Σημειώσεις για το μάθημα

ΣΤΑΤΙΣΤΙΚΗ – ΠΟΣΟΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Παπάνα Αγγελική

http://users.auth.gr/~agpapana/StatLogistics

E-mail: papanagel@yahoo.gr, agpapana@gen.auth.gr

Α.Τ.Ε.Ι. Θεσσαλονίκης
ΠΑΡΑΡΤΗΜΑ ΚΑΤΕΡΙΝΗΣ
Τμήμα Τυποποίησης και Διακίνησης Προϊόντων (Logistics)

Κεφάλαιο 1

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

1.1 Εισαγωγή

Η **στατιστική** είναι ένας κλάδος που βοηθάει στην μελέτη και κατανόηση φαινομένων ή ιδιοτήτων πολυπληθών ομάδων. Η εφαρμογή της εκτείνεται σε πολλούς κλάδους της ανθρώπινης δραστηριότητας, π.χ. πολιτική, οικονομία. Η **περιγραφική στατιστική** είναι ο κλάδος της στατιστικής που ασχολείται με την οργάνωση, συλλογή και παρουσίαση ενός συνόλου δεδομένων.

Αν έχουμε ένα σύνολο και θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους, το σύνολο αυτό το ονομάζουμε **πληθυσμό**. Τα στοιχεία του πληθυσμού τα ονομάζουμε **μονάδες** ή **άτομα**. **Απογραφή** είναι η μέθοδος συλλογής αντικειμένων κατά την οποία εξετάζουμε όλα τα στοιχεία του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει. Όταν η απογραφή είναι δύσκολη ή ασύμφορη, τότε αντί να εξεταστεί όλος ο πληθυσμός, εξετάζεται ένα υποσύνολο του, το οποίο ονομάζεται **δείγμα.** Το δείγμα πρέπει να είναι *τυχαίο* και *αντιπροσωπευτικό* για την εξαγωγή αξιόπιστων συμπερασμάτων. **Δειγματοληψία** ονομάζουμε την τεχνική με την οποία γίνεται η επιλογή του δείγματος και χρησιμοποιείται.

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό τα ονομάζουμε **μεταβλητές.** Τις μεταβλητές τις συμβολίζουμε με κεφαλαία γράμματα (συνήθως **X**). Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές** της μεταβλητής και συμβολίζονται $\mathbf{x_i}$. Η εξέταση ενός δείγματος ως προς κάποιο χαρακτηριστικό των ατόμων του, μπορεί να γίνει για ποσοτικά ή ποιοτικά χαρακτηριστικά. Δηλαδή οι μεταβλητές χωρίζονται σε δύο κατηγορίες:

Ποιοτικές ή κατηγορικές λέγονται οι μεταβλητές των οποίων οι τιμές μπορούν να ταξινομηθούν σε κατηγορίες και δεν εκφράζουν απαραίτητα κάτι το μετρήσιμο (π.χ. ομάδα). **Ποσοτικές** λέγονται οι μεταβλητές οι οποίες παίρνουν μόνο αριθμητικές τιμές και μπορούν να ταξινομηθούν σε **διακριτές** (π.χ. αριθμός παιδιών ανά οικογένεια) ή **συνεχείς** (π.χ. βάρος).

1.1 Παρουσίαση στατιστικών δεδομένων σε πίνακες συχνοτήτων

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 **X** που παίρνει τιμές $x_1, x_2,..., x_κ$, όπου $κ < ν$. Ο φυσικός αριθμός $ν_i$, $i=1,...,κ$ που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της εξεταζόμενης μεταβλητής στο σύνολο των παρατηρήσεων ονομάζεται συχνότητα. Αν $ν_1, ν_2, \dots, ν_κ$ οι συχνότητες που αντιστοιχούν στις τιμές $x_1, x_2, \dots, x_κ$ της μεταβλητής X, τότε:

$$\boxed{\boldsymbol{ν_1 + ν_2 + ... + ν_κ = ν}}$$

Αν διαιρέσουμε την συχνότητα ν_i, $i = 1, 2, \ldots, \kappa$, με το μέγεθος του δείγματος προκύπτει η σχετική συχνότητα f_i της τιμής x_i:

$$\boxed{f_i = \frac{\nu_i}{\nu}}$$

Για την σχετική συχνότητα ισχύουν οι σχέσεις:

$\boxed{0 \leq f_i \leq 1}$ για $i = 1, 2, \ldots, \kappa$ και

$\boxed{f_1 + f_2 + \ldots + f_\kappa = 1}$

Σημείωση: Συνήθως τη σχετική συχνότητα την εκφράζουμε επί τις εκατό οπότε συμβολίζουμε $f_i\%$ και ισχύει: $f_i\% = 100\, f_i$.

