**ΤΜΗΜΑ ΝΟΣΗΛΕΥΤΙΚΗΣ**

**Σημειώσεις Βιοστατιστικής**

**Εαρινό Εξάμηνο 2013-2014**

**Μ. Λαβδανίτη**

**Εισαγωγή**

* Η στατιστική, όπως προκύπτει και από την ετυμολογία της λέξης (status=κράτος) συνδεόταν αρχικά με την λειτουργία του κράτους και για αυτό το λόγο περιοριζόταν σε πληροφορίες
* Μεσαίωνας (απογραφές )
* Εκκλησιαστικά κείμενα (Παλαιά Διαθήκη)
* Ορισμός

***Στατιστική,*** είναι ένας κλάδος των Μαθηματικών ο οποίος βασίζεται σε ένα σύνολο αρχών και μεθοδολογιών για:

* Το σχεδιασμό της διαδικασίας συλλογής δεδομένων **(σχεδιασμός πειραμάτων)**
* Τη συνοπτική και αποτελεσματική παρουσίασή τους **(περιγραφική στατιστική)**
* Την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων. (***στατιστική συμπερασματολογία)***

**ΟΡΙΣΜΟΙ**

* ***Πληθυσμός*** : ονομάζεται το σύνολο ατόμων ή αντικειμένων ή άλλων οντοτήτων για τα οποία ενδιαφερόμαστε να βγάλουμε συμπεράσματα σε σχέση με κάποιες ιδιότητες που αφορούν στα στοιχεία του.
* Ο πληθυσμός πρέπει να είναι καλά ορισμένος , να περιγράφεται δηλαδή με κάποιες ιδιότητες των στοιχείων που τον αποτελούν , με τέτοιο τρόπο ώστε να μπορεί κάποιος να αποφανθεί αν κάποιο στοιχείο είναι μέλος του (ικανοποιεί τις ιδιότητες) ή όχι (δεν ικανοποιεί τις ιδιότητες)
* ***Άτομα*** του πληθυσμού ονομάζονται τα στοιχεία του πληθυσμού.
* ***Απογραφή*** λέγεται η μέθοδος συλλογής στατιστικών δεδομένων για κάποιο χαρακτηριστικό ενός πληθυσμού για το οποίο εξετάζονται όλα τα άτομα του πληθυσμού.
* ***Δείγμα*** ονομάζεται κάθε υποσύνολο του πληθυσμού.
* ***Μέγεθος του δείγματος*** ονομάζεται το πλήθος των στοιχείων του δείγματος.
* **Μεταβλητές** (μεγέθη, χαρακτηριστικά, Variables) του πληθυσμού ονομάζονται τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό.

Διακρίνονται σε

* **Ποιοτικές** (qualitative or categorical)
* **Ποσοτικές** (quantitative or numerical)

**Μεταβλητές**

* **Ποιοτικές μεταβλητές** είναι εκείνες που δεν επιδέχονται αριθμητικές μετρήσεις αλλά περιγράφονται οι κατηγορίες στις οποίες ταξινομούνται οι παρατηρήσεις όπως π.χ. το φύλο, το χρώμα της ίριδας, η παρουσία ή απουσία ενός παθολογικού χαρακτηριστικού

Διακρίνονται σε

* **Κατηγορικές** ή ονομαστικές (nominal): έχει σημασία μόνο οι διαφορετικές τιμές που μπορεί να πάρει
* **Διάταξης** μπορούμε να ορίσουμε για μεταβλητές μια σχέση διάταξης (ordinal)

**Ποσοτικές** είναι εκείνες οι μεταβλητές οι οποίες επιδέχονται αριθμητικές μετρήσεις όπως τα φυσικά μεγέθη (ανάστημα, βάρος), τα βιολογικά μεγέθη (χοληστερόλη, σάκχαρο)

**Κατηγορίες**

* **Ασυνεχείς :** είναι δυνατό να λάβουν μόνο ορισμένες αριθμητικές τιμές
* **Συνεχείς:** παίρνουν όλες τις τιμές των πραγματικών αριθμών

**Πίνακες**

* Ένας πίνακας (table) είναι πιθανά ο απλούστερος τρόπος να συνοψιστούν τα δεδομένα μας και μπορεί να χρησιμοποιηθεί για όλους τους τύπους δεδομένων.
* Τα δεδομένα παρουσιάζονται σύντομα και με σαφήνεια σε πίνακες συχνοτήτων, δηλαδή πίνακες όπου αναγραφονται σε στήλες οι τιμές της μεταβλητής, οι συχνότητες εμφάνισης κάθε τιμής της μεταβλητής, κτ.λ.
* Το πλήθος των δεδομένων ενός δείγματος που εξετάζονται για την μελέτη ενός φαινομένου ονομάζεται **μέγεθος του δείγματος και συμβολίζεται ν.**
* **Έστω λοιπόν ένα δείγμα μεγέθους ν και η μεταβλητή Χ που παίρνει τιμές x1, x2,..., xκ, όπου κ < ν.**
* *Ο φυσικός αριθμός νi, i =1,...,κ που δείχνει πόσες φορές εμφανίζεται η τιμή xi της εξεταζόμενης μεταβλητής στο σύνολο των παρατηρήσεων* ***ονομάζεται συχνότητα.***
* Αν διαιρέσουμε την συχνότητα **νi, *i = 1, 2, … ,κ,*** με το μέγεθος του δείγματος προκύπτει η σχετική συχνότητα fi της τιμής xi: fi =vi/v
* Συνήθως τη σχετική συχνότητα την εκφράζουμε επί τις εκατό οπότε συμβολίζουμε **fi%** και ισχύει: fi%=100 fi.
* Η αθροιστική σχετική συχνότητα (cumulative relative frequency) ενός διαστήματος είναι το ποσοστό του συνολικού αριθμού παρατηρήσεων που έχουν τιμή κάτω από ή ίση με το άνω όριο του διαστήματος.
* Η αθροιστική σχετική συχνότητα υπολογίζεται προσθέτοντας τις σχετικές συχνότητες για το συγκεκριμένο διάστημα και όλα τα προηγούμενα διαστήματα. Με αυτόν τον τρόπο, για την ομάδα ανδρών ηλικίας 25 έως 35 ετών στον πίνακα 2.6 , η αθροιστική σχετική συχνότητα του δεύτερου διαστήματος είναι 1,2+ 14,1=15,3%. Παρομοίως η αθροιστική σχετική συχνότητα του τρίτου διαστήματος είναι 1,2+ 14,1+41,4=56,7%.

