ΣΥΜΠΕΡΑΣΜΑΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Δοκιμασίες (ή δοκιμές ή έλεγχοι) σημαντικότητας (significance tests)

Η έννοια της *στατιστικά σημαντικής διαφοράς* (statistically significant difference):

Ο *σημαντικός δεν έχει σχέση με μεγέθη*, δηλ. δεν σημαίνει μεγάλος, υπερβολικός. Σημαντική διαφορά σημαίνει διαφορά οφειλόμενη σε συστηματικά αίτια, σε αντίθεση με τη μη σημαντική διαφορά που οφείλεται σε τυχαία αίτια.

Από τις πιο σημαντικές ιδιότητες μιας αναλυτικής μεθόδου είναι η έλλειψη συστηματικών σφαλμάτων. Η ιδιότητα αυτή μπορεί να εκτιμηθεί με την ανάλυση (μέρους) ενός δείγματος με γνωστή συγκέντρωση αναλύτη. Ωστόσο, και το τυχαίο σφάλμα μπορεί να δώσει τιμές που αποκλίνουν από την πραγματική.

Για να αποφανθούμε αν η διαφορά της μετρούμενης από την πραγματική τιμή οφείλεται σε τυχαία σφάλματα, εφαρμόζουμε μια **δοκιμασία σημαντικότητας**.

Σκοπός, δηλαδή, είναι να επιβεβαιωθεί αν η διαφορά μεταξύ δύο αποτελεσμάτων είναι στατιστικά σημαντική ή οφείλεται σε τυχαίες διακυμάνσεις.

1. Σύγκριση μιας μέσης τιμής με μια γνωστή-πραγματική τιμή
2. Σύγκριση μεταξύ δύο μέσων τιμών
3. Δοκιμασία *t* για σύγκριση ζευγών τιμών (μονόπλευρη, αμφίπλευρη)
4. Δοκιμασία *F* για σύγκριση τυπικών αποκλίσεων

Μηδενική υπόθεση

Σύγκριση αναλυτικών αποτελεσμάτων που ελήφθησαν:

Περίπτωση 1. με την ίδια μέθοδο στα δείγματα Α και Β, για να διαπιστωθεί αν τα δείγματα περιέχουν τη μετρούμενη ουσία σε ίδια ή διαφορετική συγκέντρωση.

Περίπτωση 2. με δύο διαφορετικές μεθόδους Α και Β στο ίδιο δείγμα, για να διαπιστωθεί αν οι δύο μέθοδοι παρέχουν ίδια ή διαφορετικά αποτελέσματα.

Το αποτέλεσμα αυτών των δοκιμασιών είναι η αποδοχή ή η απόρριψη της *μηδενικής υπόθεσης* (null hypothesis, H_0). Η μηδενική υπόθεση, γενικά, δηλώνει ότι: "Οι διαφορές - αποκλίσεις οφείλονται αποκλειστικά σε τυχαία και όχι σε συστηματικά αίτια-σφάλματα". Η εναλλακτική υπόθεση (alternative hypothesis, H_a) δηλώνει το ακριβώς αντίθετο.

Ποια είναι η μηδενική υπόθεση για τις παραπάνω περιπτώσεις?

Οι μέσες τιμές είναι ίδιες, δηλ. για την Περίπτωση 1: η περιεκτικότητα των δύο δειγμάτων σε προσδιοριζόμενη ουσία είναι η ίδια, για την Περίπτωση 2: και οι δύο μέθοδοι παρέχουν τα ίδια αναλυτικά αποτελέσματα. Οι παρατηρούμενες διαφορές (αν υπάρχουν) οφείλονται καθαρά σε τυχαία σφάλματα.

Εναλλακτική υπόθεση: Οι μέσες τιμές διαφέρουν σημαντικά, δηλ. για την Περίπτωση 1: οι περιεκτικότητες των δύο δειγμάτων σε προσδιοριζόμενη ουσία είναι διαφορετικές, για την Περίπτωση 2: οι μέθοδοι παρέχουν διαφορετικά αποτελέσματα (επομένως τουλάχιστον η μία παρουσιάζει συστηματικό αναλυτικό σφάλμα).

1. Σύγκριση μιας μέσης τιμής με μια γνωστή τιμή

Η μηδενική υπόθεση (null hypothesis, H_0): Δεν υπάρχει διαφορά μεταξύ των δύο συγκρίσιμων τιμών παρά μόνο εξ αιτίας τυχαίων διακυμάνσεων.

Υποθέτοντας ότι η μηδενική υπόθεση ισχύει, η πιθανότητα ότι η παρατηρούμενη διαφορά μεταξύ της μέσης τιμής του δείγματος και της γνωστής τιμής, *μ*, προκύπτει αποκλειστικά ως αποτέλεσμα τυχαίων σφαλμάτων.

Όσο μικρότερη είναι η πιθανότητα η παρατηρούμενη διαφορά να συμβεί τυχαία, τόσο λιγότερο πιθανό είναι η μηδενική υπόθεση να ισχύει.

Ορισμός επιπέδου αξιοπιστίας

Επίπεδο αξιοπιστίας 95% ή επίπεδο σημαντικότητας $p = 0{,}05$ (ή 5%): αν η H_0 απορριφθεί, υπάρχει πιθανότητα 5 στις 100 περιπτώσεις να έχει απορριφθεί η μηδενική υπόθεση ενώ ισχύει.

Για να είμαστε πιο σίγουροι ότι λαμβάνουμε τη σωστή απόφαση μπορεί να χρησιμοποιηθεί υψηλότερο επίπεδο σημαντικότητας, συνήθως 0,01 ή 0,001 (1% ή 0,1%).

Το επίπεδο σημαντικότητας υποδεικνύεται γράφοντας, για παράδειγμα, *p* (δηλαδή πιθανότητα) = 0,05 και δίνει την πιθανότητα απόρριψης μιας ισχύουσας μηδενικής υπόθεσης.

Σημαντικό: εάν δεν απορριφθεί η μηδενική υπόθεση, δεν έχει αποδειχθεί ότι ισχύει. Μόνο ότι δεν έχει αποδειχθεί ότι είναι ψευδής.

1. Σύγκριση μιας μέσης τιμής με μια γνωστή τιμή

Η μηδενική υπόθεση (null hypothesis, H_0): Δεν υπάρχει διαφορά μεταξύ των δύο συγκρίσιμων τιμών παρά μόνο εξ αιτίας των τυχαίων διακυμάνσεων.

Επίπεδο αξιοπιστίας 95% ή επίπεδο σημαντικότητας p = 0,05 (ή 5%): αν η H_0 απορριφθεί, υπάρχει πιθανότητα 5 στις 100 περιπτώσεις να έχει απορριφθεί η μηδενική υπόθεση ενώ ισχύει.

Πρέπει να αποφανθούμε αν: μέση τιμή = πραγματική (γνωστή) τιμή - δηλαδή αν ισχύει η μηδενική υπόθεση, υπολογίζεται ο στατιστικός όρος ή **στατιστικό στοιχείο (statistic)**:

$$t = (\bar{x} - x_{\pi\rho\alpha\gamma\mu.})\sqrt{n}/s$$

Αν $|t| \rangle$ κρίσιμη (θεωρητική) τιμή, η H_0 απορρίπτεται (για n–1 βαθμούς ελευθερίας).

Παράδειγμα: Προσδιορισμός Se σε νερό πιστοποιημένο υλικό αναφοράς (CRM) συγκέντρωσης 50 ng/mL.

Βρέθηκαν οι συγκεντρώσεις: 50,4 50,7 49,1 49,0 51,1 ng/mL

Υπάρχει ένδειξη συστηματικού σφάλματος?

Μέση τιμή: 50,06 ng/mL τυπική απόκλιση: 0,95 ng/mL

Μηδενική υπόθεση H_0: δεν υπάρχει συστηματικό σφάλμα

Υπολογίζεται η t = 0,14

Κρίσιμη τιμή για 0,05 επίπεδο σημαντικότητας $t(4)$ = 2,78 > 0,14, η μηδενική υπόθεση ισχύει. Δεν υπάρχει ένδειξη συστηματικού σφάλματος.

Πίνακας Η κατανομή *t*

Τιμή *t* για ένα διάστημα εμπιστοσύνης της κρίσιμης τιμής | *t* | για τιμές *P* των βαθμών ελευθερίας	90% 0,10	95% 0,05	98% 0,02	99% 0,01
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
12	1,78	2,18	2,68	3,05
14	1,76	2,14	2,62	2,98
16	1,75	2,12	2,58	2,92
18	1,73	2,10	2,55	2,88
20	1,72	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
50	1,68	2,01	2,40	2,68
∞	1,64	1,96	2,33	2,58

Οι κρίσιμες τιμές είναι κατάλληλες για αμφίπλευρες δοκιμασίες. Για μονόπλευρες δοκιμασίες, η τιμή λαμβάνεται από τη στήλη που είναι διπλάσια της επιθυμητής τιμής *P*, π.χ. για μία μονόπλευρη δοκιμασία, $P = 0{,}05$ και 5 βαθμούς ελευθερίας, η κρίσιμη τιμή λαμβάνεται από τη στήλη για $P = 0{,}10$ και είναι ίση με 2,02.

2. Σύγκριση μεταξύ δύο μέσων τιμών

Μια αναλυτική μέθοδος δοκιμάζεται με σύγκρισή της με μια δεύτερη μέθοδο (πιθανόν μέθοδος αναφοράς).

Ho: Οι δύο μέθοδοι δίνουν τα ίδια αποτελέσματα. Δηλαδή $\bar{x}_1 = \bar{x}_2$.

Απαιτείται να ελεγχθεί αν η διαφορά των μέσων τιμών διαφέρει σημαντικά από το μηδέν.