Όταν η μεταβλητή είναι ποσοτική, ορίζουμε την **αθροιστική συχνότητα** N_i, η οποία εκφράζει το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i. Αν $x_1 < x_2 < \ldots < x_\kappa$ ισχύει:

$$\boxed{N_i = \nu_1 + \nu_2 + \ldots + \nu_i}$$

Ισχύουν: $\nu_1 = N_1, \nu_2 = N_2 - \nu_1, \ldots, \nu_\kappa = N_\kappa - N_{\kappa-1}$.

Ομοίως, ορίζεται η **αθροιστική σχετική συχνότητα** F_i, η οποία εκφράζει το ποσοστό των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i:

$$\boxed{F_i = f_1 + f_2 + \ldots + f_i}$$

Ισχύουν: $f_1 = F_1, f_2 = F_2 - F_1, \ldots, f_\kappa = F_\kappa - F_{\kappa-1}$.

Συνήθως χρησιμοποιείται η σχετική αθροιστική συχνότητα %: $F_i\% = 100\%\, F_i$.

Τα δεδομένα παρουσιάζονται σύντομα και με σαφήνεια σε **πίνακες συχνοτήτων**, δηλαδή πίνακες όπου αναγράφονται σε στήλες οι τιμές της μεταβλητής, οι συχνότητες εμφάνισης κάθε τιμής της μεταβλητής, κτ.λ.

π.χ. Έστω ρωτήθηκαν 30 μαθητές μιας τάξης ενός λυκείου σχετικά με την προτίμηση τους ως προς τις ποδοσφαιρικές ομάδες της Θεσσαλονίκης. Τα δεδομένα παρουσιάζονται στον παρκάτω πίνακα συχνοτήτων.

Πίνακας 1

Ομάδα x_i	**Συχνότητα** ν_i	**Σχετική συχνότητα** f_i	**Ποσοστά** $f_i\%$
ΠΑΟΚ	6	0,2	20
ΑΡΗΣ	12	0,4	40
ΗΡΑΚΛΗΣ	9	0,3	30
ΑΠΟΛΛΩΝ ΚΑΛΑΜΑΡΙΑΣ	3	0,1	10
Σύνολο	ν=30	**1**	**100**

Ομαδοποίηση

Όταν τα δεδομένα είανι αριθμητικά και το πλήθος των τιμών της εξεταζόμενης μεταβλητής είναι μεγάλος ή η μεταβλητή είναι συνεχής, τότε ταξινομούμε τα δεδομένα σε ένα πλήθος ομάδων που τις ονομάζουμε **κλάσεις διαστημάτων** και η διαδικασία ταξινόμησης τών δεδομένων σε κλάσεις λέγεται **ομαδοποίηση**. Οι κλάσεις είναι διαστήματα της μορφής [α, β) (διάστημα κλειστό αριστερά, ανοιχτό δεξιά). Τα άκρα του διαστήματος τα λέμε **άκρα** της κλάσης.

Κέντρο μιας κλάσης [α, β) ονομάζεται ο αριθμός **(α+β)/2**.

Πλάτος μιας κλάσης [α, β) ονομάζεται ο αριθμός **c = β - α**.

Εύρος του δείγματος είναι η διαφορά της μικρότερης από τη μεγαλύτερη παρατήρηση του δείγματος

$$R = x_{max} - x_{min}.$$

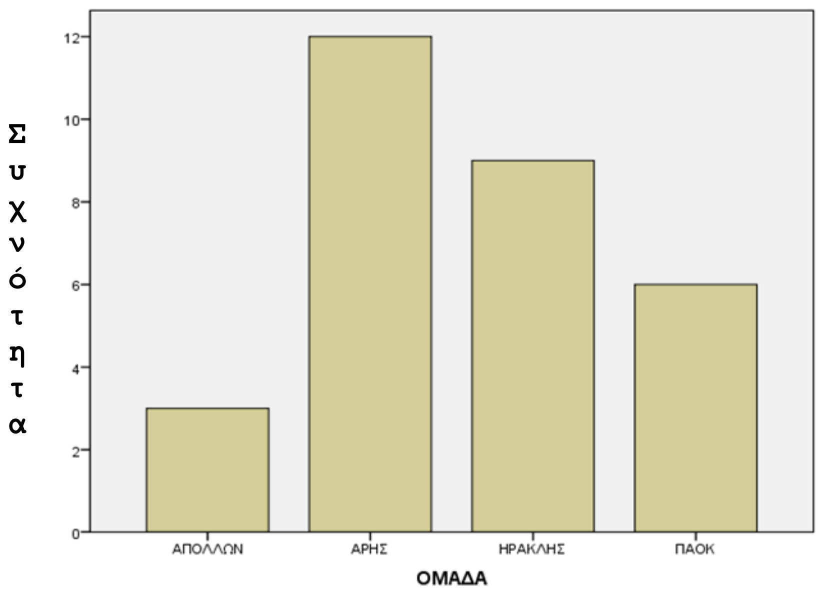
Αν **κ** συμβολίσουμε το πλήθος των κλάσεων, τότε ισχύει: $c = \frac{R}{\kappa}$ ή $\kappa = \frac{R}{c}$.

