώνεται με υπογράμμιση.

RAND[1,5]=3	1*	2	3*	4	5
RAND[2,5]=5	<u>3</u>	2*	1	4	5*
RAND[3,5]=3	<u>3</u>	<u>5</u>	1**	4	2
RAND[4,5]=5	<u>3</u>	<u>5</u>	<u>1</u>	4*	2*
RAND[5,5]=5	<u>3</u>	<u>5</u>	<u>1</u>	<u>2</u>	4**
	<u>3</u>	<u>5</u>	<u>1</u>	<u>2</u>	<u>4</u>

Στο παρακάτω πλαίσιο παρουσιάζεται, σε γλώσσα προγραμματισμού *Visual Basic*, η ρουτίνα *RandomizeList* με την οποία παράγονται οι τυχαίες μεταθέσεις των ζευγών των τιμών των δύο μεταβλητών (*linear, theta*). Η ρουτίνα βασίζεται στον αλγόριθμο απόδοσης τυχαίων ακέραιων τιμών του πίνακα 6.8.

Αλγόριθμος RandomizeList

```
Public Sub RandomizeList(p() As Single)
' Παράγει τυχαίες μεταθέσεις n διαδοχικών ακεραίων
Dim low As Integer
Dim upp As Integer
Dim i As Integer
Dim r As Integer
Dim tmp As Integer
    low = LBound(p)
    upp = UBound(p)

    Randomize
    For i = low To upp - 1
        ' Απόδοση τυχαίας τιμής στο αντικείμενο i
        r = Int((upp - i + 1) * Rnd + i)

        tmp = p(i)
        p(i) = p(r)
        p(r) = tmp
    Next i
End Sub
```

Ταξινόμηση των τιμών της γραμμικής μεταβλητής Στο παρακάτω πλαίσιο παρουσιάζεται, σε γλώσσα προγραμματισμού *Visual Basic*, η ρουτίνα *Bubblesort2* με την οποία γίνεται η ταξινόμηση των τιμών της γραμμικής μεταβλητής σε αύξουσα σειρά. Η ρουτίνα έχει δύο διανυσματικά ορίσματα, το $y(\cdot)$ που αντιστοιχεί στη γραμμική μεταβλητή (*linear*) και το $x(\cdot)$ που αντιστοιχεί στην κυκλική μεταβλητή (*theta*).¹ Δηλαδή, εκτελείται η

Bubblesort2(linear, theta).

Κατά την εκτέλεση, μετατίθενται αντίστοιχα οι τιμές της κυκλικής μεταβλητής.

¹Μπορεί να γίνει, επίσης, η αντίστροφη αντιστοίχιση, ακολουθώντας την προσέγγιση του Batschelet, σελ. 169.

Αλγόριθμος BubbleSort2

```
Public Sub BubbleSort2(y() As Single, x() As Single)
' Αύξουσα ταξινόμηση ως προς y και αντίστοιχη μετά-
θεση των x
Dim i As Integer
Dim low As Single
Dim upp As Single
Dim y_tmp As Single, x_tmp As Single
Dim swapped As Boolean
    low = LBound(y)
    upp = UBound(y) - 1

    Do
        swapped = False
        For i = low To upp
            If y(i) > y(i + 1) Then
                y_tmp = y(i)
                x_tmp = x(i)

                y(i) = y(i + 1)
                x(i) = x(i + 1)

                y(i + 1) = y_tmp
                x(i + 1) = x_tmp

                swapped = True
            End If
        Next
        upp = upp - 1
    Loop Until Not swapped
End Sub
```

Απόδοση βαθμών στην κυκλική μεταβλητή Στο παρακάτω πλαίσιο παρουσιάζεται, σε γλώσσα προγραμματισμού *Visual Basic*, η ρουτίνα *RankingList*