Αν οι τυπικές αποκλίσεις των δύο μεθόδων δεν είναι σημαντικά διαφορετικές, τότε η συνολική τυπική απόκλιση *s* και ο στατιστικός όρος *t*, υπολογίζονται

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$n_1 + n_2 - 2$, βαθμοί ελευθερίας

Παράδειγμα: Συγκρίνονται δύο μέθοδοι για τον προσδιορισμό *Cr* σε φυτικό ιστό. Πραγματοποιήθηκαν πέντε προσδιορισμοί για κάθε μέθοδο.

Μέθοδος 1: μέση τιμή 1,48 ng/g, τυπική απόκλιση 0,28 ng/g

Μέθοδος 2: μέση τιμή 2,33 ng/g, τυπική απόκλιση 0,31 ng/g

Μηδενική υπόθεση: Οι μέσες τιμές των δύο μεθόδων είναι στατιστικά ίσες.

Συνολική τυπική απόκλιση $s = 0{,}29$ και $t = 4{,}56$.

Για 8 βαθμούς ελευθερίας η κρίσιμη τιμή $t(8) = 2{,}31$ (για $P = 0{,}05$). Η μηδενική υπόθεση απορρίπτεται.

ΟΤΑΝ ΤΑ ΔΕΙΓΜΑΤΑ ΔΕΝ ΠΡΟΕΡΧΟΝΤΑΙ ΑΠΟ ΤΟΝ ΙΔΙΟ ΠΛΗΘΥΣΜΟ (ΔΗΛ. ΔΙΑΦΟΡΕΤΙΚΕΣ ΤΥΠΙΚΕΣ ΑΠΟΚΛΙΣΕΙΣ) ΥΠΑΡΧΟΥΝ ΑΛΛΟΙ ΤΡΟΠΟΙ ΥΠΟΛΟΓΙΣΜΟΥ ΚΑΙ ΣΥΓΚΡΙΣΗΣ

Πίνακας Η κατανομή *t*

Τιμή *t* για ένα διάστημα εμπιστοσύνης της κρίσιμης τιμής | *t* | για τιμές *P* των βαθμών ελευθερίας	90% 0,10	95% 0,05	98% 0,02	99% 0,01
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
12	1,78	2,18	2,68	3,05
14	1,76	2,14	2,62	2,98
16	1,75	2,12	2,58	2,92
18	1,73	2,10	2,55	2,88
20	1,72	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
50	1,68	2,01	2,40	2,68
∞	1,64	1,96	2,33	2,58

Οι κρίσιμες τιμές είναι κατάλληλες για αμφίπλευρες δοκιμασίες. Για μονόπλευρες δοκιμασίες, η τιμή λαμβάνεται από τη στήλη που είναι διπλάσια της επιθυμητής τιμής *P*, π.χ. για μία μονόπλευρη δοκιμασία, $P = 0,05$ και 5 βαθμούς ελευθερίας, η κρίσιμη τιμή λαμβάνεται από τη στήλη για $P = 0,10$ και είναι ίση με 2,02.

3. Δοκιμασία *t* για σύγκριση ζευγών (μονόπλευρη ή αμφίπλευρη)

Συχνά δύο μέθοδοι συγκρίνονται **μετρώντας δείγματα με διαφορετικές ποσότητες-συγκεντρώσεις ενός αναλύτη**. Επιθυμούμε να γνωρίζουμε αν οι μέθοδοι παράγουν στατιστικά σημαντικά διαφορετικά αποτελέσματα.

Η προηγούμενη μέθοδος σύγκρισης των μέσων τιμών δεν είναι κατάλληλη διότι *δεν διαχωρίζει τις διακυμάνσεις που προέρχονται από τη μέθοδο, από αυτές που προέρχονται από τα δείγματα.*

Η δυσκολία ξεπερνάται εξετάζοντας τη διαφορά μεταξύ των ζευγών των αποτελεσμάτων από τις δύο μεθόδους.

Υπολογίζεται ο στατιστικός όρος-στατιστικό στοιχείο *t* για *n* αποτελέσματα-ζεύγη (*d* = διαφορά)

$$t = \bar{d}\sqrt{n}/s_d$$

Σύγκριση με την θεωρητική τιμή *t* για *n* – 1 βαθμούς ελευθερίας.

Πίνακας Η κατανομή *t*

Τιμή *t* για ένα διάστημα εμπιστοσύνης της κρίσιμης τιμής | *t* | για τιμές *P* των βαθμών ελευθερίας	90% 0,10	95% 0,05	98% 0,02	99% 0,01
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
12	1,78	2,18	2,68	3,05
14	1,76	2,14	2,62	2,98
16	1,75	2,12	2,58	2,92
18	1,73	2,10	2,55	2,88
20	1,72	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
50	1,68	2,01	2,40	2,68
∞	1,64	1,96	2,33	2,58

Οι κρίσιμες τιμές είναι κατάλληλες για αμφίπλευρες δοκιμασίες. Για μονόπλευρες δοκιμασίες, η τιμή λαμβάνεται από τη στήλη που είναι διπλάσια της επιθυμητής τιμής *P*, π.χ. για μία μονόπλευρη δοκιμασία, $P = 0{,}05$ και 5 βαθμούς ελευθερίας, η κρίσιμη τιμή λαμβάνεται από τη στήλη για $P = 0{,}10$ και είναι ίση με 2,02.

Παράδειγμα: Αποτελέσματα προσδιορισμού παρακεταμόλης (mg/g) σε δέκα ταμπλέτες

Δείγμα	UV-μέθοδος	Near IR μέθοδος
1	84,63	83,15
2	84,38	83,72
3	84,08	83,84
4	84,41	84,20
5	………	………
6	………	………
7	………	………
8	………	………
9	………	………
10	84,03	84,24

Οι διαφορές μεταξύ των ζευγών των τιμών είναι:
+1,48, +0,66, +0,24, +0,21, −0,10, −0,61, −0,10, +0,09, +0,07, −0,21

$\bar{d} = 0{,}173 \quad s_d = 0{,}57$ (με τα πρόσημα)

και υπολογίζω t = 0,88. Η κρίσιμη τιμή είναι $t(9)$ = 2,26 (P = 0,05). Άρα η μηδενική υπόθεση διατηρείται (δηλαδή, οι μέθοδοι δεν δίνουν σημαντικά διαφορετικά αποτελέσματα για τα διαφορετικά δείγματα).

Η δοκιμασία t για σύγκριση ζευγών δεν απαιτεί οι επαναληψιμότητες των δύο μεθόδων να είναι ίσες αλλά οι διαφορές d να προέρχονται από κανονική κατανομή.
Η δοκιμασία t για σύγκριση ζευγών είναι ακατάλληλη για ανάλυση δειγμάτων που οι συγκεντρώσεις τους διαφέρουν πάνω από μια τάξη μεγέθους επειδή βασίζεται στην προϋπόθεση ότι τα τυχαία και συστηματικά σφάλματα είναι ανεξάρτητα της συγκέντρωσης (Απαιτείται *ανάλυση γραμμικής συσχέτισης*).

Μονόπλευρες ή αμφίπλευρες δοκιμασίες ???

Στα παραδείγματα που περιγράφηκαν ενδιέφερε η ύπαρξη διαφοράς μεταξύ δύο μέσων τιμών προς οποιαδήποτε κατεύθυνση.

Στις περισσότερες περιπτώσεις ο αναλυτής **δεν** γνωρίζει αν η διαφορά μεταξύ της πειραματικής μέσης τιμής και της τιμής αναφοράς είναι θετική ή αρνητική.

Άρα, η δοκιμασία πρέπει να καλύπτει και τις δύο πιθανότητες (**αμφίπλευρη δοκιμασία**).

Σε κάποιες άλλες περιπτώσεις, απαιτείται διαφορετικού είδους δοκιμασία (**μονόπλευρη**).

π.χ. η αύξηση της ταχύτητας μίας αντίδρασης παρουσία καταλύτη απαιτεί δοκιμασία σημαντικότητας προς τη μία (θετική) κατεύθυνση.

Τι γίνεται με τις θεωρητικές τιμές *t* ?

Για επίπεδο σημαντικότητας *P* = 0,05, η κατάλληλη τιμή για μονόπλευρη δοκιμασία βρίσκεται από τη στήλη που αντιστοιχεί σε *P* = 0,10 του αντίστοιχου πίνακα.

Πίνακας Η κατανομή *t*

Τιμή *t* για ένα διάστημα εμπιστοσύνης της κρίσιμης τιμής | *t* | για τιμές *P* των βαθμών ελευθερίας	90% 0,10	95% 0,05	98% 0,02	99% 0,01
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
12	1,78	2,18	2,68	3,05
14	1,76	2,14	2,62	2,98
16	1,75	2,12	2,58	2,92
18	1,73	2,10	2,55	2,88
20	1,72	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
50	1,68	2,01	2,40	2,68
∞	1,64	1,96	2,33	2,58

Οι κρίσιμες τιμές είναι κατάλληλες για αμφίπλευρες δοκιμασίες. Για μονόπλευρες δοκιμασίες, η τιμή λαμβάνεται από τη στήλη που είναι διπλάσια της επιθυμητής τιμής *P*, π.χ. για μία μονόπλευρη δοκιμασία, $P = 0,05$ και 5 βαθμούς ελευθερίας, η κρίσιμη τιμή λαμβάνεται από τη στήλη για $P = 0,10$ και είναι ίση με 2,02.