1.3 Γραφικές μέθοδοι για την παρουσίαση των δεδομένων

ΡΑΒΔΟΓΡΑΜΜΑ

Τα δεδομένα ενός πίνακα συχνοτήτων μπορούν να παρασταθούν γραφικά με ένα ραβδόγραμμα, όπου κάθε ράβδος παρουσιάζει τη συχνότητα (ή αθροιστική συχνότητα ή σχετική συχνότητα ή αθροιστική σχετική συχνότητα) για κάθε τιμή της μεταβλητής.

π.χ. Τα δεδομένα του **Πίνακα 1** μπορούν να παρασταθούν γραφικά με το παρακάτω ραβδόγραμμα συχντοτήτων.

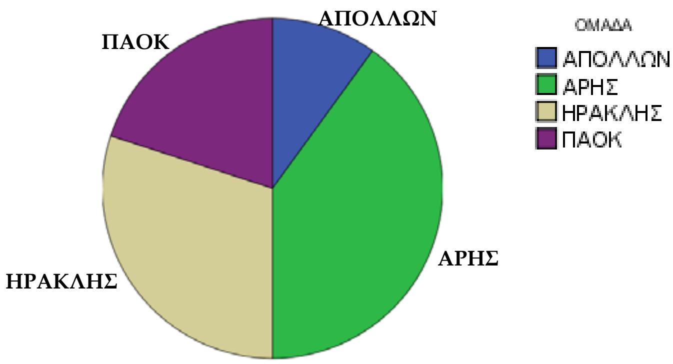


ΚΥΚΛΙΚΟ ΔΙΑΓΡΑΜΜΑ

Χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών, όταν οι διάφορες τιμές της μεταβλητής είναι λίγες. Το κυκλικό διάγραμμα περιγράφει το ποσοστό του συνολικού αριθμού παρατηρήσεων που περιέχει κάθε κατηγορία, διαιρώντας ένα κύκλο σε κυκλικούς τομείς έτσι ώστε το εμβαδόν κάθε κυκλικού τομέα να είναι ίσο με τη συχνότητα της αντίστοιχης κατηγορίας. Αν μ_i^o είναι το αντίστοιχο τόξο ενός κυκλικού τομέα στο κυκλικό διάγραμμα συχνοτήτων ,τότε:

$$\boxed{\mu_i^o = 360^0\ f_i}$$

π.χ. Τα δεδομένα του **Πίνακα 1** μπορούν να παρασταθούν γραφικά με το παρακάτω κυκλικό διάγραμμα.



ΦΥΛΛΟΓΡΑΜΜΑ

Το διάγραμμα αυτό δίνει την δυνατότητα ανασύστασης και ανάκλησης των μετρήσεων των αρχικών δεδομένων του δείγματος με ακρίβεια πράγμα το οποίο δεν επιτυγχάνεται με το ιστόγραμμα ή τους πίνακες συχνοτήτων. Χρησιμοποιείται για την επεξεργασία μέτριου αριθμού παρατηρήσεων (περίπου 150). Η παρουσίαση του σχήματος μοιάζει με εκείνου του ιστογράμματος αλλά η τεχνική κατάρτισης δεν είναι η ίδια. Το φυλλογράφημα εμφανίζει τα δεδομένα σε όλο το εύρος των παρατηρημένων μετρήσεων, παρουσιάζει την συγκέντρωση των παρατηρήσεων (συχνότητες), δείχνει την μορφή της κατανομής, εμφανίζει τυχόν ακραίες και εκτροπές παρατηρήσεις και επιτρέπει την επισήμανση της απουσίας συγκεκριμένων τιμών ή μετρήσεων.

π.χ. Δίνονται οι ηλικίες 10 ατόμων: 27 34 34 43 21 38 46 38 22 35. Να γίνει φυλλόγραμμα των δεδομένων.