**Πίνακας**  Απόλυτες και σχετικές συχνότητες επιπέδων χοληστερόλης στον ορό για 2.294 άνδρες, ΗΠΑ, 1976-1980

|  |  |  |
| --- | --- | --- |
| Επίπεδο Χοληστερόλης (mg/100ml) |  Ηλικία 25-34 |  Ηλικία 55-64 |
|  | Αριθμός ανδρών | Σχετική Συχνότητα (%) | Αριθμός ανδρών | Σχετική Συχνότητα (%) |
| 80-119120-159160-199200-239240-279280-319320-359360-399 | 131504422991153495 | 1,214,141,128,010,83,20,80,5 | 548265458281128357 | 0,43,921,937,322,910,42,90,6 |
| Σύνολο | 1.067 | 100,0 | 1.227 | 100,0 |

* Η αθροιστική σχετική συχνότητα (cumulative relative frequency) ενός διαστήματος είναι το ποσοστό του συνολικού αριθμού παρατηρήσεων που έχουν τιμή κάτω από ή ίση με το άνω όριο του διαστήματος.
* Η αθροιστική σχετική συχνότητα υπολογίζεται προσθέτοντας τις σχετικές συχνότητες για το συγκεκριμένο διάστημα και όλα τα προηγούμενα διαστήματα. Με αυτόν τον τρόπο, για την ομάδα ανδρών ηλικίας 25 έως 35 ετών στον πίνακα 2.6 , η αθροιστική σχετική συχνότητα του δεύτερου διαστήματος είναι 1,2+ 14,1=15,3%. Παρομοίως η αθροιστική σχετική συχνότητα του τρίτου διαστήματος είναι 1,2+ 14,1+41,4=56,7%.

**ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ**

Γραφικές παραστάσεις κατανομών συχνοτήτων μεταβλητών

***ΡΑΒΔΟΓΡΑΜΜΑ:***

* Το ***ραβδόγραμμα*** είναι ένα είδος γραφικής παράστασης που χρησιμοποιείται όταν έχουμε να κάνουμε γραφική παράσταση ποιοτικής μεταβλητής.
* Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες (που μοιάζουν με ράβδους) των οποίων οι βάσεις βρίσκονται πάνω στον οριζόντιο ή τον κατακόρυφο άξονα.
* Για την κατασκευή του ραβδογράμματος, σχεδιάζουμε για κάθε τιμή της μεταβλητής Χ ένα ορθογώνιο με ύψος ίσο με την αντίστοιχη συχνότητα.

***Παράδειγμα:***

Οι υποψήφιοι πρόεδροι για το κεντρικό φοιτητικό στο ΤΕΙ εκλέχτηκαν μεταξύ 100 αντιπροσώπων των τμημάτων. Οι υποψήφιοι ήταν πέντε ,οι Α,Β,Γ,Δ και Ε. Οι ψήφοι που συγκέντρωσε ο καθένας από αυτούς φαίνονται στον διπλανό πίνακα

 α)Να κατασκευαστεί ο πίνακας συχνοτήτων και να συμπληρωθεί με τις στήλες: συχνότητα, αθροιστική συχνότητα, σχετική συχνότητα και εκατοστιαία σχετική συχνότητα.

(β)Να κατασκευαστεί το ραβδόγραμμα συχνοτήτων

***Λύση***

(α) **ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Υποψήφιος***  |  | ***Σ*** |  |  |
| **Α** | 25 | 25 | 0,25 | **25** |
| **Β** | 12 | 37 | 0,12 | **12** |
| **Γ** | 4 | 41 | 0,04 | **4** |
| **Δ** | 35 | 76 | 0,35 | **35** |
| **Ε** | 24 | 100 | 0,24 | **24** |
|  |  |  |  |  |

Β

**Διάγραμμα και πολύγωνο συχνοτήτων**

* Το ***διάγραμμα συχνοτήτων*** είναι ένα είδος γραφικής παράστασης που το χρησιμοποιούμε για να παραστήσουμε τις τιμές μιας ποσοτικής μεταβλητής.
* Η κατασκευή του διαγράμματος συχνοτήτων γίνεται όπως και του ραβδογράμματος, με μόνη διαφορά ότι αντί να χρησιμοποιούμε ορθογώνια, σε κάθε τιμή *,* υψώνουμε μια κάθετη γραμμή με μήκος ίσο με την αντίστοιχη συχνότητα .Οι τιμές , είναι τοποθετημένες πάνω στον άξονα κατά αύξουσα σειρά.
* Το ***πολύγωνο συχνοτήτων*** προκύπτει αν ενώσουμε τα σημεία του διαγράμματος συχνοτήτων.

***Παράδειγμα:***

* Ο αριθμός των παιδιών που υπάρχουν σε 26 οικογένειες της Λεμεσού είναι ο παρακάτω:

2 2 4 3 2 1 1 0 2 1 2 3 4

 2 1 2 2 1 2 3 4 1 0 0 1 3

 (α)Να κατασκευαστεί ο πίνακας συχνοτήτων και να συμπληρωθεί με τις στήλες: συχνότητα, αθροιστική συχνότητα, σχετική συχνότητα και εκατοστιαία σχετική συχνότητα.

 (β)Να κατασκευάσετε το διάγραμμα και το πολύγωνο συχνοτήτων.

Λύση

(α)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Αριθμός******Παιδιών*** |  | ***Σ*** |  |  |
| **0** | 3 | 3 | 0,11 | **11** |
| **1** | 7 | 10 | 0,29 | **29** |
| **2** | 9 | 19 | 0,34 | **34** |
| **3** | 4 | 23 | 0,15 | **15** |
| **4** | 3 | 26 | 0,11 | **11** |
|  |  |  | **1** | **100** |

β)

**ΚΥΚΛΙΚΟ ΔΙΑΓΡΑΜΜΑ**

* Το ***κυκλικό διάγραμμα*** είναι ένα είδος γραφικής παράστασης που χρησιμοποιείται για να απεικονίσουμε τις τιμές μιας ποιοτικής μεταβλητής ή μιας ποσοτικής στην περίπτωση που οι τιμές της είναι σχετικά λίγες.
* Για να κατασκευάσουμε ένα κυκλικό διάγραμμα συχνοτήτων των τιμών μιας μεταβλητής Χ, αρκεί να χωρίσουμε τον κυκλικό δίσκο σε κυκλικούς τομείς, τα εμβαδά των οποίων να είναι ανάλογα προς τις αντίστοιχες συχνότητες .
* Για να βρούμε την επίκεντρη γωνία του κυκλικού τομέα που αντιστοιχεί στην τιμή της μεταβλητής Χ, χρησιμοποιούμε την σχέση:



**ΙΣΤΟΓΡΑΜΜΑ**

Το ***ιστόγραμμα*** το χρησιμοποιούμε για την παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα.