Παράδειγμα

Υπάρχει η υποψία ότι μια οξεο-βασική ογκομετρική μέθοδος εμπεριέχει σημαντικό σφάλμα λόγω του έγχρωμου δείκτη που χρησιμοποιείται. Συνεπώς, η μέθοδος έχει την τάση να δίνει αποτελέσματα με θετικό συστηματικό σφάλμα. Για να μελετήσουμε την υπόθεση, χρησιμοποιείται διάλυμα του οξέος συγκέντρωσης 0,1 M για την ογκομέτρηση 25,00 mL διαλύματος βάσεως, συγκέντρωσης 0,1 M, με τα ακόλουθα αποτελέσματα.

25,06 25,18 24,87 25,51 25,34 25,41 (mL)

Να γίνει αξιολόγηση των αποτελεσμάτων σχετικά με το αν υπάρχει ένδειξη για θετικό συστηματικό σφάλμα (positive bias).

Για τα δεδομένα ισχύει:

Μέση τιμή: 25,228 mL, τυπική απόκλιση: 0,238 mL

Υιοθετώντας τη μηδενική υπόθεση H_0, ότι δεν υπάρχει συστηματικό σφάλμα:

$x_{πραγμ}$ = 25,00 mL και

$$t = (25{,}228 - 25{,}00) \times \frac{\sqrt{6}}{0{,}238} = 2{,}35$$

Η κρίσιμη τιμή είναι t_s = 2,02 (P = 0,05, μονόπλευρη).

Η υπολογισθείσα τιμή του t είναι μεγαλύτερη από τη θεωρητική, άρα η μηδενική υπόθεση απορρίπτεται. Συνεπώς, υπάρχει ένδειξη για θετικό συστηματικό σφάλμα.

Στο προηγούμενο παράδειγμα

Παρατήρηση: Για μια αμφίπλευρη δοκιμασία (t_5 = 2,57), η μηδενική υπόθεση δε θα απορριπτόταν!

Η απόφαση για εφαρμογή μονόπλευρης ή αμφίπλευρης δοκιμασίας εξαρτάται από την υποψία ή αναμονή θετικού συστηματικού σφάλματος.

Γενικά, οι αμφίπλευρες δοκιμασίες είναι πιο συνηθισμένες απ' τις μονόπλευρες.

Το ίδιο ισχύει και για τη δοκιμασία *t* για σύγκριση ζευγών

4. Δοκιμασία *F* για σύγκριση τυπικών αποκλίσεων

Μέχρι τώρα, οι δοκιμασίες αφορούσαν σύγκριση τιμών για συστηματικά σφάλματα.

Πολύ συχνά είναι σημαντικό να συγκρίνονται οι τυπικές αποκλίσεις, δηλαδή τα τυχαία σφάλματα σειρών δεδομένων.

Υπάρχουν δύο δυνατότητες σύγκρισης:

- Η μέθοδος *A* είναι πιο επαναλήψιμη από τη μέθοδο B (μονόπλευρη δοκιμασία).

- Η μέθοδος *A* και η μέθοδος B διαφέρουν ως προς την επαναληψιμότητα (αμφίπλευρη δοκιμασία).

Η δοκιμασία *F* λαμβάνει υπόψη το λόγο των δύο μεταβλητοτήτων (ή διακυμάνσεων)

$$\frac{s_1^2}{s_2^2}$$

Ο λόγος αυτός είναι πάντοτε ≥ 1. Αν η μηδενική υπόθεση H_0 ισχύει (δηλ. οι επαναληψιμότητες είναι ίδιες), τότε ο λόγος των δύο μεταβλητοτήτων είναι ~ 1. Οι βαθμοί ελευθερίας για κάθε μεταβλητότητα είναι $n-1$.

Παράδειγμα: Δύο μέθοδοι συγκρίνονται μεταξύ τους για τον προσδιορισμό διαλυτού οξυγόνου (οκτώ προσδιορισμοί για κάθε μέθοδο):

Πρότυπη μέθοδος: 72 mg/l οξυγόνο και τυπική απόκλιση s = 3,31

Προτεινόμενη μέθοδος: 72 mg/l " και " " s = 1,51

Είναι η επαναληψιμότητα της προτεινόμενης μεθόδου σημαντικά μεγαλύτερη (δηλ. έχει μικρότερη μεταβλητότητα) από την αντίστοιχη της πρότυπης μεθόδου?

$$F = 3{,}31^2 / 1{,}51^2 = 4{,}8$$

Ενδιαφέρει αν η προτεινόμενη μέθοδος είναι πιο επαναλήψιμη από την πρότυπη μέθοδο. Εφαρμόζεται *μονόπλευρη* δοκιμασία.

Κρίσιμη τιμή $F_{7,7}$ = 3,787 (P = 0,05). Επειδή $F_{7,7}$ < 4,8, η μηδενική υπόθεση απορρίπτεται. Δηλαδή, η μεταβλητότητα της πρότυπης μεθόδου είναι σημαντικά μεγαλύτερη από αυτή της προτεινόμενης μεθόδου.

Άρα, η προτεινόμενη μέθοδος είναι πιο επαναλήψιμη (δηλ. έχει μικρότερη μεταβλητότητα) από την πρότυπη.

Πίνακας Α.3 Κρίσιμες τιμές F για μονόπλευρη δοκιμασία ($P = 0{,}05$)

v_2	v_1												
	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,745	8,703	8,660
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,858	5,803
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,619	4,558
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000	3,938	3,874
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575	3,511	3,445
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,284	3,218	3,150
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,006	2,936
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,845	2,774
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,719	2,646
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,617	2,544
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,533	2,459
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,463	2,388
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,403	2,328
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,352	2,276
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,308	2,230
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,269	2,191
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,234	2,155
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,203	2,124

v_1 = βαθμοί ελευθερίας για τον αριθμητή, v_2 = βαθμοί ελευθερίας για τον παρονομαστή.

Στο παράδειγμα για τον προσδιορισμό χρωμίου σε φυτικό ιστό υποθέσαμε ότι οι μεταβλητότητες των δύο μεθόδων δεν διαφέρουν σημαντικά.

Υπενθυμίζω:

Μέθοδος 1: μέση τιμή 1,48 ng/g   τυπική απόκλιση 0,28 ng/g

Μέθοδος 2: μέση τιμή 2,33 ng/g   τυπική απόκλιση 0,31 ng/g

Άρα $F = 0{,}31^2 / 0{,}28^2 = 1{,}23$

Δεν έχουμε σοβαρό λόγο να περιμένουμε ότι κάποια από τις δύο παρουσιάζει μεγαλύτερη μεταβλητότητα από την άλλη.

Άρα, η *αμφίπλευρη* δοκιμασία είναι πιο κατάλληλη.

$F_{4,4} = 9{,}605$ ($P = 0{,}05$) > 1,23. Συνεπώς, δεν υπάρχει σημαντική διαφορά μεταξύ των δύο μεταβλητοτήτων σε επίπεδο αξιοπιστίας 95%.

Μερικές από τις στατιστικές μεθόδους, όπως τα *t*-tests και η analysis of variance, προϋποθέτουν ομοιογένεια μεταβλητοτήτων.
Για να εκτιμηθεί η ισότητα των μεταβλητοτήτων για δύο ή περισσότερες σειρές δεδομένων μιας μεταβλητής χρησιμοποιείται, κατά περίπτωση, η **δοκιμασία Bartlett's** ή η **δοκιμασία Levene**.

Πίνακας Α.4 Κρίσιμες τιμές *F* για αμφίπλευρη δοκιμασία (P = 0,05)

v_2	v_1												
	1	2	3	4	5	6	7	8	9	10	12	15	20
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	976,7	984,9	993,1
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17
4	12,22	10,65	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,751	8,657	8,560
5	10,01	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,525	6,428	6,329
6	8,813	7,260	6,599	6,227	5,988	5,820	5,695	5,600	5,523	5,461	5,366	5,269	5,168
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,666	4,568	4,467
8	7,571	6,059	5,416	5,053	4,817	4,652	4,529	4,433	4,357	4,295	4,200	4,101	3,999
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,868	3,769	3,667
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,621	3,522	3,419
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,430	3,330	3,226
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436	3,374	3,277	3,177	3,073
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,153	3,053	2,948
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,050	2,949	2,844
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	2,963	2,862	2,756
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049	2,986	2,889	2,788	2,681
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985	2,922	2,825	2,723	2,616
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929	2,866	2,769	2,667	2,559
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880	2,817	2,720	2,617	2,509
20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,676	2,573	2,464

v_1 = βαθμοί ελευθερίας για τον αριθμητή, v_2 = βαθμοί ελευθερίας για τον παρονομαστή.

Δοκιμασία για την κανονικότητα μιας κατανομής

Πολλές στατιστικές δοκιμασίες προϋποθέτουν ότι τα δεδομένα που χρησιμοποιούνται προέρχονται από κανονικό πληθυσμό. Μια μέθοδος επιβεβαίωσης κανονικότητας είναι η δοκιμασία χ^2 (chi-squared) για περισσότερα από 50 δεδομένα.

Συνήθως, όμως, τα πειραματικά δεδομένα είναι αρκετά πιο περιορισμένα.

Ένας απλός οπτικός τρόπος είναι η κατασκευή καμπύλης αθροιστικής συχνότητας.