Διατάσουμε τα δεδομένα σε αύξουσα σειρά 21 22 27 34 34 35 38 38 43 46

Θεωρούμε ότι κάθε παρατήρηση αποτελείται από δύο τμήματα, το αρχικό ψηφίο και το επόμενο ψηφίο π.χ. 21 : 2 αρχικό, 1 επόμενο.

Κατασκευάζουμε τον παρακάτω πίνακα για όλα τα νούμερα

Κορμός	Φύλλο
2	1 2 7
3	4 4 5 8 8
4	3 6

Ως **φύλλο** κάθε στοιχείου λαμβάνεται το τελευταίο ή τα δύο τελευταία ψηφία της τιμής της παρατήρησης και ως **κορμός** το πρώτο ή τα εναπομείναντα πρώτα ψηφία.

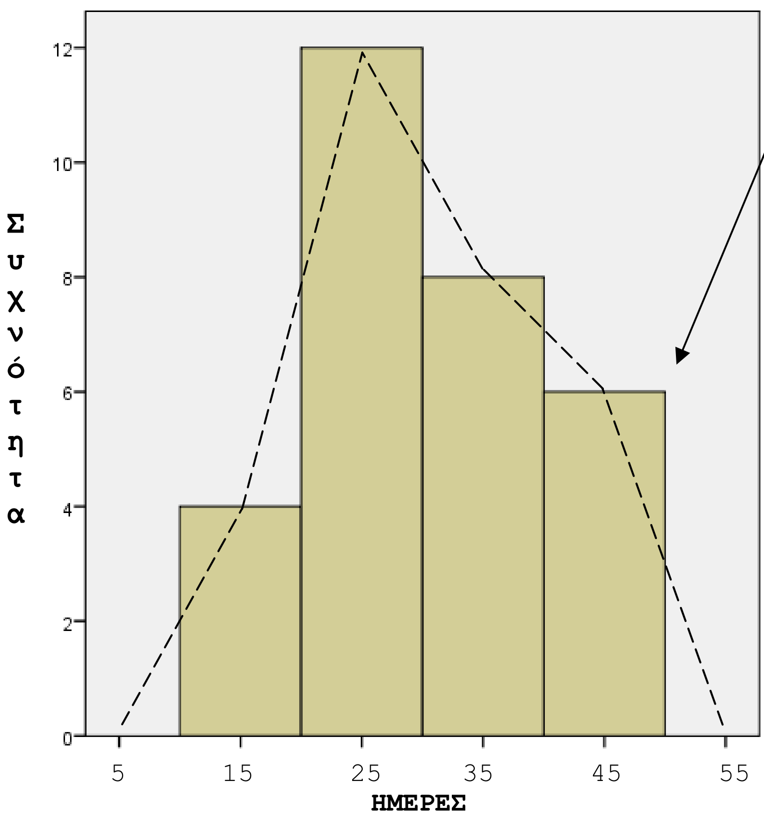
ΙΣΤΟΓΡΑΜΜΑ

Η γραφική παράσταση ενός δείγματος με ομαδοποιημένα δεδομένα γίνεται με το **ιστόγραμμα**. Στον οριζόντιο άξονα του συστήματος ορθογωνίων αξόνων σημειώνουμε τα όρια των κλάσεων. Στη συνέχεια κατασκευάζουμε διαδοχικά ορθογώνια, το καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος ισο με τη συχνότητα (ή σχετική συχνότητα) της κλάσης αυτής.

π.χ. Ο παρακάτω πίνακας δείχνει το πλήθος των λυκείων και τις αντίστοιχες ημέρες κατάληψης κατά τις μαθητικές κινητοποιήσεις 2009-2010.

Ημέρες	Κέντρο κλάσης x_i	Συχνότητα v_i
[10,20)	15	4
[20,30)	25	12
[30,40)	35	8
[40,50)	45	6
\\\\\\	Σύνολο	ν = 30

Κατασκευάζουμε τις κλάσεις ξεκινώντας από τη μικρότερη παρατήρηση και προσθέτουμε κάθε φορά το πλάτος c των κλάσεων. Κάθε παρατήρηση πρέπει να ανήκει σε μια μόνο κλάση. Αν έχουμε κλάσεις ίσου πλάτους c, και οι κεντρικές τιμές διαφέρουν κατά c.



Πολύγωνο συχνοτήτων

Κατασκευάζεται ενώνοντας με ευθείες γραμμές τα μέσα των πάνω πλευρών των ορθογωνίων του ιστογράμματος.