Οι ωριαίες απουσίες των μαθητών μιας τάξης φαίνονται στον επόμενο πίνακα.

α)Να κατασκευάσετε τον πίνακα συχνοτήτων με τις στήλες , ,Σ, και .

(β)Να βρείτε το πλήθος και το ποσοστό των μαθητών που έχουν:

 (ι)τουλάχιστον 20 ωριαίες απουσίες

 (ιι)λιγότερες από 40 ωριαίες απουσίες.

(γ)Να κατασκευάσετε το ιστόγραμμα συχνοτήτων και το πολύγωνο συχνοτήτων.

***Λύση***

(α)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Ωριαίες απουσίες*** |  |  | ***Σ*** |  |  |
|  | 5 | 8 | 8 | 0,16 | **16** |
|  | 15 | 20 | 28 | 0,4 | **40** |
|  | 25 | 15 | 43 | 0,3 | **30** |
|  | 35 | 5 | 48 | 0,1 | **10** |
|  | 45 | 2 | 50 | 0,04 | **4** |
|  |  | **50** |  | **1** | **100** |

(β) (ι) 22 μαθητές με ποσοστό 44%.

 (ιι) 48 μαθητές με ποσοστό 96%.



**Μορφές και χαρακτηριστικά των κατανοµών συχνοτήτων ποσοτικών µεταβλητών**

* Οι κατανοµές συχνοτήτων και οι αντίστοιχες καµπύλες µπορεί να διακριθούν, ανάλογα µε τη µορφή τους, σε µονοκόρυφες (unimodal) και πολυκόρυφες (multimodal).
* Στις µονοκόρυφες, οι παρατηρήσεις συγκεντρώνονται
σε κάποια αριθµητική τιµή του µετρούµενου µεγέθους, πέρα από την οποία (προς τη µια ή και τις δύο διευθύνσεις) σηµειώνεται συνεχής µείωση.
* Αντίθετα, στις πολυκόρυφες, η πτώση συχνοτήτων, που αρχίζει από κάποια κορυφή, αντιστρέφεται σε κάποιο σηµείο, µε συνέπεια τη δηµιουργία µιας νέας
κορυφής περισσότερο ή λιγότερο έκδηλης.

**Μορφές και χαρακτηριστικά των κατανοµών συχνοτήτων ποσοτικών µεταβλητών**

* Από τις µονοκόρυφες κατανοµές οι συνηθέστερες είναι οι κωδωνοειδείς
* Σ' αυτές, η κορυφή βρίσκεται ανάµεσα στα δύο άκρα της κατανοµής, σε αντιδιαστολή προς τις µονοκόρυφες κατανοµές στις οποίες η κορυφή ταυτίζεται µε το κατώτερο ή ανώτερο άκρο
* Οι κωδωνοειδείς κατανοµές µπορεί να είναι συµµετρικές ή ασύµµετρες. Ένα ιδιαίτερο είδος συµµετρικής κωδωνοειδούς κατανοµής ονοµάζεται κ α ν ο ν ι κ ή (normal ή Gaussian distribution) αι διαδραµατίζει ξεχωριστό ρόλο στη βιοστατιστική και γενικότερα στη στατιστική.
* Από τις ασύµµετρες, εκείνες που εκτείνονται δυσανάλογα προς τις µεγάλες
τιµές (προς τα δεξιά της οριζόντιας κλίµακας) ονοµάζονται θετικά λοξές
(positiνe skewness) είναι περισσότερο κοινές κατανοµές στα
βιολογικά µεγέθη, ενώ εκείνες που εκτείνονται ακριβώς αντίθετα ονοµάζονται
αρνητικά λοξές (negatiνe skewness), και είναι πιο σπάνιες.

**Κανονική κατανομή**

Η κανονική κατανομή έχει τις ακόλουθες ιδιότητες:

* Η κατανομή είναι συμμετρική περί το μ, που είναι ο μέσος όλων των δυνατών
τιμών της Χ.
* Η επικρατούσα τιμή, η διάμεσος και ο μέσος ταυτίζονται λόγω συμμετρίας
της καμπύλης της κατανομής.
* Το εύρος των τιμών της Χ είναι το σύνολο R των πραγματικών αριθμών.
* Ο οριζόντιος άξονας είναι ασύμπτωτος της καμπύλης, όταν το x → ±∞.
* Το συνολικό εμβαδόν που περικλείεται από την καμπύλη και τον άξονα των x
είναι μοναδιαίο.
* Χρησιμοποιώντας τις ιδιότητες της κανονικής κατανομής και μία σειρά από τεχνικές, μπορούμε να υπολογίσουμε εμβαδά χωρίων μεταξύ της καμπύλης και δύο ευθειών x=α και x=β.
* Γραφικές παραστάσεις κατανομών συχνοτήτων ποιοτικών μεταβλητών
* Η πιο εύχρηστη μέθοδος γραφικής παράστασης κατανομής συχνοτήτων ποιοτικών και διατάξιμων χαρακτηριστικών είναι το ιστόγραμμα
* Ο οριζόντιος άξονας αναφέρεται στο εξεταζόμενο ποιοτικό χαρακτηριστικό και ο κάθετος στον αριθμό των παρατηρήσεων
* Λέγεται και διάγραμμα στηλών, όπου οι στήλες έχουν ορισμένο ύψος, αλλά το εύρος της βάσης των στηλών





**ΜΕΤΡΑ ΚΕΝΤΡΙΚΗ ΤΑΣΕΩΣ**

**Μέση τιμή**

* **Μέσος αριθμητικός ή Μέση τιμή (Mean)** ορίζεται ως το πηλίκο των τιμών της μεταβλητής δια το πλήθος των τιμών της.
* **Συμβολισμός: μ αν αναφερόμαστε σε πληθυσμό και *x όταν*** αναφερόμαστε σε δείγμα.