Use normal probability paper to investigate whether the data below could have been drawn from a normal population:

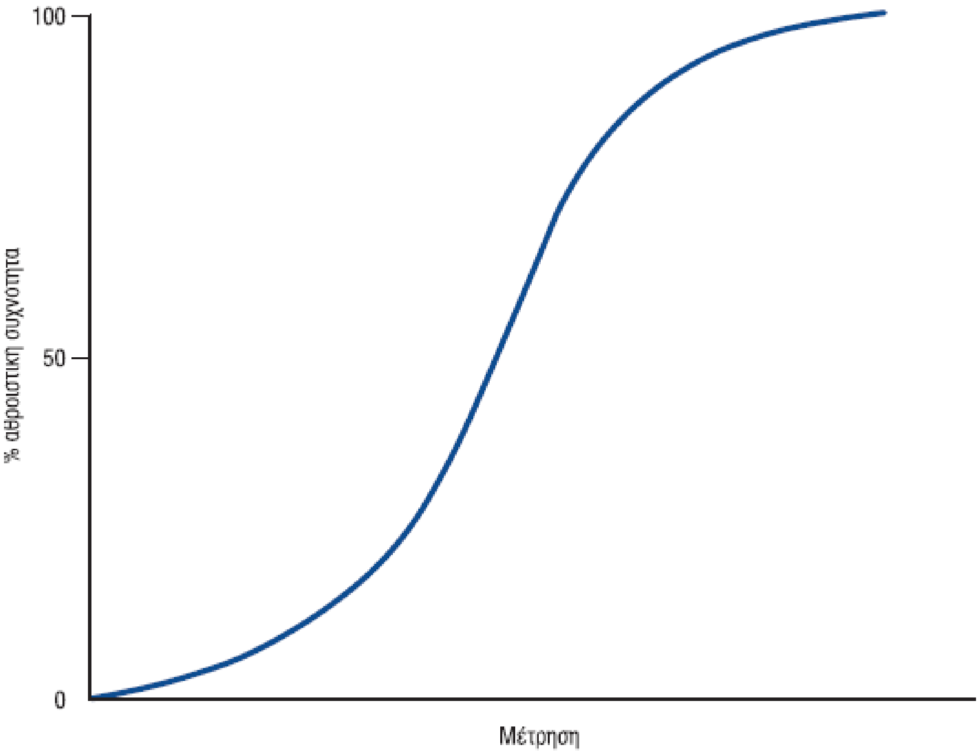
109, 89, 99, 99, 107, 111, 86, 74, 115, 107, 134, 113, 110, 88, 104

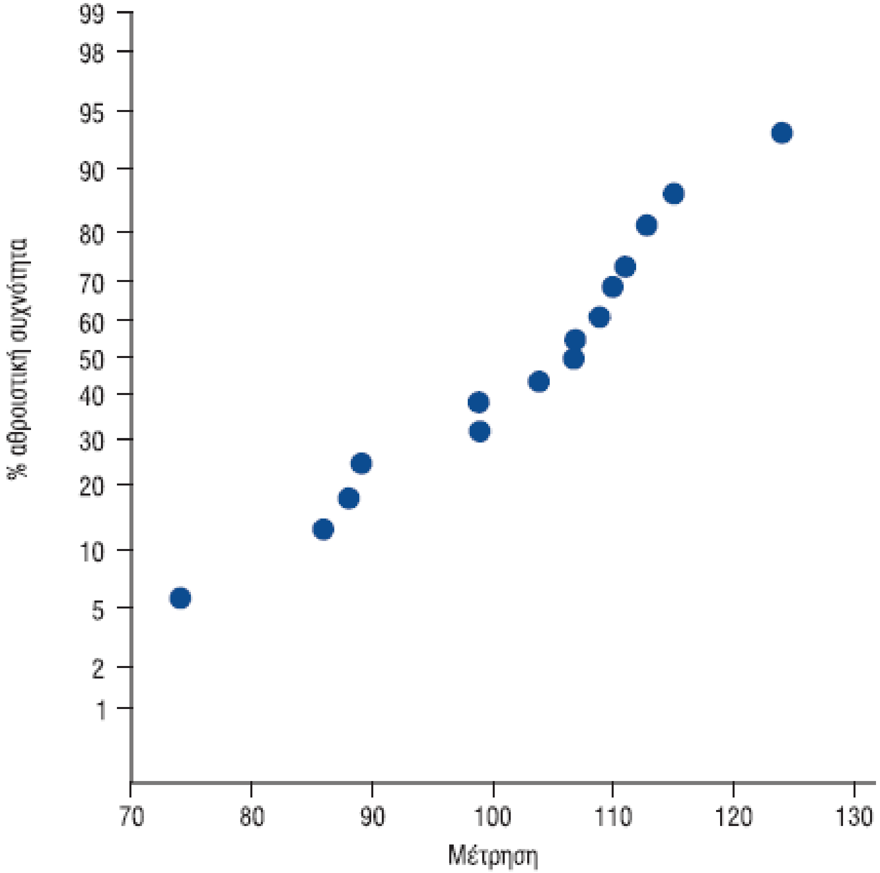
Measurement	Cumulative frequency	% Cumulative frequency
74	1	6.25
86	2	12.50
88	3	18.75
89	4	25.00
99	6	37.50
104	7	43.75
107	9	56.25
109	10	62.50
110	11	68.75
111	12	75.00
113	13	81.25
115	14	87.50
134	15	93.75

% cumulative frequency $= 100 \times$ cumulative frequency$/(n + 1)$

Η δεύτερη στήλη δίνει την αθροιστική ή συσσωρευτική συχνότητα για κάθε μέτρηση (δηλ. αριθμό μετρήσεων ίσο ή μικρότερο της μέτρησης)

Αν τα δεδομένα προέρχονται από κανονικό πληθυσμό το γράφημα της εκατοστιαίας συνολικής συχνότητας συναρτήσει των μετρήσεων παρέχει μία σιγμοειδή καμπύλη *S*.





Η μέθοδος ***Kolmogorov-Smirnov*** αποτελεί, επίσης, έναν τρόπο εκτίμησης κανονικής κατανομής.

Η συνολική κατανομή του στατιστικού δείγματος συγκρίνεται με την αθροιστική υποθετική (κανονική) κατανομή.

Στη μέθοδο Kolmogorov-Smirnov, τα δεδομένα μετασχηματίζονται με βάση την εξίσωση:

$$z = \frac{x_i - \overline{x}}{s}$$

όπου $\overline{x}$, η μέση τιμή και s η τυπική απόκλιση

Παράδειγμα

Πραγματοποιήθηκαν οκτώ τιτλοδοτήσεις και έδωσαν τα παρακάτω αποτελέσματα: 25,13, 25,02, 25,11, 25,07, 25,03, 24,97, 25,14 και 25,09 mL.

Προέρχονται τα αποτελέσματα από τον ίδιο κανονικό πληθυσμό?

Μηδενική υπόθεση: τα δεδομένα προέρχονται από κανονική κατανομή

Υπολογίζεται η μέση τιμή $\bar{x}$ = 25,07 και η τυπική απόκλιση s = 0,059.

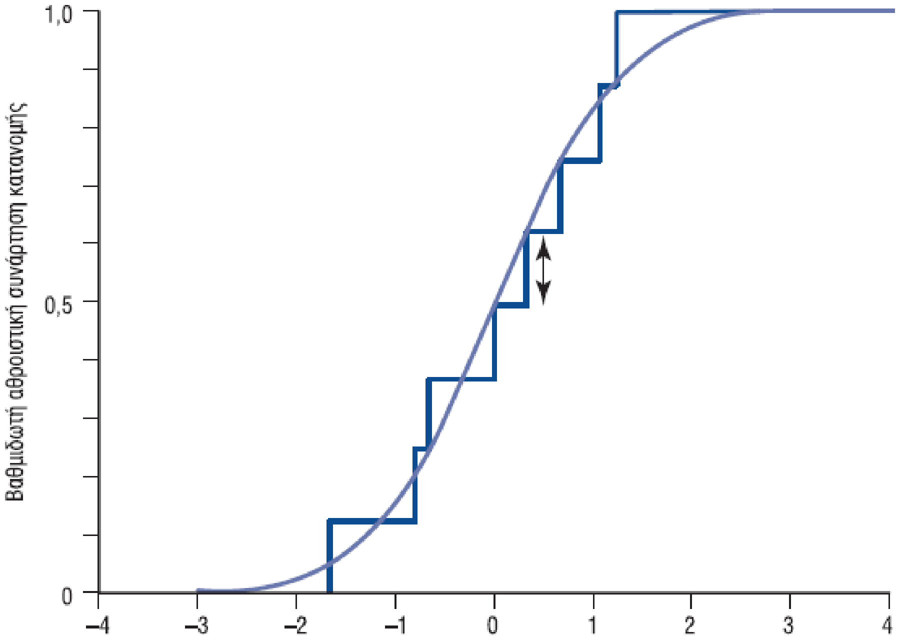
Ακολούθως, υπολογίζονται οι τιμές z από τη σχέση $$z = \frac{x_i - \bar{x}}{s}$$

$$z = (x_i - 25{,}07) / 0{,}059$$

Οι οκτώ τιμές μετασχηματίζονται σε 1,01, –0,84, 0,67, 0, –0,67, –1,69, 1,18 και 0,34.

Οι τιμές διευθετούνται κατά αύξοντα αριθμό και σχηματίζεται ένα βηματικό γράφημα με ύψος βήματος $1/n$, όπου n είναι ο αριθμός των μετρήσεων. Σ' αυτή την περίπτωση είναι $1/n = 1/8 = 0{,}125$.

Δοκιμασία Kolmogorov για κανονική κατανομή



Βρίσκω τη μέγιστη διαφορά μεταξύ της υποθετικής (θεωρητικής) και της πειραματικής τιμής, που είναι 0,132 για z = 0,34. Για n = 8 και P = 0,05, η κρίσιμη τιμή είναι 0,288. Επειδή 0,132 < 0,288,

η *μηδενική υπόθεση* ισχύει και συνεπώς, τα δεδομένα προέρχονται από κανονική κατανομή.