1.4 Περιγραφικά μέτρα στατιστικών δεδομένων

Ο πίνακας συχνοτήτων και οι γραφικές παραστάσεις των δεδομένων μας δίνουν μια συνοπτική εικόνα των δεδομένων. Τα περιγραφικά στατισικά μέτρα είναι αριθμητικά μέτρα που χαρακτηρίζουν ποσοτικά την κατανομή της τυχαίας μεταβλητής, και υπολογίζονται από τις παρατηρήσεις του δείγματος.

1.4.1 ΜΕΤΡΑ ΘΕΣΗΣ

Μέτρα θέσης ονομάζουμε κάποιες χαρακτηριστικές τιμές (αριθμητικά μεγέθη) που δείχνουν τη θέση του «κέντρου» των παρατηρήσεων. Τα κυριότερα μέτρα θέσης είναι η μέση τιμή, η διάμεσος και η επικρατούσα τιμή.

ΜΕΣΗ ΤΙΜΗ

● Έστω μια μεταβλητής **X** με τιμές $x_1,\dots,x_\nu$, όπου **ν** είναι το μέγεθος του δείγματος. Η μέση τιμή των παρατηρήσεων (**δειγματική μέση τιμή**) είναι το πηλίκο του αθροίσματος των παρατηρήσεων δια του πλήθους των παρατηρήσεων, δηλ.

$$\bar{x} = \frac{\text{άθροισμα παρατηρήσεων}}{\text{πλήθος παρατηρήσεων}} = \frac{x_1+x_2+\dots+x_\nu}{\nu} = \frac{\sum_{i=1}^{\nu} x_i}{\nu}$$

● Αν οι τιμές μιας μεταβλητής **X** είναι $x_1,\dots,x_\kappa$ με αντίστοιχες συχνότητες $\nu_1,\dots,\nu_\kappa$, τότε η μέση τιμή των παρατηρήσεων είναι

$$\bar{x} = \frac{x_1\nu_1+\dots+x_k\nu_\kappa}{\nu_1+\dots+\nu_\kappa} = \frac{\sum_{i=1}^{\kappa} x_i\nu_i}{\sum_{i=1}^{\kappa} \nu_i} = \frac{\sum_{i=1}^{\kappa} x_i\nu_i}{\nu}$$

● Αν τα δεδομένα είναι ομαδοποιημένα, τότε $x_1,\dots,x_\kappa$ είναι τα κέντρα των κλάσεων.

ΔΙΑΜΕΣΟΣ

Διάμεσος ενός δείγματος ν παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως μεσαία παρατήρηση, όταν το ν είναι περιττός αριθμός ή ως μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων, όταν το ν είναι άρτιος αριθμός.

$$\delta = \begin{cases} x_{(\frac{\nu+1}{2})} & ,\ \nu \text{ περιττός} \\ \dfrac{x_{(\frac{\nu}{2})} + x_{(\frac{\nu}{2}+1)}}{2} & ,\ \nu \text{ άρτιος} \end{cases}$$

όπου $x_{(i)}$: η i-παρατήρηση αφού διαταχθούν όλες οι παρατηρήσεις σε αύξουσα σειρά.

ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ Η ΚΟΡΥΦΗ

Επικρατούσα τιμή ονομάζεται η παρατήρηση με την μεγαλύτερη συχνότητα.

1.4.2. ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Μέτρα διασποράς ονομάζουμε τα μέτρα (αριθμητικά μεγέθη) που δείχνουν πως κατανέμονται οι τιμές του δείγματος γύρω από τις «κεντρικές τιμές». Διασπορά ονομάζουμε τη συγκέντρωση ή την απομάκρυνση των στατιστικών δεδομένων γύρω από μια κεντρική τιμή.

ΕΥΡΟΣ Εύρος R = μεγαλύτερη παρατήρηση – μικρότερη παρατήρηση

ΔΙΑΣΠΟΡΑ ή ΔΙΑΚΥΜΑΝΣΗ

● Έστω $x_1,\dots,x_\nu$ οι παρατηρήσεις μιας μεταβλητής **X** και $\overline{x}$ η μέση τιμή τους. Η **δειγματική διακύμανση ή διασπορά** δίνεται από τον τύπο:

$$s^2 = \frac{(x_1-\overline{x})^2+\dots+(x_\nu-\overline{x})^2}{\nu-1} = \frac{\sum_{i=1}^{\nu}(x_i-\overline{x})^2}{\nu-1}$$

● Έστω οι τιμές $x_1,\dots,x_\kappa$ μιας μεταβλητής **X** έχουν συχνότητες $\nu_1,\dots,\nu_\kappa$ και έστω $\overline{x}$ η μέση τιμή τους. Η διακύμανση ή διασπορά δίνεται από τον τύπο:

$$s^2 = \frac{\sum_{i=1}^{\kappa}(x_i-\overline{x})^2\nu_i}{\nu-1}$$

● Αν τα δεδομένα είναι ομαδοποιημένα, τότε $x_1,\dots,x_\kappa$ είναι τα κέντρα των κλάσεων.