¯Χ=Σx/n

**Βασικά χαρακτηριστικά**

* Ορίζεται μονοσήμαντα δηλαδή σε κάθε σύνολο στοιχειών αντιστοιχεί ένας και μόνο αριθμητικός μέσος
* Έχει απλή ερμηνεία ως ο μέσος όλων των παρατηρήσεων
* Υπολογίζεται εύκολα
* Είναι αντιπροσωπευτικός του συνόλου των παρατηρήσεων αφού όλες υπεισέρχονται τον υπολογισμό του
* Πολλοί στατιστικοί έλεγχοι έχουν ως βάση τον αριθμητικό μέσο
* Επηρεάζεται πολύ από τις ακραίες παρατηρήσεις

**Διάμεσος**

* Είναι η τιμή εκείνη η οποία χωρίζει το σύνολο των τιμών της έτσι, ώστε το πλήθος των τιμών που είναι μικρότερες ή ίσες από τη διάμεσο να είναι ίσες με το πλήθος των τιμών της που είναι μικρότερες από αυτήν

* Εάν το πλήθος των τιμών είναι άρτιο τότε τη διάταξη υπάρχουν δύο κεντρικές παρατηρήσεις .ο διάμεσος ορίζεται ως ο μέσος όρος των δύο κεντρικών παρατηρήσεων
* Εάν το πλήθος των τιμών είναι περιττό τότε η διάμεσος τους είναι η κεντρική τιμή τους αν αυτές διαταχθούν κατά τάξη αύξοντος ή φθίνοντος μεγέθους

**Βασικά χαρακτηριστικά διάμεσου**

* Ορίζεται μονοσήμαντα δηλαδή σε κάθε σύνολο στοιχειών αντιστοιχεί ένας και μόνο διάμεσος
* Έχει απλή ερμηνεία αφού είναι τιμή που αφήνει αριστερά το 50% των μικρότερων τιμών και δεξιά της το άλλο 50% των μεγαλύτερων τιμών της
* Υπολογίζεται εύκολα αρκεί να διατάξουμε τις τιμές κατά τάξη μεγέθους
* Δεν επηρεάζεται από τις ακραίες τιμές
* Οι περισσότεροι στατιστικοί έλεγχοι βασίζονται στον αριθμητικό μέσο και δεν χρησιμοποιούν τη διάμεσο

**Επικρατούσα τιμή**

* Επικρατούσα τιμή είναι αυτή που εμφανίζεται τις περισσότερες φορές
* Οι επικρατούσες τιμές διακρίνονται εύκολα σε ένα πολύγωνο συχνοτήτων , σε ένα ιστόγραμμα
* είναι περιγραφικό μέτρο ιδιαίτερα χρήσιμο για την περιγραφή ποιοτικών στοιχείων, στα οποία ο μέσος και ο διάμεσος δεν ορίζονται

**Βασικά χαρακτηριστικά επικρατούσας τιμής**

* Δεν ορίζονται μονοσήμαντα αφού σε ένα σύνολο τιμών μπορούμε νε έχουμε έναν αριθμό από επικρατούσες τιμές ή και καμία
* Η επικρατούσα τιμή σε ένα τυχαίο δείγμα είναι η καλύτερη εκτιμήτρια της επικρατούσας τιμής στον πληθυσμό
* Ως μέτρο κεντρικής τάσεως, η επικρατούσα τιμή επηρεάζεται από την ασυμμετρία λιγότερο από τον μέσο και το διάμεσο
* Επηρεάζεται από τον τρόπο δειγματοληψίας και από τον τρόπο ομαδοποίησης των στοιχείων

**Σελίδες 15-18 βιβλίου** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)

**Μέτρα διασποράς και μεταβλητότητας**

* Εύρος μεταβολής
* Η τεταρτημοριακή απόκλιση
* Η διακύμανση και η σταθερή απόκλιση
* Το εύρος μεταβολής

Το εύρος μεταβολής των τιμών της Χ είναι η διαφορά της μικρότερης τιμής Xmin από την μεγαλύτερη Xmax και συμβολίζεται με **R**

R=Xmax - Xmin

**Η τεταρτημοριακή απόκλιση**

* Αν Q1 είναι το πρώτο τεταρτημόριο και Q3 είναι το τρίτο , τότε στο διάστημα Q3- Q1 το 50% των κεντρικών ή των πιο φυσιολογικών τιμών της Χ. Το υπόλοιπο 50% των τιμών της Χ βρίκεται έξω από το διάστημα αυτό και μάλιστα 25% είναι μικρότερες από το Q1 και το 25% είναι μεγαλύτερες από το Q3
* Η ποσότητα **Qτ= Q3- Q1/2** ονομάζεται **τεταρτημοριακή απόκλιση**

**Διακύμανση- σταθερή απόκλιση**

* Τα πλέον συχνά χρησιμοποιούμενα μέτρα διασποράς είναι η **διακύμανση ή διασπορά (variance) και η σταθερή απόκλιση (standard deviation).**
* Η **διακύμανση είναι ο μέσος αριθμητικός των τετραγώνων των** διαφορών των τιμών μιας μεταβλητής από το μέσο αριθμητικό της.
* Συμβολίζεται με σ2 όταν αναφερόμαστε σε πληθυσμό και με s2 για δείγμα.
* Επειδή η διακύμανση εκφράζεται μέσω του τετραγώνου της μεταβλητής, γι’ αυτό παίρνουμε τη θετική τετραγωνική ρίζα της διακύμανσης που ονομάζεται **σταθερή απόκλιση και η οποία** εκφράζεται με τις ίδιες μονάδες μέτρησης με τη μονάδα μέτρησης της μεταβλητής.

**Σελίδες 21-24 βιβλίου** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)

**ΑΝΑΛΥΣΗ ΠΟΣΟΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

**Πιθανό σφάλμα**

* Η μέση τιμή ενός δείγματος ποσοτικών παρατηρήσεων αποτελεί κατά προσέγγιση εκτίμηση της πραγματικής μέσης τιμής του γενικότερου συνόλου από το οποίο προήλθε το δείγμα
* Εξαιτίας των αναπόφευκτων τυχαίων διακυμάνσεων η πραγματική μέση τιμή ενδέχεται να είναι μικρότερη ή μεγαλύτερη από εκείνη που βρέθηκε
* Μέτρο της ενδεχόμενης απόστασης της μέσης τιμής του δείγματος από την αντίστοιχη πραγματική είναι το πιθανό σφάλμα της μέσης τιμής του δείγματος (πιθανό σφάλμα-πρότυπο σφάλμα=Standard Error=SE ).
* Το πιθανό σφάλμα προκύπτει διαιρώντας τη σταθερή απόκλιση της κατανομής με την τετραγωνική ρίζα του αριθμού των παρατηρήσεων του δείγματος

**SE=S/√n**

* Όταν το πιθανό σφάλμα της μέσης τιμής ενός δείγματος είναι μικρό, τότε η ακρίβεια της εκτίμησης της αντίστοιχης πραγματικής μέσης τιμής είναι μεγάλη και αντίστροφα
* Από τον ορισμό προκύπτει ότι όταν ο αριθμός των παρατηρήσεων n είναι μεγάλος , το πιθανό σφάλμα είναι μικρό και αντίστροφα

