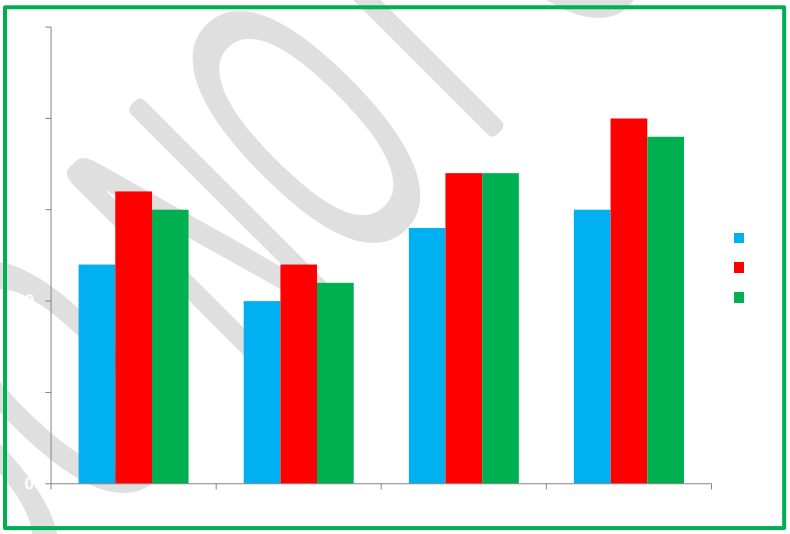
Τμήμα Γεωπονίας ΑΠΘ

Εργαστήριο “Γεωργίας”

Τετράδιο Ασκήσεων Βιομετρίας (Ι)



Γεώργιος Μενεξές

Αναπληρωτής Καθηγητής

Θεσσαλονίκη, Ιούλιος 2021

DO NOT COPY

Άσκηση 1

Α) Στον Πίνακα 1 δίνονται 100 τιμές ενός βιολογικού δείκτη (κίτρινη επισήμανση). Θεωρούμε ότι οι 100 μετρήσεις αποτελούν τον **πληθυσμό** που μας ενδιαφέρει να μελετήσουμε. Να υπολογίσετε τα παρακάτω:

1. Τη μικρότερη τιμή.
2. Τη διάμεση τιμή.
3. Τη μεγαλύτερη τιμή.
4. Το εύρος των τιμών.
5. Τα ποσοστιαία σημεία Q25, Q50 και Q75 (τεταρτημόρια) της κατανομής.
6. Το ενδοτεταρτημοριακό εύρος της κατανομής.
7. Το ήμι-ενδοτεταρτημοριακό της κατανομής.
8. Τον αριθμητικό μέσο όρο.
9. Τον αρμονικό μέσο όρο.
10. Την παραλλακτικότητα/διακύμανση/διασπορά.
11. Την τυπική απόκλιση.
12. Τον συντελεστή CV (%).
13. Τον συντελεστή ασυμμετρίας της κατανομής.
14. Τον συντελεστή κύρτωσης της κατανομής.

Β) Επιπροσθέτως,

1. Να απεικονίσετε την κατανομή των 100 μετρήσεων με Ιστόγραμμα (7 κλάσεις/διαστήματα) και Θηκοδιάγραμμα (box plot).
2. Να πραγματοποιήσετε όλους τους απαραίτητους ελέγχους για να διαπιστώσετε αν οι τιμές προσαρμόζονται ικανοποιητικά στην Κανονική Κατανομή.
3. Να πραγματοποιήσετε έλεγχο αν οι τιμές 32,51 και 82,48 είναι outliers.
4. Να μετασχηματίσετε τις 100 αρχικές έτσι ώστε ο μέσος όρος των μετασχηματισμένων τιμών να είναι ίσος με 100 και η τυπική απόκλιση ίση με 20.
5. Στις 100 αρχικές τιμές να "εισάγετε" τυχαία αβεβαιότητα (uncertainty) της τάξης του $\pm$ 15 και να υπολογίσετε τον μέσο όρο, την τυπική απόκλιση και το τυπικό σφάλμα στην εκτίμηση του μέσου όρου. Πως επηρεάζει η αβεβαιότητα τις εκτιμήσεις των 3 στατιστικών δεικτών;
6. Στις 100 αρχικές τιμές να εισάγετε τυχαίο "θόρυβο" ($\pm$ noise) από Ομοιόμορφη Κατανομή στο διάστημα [0, 1] και να υπολογίσετε τον μέσο όρο, την τυπική απόκλιση και το τυπικό σφάλμα στην εκτίμηση του μέσου όρου. Πως επηρεάζει ο "θόρυβος" τις εκτιμήσεις των 3 στατιστικών δεικτών;
7. Στις 100 αρχικές τιμές να εισάγετε τυχαίο "θόρυβο" ($\pm$ noise) από την Τυποποιημένη Κανονική Κατανομή και να υπολογίσετε τον μέσο όρο, την τυπική απόκλιση και το τυπικό σφάλμα στην εκτίμηση του μέσου όρου. Πως επηρεάζει ο "θόρυβος" τις εκτιμήσεις των 3 στατιστικών δεικτών;
8. Ποια τιμή πρέπει να προσθέσετε σε κάθε μία από τις 100 αρχικές τιμές ώστε ο συντελεστής CV να γίνει ίσος με 4% (0,04).

DO NOT COPY

Πίνακας 1

ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές
1	47,00	21	46,73	41	35,56	61	56,94	81	72,06
2	37,22	22	46,30	42	41,53	62	53,23	82	64,44
3	52,44	23	63,43	43	34,78	63	40,60	83	63,04
4	62,76	24	49,15	44	46,37	64	47,59	84	51,13
5	61,98	25	48,14	45	49,68	65	51,32	85	50,02
6	67,33	26	44,87	46	50,28	66	55,58	86	54,54
7	28,16	27	69,72	47	46,77	67	51,39	87	49,74
8	47,66	28	58,66	48	71,95	68	40,89	88	39,45
9	60,95	29	73,76	49	32,58	69	68,85	89	32,25
10	39,13	30	43,45	50	42,64	70	54,87	90	58,28
11	43,10	31	66,61	51	24,22	71	50,72	91	54,44
12	33,10	32	33,88	52	64,48	72	58,30	92	56,18
13	31,53	33	55,39	53	37,20	73	58,62	93	52,13
14	40,22	34	59,02	54	43,46	74	43,63	94	39,73
15	42,26	35	69,19	55	57,58	75	40,77	95	62,38
16	28,82	36	49,15	56	54,67	76	61,11	96	46,89
17	44,32	37	44,76	57	58,75	77	37,99	97	41,60
18	45,96	38	56,75	58	55,96	78	34,41	98	41,79
19	51,35	39	46,19	59	36,28	79	57,11	99	45,71
20	46,35	40	57,58	60	38,84	80	56,38	100	45,47

Άσκηση 2

Α) Με τη βοήθεια πίνακα τυχαίων αριθμών ή με τη βοήθεια γεννήτριας τυχαίων αριθμών, από τις 100 τιμές του Πίνακα 1, να επιλέξετε 5 δείγματα μεγέθους 31. Για κάθε δείγμα να υπολογίσετε τα παρακάτω:

1. Τον αριθμητικό μέσο όρο.
2. Την παραλλακτικότητα/διακύμανση/διασπορά.
3. Την τυπική απόκλιση.
4. Το τυπικό σφάλμα στην εκτίμηση του μέσου όρου.
5. Τον συντελεστή CV (%).
6. Ένα 95% διάστημα εμπιστοσύνης (όρια εμπιστοσύνης) για το μέσο όρο του πληθυσμού.
7. Ένα 90% διάστημα εμπιστοσύνης (όρια εμπιστοσύνης) για το μέσο όρο του πληθυσμού.
8. Τα παραπάνω διαστήματα εμπιστοσύνης να υπολογιστούν **και** με τη μέθοδο προσομοίωσης Bootstrap (1000 κύκλοι δειγματοληψίας).