με την οποία αποδίδονται βαθμοί (*ranks*) στην κυκλική μεταβλητή. Η ρουτίνα έχει τρία διανυσματικά ορίσματα, το $x(\cdot)$ που αντιστοιχεί στην κυκλική μεταβλητή (*theta*), το $y(\cdot)$ που αντιστοιχεί στους δείκτες (*index*) της γραμμικής μεταβλητής και το $r(\cdot)$ που αντιστοιχεί στους κυκλικούς βαθμούς (*ranks*). Αρχικά, γίνεται ταξινόμηση των τιμών της κυκλικής μεταβλητής σε αύξουσα σειρά με την εκτέλεση της ρουτίνας *Bubblesort2*(x, y) έτσι, ώστε να μπορούν να αποδοθούν οι βαθμοί. Σε περίπτωση ισοτιμίας, ο βαθμός προκύπτει ως ο αριθμητικός μέσος των ίσων τιμών. Οι βαθμοί καταχωρίζονται στο διάνυσμα $r(\cdot)$ μέσω της μεταβλητής *rank* και της βοηθητικής μεταβλητής *count* (ο μετρητής των στοιχείων της ισοτιμίας) βάσει της σχέσης

$$rank = i + count/2 - 1/2$$

η οποία προκύπτει από την

$$rank = (1/count) \times [i + (i+1) + \dots + (i+k-1)] = \dots = i + count/2 - 1/2$$

καθώς η ποσότητα μέσα στις αγκύλες είναι το άθροισμα των k πρώτων όρων αριθμητικής προόδου με πρώτο όρο $\alpha_1 = i$ και διαφορά $\omega = 1$. Δηλαδή, εκτελείται η

$$RankingList(theta, index, ranks).$$

Κατά την εκτέλεση της ρουτίνας, διατηρείται η αντιστοίχιση με τις τιμές της γραμμικής μεταβλητής. Τα ζεύγη των τιμών των παρατηρήσεων μπορούν να αναδιαταχθούν σε αύξουσα σειρά ως προς τις τιμές της γραμμικής μεταβλητής και να εμφανίζονται κατ' αυτό τον τρόπο με την εκτέλεση της ρουτίνας