**Πιθανό σφάλμα**

* Αν Δ είναι η διαφορά της μέσης τιμής Χ από την ορισμένη σταθερή τιμή Α, τότε η διαφορά αυτή Δ=Α-x όταν η τιμή του κριτηρίου t (t –value) είναι μεγαλύτερη σε απόλυτη τιμή από την οριακή τιμή που επιλέχτηκε

**Δ/SE=Δ/S/√n**

Αν συγκρίνονται δύο μέσες τιμές οι οποίες προέρχονται από δύο δείγματα και υπόκεινται και οι δύο σε τυχαία διακύμανση, τότε η διαφορά τους διαιρείται με το πιθανό σφάλμα της διαφοράς . Το μέγεθος αυτό συμβολίζεται ως

**SEΔ=√SE12+ SE22**

**t-value**

* Η τιμή του κριτηρίου στην περίπτωση αυτή είναι:

**t-value =Δ/ SEΔ**

* Όταν συγκρίνονται δύο μέσες τιμές, οι βαθμοί ελευθερίας για την αξιολόγηση του κριτηρίου t είναι: n1+n2-2
* Προϋποθέσεις εφαρμογής του κριτηρίου είναι η κατά προσέγγιση κανονικότητα των κατανομών και η κατά προσέγγιση ισότητα των σταθερών αποκλίσεων των κατανομών αυτών

**Παράδειγμα εφαρμογής της δοκιμασίας t-test**

* Δίνονται οι κατανομές των συχνοτήτων των τιμών της χοληστερόλης και του ορού 35 υγιών και 29 ατόμων με έμφραγμα μυοκαρδίου. Να βρεθεί η μέση τιμή, η σταθερή απόκλιση και το πιθανό σφάλμα της μέσης τιμής κάθε μιας από τις δύο κατανομές και να αξιολογηθεί στατιστικά η διαφορά των δύο μέσων τιμών και η δοκιμασία t-test

**Σελίδες 31-39 βιβλίου**

**Όρια αξιοπιστίας και διάστημα εμπιστοσύνης**

* Το διάστημα που καθορίζεται από την προσθαφαίρεση 1,96\*SE, δηλαδή περίπου δύο πιθανών σφαλμάτων στην ευρεθείσα τιμή , περιλαμβάνει την αντίστοιχη πραγματική με πιθανότητα 95%
* Το διάστημα x±1,96\*SE λέγεται το κατά 95% διάστημα εμπιστοσύνης και τα όρια του αντίστοιχα τα κατά 95% όρια αξιοπιστίας
* Όταν αναγράφεται απλά το διάστημα αξιοπιστίας ή όρια αξιοπιστίας υπονοούνται τα κατά 95% αντίστοιχα δεδομένου ότι η κατά 95% εμπιστοσύνη (P˜0,05) αποτελεί το συμβατικό αποδεκτό όριο στατιστικής σημαντικότητας για τους περισσότερους συγγραφείς και ερευνητές

**ΠΡΟΥΠΟΘΕΣΕΙΣ ΕΦΑΡΜΟΓΗΣ Τ-ΤΕΣΤ ΕΙΔΗ ΔΕΙΓΜΑΤΩΝ**

Προϋποθέσεις εφαρμογής t-test

Για την εφαρμογή της δοκιμασίας t-test απαιτείται η εκπλήρωση κάποιων προϋποθέσεων όπως

* Η κατά προσέγγιση κανονικότητα των κατανομών που συγκρίνονται και η
* Κατά προσέγγιση ισότητα των σταθερών αποκλίσεων των κατανομών αυτών

Αντίθετα η δοκιμασία t-test δεν είναι εφαρμόσιμη όταν :

* Οι κατανομές είναι από τη φύση τους πολυκόρυφες
* Οι κατανομές είναι ασύμμετρες και ο ολικός αριθμός των παρατηρήσεων είναι μικρός
* ο αριθμός των παρατηρήσεων είναι μικρός και η μορφή των κατανομών άγνωστη
* Ο αριθμός των παρατηρήσεων είναι σχετικά μεγάλος και οι κατανομές κωδωνοειδείς αλλά εξαιρετικά ασύμμετρες

**Σε όσες περιπτώσεις δεν επιτρέπεται η εφαρμογή της δοκιμασίας t-test σε ποσοτικά χαρακτηριστικά, μπορεί να εφαρμοστούν οι μη παραμετρικές δοκιμασίες**

**Παραμετρικές-μη παραμετρικές δοκιμασίες**

* Οι παραμετρικές δοκιμασίες π.χ. t-testέχουν προϋποθέσεις για την κατανομή των μεταβλητών στους πληθυσμούς από τους οποίους προέρχονται ενώ οι μη –παραμετρικές μέθοδοι (όπως το test Mann-Wintey) δεν έχουν προϋποθέσεις για την κατανομή των μεταβλητών
* Για κάθε παραμετρικό test υπάρχει ένα μη παραμετρικό ανάλογο του, το οποίο είναι πιο γενικό με την έννοια ότι μπορεί να χρησιμοποιηθεί σε περισσότερες περιπτώσεις
* Π.χ. για μια ποιοτική μεταβλητή μπορεί να χρησιμοποιηθεί το test Mann-Wintey ενώ το t-test μπορεί να χρησιμοποιηθεί μόνο για ποσοτικές μεταβλητές που ακολουθούν κανονική κατανομή
* Τα παραμετρικά test είναι ακριβέστερα από τα αντίστοιχα μη παραμετρικά (δίνουν δηλαδή μεγαλύτερης ακρίβειας αποτελέσματα)

**Παραμετρικές-μη παραμετρικές δοκιμασίες**

* Μειονεκτήματα των μη παραμετρικών δοκιμασιών είναι η μειωμένη ισχύς τους και το γεγονός ότι δεν επιδέχονται όλες τις εφαρμογές και τις γενικεύσεις τις οποίες επιδέχονται οι παραμετρικές δοκιμασίες
* Μεταξύ των μη παραμετρικών δοκιμασιών περιλαμβάνονται και εκείνες οι οποίες εφαρμόζονται σε ποιοτικά χαρακτηριστικά

**Διάκριση δειγμάτων**

* Δύο ή περισσότερα δείγματα χαρακτηρίζονται ως **ανεξάρτητα** όταν η πραγματοποίηση των μετρήσεων στο ένα δεν εξαρτάται από την πραγματοποίηση στο άλλο
* Αν μετρήσουμε την ίδια μεταβλητή στα ίδια υποκείμενα σε διαφορετικές χρονικές στιγμές (π.χ. αρτηριακή πίεση πριν και μετά από 10 λεπτά τρέξιμο) ή καταστάσεις τα δύο δείγματα που προκύπτουν είναι μεταξύ τους συνδεδεμένα και λέγονται **εξαρτημένα**