Β) Επιπροσθέτως,

DO NOT COPY

1. Να βρείτε πόσα από τα παραπάνω 10 διαστήματα εμπιστοσύνης περιλαμβάνουν τον πραγματικό μέσο όρο των 100 μετρήσεων;
2. Να υπολογίσετε πόσα είναι όλα τα δυνατά δείγματα μεγέθους 31 από τον πληθυσμό των 100 μετρήσεων;
3. Να δημιουργήσετε 5 μεταβλητές, η κάθε μία με 50 τιμές, από Κανονική Κατανομή με μέσο όρο 100 και τυπική απόκλιση 15.
4. Να δημιουργήσετε 5 μεταβλητές, η κάθε μία με 50 τιμές, από Ομοιόμορφη Κατανομή στο διάστημα [40, 60].
5. Να δημιουργήσετε 5 μεταβλητές, η κάθε μία με 50 τιμές, από Ομοιόμορφη Κατανομή στο διάστημα [0, 1].
6. Να δημιουργήσετε 5 μεταβλητές, η κάθε μία με 50 τιμές, από Ομοιόμορφη Κατανομή με τιμές 0 ή 1.
7. Να υπολογίσετε πόσα είναι όλα τα δυνατά δείγματα μεγέθους 100 από έναν πληθυσμό 1000 μετρήσεων;

Άσκηση 3

Με βάση τα στοιχεία του Πίνακα 1 να υπολογίσετε τα παρακάτω:

1. Το ποσοστό των τιμών που είναι μεγαλύτερες από την τιμή 40.
2. Το ίδιο ποσοστό να εκτιμηθεί από τυχαίο δείγμα μεγέθους 31.
3. Από το τυχαίο δείγμα μεγέθους 31, να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης (όρια εμπιστοσύνης) για το πραγματικό ποσοστό των τιμών που είναι μεγαλύτερες από την τιμή 40.
4. Περιλαμβάνει το διάστημα αυτό το πραγματικό ποσοστό των τιμών που είναι μεγαλύτερες από την τιμή 40;
5. Να απαντήσετε στο ερώτημα 4 κατασκευάζοντας ένα 95% διάστημα εμπιστοσύνης (όρια εμπιστοσύνης) για το πραγματικό ποσοστό των τιμών που είναι μεγαλύτερες από την τιμή 40 με τη μέθοδο προσομοίωσης bootstrap (1000 κύκλοι δειγματοληψίας).

Άσκηση 4

Στον Πίνακα 2 δίνονται οι τιμές ενός τυχαίου δείγματος 100 μετρήσεων (κίτρινη επισήμανση) ενός φυσιολογικού χαρακτηριστικού του φυτικού είδους XXXX.

1. Να βρεθεί η Κατανομή Πιθανότητας του χαρακτηριστικού.
2. Ποιον ή ποιους στατιστικούς ελέγχους καλής προσαρμογής θα χρησιμοποιήσετε;
3. Να εκτιμήσετε τις παραμέτρους της κατανομής.
4. Ποια είναι η φυσική ερμηνεία των παραμέτρων;
5. Σε ποιες περιπτώσεις (προβλήματα, φαινόμενα) χρησιμοποιείται η συγκεκριμένη κατανομή;

DO NOT COPY

Πίνακας 2

ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές	ΑΑ	Τιμές
1	350	21	242	41	295	61	249	81	263
2	237	22	233	42	120	62	270	82	168
3	130	23	398	43	167	63	150	83	223
4	79	24	410	44	167	64	241	84	236
5	164	25	219	45	250	65	209	85	215
6	265	26	200	46	220	66	230	86	255
7	350	27	265	47	278	67	163	87	195
8	113	28	221	48	186	68	128	88	206
9	127	29	193	49	232	69	192	89	244
10	105	30	200	50	140	70	183	90	166
11	223	31	200	51	115	71	150	91	132
12	101	32	126	52	135	72	256	92	240
13	142	33	200	53	264	73	134	93	184
14	85	34	274	54	236	74	300	94	228
15	210	35	192	55	170	75	190	95	221
16	230	36	212	56	170	76	216	96	218
17	181	37	112	57	237	77	110	97	132
18	191	38	65	58	120	78	110	98	188
19	266	39	223	59	230	79	139	99	188
20	216	40	191	60	177	80	132	100	263

Άσκηση 5

1. Μπορείτε να δημιουργήσετε ακριβώς τα ίδια 3 τυχαία δείγματα από μια κατανομή;
2. Να δημιουργήσετε δύο τυχαίες μεταβλητές, με 50 τιμές η κάθε μία, ώστε να είναι γραμμικά ασυσχέτιστες. Στη συνέχεια, να τροποποιήσετε τις τιμές των δύο μεταβλητών ώστε να δημιουργήσετε 3 νέα ζεύγη μεταβλητών που να έχουν συντελεστή γραμμικής συσχέτισης (*r* του Pearson) ίσο (περίπου) με 0,30, 0,85 και -0,60 αντίστοιχα.
3. Να δημιουργήσετε 50 τυχαίους αριθμούς με ακρίβεια 7 δεκαδικών. Στη συνέχεια, να κάνετε στρογγυλοποίηση των αριθμών στα 5, 3 και 0 δεκαδικά ψηφία. Στις αρχικές 50 τιμές να κάνετε αποκοπή στα 5, 3 και 0 δεκαδικά ψηφία. Σε κάθε περίπτωση να υπολογίσετε τον αριθμητικό μέσο όρο, την τυπική απόκλιση και το τυπικό σφάλμα στην εκτίμηση του μέσου όρου. Τι παρατηρείτε σχετικά με την επίδραση της στρογγυλοποίησης ή της αποκοπής στην εκτίμηση των 3 στατιστικών δεικτών;
4. Με τη βοήθεια του γραμμικού υποδείγματος Y=2,698+12,457X (Y=a+bX) να εκτιμήστε την τιμή της εξαρτημένης μεταβλητής Y όταν η ανεξάρτητη μεταβλητή X πάρει την τιμή 122,489. Να κάνετε το ίδιο, αλλά με στρογγυλοποίηση των συντελεστών a και b της ευθείας, στα 2, 1 και 0 δεκαδικά ψηφία. Να κάνετε το ίδιο με αποκοπή των τιμών των συντελεστών a και b της ευθείας στα 2, 1 και 0 δεκαδικά ψηφία. Στο αρχικό υπόδειγμα, να εκτιμήσετε την τιμή της Y για X=122,49, X=122,5, X=122 και X=123. Σε κάθε περίπτωση η

DO NOT COPY

εκτίμηση της *Y* να γίνει με 3 δεκαδικά ψηφία. Τι παρατηρείτε σχετικά με την επίδραση της στρογγυλοποίησης ή της αποκοπής στην εκτίμηση της *Y*;

5. Στο γραμμικό υπόδειγμα *Y*=2,698+12,457*X* (*Y*=a+b*X*) ας υποθέσουμε ότι ένα 95% διάστημα εμπιστοσύνης (όρια εμπιστοσύνης) για τον σταθερό όρο a είναι το [2,002, 3,417] και ένα 95% διάστημα εμπιστοσύνης για τον συντελεστή παλινδρόμησης (κλίση) b είναι το [11,005, 13,998]. Αν η ανεξάρτητη μεταβλητή *X* παίρνει τιμές στο διάστημα [10, 20] και αν υποθέσουμε ότι οι "λογικές" τιμές για την εξαρτημένη μεταβλητή *Y* αναμένουμε να βρίσκονται στο διάστημα [150, 270] τότε να μελετηθεί η καταλληλότητα του υποδείγματος ως προς τις πιθανές τιμές των a, b και *X* σε σχέση με τις εκτιμώμενες τιμές της *Y*. Αν υποθέσουμε ότι οι βέλτιστες τιμές της *Y* είναι στο διάστημα 200 ± 5 τότε να εντοπίσετε τιμές (ή εύρη τιμών) των a, b και *X* ώστε να βελτιστοποιείται η *Y*;