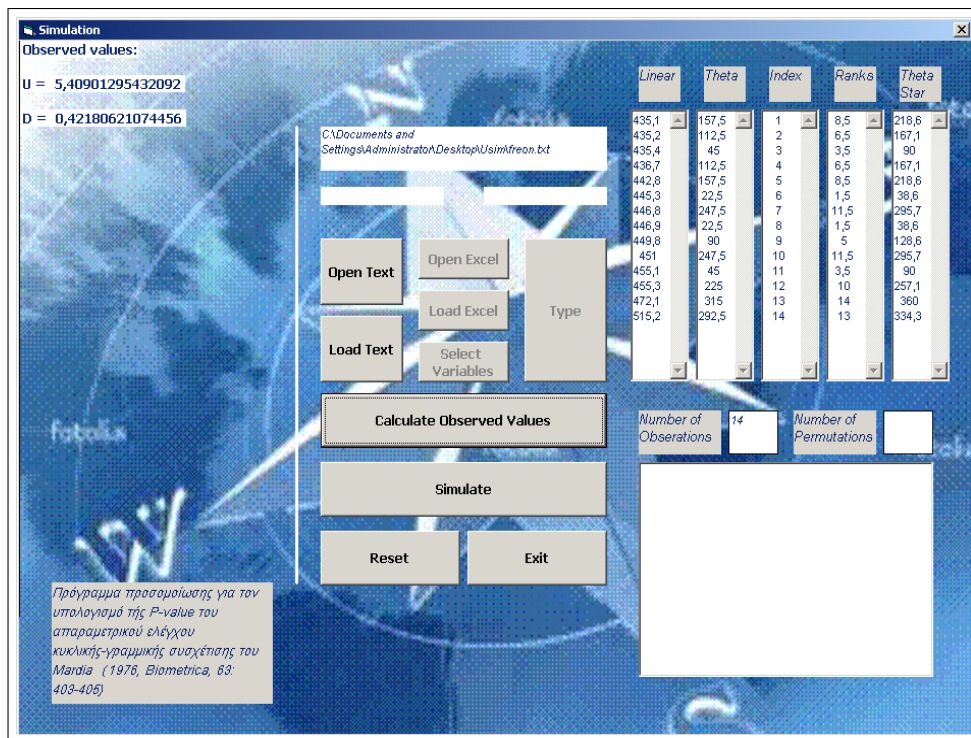
$$Bubblesort2(index, ranks).$$

Ρουτίνα βαθμονόμησης των τιμών της κυκλικής μεταβλητής Στο παρακάτω πλαίσιο παρουσιάζεται η ρουτίνα βαθμονόμησης των τιμών της κυκλικής μεταβλητής. Ο αλγόριθμος αποδίδει βαθμούς στην κυκλική μεταβλητή διατηρώντας την αντιστοίχιση με τις τιμές της γραμμικής μεταβλητής.

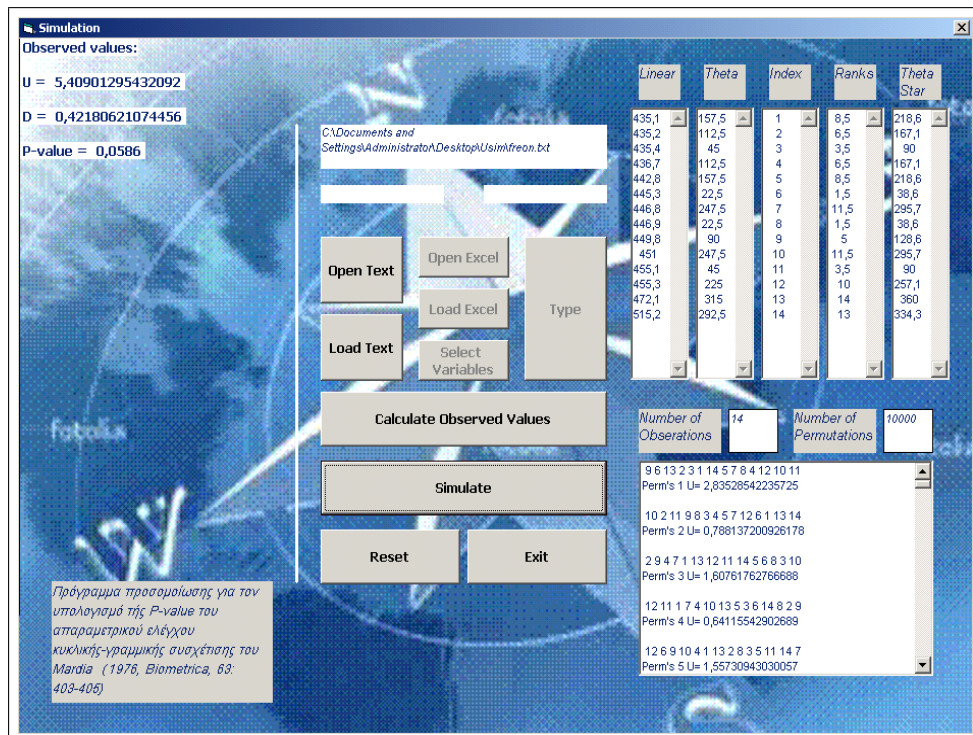
Αλγόριθμος RankingList

```
Public Sub RankingList(x() As Single, y() As Single,
r() As Single)
' Δίνει βαθμούς (ranks) στο διάνυσμα x (ως αριθμητι-
κούς μέσους των ίσων τιμών) και κρατάει την αντιστοι-
χισή με το y. Οι βαθμοί καταχωρίζονται στο r.
Dim i As Integer, j As Integer, k As Integer