Critical two-tailed values at $P = 0.05$

n	Critical values
3	0.376
4	0.375
5	0.343
6	0.323
7	0.304
8	0.288
9	0.274
10	0.262
11	0.251
12	0.242
13	0.234
14	0.226
15	0.219
16	0.213
17	0.207
18	0.202
19	0.197
20	0.192

Δοκιμασία Shapiro-Wilk

Στο **κριτήριο Shapiro-Wilk** οι m τιμές του δείγματος διατάσσονται κατά αύξουσα σειρά και ακολούθως υπολογίζεται η ποσότητα

$$W = \frac{\left(\sum_{i=1}^{k} a_i (x_{m-i+1} - x_i)\right)^2}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

όπου k = m/2 ή k = (m-1)/2 ανάλογα αν ο m είναι άρτιος ή περιττός, ενώ οι σταθερές a_i λαμβάνονται από πίνακες.

H_0: Το δείγμα προέρχεται από κανονικό πληθυσμό,

H_a: Το δείγμα δεν προέρχεται από κανονικό πληθυσμό

Η δοκιμασία αυτή είναι μάλλον πιο περίπλοκη, καθώς ο υπολογισμός χρησιμοποιεί τόσο τις μεμονωμένες μετρήσεις όσο και τις διατεταγμένες μεμονωμένες μετρήσεις.

Εναλλακτικά, η **δοκιμασία Anderson-Darling**, που είναι επίσης μία ισχυρή δοκιμασία, απαιτεί όμως δείγματα με πλήθος τιμών μεγαλύτερο από 6.

Βήματα στη δοκιμασία (δοκιμή) σημαντικότητας

Συμπερασματικά, περισσότερες δοκιμασίες σημαντικότητας βασίζονται στα ίδια γενικά βήματα:

1. Αναφέρετε τη μηδενική υπόθεση σχετικά με τα δεδομένα που θέλετε να ελέγξετε.

2. Δηλώστε την εναλλακτική υπόθεση, που θα ληφθεί ως αληθής εάν η μηδενική υπόθεση απορριφθεί.

3. Ελέγξτε την κατανομή των δεδομένων. Μέχρις εδώ, όλες οι δοκιμασίες προϋποθέτουν ότι τα δεδομένα ακολουθούν (περίπου) κανονική κατανομή.

4. Επιλέξτε την κατάλληλη δοκιμασία.

5. Επιλέξτε το επίπεδο σημαντικότητας για τη δοκιμασία.

6. Επιλέξτε τον αριθμό των «ουρών» για τη δοκιμασία. Αυτό συνήθως προκύπτει άμεσα από την ακριβή μορφή της εναλλακτικής υπόθεσης.

7. Υπολογίστε το στατιστικό στοιχείο της δοκιμασίας.

8. Λάβετε την κρίσιμη τιμή για τη δοκιμασία.

9. Συγκρίνετε το στατιστικό στοιχείο της δοκιμασίας με την κρίσιμη τιμή.*

Βήματα στη δοκιμασία (δοκιμή) σημαντικότητας

*** Σύγκριση του στατιστικού στοιχείου της δοκιμασίας με την κρίσιμη τιμή**
Για να αποφασιστεί εάν η μηδενική υπόθεση πρέπει να γίνει αποδεκτή ή να απορριφθεί, το στατιστικό στοιχείο της δοκιμασίας συγκρίνεται με την κατάλληλη κρίσιμη τιμή για τη δοκιμασία.
Εάν το υπολογιζόμενο στατιστικό στοιχείο υπερβαίνει την κρίσιμη τιμή στο επιλεγμένο επίπεδο εμπιστοσύνης, η μηδενική υπόθεση απορρίπτεται και το αποτέλεσμα περιγράφεται ως στατιστικά σημαντικό.

Βήματα στη δοκιμασία (δοκιμή) σημαντικότητας

Η χρήση τιμών *p*

Εάν χρησιμοποιείται ένα στατιστικό λογισμικό για τις δοκιμασίες σημαντικότητας, το λογισμικό δίνει μια τιμή *p*, μαζί με το στατιστικό στοιχείο και την κρίσιμη τιμή.

Η τιμή *p* αντιπροσωπεύει την πιθανότητα μια τιμή του στατιστικού στοιχείου να είναι μεγαλύτερη ή ίση με την κρίσιμη τιμή εάν η μηδενική υπόθεση ισχύει.

Εάν αυτή η πιθανότητα είναι χαμηλή, το παρατηρούμενο αποτέλεσμα είναι απίθανο να προκύψει τυχαία, ή (θυμίζω) όσο μικρότερη είναι η πιθανότητα η παρατηρούμενη διαφορά να συμβεί τυχαία, τόσο λιγότερο πιθανό είναι η μηδενική υπόθεση να ισχύει.

Κατά συνέπεια, μια τιμή *p* μικρότερη από το επίπεδο σημαντικότητας που χρησιμοποιήθηκε για τη δοκιμή σημαντικότητας υποδεικνύει ότι είναι απίθανο να είχαν ληφθεί τα αποτελέσματα εάν η μηδενική υπόθεση ίσχυε. Επομένως, η μηδενική υπόθεση πρέπει να απορριφθεί.

Οι απλοί κανόνες για την ερμηνεία των τιμών *p* μπορούν να συνοψιστούν ως εξής:

- **Χαμηλή τιμή *p***: το αποτέλεσμα της δοκιμασίας είναι στατιστικά σημαντικό. Απορρίψτε τη μηδενική υπόθεση.
- **Υψηλή τιμή *p***: το αποτέλεσμα της δοκιμασίας δεν είναι στατιστικά σημαντικό. Αποδεχτείτε τη μηδενική υπόθεση.

Παράδειγμα

Να εξετασθεί αν η διάσπαση μιας ουσίας είναι στατιστικά σημαντική όταν σε ένα διάλυμά της συγκέντρωσης 10 mg/L, μια μέρα από την παρασκευή του διαλύματος έγιναν 10 μετρήσεις που έδωσαν τιμές:

9,4 9,5 9,7 10,2 10,1 10,0 8,9 9,4 10,1 8,8

Οι υποθέσεις H_0 και H_a στο συγκεκριμένο πρόβλημα είναι H_0: $\bar{x}$ = 10 και H_a: $\bar{x}$ < 10 επειδή θέλουμε να ελέγξουμε αν διασπάται η ουσία και συνεπώς, αν η μέση τιμή είναι μικρότερη από 10.

1. Εφαρμογή δοκιμασίας κανονικότητας. Το δείγμα δεν παρουσιάζει στατιστικά σημαντικές διαφοροποιήσεις από την κανονικότητα.

2. Άρα, μπορούμε να εφαρμόσουμε τη δοκιμασία *t* της μέσης τιμής (κατά προτίμηση μονόπλευρη).

Λογισμικό ChemStat

9,4	**One Sample tests**				
9,5					
9,7	**Anderson-Darling Normality test:**				
10,2					
10,1	p-value=	0,336049	Normality may be assumed		
10	**Parametric t-test**				
8,9	Mean=	9,61		Var=	0,249889
9,4	2 tailed t-test:				
10,1	t-value=	2,467125			
8,8	p-value=	0,035738	Null hypothesis, mean = 10, maybe rejected at level 0.05		
	2 tailed Permutation/Bootstrap tests				
	Iterations=	10000			
	p(permutation)=	0,0368	Null hypothesis, mean = 10 maybe rejected at level 0.05		
	p(bootstrap)=	0,0447	Null hypothesis, mean = 10 maybe rejected at level 0.05		
	Elapsed time =	0,003 min			

Παράδειγμα

Να εξετασθεί αν η διάσπαση μιας ουσίας είναι στατιστικά σημαντική όταν σε ένα διάλυμά της συγκέντρωσης 10 mg/L, μια μέρα από την παρασκευή του διαλύματος έγιναν 10 μετρήσεις που έδωσαν τιμές:

9,4 9,5 9,7 10,2 10,1 10,0 8,9 9,4 10,1 8,8

Επειδή το πρόγραμμα εκτελεί πάντα αμφίπλευρο έλεγχο, πρέπει την τιμή *p* που παίρνουμε να τη διαιρέσουμε δια 2. Συνεπώς, έχουμε p-value = 0,036/2 = 0,01787 < 0,05 που δείχνει ότι η πιθανότητα η διαφορά της μέσης τιμής του δείγματος x = 9,61 από την τιμή 10 να οφείλεται σε τυχαίο σφάλμα είναι μικρότερη από 1,8% (1,787).

Λογισμικό ChemStat

9,4	**One Sample tests**				
9,5					
9,7	**Anderson-Darling Normality test:**				
10,2					
10,1	p-value=	0,336049	Normality may be assumed		
10	**Parametric t-test**				
8,9	Mean=	9,61		Var=	0,249889
9,4	2 tailed t-test:				
10,1	t-value=	2,467125			
8,8	p-value=	0,035738	Null hypothesis, mean = 10, maybe rejected at level 0.05		
	2 tailed Permutation/Bootstrap tests				
	Iterations=	10000			
	p(permutation)=	0,0368	Null hypothesis, mean = 10 maybe rejected at level 0.05		
	p(bootstrap)=	0,0447	Null hypothesis, mean = 10 maybe rejected at level 0.05		
	Elapsed time =	0,003 min			

Καταλήγουμε στο συμπέρασμα ότι το διάλυμα της ουσίας δεν είναι σταθερό.

Μη παραμετρικές δοκιμασίες (non-parametric tests)

Οι στατιστικές δοκιμασίες που περιγράφηκαν **προϋποθέτουν ότι τα δεδομένα ακολουθούν κανονική κατανομή (παραμετρικές δοκιμασίες, parametric tests)**.

Το *Θεώρημα Κεντρικού Ορίου* δείχνει ότι *η κατανομή των μέσων τιμών είναι περίπου κανονική ακόμη κι όταν η κατανομή του αρχικού πληθυσμού είναι εντελώς διαφορετική*.