Άσκηση 6

Στον Πίνακα 3 παρουσιάζονται τα αποτελέσματα ενός πειράματος που αφορά στη σύγκριση της αποτελεσματικότητας δύο φαρμάκων που χορηγήθηκαν σε δύο ομάδες ασθενών (*n*=6 ασθενείς σε κάθε ομάδα). Οι τιμές εκφράζουν την αποτελεσματικότητα του φαρμάκου σε μια κλίμακα 0-10 (0: καθόλου αποτελεσματικό έως 10: Πάρα πολύ αποτελεσματικό). Να υπολογίσετε τη διαφορά των μέσων όρων αποτελεσματικότητας των δύο φαρμάκων (Φάρμακο Β - Φάρμακο Α). Στη συνέχεια, με τυχαία διαδικασία να μεταθέσετε (permute), χωρίς επανατοποθέτηση, τις 12 τιμές στις δύο ομάδες 10 φορές. Φροντίστε το πλήθος των τιμών στις δύο ομάδες να είναι σχετικά ισορροπημένο, δηλαδή να μην δημιουργηθεί ομάδα με λιγότερες από 4 μετρήσεις. Συνεπώς, θα πρέπει να δημιουργήσετε 10 νέους πίνακες. Με τα στοιχεία του κάθε πίνακα να υπολογίσετε τις διαφορές των μέσων όρων αποτελεσματικότητας των δύο φαρμάκων (Φάρμακο Β - Φάρμακο Α). Σε πόσες περιπτώσεις από τις 10, το πρόσημο της διαφοράς είναι θετικό γεγονός που σημαίνει ότι κατά μέσο όρο το φάρμακο Β είναι πιο αποτελεσματικό από το Α;

Πίνακας 3

Φάρμακο Α	Φάρμακο Β
3	6
2	6
1	9
8	1
7	2
6	9

Άσκηση 7

Στον Πίνακα 4 παρουσιάζεται η κοινή κατανομή συχνοτήτων δύο ποιοτικών μεταβλητών *A* και *K*. Οι γραμμές του πίνακα αντιστοιχούν σε 5 ανθοκομικά είδη (A1-A5) και οι στήλες σε 4 κλάσεις ποιότητας γύρης (K1-K4, η τιμή K1 αντιστοιχεί στην άριστη ποιότητα). Τα δεδομένα

DO NOT COPY

συγκεντρώθηκαν για 200 φυτά με μέθοδο της τυχαίας δειγματοληψίας. Θεωρείστε το δείγμα "μεγάλο" και αντιπροσωπευτικό των αντίστοιχων πληθυσμών.

1. Να εξεταστεί αν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.
2. Πόσο ισχυρή είναι η συσχέτιση (αν υπάρχει);
3. Πως αιτιολογείται η συσχέτιση (αν υπάρχει);
4. Αν υποθέσουμε ότι ο στόχος της έρευνας είναι να γίνει σύγκριση των 5 ανθοκομικών ειδών ως προς την κατανομή των κλάσεων ποιότητας γύρης, τότε με βάση τα δεδομένα του Πίνακα 4 να κατασκευαστεί κατάλληλο στατιστικό διάγραμμα που να διευκολύνει την οπτικοποίηση των επιθυμητών συγκρίσεων.

Πίνακας 4

	K1	K2	K3	K4	Σύνολα
A1	11	6	9	14	40
A2	7	6	7	9	29
A3	14	5	7	11	37
A4	11	4	7	20	42
A5	22	2	12	16	52
Σύνολα	65	23	42	70	200

5. Να κατασκευαστεί ο πίνακας με τα "προφίλ" των γραμμών του Πίνακα 4 (σχετικές συχνότητες επί του συνόλου γραμμής).
6. Να κατασκευαστεί ο πίνακας με τα "προφίλ" των στηλών του Πίνακα 4 (σχετικές συχνότητες επί του συνόλου στήλης).
7. Να κατασκευαστεί ο πίνακας αντιστοιχιών του Πίνακα 4 (σχετικές συχνότητες επί του γενικού συνόλου).
8. Να υπολογιστεί η περιθώρια κατανομή σχετικών συχνοτήτων της μεταβλητής *A*.
9. Να υπολογιστεί η περιθώρια κατανομή σχετικών συχνοτήτων της μεταβλητής *K*.
10. Να υπολογιστεί η ποσότητα (**εντροπία** της μεταβλητής *K* σε nats):

    $S(K)= -\sum p(Kj)ln(p(Kj))$, $j=1\ldots4$ και $p(Kj)$ είναι η πιθανότητα του ενδεχομένου Kj.
11. Να υπολογιστεί η ποσότητα (**εντροπία** της μεταβλητής *A* σε nats):

    $S(A)= -\sum p(Ai)ln(p(Ai))$, $i=1\ldots5$ και $p(Ai)$ είναι η πιθανότητα του ενδεχομένου Ai.
12. Να υπολογιστεί η ποσότητα (**κοινή εντροπία** των μεταβλητών *K* και *A* σε nats)::

    $S(A, K)= -\sum p(Ai\cap Kj)ln(p(Ai\cap Kj))$, $i=1\ldots5$ και $j=1\ldots4$ και $p(Ai\cap Kj)$ είναι η πιθανότητα ταυτόχρονης πραγματοποίησης των ενδεχομένων Ai και Kj.
13. Να υπολογιστεί η ποσότητα (**πληροφοριακή ή εντροπική αλληλεξάρτηση** των μεταβλητών *A* και *K*):

    $$\Pi_{KA}= \frac{S(K)+S(A)}{S(A,K)} -1$$

DO NOT COPY

14. Οι υπολογισμοί στα ερωτήματα 6 έως 9 να γίνουν κάνοντας χρήση του $\log_2$ (εντροπία σε bits) αντί του ln.

Άσκηση 8

Τα ερωτήματα της άσκησης 7 να απαντηθούν με βάση τα δεδομένα του Πίνακα 5.

Πίνακας 5

	K1	K2	K3	K4	Σύνολα
A1	16	0	104	16	136
A2	0	0	4	8	12
A3	12	4	8	8	32
A4	4	0	8	12	24
A5	0	0	0	4	4
Σύνολα	32	4	124	48	208

Επιπροσθέτως,

1. Ποια ανθοκομικά είδη διαφοροποιούνται στατιστικά σημαντικά, σε επίπεδο σημαντικότητας α=0,05 (p≤0,05), ως προς την κατανομή των κλάσεων ποιότητας γύρης;
2. Αν θεωρήσουμε ως ανεξάρτητη μεταβλητή την *A* (ανθοκομικά είδη) και εξαρτημένη την *K* (κλάσεις ποιότητας γύρης), πόσο βελτιώνεται η πρόβλεψη των τιμών της *K* γνωρίζοντας τις τιμές της *A* ή ισοδύναμα ποια είναι η αναλογική μείωση του σφάλματος πρόβλεψης της *K* γνωρίζοντας την *A*;
3. Σε ποια περίπτωση (Πίνακας 4 *vs* Πίνακας 5) η πληροφοριακή ή εντροπική αλληλεξάρτηση μεταξύ των δύο μεταβλητών είναι ισχυρότερη;
4. Για κάθε κελί του Πίνακα 5 να υπολογιστούν δύο τουλάχιστον μέτρα (δείκτες) απόκλισης από την (υποθετική) κατάσταση ανεξαρτησίας των δύο μεταβλητών *A* και *K*.