Dim count As Integer
Dim rank As Single
    i = 1
    rank = 0

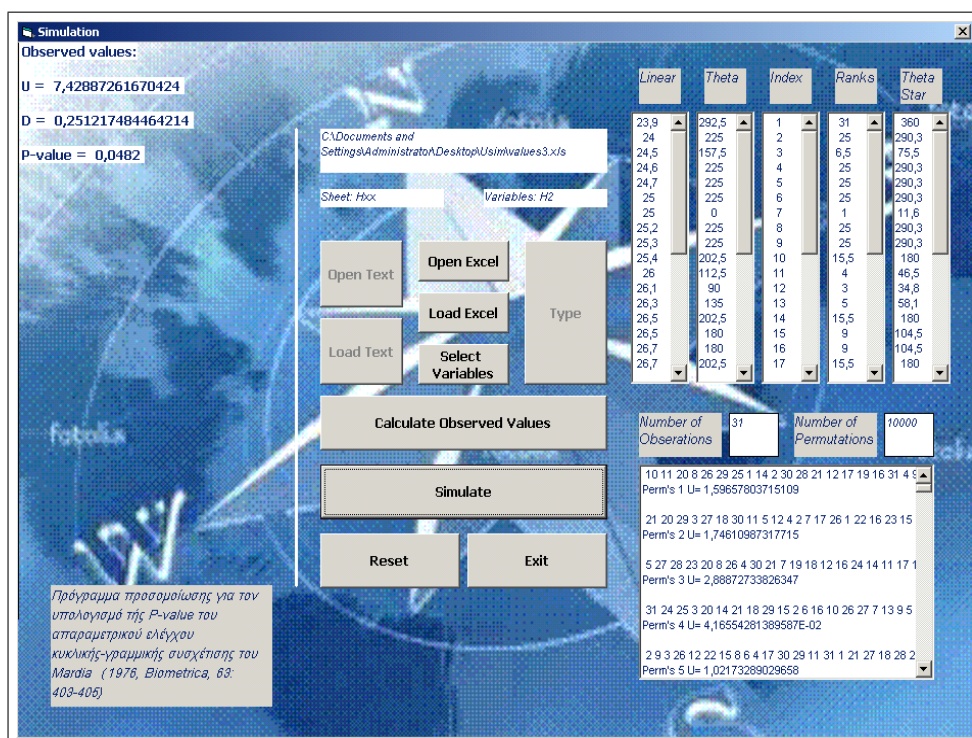
    BubbleSort2 x, y
    Do While i <= n
        count = 0
        For j = i To n
            If x(j) = x(i) Then
                count = count + 1
            End If
        Next j
        ' rank = (1/count) * [i + (i+1) + ... + (i+k-
1)]=...= i + count/2 - 1/2 (η ποσότητα μέσα στις αγκύ-
λες λογίζεται ως αθροισμα των k πρώτων όρων αριθμητι-
κής προόδου με πρώτο όρο α1=i και ω=1)
        rank = i + count / 2 - 1 / 2
        For k = i To i + count - 1
            r(k) = rank
        Next k
        i = i + count
    Loop
End Sub
```