**Μεταβλητή**

* Κάθε μεταβλητή μπορεί να είναι ή να χρησιμοποιηθεί ως ανεξάρτητη ή εξαρτημένη στην ανάλυση μας
* Μια μεταβλητή είναι εξαρτημένη όταν τα συμπεράσματα για αυτήν προκύπτουν από τη μελέτη μιας ή περισσότερων άλλων μεταβλητών (ανεξάρτητων) που επίσης έχουν καταγραφεί στα ίδια υποκείμενα
* Είναι συνηθισμένο το γεγονός να καθορίζουμε συγκεκριμένες τιμές για κάποια ανεξάρτητη μεταβλητή (όπως για παράδειγμα το φύλο του υποκειμένου, η δοσολογία ενός φαρμάκου) και να μελετάμε τη συμπεριφορά των εξαρτημένων μεταβλητών ως προς αυτή την τιμή

**Η Δοκιμασία t-test για παρατηρήσεις κατά ζεύγη**

* Οι παρατηρήσεις των δύο συγκρινόμενων ομάδων εμφανίζουν ατομική αντιστοιχία και οι συγκρίσεις μπορούν να γίνονται ανά ζεύγη π.χ εάν συγκρίνεται η αποτελεσματικότητα δύο αναλγητικών είναι σκόπιμο να χορηγούνται και τα δύο αναλγητικά σε διαφορετικό χρόνο βέβαια και να εξετάζεται η διαφορά αποτελεσματικότητας των φαρμάκων

**Η Δοκιμασία t-test για παρατηρήσεις κατά ζεύγη**

* Σε 12 άτομα χορηγήθηκε φάρμακο το οποίο θεωρείται ότι έχει υποτασικές ιδιότητες. Η συστολική πίεση των ατόμων αυτών πριν και μετά τη χορήγηση του φαρμάκου αναγράφεται στον πίνακα. Ζητείται να αξιολογηθεί στατιστικά η ενδεχόμενη υποτασική δραστηριότητα του φαρμάκου

**Σελίδες 44-48 του βιβλίου** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)

**Έλεγχοι υποθέσεων**

* Υπάρχει ένα φάρμακο με το οποίο εάν γίνει θεραπεία, εμφανίζονται τα πρώτα αποτελέσματα σε 10 ημέρες. Ανακαλύπτεται ένα καινούργιο φάρμακο για την ασθένεια και η κατασκευαστική εταιρεία ισχυρίζεται ότι φέρνει αποτελέσματα σε συντομότερο χρονικό διάστημα
* Αν λοιπόν για , ο μέσος όρος για το πρώτο είναι μ=10 ημέρες , χρειάζεται να ελεγχθεί για το δεύτερο φάρμακο ο μέσος χρόνος είναι μικρότερος (μ<10)

**Έλεγχοι υποθέσεων**

* Η διαδικασία που ακολουθείται λέγεται έλεγχος υπόθεσης
* Η υπόθεση ότι μ=10 λέγεται μηδενική υπόθεση (Ηο)
* Η υπόθεση ότι μ<10 που ελέγχεται ως προς την Ηο λέγεται εναλλακτική υπόθεση
* Σε κάθε υπόθεση μηδέν (Ηο) αντιστοιχεί η λεγόμενη εναλλακτική υπόθεση (Η1) η οποία αποτελεί τη λογική άρνηση της Ηο
* Η απόφαση να γίνει δεκτή η Ηο ή να απορριφθεί στηρίζεται σε ένα στατιστικό που υπολογίζεται από τα δεδομένα του δείγματος

**Διαδικασία ελέγχου υποθέσεων**

* Ορίζεται μηδενική υπόθεση Ηο
* Ορίζεται η εναλλακτική υπόθεση Η1
* Ελέγχεται το κατάλληλο στατιστικό του ελέγχου
* Υπολογίζεται το στατιστικό από τα δεδομένα του δείγματος
* Συγκρίνεται η τιμή του στατιστικού με την κριτική τιμή την οποία παίρνουμε από τους στατιστικούς πίνακες
* **Αν η τιμή του στατιστικού είναι μικρότερη από την κριτική τιμή, η μηδενική υπόθεση δεν μπορεί να απορριφθεί και γίνεται δεκτή**
* **Αν η τιμή του στατιστικού είναι μεγαλύτερη από την κριτική τιμή , η μηδενική υπόθεση απορρίπτεται και γίνεται δεκτή η εναλλακτική**

**Παράδειγμα 1**

* Μια μελέτη διεξήχθη για να καθορισθεί πόσο το κάπνισμα της μητέρας έχει κάποια επίδραση την περιεκτικότητα των μετάλλων και στα οστά των υγιών νεογέννητων. Ένα δείγμα 77 βρεφών , των οποίων οι μητέρες κάπνιζαν κατά τη διάρκεια της εγκυμοσύνης έχει μέση περιεκτικότητα 0,098 gr/cm3 και τυπική απόκλιση 0,026 gr/cm3.ένα δείγμα 161 βρεφών , των οποίων οι μητέρες δεν κάπνιζαν έχει μέση περιεκτικότητα 0,095 gr/cm3 και τυπική απόκλιση 0 ,025 gr/cm3. να κάνετε τον έλεγχο των υποθέσεων για την ισότητα των τιμών αυτών

**ΛΥΣΗ**

**ΔΕΔΟΜΕΝΑ**

*ΔΕΙΓΜΑ 1*

N1=77

Μέση τιμή =0,098

S12=0,0262

*ΔΕΙΓΜΑ 2*

N2=161

Μέση τιμή =0,095

S12=0,0252

**¨**

**ΥΠΟΘΕΣΕΙΣ**

Η0: μ1=μ2

Η1: μ1=μ2

**ΣΤΑΤΙΣΤΙΚΟ**

Επειδή τα δείγματα είναι μεγάλα >30, το στατιστικό είναι:

t= 0,843109

**KΡΙΤΙΚΗ ΤΙΜΗ: 1,96**

**ΑΠΟΦΑΣΗ:** επειδή 0,843109<1,96, η μηδενική υπόθεση δεν μπορεί να απορριφθεί και γίνεται δεκτή. Επομένως οι μέσες τιμές των δύο δειγμάτων δεν διαφέρουν στατιτικά

**Σφάλματα**

Τα αποτελέσματα μιας στατιστικής δοκιμασίας ενδέχεται να οδηγήσουν σε δύο είδη σφαλμάτων