Άσκηση 9

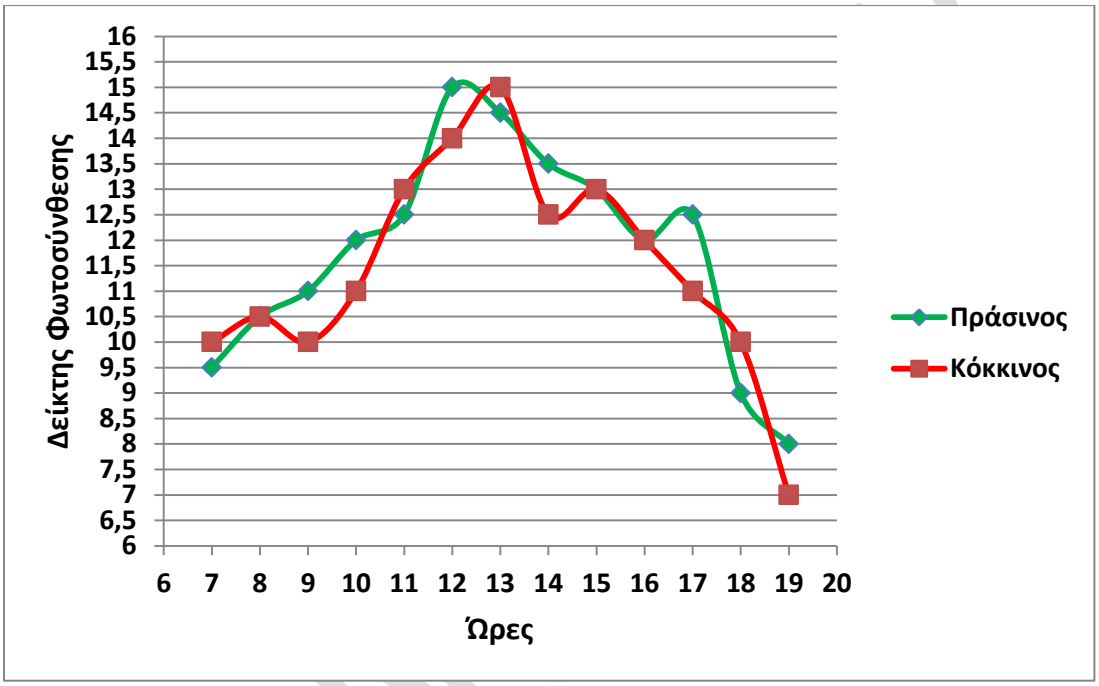
Στο Διάγραμμα 1 παρουσιάζεται η ωριαία διακύμανση (από 7:00 έως 17:00) ενός δείκτη φωτοσύνθεσης για δύο ποικιλίες βασιλικού (πράσινος και κόκκινος).

1. Σε ποιον από τους δύο βασιλικούς ο δείκτης φωτοσύνθεσης έχει τη μεγαλύτερη διακύμανση; Ποιος βασιλικός παρουσιάζει εν γένει το μεγαλύτερο δείκτη φωτοσύνθεσης κατά τη διάρκεια της ημέρας; Υπόδειξη: Να υπολογίσετε με μεθόδους της Αριθμητικής Ανάλυσης την επιφάνεια (εμβαδό) κάτω από την καμπύλη (Area Under the Curve-AUC) για κάθε βασιλικό.
2. Μελετώντας το διάγραμμα, ποια ώρα της ημέρας φαίνεται να έχει μέγιστο ο δείκτης φωτοσύνθεσης για κάθε βασιλικό;
3. Προσαρμόστε στα δεδομένα για κάθε βασιλικό σε κατάλληλο υπόδειγμα (μαθηματικό μοντέλο) παλινδρόμησης και γράψτε τη μαθηματική έκφραση του κάθε υποδείγματος.

DO NOT COPY

4. Να πραγματοποιήσετε όλους τους σχετικούς ελέγχους των προϋποθέσεων εγκυρότητας των αντίστοιχων υποδειγμάτων παλινδρόμησης.
5. Για ποιον βασιλικό τα δεδομένα έχουν πιο καλή προσαρμογή στο αντίστοιχο υπόδειγμα παλινδρόμησης;
6. Με βάση τα μαθηματικά υποδείγματα, να υπολογίσετε την επιφάνεια (εμβαδό) κάτω από την καμπύλη (Area Under the Curve-AUC) για κάθε βασιλικό. Να σχεδιάστε και τις αντίστοιχες γραφικές παραστάσεις των υποδειγμάτων σε κοινό διάγραμμα.

Διάγραμμα 1



7. Με βάση τα μαθηματικά υποδείγματα, ποια ώρα της ημέρας ο δείκτης φωτοσύνθεσης αναμένεται να έχει μέγιστο και ποια ώρα ελάχιστο για κάθε βασιλικό; Ποιες είναι οι τιμές μεγίστου; Υπάρχει συμφωνία με τις αντίστοιχες τιμές για το μέγιστο που υπολογίσατε στο ερώτημα 2. Σε κάθε περίπτωση, πόση είναι η απόκλιση μεταξύ τους;
8. Υπάρχει διαφορά μεταξύ των δύο εκτιμήσεων της AUC όπως αυτήν την υπολογίσατε στο ερώτημα 1 και στο ερώτημα 6; Πόση είναι η απόκλιση;
9. Να γίνει εκτίμηση του δείκτη φωτοσύνθεσης κατά την χρονική στιγμή 12:30 και για τους δύο βασιλικούς. Το ίδιο και για την χρονική στιγμή 17:30.
10. Αν στα δεδομένα και των δύο βασιλικών "εισάγουμε" τυχαία αβεβαιότητα της τάξης του ±0,5 να απαντήσετε στα ερωτήματα 3 έως 9. Τι παρατηρείτε;
11. Με βάση τα μαθηματικά υποδείγματα που προσδιορίσατε χρησιμοποιώντας τα αρχικά δεδομένα από το Διάγραμμα 1 (και τα αριθμητικά αποτελέσματα των αντίστοιχων στατιστικών αναλύσεων) να κατασκευάσετε και να απεικονίσετε τεχνητά δεδομένα (προσομοίωση) για ένα φυτό πράσινου και ένα φυτό κόκκινου βασιλικού.

12. Στα δεδομένα του Διαγράμματος 1 να προσαρμόσετε τις αντίστοιχες καμπύλες Loess και να περιγράψετε την ωριαία διακύμανση του δείκτη φωτοσύνθεσης για κάθε βασιλικό.

Άσκηση 10 (ερωτήματα 1-7 προσαρμογή από Αποστολόπουλος, 1991, σ.218)

Μία στρατιωτική μονάδα πρόκειται να δεχθεί για κατάταξη 10000 στρατευσίμους και θέλει να παραγγείλει τον αναγκαίο αριθμό στολών. Οι στολές, με βάση το ανάστημα (ύψος) των στρατευσίμων, διαιρούνται σε τρία μεγέθη ως εξής:

Μέγεθος 1: Για ανάστημα ≤165 cm

Μέγεθος 2: Για ανάστημα μεταξύ 165 και 182 cm

Μέγεθος 3: Για ανάστημα ≥182 cm

Είναι γνωστό (από παλαιότερες έρευνες) ότι το ανάστημα των στρατευσίμων ακολουθεί Κανονική Κατανομή με μέσο όρο μ=175 cm και τυπική απόκλιση σ=4 cm.