Σχήμα 6.1: Εφαρμογή του προγράμματος προσομοίωσης για τον υπολογισμό της P -value του απαραμετρικού ελέγχου του Mardia (1976, Biometrika, 63: 403–405) για τα δεδομένα της μελέτης των Solow et al. (1988, p. 225). Η εισαγωγή των δεδομένων μπορεί να γίνει από αρχείο κειμένου, υπολογιστικό φύλλο ή από το πληκτρολόγιο. Δεξιά, φαίνονται οι λίστες με τις τιμές της γραμμικής μεταβλητής (πρώτη στήλη — *Linear*) *Freon-12 concentration (ppt)* σε αύξουσα σειρά (τρίτη στήλη — *Index*), οι αντίστοιχες τιμές της κυκλικής μεταβλητής (δεύτερη στήλη — *Theta*) *wind direction ($^{\circ}N$)* στις οποίες έχουν αποδοθεί βαθμοί (τέταρτη στήλη — *Ranks*) και οι τροποποιημένες τιμές της κυκλικής μεταβλητής (κυκλικοί βαθμοί) (πέμπτη στήλη — *Theta Star*).



Σχήμα 6.2: Εφαρμογή του προγράμματος προσομοίωσης για τα δεδομένα της μελέτης των Solow et al. (1988, p. 225). Μετά την εισαγωγή των δεδομένων και την προετοιμασία για την εφαρμογή του ελέγχου (βλ. προηγούμενη εικόνα 6.1), υπολογίζονται οι τιμές του στατιστικού U και του κανονικοποιημένου αντιστοίχου του D . Κατόπιν, εφαρμόζεται η ρουτίνα παραγωγής 10.000 τυχαίων μεταθέσεων των τιμών της γραμμικής μεταβλητής (οι οποίες έχουν αντιστοιχιστεί με τους δείκτες i – στήλη *Index*), με τις τιμές της κυκλικής μεταβλητής να παραμένουν σταθερές. Ο υπολογισμός της P -value γίνεται από την εμπειρική κατανομή του στατιστικού U που προκύπτει από αυτή τη διαδικασία. Κάτω δεξιά εμφανίζεται η λίστα με τις παραχθείσες μεταθέσεις και την τιμή του στατιστικού U για κάθε μία από αυτές. Όπως σημειώνουν και οι συγγραφείς, η τιμή P -value κινείται στην περιοχή του 0.06.



Σχήμα 6.3: Εφαρμογή του προγράμματος προσομοίωσης για τον έλεγχο συσχέτισης των μεταβλητών TEMPER_H2 και WINDIR_H2.

Βιβλιογραφία

- E. Batschelet. "*Circular Statistics in Biology*". Academic Press, 1981.
- N.I. Fisher and A.J. Lee. "regression models for an angular response". *Biometrics*, 48:665–667, 1992.
- R.A. Johnson and T. Wehrly. "measures and models for angular correlation and angular–linear correlation". *Journal of the Royal Statistical Society*, pages 222–229, 1977.
- I.G. Liddell and J.K. Ord. "linear–circular correlation coefficients: Some further results". *Biometrika*, 65:448–450, 1978.
- K. Mardia. "linear–circular correlation coefficients and rhythmometry". *Biometrika*, 63:403–405, 1976.
- K. Mardia and P. Jupp. "*Directional Statistics*". John Wiley & Sons, 2000.
- K.V. Mardia and T.W. Sutton. "a model for cylindrical variables with applications". *Journal of the Royal Statistical Society*, 1978.
- S.B. Maurer and A. Ralston. "*Discrete Algorithmic Mathematics*". A K Peters, 2004.
- A. Solow, J. Bullister, and S. Nevison. "an application of circular–linear correlation analysis to the relationship between freon concentration and wind direction in woods hole massachusetts". *Environmental Monitoring and Assessment*, 10: 219–228, 1988.
- J. Zar. "*Biostatistical Analysis*". Prentice–Hall, 1999.

Ευρετήριο

- chi-squared test, 57
chi-squared test (two-sample), 88
Cox's test, 62
- Hodges-Ajne test, 50
Hotelling's one-sample test, 119
Hotelling's two-sample test, 122
- Kuiper's test, 59
Kuiper's test (two-sample), 89
- Mardia's two-sample test, 123
Mardia-Watson-Wheeler test, 79
modified Hodges-Ajne test, 52
Moore's test, 120
- range test, 55
rank-sum test, 98
Rao's spacing test, 53
Rayleigh test, 41
run test, 99
- V test, 46
- Watson's U^2 test, 61
Watson's U^2 test (two-sample), 91
Watson-Williams test, 67
- γενικός μέσος k μέσων γωνιών, 111
γραμμικό ιστόγραμμα, 5
γωνιακή απόκλιση, 17, 101
γωνιακή διακύμανση, 17
- δειγματική κυκλική διασπορά, 20
διάμεση κατεύθυνση, 15
διόρθωση για ομαδοποιημένα δεδομένα,
19
ελλειπτική περιοχή εμπιστοσύνης, 115
κατανομή von Mises, 29
κυκλικά δεδομένα, xxiii
κυκλικές κατανομές, 25
κυκλική–γραμμική συσχέτιση, 133
κυκλική–κυκλική συσχέτιση, 129
κυκλικό διάγραμμα διασποράς, 1
κυκλικό διμεταβλητό ιστόγραμμα, 5
κυκλικό ιστόγραμμα, 4
μέση κατεύθυνση, 6
μέσο διάνυσμα, 6
μέτρο της συγκέντρωσης, 16
ομοιόμορφη κατανομή, 27
- παράγοντας επαναπατρισμού, 21
ροδόγραμμα, 2