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

Η **τυπική απόκλιση** είναι η θετική τετραγωνική ρίζα της διακύμανσης $s=\sqrt{s^2}$.

ΤΕΤΑΡΤΗΜΟΡΙΑ

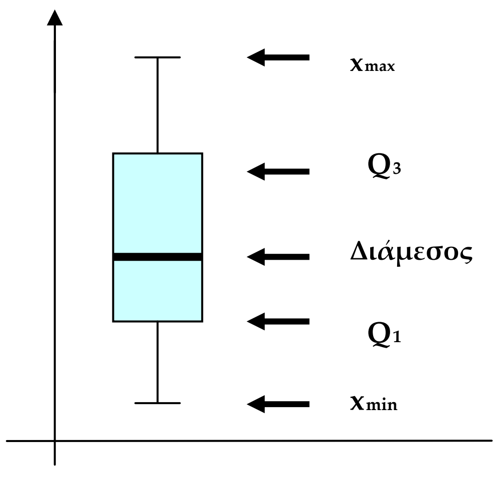
1ο Τεταρτημόριο **3ο Τεταρτημόριο**

$$Q_1 = \begin{cases} x_{(\frac{\nu+1}{4})} & ,\text{αν } \dfrac{\nu+1}{4} \text{ ακέραιος} \\ x_{(\kappa)}+\alpha x_{(\kappa+1)} & ,\text{ αλλιώς} \end{cases} \quad \text{και} \quad Q_3 = \begin{cases} x_{(3\frac{\nu+1}{4})} & ,\text{αν } 3\dfrac{\nu+1}{4} \text{ ακέραιος} \\ x_{(\kappa)}+\alpha x_{(\kappa+1)} & ,\text{ αλλιώς} \end{cases}$$

Αν το (ν+1)/4 δεν είναι ακέραιος, τότε $x_{(\kappa)}$ είναι η παρατήρηση που βρίσκεται στην αμέσως μικρότερη θέση, $x_{(\kappa+1)}$ είναι η παρατήρηση που βρίσκεται στην αμέσως μεγαλύτερη θέση, και α το δεκαδικό μέρος του (ν+1)/4.

ΘΗΚΟΓΡΑΜΜΑ

Το θηκόγραμμα είναι μια γραφική παράσταση που παριστάνει την κατανομή των δεδομένων και συγκεκριμένα αναπαριστά τη διάμεσο, τα τεταρτημόρια και την μέγιστη και ελάχιστη παρατήρηση του δείγματος.



1.5 Περιγραφική στατιστική δύο μεταβλητών

Πίνακες συνάφειας: χρησιμοποιείται για να περιγράψει τη σχέση δυο μεταβλητών. Καταγράφει τη συχνότητα ή την σχετική συχνότητα για την κάθε παρατήρηση των τιμών των δυο μεταβλητών.

Ποιοτικές μεταβλητές

Πχ. Ένα δείγμα 180 αναγνωστών ρωτήθηκε ποιά εφημερίδα διαβάζει και τι μορφωτικού επιπέδου είναι. Οι εφημερίδες που περιλαμβάνονταν στην έρευνα ήταν η Ελευθεροτυπία (1), ο Φίλαθλος (2) και η Εσπρέσο (3), ενώ το μορφωτικό τους επίπεδο ήταν απόφοιτος γυμνασίου (1), ή λυκείου (2), πτυχιούχος (3) ή κάτοχος μεταπτυχιακού (4) .

Πίνακας συνάφειας – με συχνότητες

Μορφωτικό επίπεδο / Εφημερίδα	Απόφοιτος γυμνασίου	Απόφοιτος λυκείου	Πτυχιούχος	Κάτοχος μεταπτυχ.	Σύνολο
Ελευθεροτυπία	3	14	30	11	58
Φίλαθλος	20	20	23	14	77
Εσπρέσο	17	16	7	5	45
Σύνολο	40	50	60	30	180

Πίνακας συνάφειας – με σχετικές συχνότητες

Εφημερίδα	Απόφοιτος γυμνασίου	Απόφοιτος λυκείου	Πτυχιούχος	Κάτοχος μεταπτυχιακού
Ελευθεροτυπία	3/40=0,075	14/50=0,28	30/60=0,5	11/30=0,36
Φίλαθλος	20/40=0,5	20/50=0,4	23/60=0,38	14/30=0,46
Εσπρέσο	17/40=0,425	16/50=0,32	7/60=0,12	5/30=0,18
Σύνολο	1	1	1	1