1. Πόσες στολές από κάθε μέγεθος θα πρέπει να παραγγείλει η μονάδα;
2. Κάτω ποιου αναστήματος βρίσκεται το 10% των στρατευσίμων;
3. Άνω ποιου αναστήματος βρίσκεται το 15% των στρατευσίμων;
4. Έστω ότι η μονάδα υποδέχεται τους πρώτους 30 στρατευσίμους. Πόσοι από αυτούς αναμένεται να χρειάζονται στολή με μέγεθος 3;
5. Πόσο πιθανό είναι το πολύ οι μισοί (15) να χρειάζονται στολή με μέγεθος 3;
6. Έστω ότι η μονάδα υποδέχεται τους πρώτους 300 στρατευσίμους. Πόσοι από αυτούς αναμένεται να χρειάζονται στολή με μέγεθος 2.
7. Πόσο πιθανό είναι το πολύ οι μισοί (150) να χρειάζονται στολή με μέγεθος 2;
8. Στον Πίνακα 6 παρουσιάζονται τα αποτελέσματα μιας πρόσφατης μελέτης αναφορικά με το ανάστημα (ύψος) των στρατευσίμων, από τυχαίο και αντιπροσωπευτικό δείγμα 100 ατόμων. Με τα νέα δεδομένα, να απαντήσετε στα ερωτήματα 1-7.
9. Σε παλαιότερη έρευνα, 100000 στρατεύσιμοι υποβλήθηκαν σε έναν διαγνωστικό έλεγχο (τεστ) για να διαπιστωθεί αν πάσχουν από μια μεταδοτική νόσο. Ο διαγνωστικός έλεγχος έχει **ευαισθησία** 0,7333 και **ειδικότητα** 0,9715. Ο **επιπολασμός** της νόσου στον γενικό πληθυσμό είναι 0,000093. 10000 στρατεύσιμοι ανιχνεύθηκαν θετικοί στη νόσο. Πόσα από τα 10000 θετικά τεστ καταδεικνύουν πραγματικά περιστατικά της νόσου;

   Βοήθεια: **ευαισθησία** ενός διαγνωστικού ελέγχου είναι η πιθανότητα **σωστής θετικής** διάγνωσης της ασθένειας (δηλαδή, η πιθανότητα το τεστ να είναι θετικό και το άτομο να έχει όντως τη νόσο) ενώ **ειδικότητα** είναι αντίστοιχα η πιθανότητα **σωστής αρνητικής** διάγνωσης (δηλαδή, η πιθανότητα το τεστ να είναι αρνητικό και το άτομο όντως να μην έχει τη νόσο). **Επιπολασμός** είναι η εκ των προτέρων πιθανότητα εμφάνισης της νόσου. Συμβουλευτείτε και την Εικόνα 1.
10. Η στρατιωτική μονάδα υποδέχθηκε τους 10000 στρατευσίμους και τους έδωσε στολές. Οι 10000 στρατεύσιμοι υποβλήθηκαν και στον διαγνωστικό έλεγχο για να διαπιστωθεί αν

DO NOT COPY

είναι φορείς/νοσούν της μεταδοτικής νόσου. 1000 από τα τεστ έδειξαν θετικό αποτέλεσμα. Πόσοι από τους 1000 στρατευσίμους νοσούν πραγματικά;

11. Να απαντήσετε στα ερωτήματα 9 και 10 αν ο επιπολασμός της νόσου είναι 0,02, η ευαισθησία του τεστ είναι 0,938 και η ειδικότητα 0,964.

Πίνακας 6

ΑΑ	Τιμές	**ΑΑ**	**Τιμές**	**ΑΑ**	**Τιμές**	**ΑΑ**	**Τιμές**	**ΑΑ**	**Τιμές**
1	179,24	21	185,46	41	181,06	61	181,92	81	184,53
2	184,33	22	189,85	42	172,74	62	174,23	82	174,45
3	173,46	23	179,64	43	180,46	63	168,96	83	194,50
4	176,37	24	176,47	44	171,76	64	181,28	84	182,44
5	191,59	25	185,22	45	182,60	65	172,90	85	176,75
6	179,25	26	182,61	46	182,86	66	181,10	86	172,22
7	172,40	27	179,34	47	177,48	67	182,76	87	175,27
8	189,41	28	177,49	48	185,15	68	180,71	88	184,91
9	178,42	29	178,50	49	167,53	69	173,29	89	175,30
10	180,54	30	175,86	50	176,88	70	170,82	90	172,65
11	186,30	31	182,05	51	186,15	71	187,88	91	180,44
12	179,27	32	173,55	52	167,63	72	187,38	92	183,16
13	175,59	33	184,82	53	181,52	73	182,06	93	175,43
14	192,81	34	176,32	54	179,91	74	172,27	94	171,63
15	180,11	35	176,06	55	182,03	75	183,89	95	173,11
16	189,16	36	178,65	56	171,58	76	185,85	96	189,80
17	172,91	37	188,14	57	182,53	77	181,37	97	182,76
18	175,98	38	171,93	58	171,70	78	178,63	98	175,06
19	170,03	39	176,26	59	191,10	79	181,90	99	170,52
20	186,96	40	186,14	60	178,81	80	182,76	100	190,90

Εικόνα 1

DO NOT COPY

Άσκηση 11

Είναι γνωστό ότι η απόδοση της ποικιλίας Α ενός φυτικού είδους ZZZZ ακολουθεί Κανονική Κατανομή με μέσο όρο μ=100 και τυπική απόκλιση σ=30. Είναι γνωστό ότι η απόδοση της ποικιλίας Β του ιδίου φυτικού είδους ακολουθεί Κανονική Κατανομή με μέσο όρο μ=120 και τυπική απόκλιση σ=25. Συνεπώς, είναι **βέβαιο** ότι οι δύο ποικιλίες κατά μέσο όρο διαφέρουν ως προς τη μέση απόδοση (μ) και ότι η ποικιλία Β είναι πιο αποδοτική από την Α. Επιπροσθέτως, είναι **βέβαιο** ότι διαφοροποιείται και η παραλλακτικότητα (σ^2) της απόδοσης μεταξύ των δύο ποικιλιών.

1. Για κάθε ποικιλία, να κατασκευάστε 50 τυχαία δείγματα μεγέθους 5, 50 μεγέθους 10 και 50 μεγέθους 20, από τις αντίστοιχες κατανομές (συνολικά 150 δείγματα).
2. Σε κάθε περίπτωση, με το *t*-test και το Mann-Whitney test να καταγράψετε πόσες φορές επαληθεύεται το γεγονός ότι οι δύο ποικιλίες διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05 και μάλιστα ότι η ποικιλία Β είναι πιο αποδοτική από την Α. Να κάνετε το ίδιο και για τις παραλλακτικότητες.

   <u>Υπόδειξη</u>: Η παρατηρούμενη στάθμη σημαντικότητας (*p*-value) του ελέγχου Mann-Whitney να υπολογιστεί με τη μέθοδο προσομοίωσης Monte-Carlo (10.000 κύκλοι δειγματοληψίας). Το *t*-test να πραγματοποιηθεί **και** με τη μέθοδο προσομοίωσης bootstrap (1.000 κύκλοι δειγματοληψίας).

3. Για κάθε έναν από τους ελέγχους (*t*-test και Mann-Whitney test) που θα πραγματοποιήσετε να υπολογίσετε ενδεικτικά και την αντίστοιχη *post hoc* ισχύ (power, γ). Ειδικότερα, για κάθε κλάση μεγέθους δείγματος, επιλέξτε 3 ελέγχους με στατιστικά ($p \leq 0,05$) και 2 ελέγχους με μη στατιστικά σημαντικά ($p > 0,05$) αποτελέσματα.
4. Με βάση τη διαθέσιμη πληροφορία για την κατανομή της απόδοσης των δύο ποικιλιών, πόσο θα έπρεπε να είναι το ελάχιστο μέγεθος δείγματος για κάθε ποικιλία, ώστε το *t*-test και το Mann-Whitney test να αναδείξουν τη διαφορά των δύο ποικιλιών ως στατιστικά σημαντική σε επίπεδο σημαντικότητας α=0,05 και *a priori* επίπεδο ισχύος γ=0,80;
5. Να απαντήσετε στο ερώτημα 4 για α=0,05 και γ=0,90.
6. Σε ποια γενικά συμπεράσματα καταλήξατε;
7. Να απαντήσετε στα ερωτήματα 1, 2 (χωρίς τον έλεγχο για τις παραλλακτικότητες) και 6 θεωρώντας ότι η πραγματική απόδοση της ποικιλίας Α ακολουθεί συνεχή Ομοιόμορφη Κατανομή με τιμές στο διάστημα [90, 110] και η πραγματική απόδοση της ποικιλίας Β ακολουθεί συνεχή Ομοιόμορφη Κατανομή με τιμές στο διάστημα [110, 130].