* **Σφάλμα τύπου Ι** (απόρριψη μηδενικής υπόθεσης όταν αυτή ισχύει)
* **Σφάλμα τύπου ΙΙ** (μη απόρριψη της μηδενικής αποδοχής όταν αυτή δεν ισχύει)

**ΑΝΑΛΥΣΗ ΠΟΙΟΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

* Διαξονική ταξινόμηση των ποιοτικών παρατηρήσεων
* Όταν οι παρατηρήσεις έχουν ποιοτικό χαρακτήρα δεν είναι δυνατές μετρήσεις και το t-test δεν είναι εφαρμόσιμο
* Τα μέλη ενός συνόλου μπορεί να ταξινομηθούν με δύο ή περισσότερους τρόπους είτε διαδοχικά είτε ταυτόχρονα. Η ταυτόχρονη ταξινόμηση με δύο τρόπους καλείται ταξινόμηση δύο διευθύνσεων ή διαξονική
* ( two-way classification)
* H διαξονική διάταξη αποτελεί προϋπόθεση εφαρμογής της στατιστικής δοκιμασίας Χ2 για τον έλεγχο της συσχέτισης των ποιοτικών χαρακτηριστικών

**δοκιμασία Χ2**

* Υπάρχει στατιστικά σημαντική σχέση μεταξύ επαγγέλματος και ανάπτυξης καρκίνου της ουροδόχου κύστης;
* Σε κάθε πίνακα υπάρχουν τόσες στήλες όσες και οι κατηγορίες της οριζόντιας ταξινόμησης και τόσες σειρές όσες και οι κατηγορίες της κάθετης ταξινόμησης.
* Κ : ο αριθμός των στηλών
* L: ο αριθμός των σειρών
* Σε κάθε διαξονικό πίνακα ο αριθμός των κελιών ή συνδυασμός κατηγοριών που υπάρχουν ισούται με Κ\* L
* Σε καθένα από τα Κ\* L κελιά αντιστοιχεί μια παρατηρηθείασα συχνότητα που παριστάνεται με το σύμβολο Ο
* Ο υπολογισμός των αναμενόμενων συχνοτήτων που παριστάνεται με το σύμβολο Ε υπολογίζεται από τον τύπο
* Ε=(αντίστοιχο οριζόντιο σύνολο) \* (αντίστοιχο κάθετο σύνολο)/(γενικό σύνολο)
* Οι βαθμοί ελευθερίας των διαξονικών πινάκων είναι
* (Κ-1)\* (L -1)
* Δοκιμασία Χ2  σε δύο δυαδικές ταξινομήσεις
* Χ2=(ad-bc)2\*n/(a+b)\*(c+d)\*(a+c)\* (b+d)
* Ένας βαθμός ελευθερίας

**Τετράπτυχος πίνακας**

|  |
| --- |
| Χαρακτηριστικό πρώτο |
| Χαρακτηριστικό δεύτερο  |  | Παρόν  | Απόν  | Σύνολο  |
| Παρόν | a | b | a+b |
|  | Απόν | c | d | c+d |
|  | Σύνολο | a+c | b+d | a+b+c+d=n |

**Σελίδες 55-60 και 64-68 βιβλίου** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)

**ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΕΞΑΡΤΗΣΗ**

* Η συσχέτιση δύο μεγεθών (και η εξάρτηση) θεωρείται θετική όταν σε υψηλές τιμές του ενός ποσοτικού χαρακτηριστικού αντιστοιχούν συνήθως υψηλές τιμές του άλλου, αρνητική δε όταν σε υψηλές τιμές του ενός αντιστοιχούν συνήθως χαμηλές τιμές του άλλου
* Ο υπολογισμός του παραμετρικού συντελεστή συσχέτισης κατά Pearson προϋποθέτει ότι, οι κατανομές συχνοτήτων και των δύο μεταβλητών είναι ακριβώς ή κατά προσέγγιση κανονικές.
* Αντίθετα, η κοινή μέθοδος διερεύνησης της στατιστικής εξάρτησης προϋποθέτει κανονική κατανομή συχνοτήτων (ακριβώς ή κατά προσέγγιση)μόνο της εξαρτημένης μεταβλητής ενώ οι τιμές της ανεξάρτητης μεταβλητής μπορεί να προσδιορίζονται αυθαίρετα ή να εμφανίζουν κατανομή συχνοτήτων οποιασδήποτε μορφής.
* Όταν η κατανομή συχνοτήτων μιας ή και των δύο μεταβλητών δεν είναι κανονική η συσχέτιση ανάμεσα τους μπορεί να διερευνηθεί με τις μη παραμετρικές μεθόδους στατιστικής εξάρτησης

**Διάγραμμα συσχέτισης**

* Το διάγραμμα συσχέτισης λέγεται επίσης και στικτόγραμμα αποτελεί απλή αλλά και πολλή χρήσιμη μέθοδος μελέτης της συσχέτισης των δύο μεταβλητών
* Αν πρόκειται για πρόβλημα εξάρτησης τότε η ανεξάρτητη μεταβλητή τοποθετείται συνήθως στον οριζόντιο άξονα (τον άξονα των χ) και η εξαρτημένη στον κάθετο άξονα (τον άξονα των Y). Και οι δύο άξονες μπορούν να αρχίσουν από οποιαδήποτε τιμή και όχι υποχρεωτικά από το μηδέν

**Η μέθοδος διερεύνησης της στατιστικής συσχέτισης δύο ποσοτικών χαρακτηριστικών**

* Η διερεύνηση της συσχέτισης δύο ποσοτικών χαρακτηριστικών γίνεται συνήθως με τον προσδιορισμό του (παραμετρικού κατά pearson) συντελεστής συσχέτισης
* Ο συντελεστής συσχέτισης συμβολίζεται με το σύμβολο r και αποτελεί μέτρο του βαθμού συσχέτισης των δύο μεταβλητών
* Αν η κατανομή τω τιμών μιας μεταβλητής είναι κανονική αλλά οι τιμές της άλλης έχουν επιλεγεί αυθαίρετα ή κατανέμονται κατά τρόπο μη κανονικό ο συντελεστής συσχέτισης μπορεί να υπολογιστεί αλλά η σημασιολόγηση του είναι δυσκολότερη
* Αν και οι δύο μεταβλητές δεν κατανέμονται κανονικά ο παραμετρικός συντελεστής συσχέτισης δεν μπορεί να εφαρμοστεί και χρησιμοποιείται μη παραμετρική μεθοδολογία