Ποσοτικές μεταβλητές

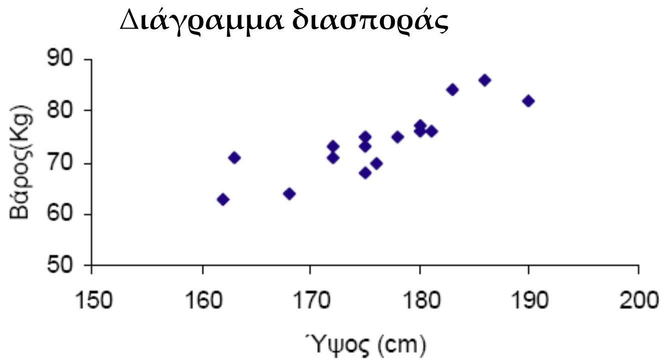
Επιλέγουμε ένα τυχαίο δείγμα και σε κάθε μονάδα του δείγματος μελετάμε δύο ή περισσότερα χαρακτηριστικά. Στην περίπτωση αυτή μας ενδιαφέρει πως οι μεταβλητές συσχετίζονται. π.χ. εξετάζουμε σε ποιο βαθμό η τιμή της αγοράς ενός σπιτιού συσχετίζεται με το μέγεθος του σπιτιού.

- **Καθορισμός μεταβλητών** : ανεξάρτητη μεταβλητή Χ (μέγεθος σπιτιού)
εξαρτημένη μεταβλητή Υ (τιμή σπιτιού)

- **Διάγραμμα διασποράς**: Αναπαράσταση των ζευγών των παρατηρήσεων (x, y) σε ένα διάγραμμα. Η ανεξάρτητη μεταβλητή Χ τοποθετείται στον οριζόντιο άξονα και η μεταβλητή Υ στον άλλον άξονα.

Π.χ. ΄Υψος και βάρος 16 φοιτητών

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ύψος (cm)	183	162	172	181	180	168	176	180	190	175	178	175	186	172	175	163
Βάρος (Kg)	84	63	71	76	77	64	70	76	82	68	75	73	86	73	75	71



Από το διάγραμμα διασποράς φαίνεται ότι οι φοιτητές στο δείγμα που έχουν μεγαλύτερο ύψος έχουν και μεγαλύτερο βάρος. Φαίνεται, δηλαδή, να υπάρχει μια ανάλογη σχέση μεταξύ του ύψους και του βάρους των φοιτητών.

Συντελεστής γραμμικής συσχέτισης PEARSON

Στατιστικό μέτρο που χρησιμοποιείται για να εξετάσουμε πόσο ισχυρή είναι η συσχέτιση δυο μεταβλητών. Ορίζεται ως

$$r = \frac{s_{xy}}{s_x s_y}$$

όπου s_x, s_y είναι δειγματικές τυπικές αποκλίσεις των μεταβλητών Χ και Υ αντίστοιχα και s_{xy} είναι η **δειγματική συνδιαπορά**:

$$s_{xy} = \frac{\sum_{i=1}^{\nu}(x_i - \overline{x})(y_i - \overline{y})}{\nu - 1} = \frac{\sum_{i=1}^{\nu} x_i y_i - \nu \overline{x}\overline{y}}{\nu - 1}.$$

• Παίρνει τιμές στο κλειστό διάστημα **[-1 , 1].**

• Αν $r = 1$ ή $r = -1$ υπάρχει **τέλεια γραμμική** συσχέτιση.

• Αν $-0,3 \le r < 0,3$ **δεν υπάρχει γραμμική** συσχέτιση.

• Αν $-0,5 < r \le -0,3$ ή $0,3 \le r < 0,5$ υπάρχει **ασθενής γραμμική** συσχέτιση.

• Αν $-1 < r \le -0,8$ ή $0,8 \le r < 1$ υπάρχει **πολύ ισχυρή γραμμική** συσχέτιση.

• Το πρόσημο του r καθορίζει το είδος, μόνο, της συσχέτισης (θετική ή αρνητική), δηλαδή αν αύξηση της μεταβλητής Χ αντιστοιχεί σε αύξηση ή μείωση της Υ.