Άσκηση 12

1. Να επιβεβαιώσετε ή να απορρίψετε τον ισχυρισμό ότι σε μια πόλη 1.000.000 κατοίκων υπάρχουν τουλάχιστον δύο άτομα τα οποία έχουν τον ίδιο αριθμό τριχών στο κεφάλι.
2. Να προτείνετε μεθοδολογία που να επιτρέπει τον υπολογισμό του πλήθους (περίπου) των ψαριών που υπάρχουν σε μια μικρή λίμνη. Διατυπώστε και τις κατάλληλες προϋποθέσεις εφαρμογής της προτεινόμενης μεθοδολογίας.

DO NOT COPY

3. Να ελεγχθεί αν η παρακάτω ακολουθία (σειρά εμφάνισης) συμβόλων είναι τυχαία (random):
   FFFMMFFFFMFMMMFFFFMMFFFMM
   Να κάνετε το ίδιο και για τις παρακάτω ακολουθίες:
   MMFFFFMFFMMFMMMMFFMMFMFFMMMMFFMFMMFFMMMF
   2 1 2 1 2 1 2 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1 2 1 2 1 2 1
4. Να δείξετε εμπειρικά με ένα διάγραμμα ότι ο λόγος δύο διαδοχικών όρων (F_{n+1}/F_n) της ακολουθίας Fibonacci συγκλίνει στον αριθμό $\varphi=\frac{1}{2}(1+\sqrt{5})$ (χρυσή τομή, χρυσή αναλογία). Από ποιον όρο (n) της ακολουθίας και μετά ο λόγος είναι ίσος με 1,618034 (με ακρίβεια 6 δεκαδικών); <u>Βοήθεια</u>: η ακολουθία Fibonacci είναι 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, …, δηλαδή ισχύει $F_0=1$, $F_1=1$ και $F_n=F_{n-2}+F_{n-1}$ για n>1, όπου n φυσικός αριθμός.
5. Έστω ότι σε έναν γαλαξία υπάρχουν *n* πλανήτες και ότι η πιθανότητα ένας από αυτούς να φιλοξενεί νοήμονα ζωή είναι 1/*n* (στον Γαλαξία μας γνωρίζουμε έναν τουλάχιστον τέτοιον πλανήτη). Να υπολογιστεί η πιθανότητα να υπάρχουν δύο ή περισσότεροι πλανήτες (δηλαδή τουλάχιστον δύο) που να φιλοξενούν νοήμονα ζωή. Τι σκέφτεστε, η πιθανότητα είναι "μεγάλη" ή "μικρή";
6. Σε μία πόλη 1.000.000 κατοίκων θέλουμε να εκτιμήσουμε τον αριθμό των ατόμων που είναι φορείς μιας μολυσματικής ασθένειας. Γνωρίζουμε ότι το ποσοστό των φορέων της ασθένειας αυτής κυμαίνεται μεταξύ 3 και 5%. Θέλουμε το σφάλμα της εκτίμησης του συνολικού αριθμού των φορέων να είναι μικρότερο των 500 ατόμων με σιγουριά 0,95 (95%). Πόσο θα πρέπει να είναι το ελάχιστο μέγεθος δείγματος, ώστε να επιτευχθεί η επιθυμητή ακρίβεια (θεωρείστε ότι το δείγμα θα συγκεντρωθεί με τη μέθοδο της απλής τυχαίας δειγματοληψίας).

Άσκηση 13

1. Οι παρακάτω παραμετρικές εξισώσεις περιγράφουν τη συμπεριφορά ενός βιολογικού συστήματος, όπου X και Y πραγματικοί αριθμοί:

   $$X_{n+1}=\sin(aY_n)-\cos(bX_n)$$

   $$Y_{n+1}=\sin(cX_n)-\cos(dY_n)$$

   Να εντοπίσετε εμπειρικά κάποιες τιμές των παραμέτρων a, b, c και d για τις οποίες το σύστημα έχει σχετικά προβλέψιμη συμπεριφορά και κάποιες τιμές για τις οποίες η συμπεριφορά είναι "χαοτική". <u>Υπόδειξη</u>: Αρχικά θέστε a=b=c=d=0 και στη συνέχεια κάντε δοκιμές για θετικές ακέραιες τιμές των 4 παραμέτρων. Κατόπιν, κάντε δοκιμές και με πραγματικές τιμές. Θα σας διευκολύνει πάρα πολύ αν κατασκευάσετε το διάγραμμα διασποράς (scatter plot) μεγάλου αριθμού (πχ 1.000) ζευγών σημείων (X, Y).
2. Σε σχέση με το προηγούμενο ερώτημα. Επιλέξτε ένα από τα διαγράμματα διασποράς (spot) όπου το σύστημα είναι σε κατάσταση "χαοτική" και επί αυτού σχεδιάστε την καμπύλη Loess. Μπορείτε να αναγνωρίσετε κάποιο πρότυπο (μεταβάλλοντας κατάλληλα τις παραμέτρους της καμπύλης, δηλαδή το % προσαρμογής σημείων και τον πυρήνα-kernel);

DO NOT COPY

Αν ναι, να προτείνεται ένα μαθηματικό υπόδειγμα που να συνδέει την μεταβλητή X με την μεταβλητή Y. Για το προτεινόμενο υπόδειγμα να πραγματοποιήσετε όλους τους σχετικούς ελέγχους των προϋποθέσεων εγκυρότητας.

3. Το παρακάτω σύστημα εξισώσεων περιγράφει μια συνιστώσα της συμπεριφοράς ενός βιολογικού φαινομένου, όπου X και Y πραγματικοί αριθμοί:

$$1X+2Y=4$$
$$2X+3{,}999Y=7{,}999$$

Να λύσετε το σύστημα ως προς X και Y. Να κάνετε το ίδιο και για τα συστήματα:

$$1X+2Y=4{,}001$$
$$2X+3{,}999Y=7{,}998$$

και

$$1{,}001X+2{,}001Y=4{,}001$$
$$2{,}001X+3{,}998Y=7{,}999$$

Σε τι γενικά συμπεράσματα καταλήγετε;

Άσκηση 14

1. Ας υποθέσουμε ότι το πάχος ενός φύλλου χαρτιού είναι 1 mm (οι διαστάσεις του χαρτιού είναι αδιάφορες). Πόσες φορές πρέπει να διπλώσουμε το χαρτί ώστε το πάχος του να ξεπερνάει την απόσταση Γης Σελήνης;
2. Δίνεται η παρακάτω εξίσωση (logistic map):

$$X_{n+1}=aX_n(1-X_n), \text{ με } X\in[0, 1] \text{ και } a\in[0, 4]$$

όπου a είναι ο μέγιστος ρυθμός αύξησης ενός πληθυσμού (αύξηση = γεννήσεις - θάνατοι) και X_n είναι ο λόγος (το ποσοστό) του μεγέθους του υπάρχοντος πληθυσμού, σε ένα περιβάλλον κατά την χρονική στιγμή n, προς το συνολικό μέγεθος του πληθυσμού που το συγκεκριμένο περιβάλλον μπορεί να συντηρήσει. Δηλαδή, η υπόθεση είναι ότι το μέγεθος ενός πληθυσμού την χρονική στιγμή n+1 εξαρτάται με κάποιον τρόπο (είναι συνάρτηση) από το μέγεθος του πληθυσμού κατά την χρονική στιγμή n. Να δείξετε εμπειρικά με τη βοήθεια κατάλληλων διαγραμμάτων ότι α) για μικρές τιμές της παραμέτρου a ο πληθυσμός συγκλίνει προς μια σταθερή κατάσταση και β) καθώς το a αυξάνεται αρχίζουν να εμφανίζονται ταλαντώσεις αρχικά μεταξύ δύο διαφορετικών τιμών, έπειτα μεταξύ 4, μετά 8, μετά 16, και ούτω καθεξής. Να προσδιορίσετε εμπειρικά με ακρίβεια 4 δεκαδικών ψηφίων την τιμή του a πέραν της οποίας οι κανονικές ταλαντώσεις σταματούν και το σύστημα συμπεριφέρεται χαοτικά. <u>Υπόδειξη</u>: αρχικά μελετείστε τη συμπεριφορά της συνάρτησης για n έως 50 και στη συνέχεια για n έως 1000.