*Ω*στόσο, το θεώρημα δεν ισχύει (πάντα) για μικρές σειρές δεδομένων (τρία ή τέσσερα δεδομένα) που συνήθως χρησιμοποιούνται στην αναλυτική χημεία.

Οι δοκιμασίες που δεν προϋποθέτουν κανονική κατανομή των δεδομένων λέγονται *μη παραμετρικές* (**non-parametric**) ή ελεύθερες κατανομής (**distribution-free**).

Στις μη παραμετρικές δοκιμασίες, ως μέτρο θέσης, χρησιμοποιείται η *διάμεση τιμή* αντί της *μέσης τιμής* (το *ενδοτεταρτομοριακό εύρος* αντικαθιστά την τυπική απόκλιση).

Δοκιμασία προσήμων (sign test)

Είναι η πιο απλή από όλες τις μη παραμετρικές στατιστικές δοκιμασίες. Μπορεί να εφαρμοστεί με διαφορετικούς τρόπους.

Παράδειγμα

Ένα φαρμακευτικό σκεύασμα θεωρείται ότι περιέχει 8,0% από τη δραστική ουσία. Διαφορετικές παρτίδες βρέθηκαν ότι περιέχουν 7,3, 7,1, 7,9, 9,1, 8,0, 7,1, 6,8 και 7,3% από το συστατικό. Είναι τα αποτελέσματα αυτά συνεπή με τις προδιαγραφές του κατασκευαστή? (Δηλαδή, υπάρχει διαφορά μεταξύ της θεωρητικής τιμής και της πειραματικής τιμής?)

Το *t*-test προϋποθέτει κανονική κατανομή των δεδομένων από τα οποία υπολογίζονται η μέση τιμή και η τυπική απόκλιση.

Η *δοκιμασία προσήμων* αποφεύγει την υπόθεση αυτή και είναι πιο εύκολο να εκτελεστεί.

Χρησιμοποιούνται οι ίδιες βασικές αρχές, όπως και σε άλλες δοκιμασίες σημαντικότητας:
<u>ορίζεται η μηδενική υπόθεση</u>

στην περίπτωση του παραδείγματος, *η μηδενική υπόθεση είναι ότι τα δεδομένα προέρχονται από ένα πληθυσμό με διάμεση τιμή (median) του συστατικού 8,0%.*

<u>προσδιορίζεται η πιθανότητα *p* απόκτησης των πειραματικών δεδομένων</u>

<u>η μηδενική υπόθεση απορρίπτεται εάν η πιθανότητα *p* είναι μικρότερη από ένα ορισμένο κρίσιμο επίπεδο</u>.

<u>Εφαρμογή της δοκιμασίας προσήμων στο προηγούμενο παράδειγμα</u>
Η υποτιθέμενη περιεκτικότητα (δηλ. 8,0%), αφαιρείται από κάθε πειραματική τιμή με τη σειρά και λαμβάνεται υπόψη το *πρόσημο* κάθε αποτελέσματος (θυμίζω … 7,3, 7,1, 7,9, 9,1, 8,0, 7,1, 6,8 και 7,3%)

Οι τιμές που είναι ίσες με τη διάμεσο αγνοούνται εντελώς.

Στην περίπτωση αυτή έχουμε ουσιαστικά επτά πειραματικές τιμές, έξι από αυτές χαμηλότερες από τη διάμεσο με πρόσημο αρνητικό και μία υψηλότερη από την διάμεσο με πρόσημο θετικό.

Διωνυμική κατανομή με παραμέτρους *n* και *p*, είναι η διακριτή κατανομή πιθανοτήτων ενός αριθμού επιτυχιών σε μια αλληλουχία *n* ανεξάρτητων πειραμάτων της μορφής ναι/όχι, καθένα από τα οποία έχει πιθανότητα επιτυχίας *p*.

Για να ελέγξουμε αν η επικράτηση του αρνητικού προσήμου (στο παράδειγμά μας) είναι σημαντική χρησιμοποιούμε το *διωνυμικό θεώρημα*. Το θεώρημα αυτό δείχνει ότι η πιθανότητα να λάβουμε *r* αρνητικά (ή θετικά) πρόσημα από ένα σύνολο *n* προσήμων δίνεται από τη σχέση

$$P(r) = {}^{n}C_{r}\, p^{r} q^{(n-r)}$$

όπου ${}^{n}C_{r}$ ο διωνυμικός συντελεστής, δηλ. ο αριθμός των συνδυασμών των στοιχείων (δηλ. προσήμων) *r* από ένα σύνολο *n* στοιχείων

(υπολογιζόμενο ως ${}^{n}C_{r} = n!/r!(n - r)!$)

p η πιθανότητα ένα αποτέλεσμα να πάρει αρνητικό πρόσημο και
q η πιθανότητα ένα αποτέλεσμα να μην πάρει αρνητικό πρόσημο, δηλαδή

$$q = 1 - p$$

Δεδομένου ότι η διάμεσος ορίζεται έτσι ώστε το ήμισυ των πειραματικών δεδομένων να βρίσκονται πάνω από αυτό και το άλλο ήμισυ κάτω από αυτό, είναι σαφές ότι, αν η διάμεσος είναι 8,0, στην περίπτωση αυτή αμφότερα τα *p* και *q* πρέπει να είναι 1/2. Χρησιμοποιώντας την Εξίσωση

$$P(r) = {}^{n}C_{r}\,p^{r}q^{(n-r)}$$

βρίσκουμε την πιθανότητα να λάβουμε έξι αρνητικά πρόσημα

$$P(6) = {}^{7}C_{6} \times (\tfrac{1}{2})^{6} \times \tfrac{1}{2} = 7 \times (\tfrac{1}{2})^{6} \times \tfrac{1}{2} = 7/128$$

Ομοίως, μπορούμε να υπολογίσουμε ότι η πιθανότητα να λάβουμε επτά αρνητικά πρόσημα είναι

$$P(7), \text{ is } 1 \times (\tfrac{1}{2})^{7} \times 1 = 1/128$$

Ως εκ τούτου, συνολικά, η πιθανότητα να πάρουμε έξι ή περισσότερα αρνητικά πρόσημα στο πείραμά μας είναι **8/128**.

Ζητάμε να δούμε αν τα δεδομένα μας αυτά *διαφέρουν σημαντικά* από τη διάμεση τιμή. Έτσι, θα εκτελεστεί μια αμφίπλευρη δοκιμασία.

Άρα, όταν επτά αποτελέσματα λαμβάνονται τυχαία, πρέπει να βρούμε την πιθανότητα απόκτησης έξι ή περισσότερων όμοιων προσήμων (δηλαδή 6 θετικά ή 6 αρνητικά πρόσημα). Αυτή η πιθανότητα *p* αντιστοιχεί σε 2×8/128 = 16/128 =0,125.

Στη συνέχεια συγκρίνουμε την τιμή αυτή με το *P*=0,05,
δηλαδή πραγματοποιούμε τη δοκιμή σε επίπεδο σημαντικότητας 5% (το κρίσιμο επίπεδο πιθανοτήτων που χρησιμοποιείται συνήθως).

Δεδομένου ότι 0,125 > 0,05, η μηδενική υπόθεση, ότι τα δεδομένα προέρχονται από έναν πληθυσμό με διάμεση τιμή 8,0, *δεν* μπορεί να απορριφθεί (δηλαδή ισχύει).

Η δοκιμασία προσήμων μπορεί επίσης να εφαρμοστεί ως μια εναλλακτική μη παραμετρική δοκιμασία *t* για σύγκριση ζευγών τιμών.

Παράδειγμα

Αν οκτώ δείγματα αναλύονται με δύο μεθόδους *Α* και *Β*, μπορούμε να ελέγξουμε αν οι δύο μέθοδοι δίνουν στατιστικά διαφορετικά αποτελέσματα.

Μηδενική υπόθεση H_0 : οι δύο μέθοδοι δεν δίνουν (στατιστικά) σημαντικά διαφορετικά αποτελέσματα.

Πρακτικά, αυτό σημαίνει ότι η πιθανότητα να εμφανιστεί ένα θετικό (ή αρνητικό) πρόσημο για κάθε διαφορά είναι $\frac{1}{2}$. Ο αριθμός των θετικών ή αρνητικών προσήμων που πραγματικά λαμβάνεται μπορεί να συγκριθεί με την πιθανότητα που εξάγεται από την εξίσωση

$$P(r) = {}^{n}C_{r}\, p^{r} q^{(n-r)}$$

Παράδειγμα

Οι συγκεντρώσεις της ανοσογλοβουλίνης (σε mg/mL) στο αίμα δέκα δοτών μετρήθηκαν με δύο μεθόδους με τα ακόλουθα αποτελέσματα:

Δότης:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Μέθοδος Α:	1,3	1,5	0,7	0,9	1,0	1,1	0,8	1,8	0,4	1,3
Μέθοδος Β:	1,1	1,6	0,5	0,8	0,8	1,0	0,7	1,4	0,4	0,9

Είναι στατιστικά σημαντικές οι διαφορές των δύο μεθόδων με βάση τα παραπάνω αποτελέσματα?

Αν οι τιμές της μεθόδου Β αφαιρεθούν από τις αντίστοιχες της μεθόδου *Α*, τα πρόσημα που προκύπτουν είναι + - + + + + + + 0 +.
Στην πραγματικότητα έχουμε εννιά αποτελέσματα, οκτώ με πρόσημο + κι ένα με πρόσημο -.
Η πιθανότητα να είναι όμοια τα οκτώ πρόσημα από το σύνολο των εννιά είναι *P* = 0,04.