**Ιδιότητες συντελεστή συσχέτισης**

* Ο συντελεστής συσχέτισης αποτελεί μέτρο της ευθύγραμμης συσχέτισης
* Ο συντελεστής συσχέτισης είναι καθαρός αριθμός δηλαδή στερείται φυσικών διαστάσεων και μονάδων
* Ο συντελεστής συσχέτισης μπορεί να παίρνει τιμές από -1 έως +1. η τιμή -1 δηλώνει ότι υπάρχει απόλυτη αρνητική ευθύγραμμη συσχέτιση , η τιμή +1 ότι υπάρχει απόλυτα θετική ευθύγραμμη συσχέτιση

**Σελ 119-125 βιβλίου** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)

**ΕΞΑΡΤΗΣΗ ΚΑΙ ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΔΟΚΙΜΑΣΙΕΣ**

**Πολλαπλή γραμμική εξάρτηση**

* Το αντικείμενο της ανάλυσης εστιάζεται τον τρόπο με τον οποίο μια μεταβλητή (η εξαρτημένη) επηρεάζεται ή εξαρτάται από πολλές άλλες (τις ανεξάρτητες).
* Οι ανεξάρτητες μεταβλητές μπορεί να είναι ποσοτικές, διατάξιμες ή ποιοτικές
* Αν η εξαρτημένη μεταβλητή είναι ποσοτική τότε ενδείκνυται η μέθοδος της πολλαπλής γραμμικής εξάρτησης (multiple linear regression)
* Πολλαπλή γραμμική εξάρτηση

H πολλαπλή εξάρτηση εκφράζεται με τη σχέση:

ŷ=a+b1x1+……+bkxk

Όπου y είναι η εξαρτημένη μεταβλητή, ŷ η προβλεπόμενη τιμή της, x1, x2, x3…. xk είναι οι ανεξάρτητες μεταβλητές και α η σταθερά της εξίσωσης. Οι υπολογιζόμενες ποσότητες b1, b2, b3….. bk είναι οι συντελεστές μερικής εξάρτησης

* Προϋποθέσεις εφαρμογής της πολλαπλής γραμμικής εξάρτησης
* Η εξαρτημένη μεταβλητή θα πρέπει να έχει κανονική κατανομή για κάθε συνδυασμό τιμών των ανεξάρτητων μεταβλητών

**Λογαριθμιστική εξάρτηση**

* Στην περίπτωση που η εξαρτημένη μεταβλητή είναι ποιοτική με δύο πιθανά αποτελέσματα και επομένως δεν είναι δυνατή η εφαρμογή της πολλαπλής γραμμικής εξάρτησης συνήθως εφαρμόζεται η λογαριθμιστική εξάρτηση
* Η εξαρτημένη μεταβλητή κωδικοποιείται με δύο τιμές: με 0 αν δεν υπάρχει η κατάσταση που μελετάται και με 1 εάν υπάρχει. Η μέση τιμή της εξαρτημένης μεταβλητής είναι η αναλογία των ατόμων που έχουν την κατάσταση.

**Παραμετρικές δοκιμασίες**

* **Παραμετρικές** δοκιμασίες είναι εκείνες που εφαρμόζονται σε ποσοτικά δεδομένα ή με κάποιον άλλο τρόπο που επαρκώς προσδιορίζεται με μια σειρά παραμέτρων (όπως η μέση τιμή και η σταθερή απόκλιση)
* Τέτοιες δοκιμασίες είναι το t-test και ο παραμετρικός συντελεστής συσχέτισης του Pearson και οι διάφορες τεχνικές ανάλυσης της μεταβλητότητας
* **Μη παραμετρικές δοκιμασίες** εννοούνται οι δοκιμασίες οι οποίες εφαρμόζονται σε ποσοτικά μεγέθη που δεν κατανέμονται κανονικά καθώς και σε ιεραρχικώς διατάξιμα χαρακτηριστικά

**Πλεονεκτήματα μη παραμετρικών δοκιμασιών**

* Είναι δυνατό να εφαρμοστούν σε ποσοτικά χαρακτηριστικά των οποίων οι κατανομές είναι έκδηλα μη κανονικές ανεξάρτητα από τον αριθμό των παρατηρήσεων
* Είναι δυνατό να εφαρμοστούν σε ποσοτικά χαρακτηριστικά όταν ο αριθμός των παρατηρήσεων είναι περιορισμένος και οι κατανομές είναι άγνωστες ή φαίνονται να εκτρέπονται από τη μορφή της κανονικής κατανομής
* Οι μη παραμετρικές μέθοδοι είναι δυνατό να εφαρμοστούν σε ιεραρχικώς διατάξιμα χαρακτηριστικά.
* Οι μη παραμετρικές δοκιμασίες είναι εξαιρετικά εύχρηστες και πολύ απλούστερες από τις παραμετρικές όσον αφορά τους απαιτούμενους αριθμητικούς υπολογισμούς

**Μειονεκτήματα μη παραμετρικών δοκιμασιών**

* Όταν οι παραμετρικές δοκιμασίες είναι δυνατό να εφαρμοστούν τότε αυτές διαθέτουν μεγαλύτερη ισχύ από εκείνη των αντίστοιχων μη παραμετρικών
* Οι μη παραμετρικές μέθοδοι δεν είναι δυνατόν να εφαρμοστούν σε σύνθετες στατιστικές αναλύσεις. Επιπλέον, ο υπολογισμός των ορίων αξιοπιστίας μιας διαφοράς σε μη παραμετρικές δοκιμασίες είναι δύσκολος
* Οι κυριότερες μη παραμετρικές δοκιμασίες
* Η δοκιμασία των σημείων η οποία αντιστοιχεί η οποία αντιστοιχεί στην παραμετρική δοκιμασία κατά ζεύγη t-test
* η δοκιμασία του Wilcoxon για παρατηρήσεις κατά ζεύγη η οποία αντιστοιχεί επίσης στην παραμετρική επίσης την παραμετρική δοκιμασία κατά ζεύγη t-test
* Η δοκιμασία του Wilcoxon για παρατηρήσεις χωρίς αντιστοιχία η οποία αντιστοιχεί στην παραμετρική δοκιμασία t-test
* Ο βαθμός συσχέτισης μεταξύ ποσοτικών χαρακτηριστικών που δεν κατανέμονται κανονικά ή μεταξύ χαρακτηριστικών ιεραρχικά διατάξιμων μπορεί να ελεγχθεί με το μη παραμετρικό συντελεστή συσχέτισης Spearman ο οποίος αντιστοιχεί στον συντελεστή συσχέτισης Pearson

**Σελ 155-159, 161,165,168** (ΒΙΟΣΤΑΤΙΤΙΚΗ, ΤΡΙΧΟΠΟΥΛΟΣ-ΤΖΩΝΟΥ-ΚΑΤΣΟΥΓΙΑΝΝΗ)