DO NOT COPY

Άσκηση 15

Η ετήσια συνολική κατανάλωση ενός προϊόντος (Y σε τόνους) σε μια μικρή χώρα για τα έτη 2011-2019 παρουσιάζεται στον Πίνακα 7. Θέλουμε να προσαρμόσουμε τα δεδομένα στο μαθηματικό υπόδειγμα:

$$Y_t=\gamma\alpha^{(t-2000)}$$

1. Να γίνει γραφική παράσταση των δεδομένων (διάγραμμα διασποράς- scatter plot). Να σχεδιάσετε και την καμπύλη Loess.
2. Να υπολογίσετε τους εκτιμητές ελαχίστων τετραγώνων των συντελεστών α και γ.
3. Να πραγματοποιήσετε όλους τους σχετικούς ελέγχους των προϋποθέσεων εγκυρότητας του υποδείγματος. Να κάνετε έλεγχο και αν τα υπόλοιπα (residuals) του υποδείγματος έχουν αυτοσυσχέτιση.
4. Στο διάγραμμα διασποράς των δύο μεταβλητών (t, Y_t) να σχεδιάσετε την καμπύλη του υποδείγματος.
5. Να υπολογίσετε με μεθόδους της Αριθμητικής Ανάλυσης την επιφάνεια (εμβαδό) κάτω από την καμπύλη του υποδείγματος (Area Under the Curve-AUC).
6. Με βάση το μαθηματικό υπόδειγμα, να υπολογίσετε την επιφάνεια (εμβαδό) κάτω από την καμπύλη (Area Under the Curve-AUC).
7. Προσαρμόζονται καλά τα δεδομένα στο υπόδειγμα; Αιτιολογείστε την απάντησή σας με βάση αριθμητικά και διαγραμματικά αποτελέσματα.
8. Να γίνει πρόβλεψη της κατανάλωσης για τα έτη 2020 και 2021.
9. Να γίνει εκτίμηση των σφαλμάτων για τις δύο προβλέψεις του ερωτήματος 4.
10. Με βάση το αρχικό προτεινόμενο υπόδειγμα να εκτιμήσετε την κατανάλωση για την δεκαετία 2022-2031.
11. Στη συνέχεια, να "εισάγετε" τυχαία στις εκτιμήσεις αβεβαιότητα της τάξης του ±60.
12. Με τα νέα δεδομένα να απαντήσετε στα ερωτήματα 1-7 τόσο για την περίοδο 2022-2031 όσο και για την περίοδο 2011-2031.

Πίνακας 7

Έτος (t)	Κατανάλωση (Y_t)
2011	97
2012	110
2013	147
2014	180
2015	230
2016	269
2017	406
2018	485
2019	615

Άσκηση 16

Σε δυο ομάδες ασθενών (12 ασθενείς σε κάθε ομάδα) με ψυχολογικά προβλήματα (ασθένεια Α και ασθένεια Β) ερευνήθηκαν οι τάσεις αυτοκτονίας (Πίνακας 8). Να γίνει έλεγχος αν υπάρχει

DO NOT COPY

σχέση μεταξύ ασθένειας και τάσης για αυτοκτονία με τον ακριβή έλεγχο του Fisher (Fisher's Exact Test) αλλά και με τον έλεγχο ανεξαρτησίας X^2 σε επίπεδο σημαντικότητας α=**0,10**. Συμφωνούν τα συμπεράσματα που προκύπτουν από τους δύο ελέγχους; Αν όχι, γιατί; Σε ποιον από τους δύο ελέγχους θα στηρίξετε τα συμπεράσματα της μελέτης;

Πίνακας 8

	Ασθένεια Α	Ασθένεια Β
Με τάση αυτοκτονίας	2	6
Χωρίς τάση αυτοκτονίας	10	6

Άσκηση 17

Στον Πίνακα 9 δίνονται 100 ζεύγη τιμών δύο μεταβλητών Χ1 και Χ2 (τα δεδομένα ελήφθησαν με τυχαία δειγματοληψία).

Πίνακας 9

X1	**X2**	**X1**	**X2**	**X1**	**X2**
102,503	141,502	38,229	76,724	96,519	118,690
110,307	143,387	42,433	94,397	157,238	202,060
120,286	176,975	104,050	158,508	68,864	113,481
96,054	121,390	80,676	152,527	86,040	143,717
124,905	182,333	64,143	136,691	86,460	134,832
98,256	123,819	81,179	115,111	102,973	127,043
75,807	83,511	143,280	222,891	133,977	148,506
79,685	115,520	116,876	155,240	107,080	151,728
137,929	209,379	84,849	160,399	71,207	137,035
90,418	146,677	79,028	191,042	69,888	115,681
92,357	170,400	96,355	131,857	90,878	140,613
74,398	124,928	104,440	124,147	95,827	157,034
197,920	249,868	103,659	95,089	67,120	129,693
163,800	178,416	131,926	162,956	110,103	151,323
115,078	145,903	126,699	146,538	111,650	150,707
146,891	178,773	87,582	123,339	72,276	142,462
104,649	182,442	96,290	162,016	125,770	136,026
128,906	154,749	66,472	106,787	125,447	134,954
106,360	122,592	57,938	99,204	76,820	122,268
98,339	133,358	88,526	94,063	93,401	109,155
92,296	144,674	111,952	154,537	132,481	165,246
143,009	173,735	109,936	152,187	121,690	187,267
55,728	91,931	92,729	102,281	73,735	135,421
104,454	180,461	133,225	174,195	65,365	152,150
90,485	109,601	129,567	226,324	93,863	161,522

DO NOT COPY

151,091	151,230	98,720	169,702	117,878	123,095
111,637	122,114	87,014	108,961	132,855	213,129
118,056	153,201	161,771	153,755	84,131	146,528
117,738	162,689	87,744	99,964	154,486	198,402
70,980	144,650	52,460	113,091	96,978	122,496
152,699	172,132	120,419	208,993	167,062	205,624
157,757	173,440	59,137	122,344	76,958	110,190
160,649	207,913	94,413	136,269		
126,570	148,540	78,092	139,819		

1. Να κατασκευάσετε το διάγραμμα διασποράς των δύο μεταβλητών (πχ το X1×X2).
2. Από το διάγραμμα διασποράς, να κάνετε μια εμπειρική εκτίμηση για την κατεύθυνση, το είδος και την ένταση της συσχέτισης μεταξύ των δύο μεταβλητών.
3. Να κάνετε έλεγχο σχετικά με το εάν υπάρχουν μονομεταβλητά και πολυμεταβλητά (2D) παράτυπα σημεία (outliers).
4. Να κάνετε έλεγχο αν οι δύο μεταβλητές ακολουθούν η κάθε μία Κανονική Κατανομή και να εκτιμήσετε τις αντίστοιχες παραμέτρους.
5. Να κάνετε έλεγχο αν οι δύο μεταβλητές ακολουθούν τη Διδιάστατη Κανονική Κατανομή και να εκτιμήσετε τις παραμέτρους αυτής.
6. Να υπολογίσετε τη συνδυασπορά (covariance) των δύο μεταβλητών.
7. Να υπολογίσετε τον συντελεστή γραμμικής συσχέτισης (*r* του Pearson) μεταξύ των δύο μεταβλητών.
8. Πως συνδέεται ο συντελεστής συσχέτισης των δύο μεταβλητών με τη συνδυασπορά τους;
9. Να αξιολογήσετε τη στατιστική σημαντικότητα και την ένταση/βαθμό του συντελεστή συσχέτισης.