Έτσι, για επίπεδο σημαντικότητας P = 0,05, η μηδενική υπόθεση H_0 *απορρίπτεται: τα αποτελέσματα είναι σημαντικά διαφορετικά.*

Δοκιμασία *U* Mann–Whitney

Είναι η μη-παραμετρική δοκιμασία ισοδύναμη της δοκιμασίας *t* μέσων τιμών δύο δειγμάτων χωρίς να απαιτείται η υπόθεση γύρω από την κατανομή του πληθυσμού.

Η δοκιμασία ***U* Mann–Whitney** είναι απλή στην κατανόηση και εκτέλεση, η εφαρμογή της οποίας αποδεικνύεται πιο εύκολα με ένα παράδειγμα.

Παράδειγμα

Ένα δείγμα ορού αίματος αναλύθηκε για τη συγκέντρωσή του σε αλβουμίνη, με μια χρωματομετρική μέθοδο με τη χρήση χρωστικής και πέντε διαδοχικές μετρήσεις που έδωσαν τιμές 3,8, 4,2, 4,7, 3,5 και 4,5 g/100 mL. Μετά την εφαρμογή μιας νέας θεραπείας σχεδιασμένης για τη μείωση του επιπέδου της αλβουμίνης, το δείγμα αναλύθηκε ξανά με την ίδια διαδικασία, με πέντε διαδοχικές μετρήσεις, που έδωσαν τιμές 1,7, 3,7, 2,0, 3,9 και 3,0 g/100 mL. Υπάρχει κάποια ένδειξη ότι η νέα μέθοδος θεραπείας προκάλεσε σημαντική μείωση στα επίπεδα της αλβουμίνης;

Ορός χωρίς τη θεραπεία	Υψηλότερες τιμές στον ορό μετά τη θεραπεία	Αριθμός υψηλότερων τιμών
3,8	3,9	1
4,2	-	0
4,7	-	0
3,5	3,7, 3,9	2
4,5	-	0

Ορός χωρίς τη θεραπεία	Υψηλότερες τιμές στον ορό μετά τη θεραπεία	Αριθμός υψηλότερων τιμών
3,8	3,9	1
4,2	-	0
4,7	-	0
3,5	3,7, 3,9	2
4,5	-	0

Θεωρούμε ότι η συγκέντρωση της αλβουμίνης του ορού μετά τη θεραπεία, πρέπει, αν μη τι άλλο, να είναι *χαμηλότερη* από αυτή του ορού που δεν έχει υποβληθεί σε θεραπεία. Επομένως, είναι κατάλληλη μια μονόπλευρη δοκιμασία.

Αναμένουμε ότι: ο αριθμός των περιπτώσεων στις οποίες ο ορός που έχει υποβληθεί σε θεραπεία δίνει υψηλότερη τιμή από έναν ορό που δεν έχει υποβληθεί σε θεραπεία να είναι μικρός.

Ορός χωρίς τη θεραπεία	Υψηλότερες τιμές στον ορό μετά τη θεραπεία	Αριθμός υψηλότερων τιμών
3,8	3,9	1
4,2	-	0
4,7	-	0
3,5	3,7, 3,9	2
4,5	-	0

Το σύνολο των αριθμών στην τρίτη στήλη, σε αυτήν την περίπτωση, είναι 3 και είναι το **στατιστικό στοιχείο της δοκιμασίας**.

Οι κρίσιμες τιμές που οδηγούν στην απόρριψη της μηδενικής υπόθεσης είναι εκείνες που είναι *μικρότερες ή ίσες* των αριθμών που παρουσιάζονται στον πίνακα.

Mann–Whitney U-test. Critical values for U or the lower of T_1 and T_2 at $P = 0.05$

n_1	n_2	One-tailed test	Two-tailed test
3	3	0	NA
3	4	0	NA
3	5	1	0
3	6	2	1
4	4	1	0
4	5	2	1
4	6	3	2
4	7	4	3
5	5	4	2
5	6	5	3
5	7	6	5
6	6	7	5
6	7	8	6
7	7	11	8

The null hypothesis can be rejected when U or the lower T value is less than or equal to the tabulated value. NA indicates that the test cannot be applied.

Για μια μονόπλευρη δοκιμασία για P = 0,05, με πέντε μετρήσεις σε κάθε περίπτωση, το στατιστικό στοιχείο πρέπει να είναι ≤ 4 εάν η μηδενική υπόθεση πρόκειται να απορριφθεί.

Στο παράδειγμά μας μπορούμε, έτσι, να απορρίψουμε την H_0, δηλαδή, η θεραπεία, πιθανώς, μειώνει το επίπεδο της πρωτεΐνης στον ορό.

Παράδειγμα

Στον πίνακα δίνονται τα ολικά αιωρούμενα σωματίδια (total suspended particles,TSP) σε μg/m^3 στον αέρα, στους σταθμούς Παγκράτι και Κατεχάκη, τους δύο πρώτους μήνες του 2017. Να ελεγχθεί αν υπάρχει στατιστική διαφορά στη ρύπανση μεταξύ των σταθμών.

Ημερομηνία	Παγκράτι	Ημερομηνία	Κατεχάκη
10/1/2017	61,67	2/1/2017	83,49
22/1/2017	48,75	4/1/2017	37,96
26/1/2017	40,83	6/1/2017	34,71
28/1/2017	29,17	8/1/2017	62,57
30/1/2017	20,42	10/1/2017	106,11
1/2/2017	50,42	12/1/2017	101,07
5/2/2017	28,75	14/1/2017	61,85
7/2/2017	97,08	5/2/2017	32,13
9/2/2017	108,75	7/2/2017	53,94
11/2/2017	55,00	9/2/2017	136,76
13/2/2017	58,33	11/2/2017	51,79
15/2/2017	37,92	13/2/2017	62,94
17/2/2017	16,67	17/2/2017	47,14
19/2/2017	29,17	19/2/2017	41,37
23/2/2017	36,67	21/2/2017	65,56
		23/2/2017	70,44

Ανάλυση στο λογισμικό ChemStat

Ο έλεγχος της κανονικότητας είναι απαραίτητος.

Με το κριτήριο *Anderson-Darling* και για επίπεδο σημαντικότητας $p = 0{,}05$ δεν διαπιστώνονται στατιστικά σημαντικές αποκλίσεις από την κανονικότητα.

Η παραμετρική δοκιμασία για τις μέσες τιμές δείχνει ότι η μηδενική υπόθεση δεν μπορεί να απορριφθεί ($p = 0{,}085$).

Το σχετικά μικρό τους μέγεθος και το γεγονός ότι πρόκειται για περιβαλλοντικά δείγματα πρέπει να μας κάνει επιφυλακτικούς ως προς την κανονικότητα.

Το ισχυρό κριτήριο των Shapiro-Wilk στο πρώτο δείγμα διαπιστώνει στατιστικά σημαντικές αποκλίσεις από την κανονικότητα.

Εναλλακτικά, εξετάζουμε τα αποτελέσματα μη παραμετρικών ελέγχων.
Η μη παραμετρική δοκιμασία με το κριτήριο Mann-Whitney και η παραλλαγή της με τη μέθοδο Monte-Carlo με αντιμεταθέσεις δείχνουν ότι η μηδενική υπόθεση πρέπει να απορριφθεί (p-value = 0,032 και 0,031, αντίστοιχα). Αυτό σημαίνει ότι πρέπει να υπάρχει στατιστικά σημαντική διαφορά στην αέρια ρύπανση των δύο περιοχών.

Two Independent Samples tests - 2 tailed

Anderson-Darling Normality test:

p1-value=	0,062292	Sample1-Normality may be assumed
p2-value=	0,110994	Sample2-Normality may be assumed

Mean1=	47,97222	Mean2=	65,61616
Var1=	682,338	Var2=	828,6293

Test of variances with F test

p-value= 0,721449 Equality of variances may be assumed

2 tailed parametric t test:

If equality of variances may be assumed

t-value=	1,783135	
p-value=	0,085036	Null hypothesis, mean1=mean2, may be assumed at level 0.05

If equality of variances may be rejected

t-value=	1,788904	
p-value=	0,084082	Null hypothesis, mean1=mean2, may be assumed at level 0.05