   Υπόδειξη: Η στατιστική σημαντικότητα του συντελεστή συσχέτισης να αξιολογηθεί με παραμετρικό έλεγχο και με τις μεθόδους προσομοίωσης Monte-Carlo και Bootstrap.

10. Για κάθε μεταβλητή να υπολογίσετε τα αντίστοιχα z-scores.
11. Πως συνδέονται τα *z*-scores με το συντελεστή γραμμικής συσχέτισης των δύο μεταβλητών;
12. Να εκτιμήσετε την ευθεία παλινδρόμησης της X1 πάνω στη X2.

    Υπόδειξη: Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση, σχολιασμός και ερμηνεία των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.

13. Με τη μέθοδο προσομοίωσης Bootstrap να εκτιμήσετε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του υποδείγματος (a και b).
14. Να κατασκευάσετε το κατάλληλο διάγραμμα διασποράς.
15. Επί του διαγράμματος διασποράς να σχεδιάσετε την ευθεία ελαχίστων τετραγώνων καθώς και την καμπύλη Loess (30%, 50% και 70% των γειτονικών σημείων).

    Υπόδειξη: Να κατασκευαστούν 3 διαγράμματα διασποράς ανάλογα με το ποσοστό των γειτονικών σημείων βάσει των οποίων θα σχεδιαστεί η καμπύλη Loess.

16. Τι παρατηρείτε;
17. Να εκτιμήσετε την τιμή της εξαρτημένης μεταβλητής όταν η ανεξάρτητη μεταβλητή πάρει τις τιμές 262, 290 και 145. Να υπολογίσετε και τα αντίστοιχα σφάλματα στις εκτιμήσεις.

DO NOT COPY

18. Να εκτιμήσετε την ευθεία παλινδρόμησης της X2 πάνω στη X1.

    Υπόδειξη: Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση, σχολιασμός και ερμηνεία των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.

19. Με τη μέθοδο προσομοίωσης Bootstrap να εκτιμήσετε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του υποδείγματος (a και b).
20. Να κατασκευάσετε το κατάλληλο διάγραμμα διασποράς.
21. Επί του διαγράμματος διασποράς να σχεδιάσετε την ευθεία ελαχίστων τετραγώνων καθώς και την καμπύλη Loess (30%, 50% και 70% των γειτονικών σημείων).

    Υπόδειξη: Να κατασκευαστούν 3 διαγράμματα διασποράς ανάλογα με το ποσοστό των γειτονικών σημείων βάσει των οποίων θα σχεδιαστεί η καμπύλη Loess.

22. Τι παρατηρείτε;
23. Να διατυπώσετε μία τουλάχιστον γεωμετρική ερμηνεία του συντελεστή γραμμικής συσχέτισης του Pearson.
24. Να υπολογίσετε τον συντελεστή συσχέτισης των τάξεων (*rho* του Spearman) μεταξύ των δύο μεταβλητών.
25. Να αξιολογήσετε τη στατιστική σημαντικότητα και την ένταση/βαθμό του συντελεστή συσχέτισης των τάξεων.
26. Τι παρατηρείτε για τις τιμές των δύο συντελεστών συσχέτισης *r* και *rho*;
27. Να επιλέξετε τυχαία 30 ζεύγη τιμών για τις μεταβλητές X1 και X2 και να απαντήσετε στα προηγούμενα ερωτήματα.
28. Να συγκρίνετε τα ευρήματα που προέκυψαν από την ανάλυση των 100 ζευγών με αυτά που προέκυψαν από την ανάλυση των 30 τυχαίων ζευγών.

Άσκηση 18

1) Ο ποταμός Χρυσός εκβάλει σε 3 σημεία (Α, Β και Γ). Σε κάθε σημείο εκβολής ελήφθησαν 7 τυχαία δείγματα νερού από 3 διαφορετικά βάθη (επιφάνεια, βάθος 20 cm και βάθος 1 m). Σε κάθε δείγμα μετρήθηκε η συγκέντρωση του χημικού στοιχείου *XXX*. Να προτείνετε στρατηγική και μοντέλο(α) για τη στατιστική επεξεργασία των δεδομένων. Να κάνετε αναφορά και στις αντίστοιχες προϋποθέσεις.
2) Θέλουμε να αξιολογήσουμε την επίδραση 4 τύπων ελαστικών (Α, Β, Γ και Δ) στην κατανάλωση καυσίμου αυτοκίνητων. Στο πείραμα συμμετείχαν 4 όμοια αυτοκίνητα (ίδιο μοντέλο) και 4 οδηγοί. Να προτείνετε στρατηγική και μοντέλο(α) για τη στατιστική επεξεργασία των δεδομένων. Να κάνετε αναφορά και στις αντίστοιχες προϋποθέσεις.
3) Οκτώ φοιτητές ενός Τμήματος Στατιστικής διδάχθηκαν ένα στατιστικό πακέτο από 3 εισηγητές. Ο κάθε εισηγητής είχε τη δική του διδακτική προσέγγιση. Μετά το πέρας των μαθημάτων από κάθε εισηγητή οι φοιτητές εξετάστηκαν ως προς την ταχύτητα επίλυσης (σε sec) ενός προβλήματος με το στατιστικό πακέτο. Το πρόβλημα ήταν το ίδιο και για τις 3 εξετάσεις. Να προτείνετε στρατηγική και μοντέλο(α) για τη στατιστική επεξεργασία των δεδομένων. Να κάνετε αναφορά και στις αντίστοιχες προϋποθέσεις.
4) Σε δύο περιοχές (Ταξιάρχης, Πολυπόταμος) και για δύο έτη (2010, 2011) εγκαταστάθηκε πείραμα για να μελετηθεί η επίδραση τριών φυτικών ειδών (Grass, Forbs, Legumes) σε 7 ποσοτικά χαρακτηριστικά οικολογικής σημασίας. Μήπως πρέπει να δοθούν και άλλες

DO NOT COPY

διευκρινήσεις σχετικά με την εγκατάσταση του πειράματος; Να προτείνετε στρατηγική και μοντέλο(α) για τη στατιστική επεξεργασία των δεδομένων. Να κάνετε αναφορά και στις αντίστοιχες προϋποθέσεις.

5) Σε ένα κτηνιατρικό πείραμα, 36 σκύλοι χωρίστηκαν τυχαία σε δύο ομάδες. Αμέσως μετά από μια χειρουργική επέμβαση, στη μία ομάδα (ομάδα 1) χορηγήθηκε το φάρμακο Α και στην άλλη το φάρμακο Β. Έγινε καταγραφή των επιπέδων ουρίας σε 6 ισαπέχουσες χρονικές στιγμές. Με βάση την περιγραφή του πειράματος τι θα πρέπει να προσέξουν οι ερευνητές κτηνίατροι ώστε το πείραμα να είναι ορθά σχεδιασμένο; Να προτείνετε στρατηγική και μοντέλο(α) για τη στατιστική επεξεργασία των δεδομένων. Να κάνετε αναφορά και στις αντίστοιχες προϋποθέσεις.

DO NOT COPY