2 tailed Permutation test

p(permut.)= 0,0793 Equal variances are assumed

2 tailed Mann-Witney Non-Parametric test

U-value= 66 Z= -2,13475

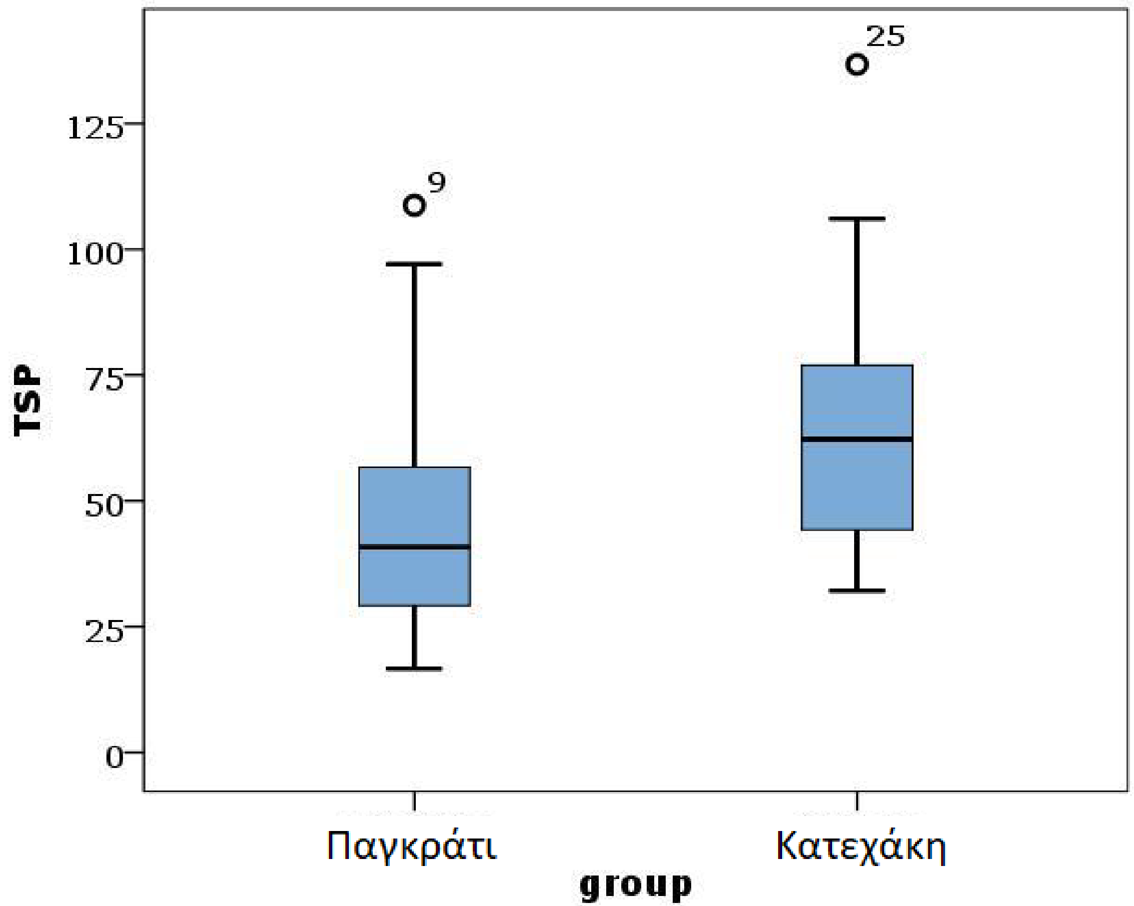
p(asymp.)= 0,032781 Null hypothesis may be rejected at level 0.05

2 tailed Mann-Whitney Permutation test

p(permut.)= 0,0312 Null hypothesis, d = 0, may be rejected at level 0.05

Monte-Carlo i	10000
Elapsed time	0,053 min

Για να δούμε σε ποια περιοχή η ρύπανση είναι μικρότερη, μπορούμε να χρησιμοποιήσουμε τα Θηκογράμματα.
Παγκράτι: διάμεσος = 40,83 μg/m^3 Κατεχάκη: διάμεσος = 62,21 μg/m^3
Συνεπώς, μπορούμε να συμπεράνουμε ότι στο Παγκράτι η αέρια ρύπανση φαίνεται να είναι μικρότερη από την ρύπανση στην Κατεχάκη (τουλάχιστον στους δύο πρώτους μήνες του 2017).



Παράδειγμα

Εμβολιάζονται 200 άνθρωποι για το νέο ιό της γρίπης και οι 185 αποκτούν ανοσία. Ζητείται να ορίσουμε το ποσοστό επιτυχιών *p* καθώς και να ορίσουμε ένα διάστημα εμπιστοσύνης 99,5 % για το παραπάνω αποτέλεσμα.

Το μικρότερο υποσύνολο πρέπει να έχει τουλάχιστον 10 παρατηρήσεις. Εδώ έχουμε 15 αποτυχίες > 10.
Εκτιμούμε τις πιθανότητες βάσει του δείγματος: $p = 185 / 200 = 0{,}925$, $q = 1-p = 0{,}075$.

Βρίσκουμε το z_0, όπως κάναμε πιο πάνω που είναι ίσο με 2,58.

Βρίσκουμε διάστημα εμπιστοσύνης: 0,925 ± 2,58 × (0,925 × 0,075/200)1/2 = 0,925 ± 0,048 = 0,877 έως 0,973.

Άλλες σημαντικές δοκιμασίες

Tukey's
Dunn's
Kruskal-Wallis η μη-παραμετρική ισοδύναμη της Ανάλυσης Διακύμανσης-Analysis of Variance (ANOVA)

Παρουσίαση στατιστικών δεδομένων σε μία ερευνητική αναφορά

Τα αποτελέσματα των στατιστικών αναλύσεων βοηθούν στην κατανόηση του αποτελέσματος της μελέτης, π.χ.

αν κάποια μεταβλητή επηρεάζει το αποτέλεσμα,

αν οι μεταβλητές σχετίζονται,

αν οι διαφορές μεταξύ των ομαδοποιημένων δεδομένων-παρατηρήσεων διαφέρουν ή όχι κλπ.

Η στατιστική θα πρέπει να χρησιμοποιείται για να τεκμηριώσει τα ευρήματα και να βοηθήσει στην εξαγωγή αντικειμενικών συμπερασμάτων.

Επομένως, όταν αναφέρετε τα στατιστικά αποτελέσματα που σχετίζονται με τη μελέτη σας, «υποτάξτε τα» στα πραγματικά αποτελέσματα.

Δομή ερευνητικής αναφοράς: Τίτλος, περίληψη, εισαγωγή, μέθοδοι-θεωρητικό υπόβαθρο, αποτελέσματα, συζήτηση, συμπεράσματα, βιβλιογραφία

Αναφορά αποτελεσμάτων περιγραφικής στατιστικής

Μέσες τιμές

1. Αναφέρετε πάντα τη μέση τιμή μαζί με ένα μέτρο μεταβλητότητας (τυπική απόκλιση ή τυπικό σφάλμα του μέσου όρου).

Δύο συνηθισμένοι τρόποι για να εκφραστεί ο μέσος όρος και η μεταβλητότητα φαίνονται παρακάτω:

"Συνολική συγκέντρωση νιτρικών ιόντων σε δείγματα νερού (n = 128) από τη λίμνη Παμβώτιδα, το Μάιο του 2004, κατά μέσο όρο 1,2 mg/L (s = 0,4 mg/L)."

s = τυπική απόκλιση

"Συνολική συγκέντρωση νιτρικών ιόντων σε δείγματα νερού (n = 128) από τη λίμνη Παμβώτιδα, το Μάιο του 2004, κατά μέσο όρο 1,2 ± 0,4 mg/L."

Ο τελευταίος αυτός τρόπος απαιτεί συγκεκριμένα να αναφέρεται στις *Μεθόδους* της εργασίας ποιο είναι το μέτρο μεταβλητότητας που αναφέρεται στο μέσο όρο.

Αναφορά αποτελεσμάτων περιγραφικής στατιστικής

Μέσες τιμές

2. Εάν τα συνοπτικά στατιστικά στοιχεία παρουσιάζονται σε γραφική μορφή (σχήμα), μπορεί απλά να αναφέρεται το αποτέλεσμα στο κείμενο ως:

"Η μέση τιμή της συγκέντρωσης νιτρικών ιόντων στη λίμνη Παμβώτιδα αυξήθηκε κατά 15% (ή δίνεται η απόλυτη αύξηση σε mg/L) μεταξύ Μαΐου και Σεπτεμβρίου 2004 (σχήμα...)"

Συχνότητες

Τα δεδομένα συχνότητας πρέπει να συνοψίζονται στο κείμενο με κατάλληλα μέτρα, όπως ποσοστά ή αναλογίες.

"Κατά την περίοδο του φθινοπώρου, το 47% των δειγμάτων νερού και το 24% των δειγμάτων του νερού των πόρων (pore water) των ιζημάτων περιείχαν συγκεντρώσεις μεγαλύτερες από 1,0 mg/L (Πίνακας ή Ραβδόγραμμα-Ιστόγραμμα)."

Αναφορά αποτελεσμάτων συμπερασματικής στατιστικής

Στο παρακάτω παράδειγμα, το βασικό αποτέλεσμα εμφανίζεται με μπλε χρώμα και το στατιστικό αποτέλεσμα, το οποίο *τεκμηριώνει* το εύρημα, είναι με κόκκινο.

"Η μέση συνολική συγκέντρωση νιτρικών ιόντων στο νερό της λίμνης Παμβώτιδας μεταξύ Μαΐου (1,2 ± 0,4 mg/L, *η* = 128) και Σεπτεμβρίου (1,6 ± 0,3 mg/L, *η* = 114) του έτους 2004 παρουσίασε σημαντική αύξηση (0,4 mg/L) (δοκιμασία *t* δύο μέσων τιμών δειγμάτων, $p < 0{,}001$)."

Σύνοψη των στατιστικών αποτελεσμάτων των στατιστικών δοκιμασιών σε γραφήματα

Εάν τα αποτελέσματα που παρουσιάζονται σ' ένα σχήμα έχουν υποστεί μια συμπερασματική δοκιμασία, είναι σκόπιμο να συνοψίζονται τα στοιχεία της δοκιμασίας αυτής σ' ένα γράφημα, για τη γρήγορη κατανόηση της σημασίας των ευρημάτων. Επιβάλλεται να συμπεριλαμβάνονται πληροφορίες στις *Μεθόδους* ή στη λεζάντα του σχήματος, για να εξηγείται το σύστημα κωδικοποίησης που χρησιμοποιείται.

Μέθοδοι περιγραφής των στατιστικών αποτελεσμάτων παρουσιάζονται παρακάτω.

Παραδείγματα: Σύγκριση ομάδων δεδομένων (δοκιμασίες *t*, ANOVA κλπ.)

Η σύγκριση των μέσων τιμών των δύο ή περισσοτέρων ομάδων απεικονίζεται συνήθως σε ραβδογράμματα (bar graphs) των μέσων τιμών με τις σχετικές γραμμές σφάλματος.