1.6 Ασκήσεις

1.6.1. Σε μια έρευνα αθλητικής εφημερίδας ρωτήθηκαν 100 φίλαθλοι ποια ομάδα νομίζουν ότι θα πάρει το πρωτάθλημα. Τα αποτελέσματα ήταν τα εξής: Άρης: 25, ΠΑΟΚ: 35, Ολυμπιακός: 15, ΑΕΚ: 10, Άλλη ομάδα : 15

α) Να γίνει πίνακας κατανομής συχνοτήτων.

β) Να βρεθεί το ποσοστό των φιλάθλων που νομίζουν ότι ο ΠΑΟΚ θα είναι πρωταθλητής.

γ) Να γίνει ραβδόγραμμα σχετικών συχνοτήτων.

1.6.2. Οι επιβάτες 10 αυτοκινήτων ήταν οι ακόλουθοι:

1 1 3 2 2 1 4 1 2 3

α) Να γίνει πίνακας συχνοτήτων και σχετικών συχνοτήτων.

β) Πόσοι οδηγοί είχαν και συνεπιβάτες;

γ) Ποιος ο αριθμός των αυτοκινήτων που είχαν τουλάχιστον 3 επιβάτες;

δ) Ποιο το ποσοστό των οδηγών που ήταν μόνοι τους στα αυτοκίνητα;

ε) Βρείτε την διάμεσο, το πρώτο και τρίτο τεταρτημόριο των παραπάνω παρατηρήσεων.

στ) Σχεδιάστε το θηκόγραμμα των παραπάνω παρατηρήσεων

ζ) Υπολογίστε το μέσο πλήθος ατόμων στα αυτοκίνητα και την τυπική απόκλιση.

1.6.3. Δίνονται οι βαθμολογίες 10 φοιτητών στο μάθημα της στατιστικής

4 3 2 5 9 8 7 6 5 1

Να υπολογίσετε:

α) την μέση τιμή, β) την τυπική απόκλιση, γ) την διάμεσο και δ) να γίνει θηκόγραμμα.

Κεφάλαιο 2

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΤΗΣ ΘΕΩΡΙΑΣ ΠΙΘΑΝΟΤΗΤΩΝ

2.1 Εισαγωγή

Πείραμα τύχης λέγεται ένα πείραμα κατά το οποίο δεν μπορούμε εκ των προτέρων να προβλέψουμε το αποτέλεσμα, παρότι φαινομενικά τουλάχιστον φαίνεται να επαναλαμβάνεται υπό τις ίδιες συνθήκες.

Δειγματικός χώρος Αν $\omega_1, \omega_2, \ldots, \omega_\kappa$ είναι τα δυνατά αποτελέσματα ενός πειράματος τύχης, το σύνολο $\Omega = \{\omega_1, \omega_2, \ldots, \omega_\kappa\}$ λέγεται δειγματικός χώρος του πειράματος.

Κάθε υποσύνολο του δειγματικού χώρου Ω λέγεται **ενδεχόμενο** ή **γεγονός**. Ένα ενδεχόμενο λέγεται **απλό** αν έχει ένα μόνο στοιχείο και **σύνθετο** αν έχει περισσότερα από ένα στοιχεία. Το **Ω** λέγεται **βέβαιο ενδεχόμενο** ενώ το $\varnothing$ λέγεται **αδύνατο ενδεχόμενο**. Όταν το αποτέλεσμα μιας εκτέλεσης του πειράματος είναι στοιχείο ενός ενδεχομένου, λέμε ότι το ενδεχόμενο **πραγματοποιείται**. Τα στοιχεία ενός ενδεχομένου λέγονται και **ευνοϊκές περιπτώσεις** για την πραγματοποίηση του.

2.2 Πράξεις με ενδεχόμενα

Τομή $A \cap B$ δυο ενδεχομένων Α και Β είναι το ενδεχόμενο που πραγματοποιείται όταν πραγματοποιούνται συγχρόνως τα Α και Β.

Ένωση $A \cup B$ δυο ενδεχομένων Α και Β είναι το ενδεχόμενο που πραγματοποιείται όταν πραγματοποιείται ένα τουλάχιστον από τα Α και Β.

Συπληρωματικό ενδεχόμενο του Α είναι το ενδεχόμενο $\overline{A}$ που πραγματοποιείται όταν δεν πραγματοποιείται το Α.

Διαφορά του Β από το Α είναι το ενδεχόμενο **Α – Β** που πραγματοποιείται όταν πραγματοποιείται το Α αλλά όχι το Β. Ισχύει: $\mathbf{A - B = A \cap B'}$.

Αμοιβαίως αποκλειόμενα ή ***ασυμβίβαστα ενδεχόμενα*** λέγονται δυο ενδεχόμενα Α και Β αν $\mathbf{A \cap B = \varnothing}$. ***Στοχαστικά ανεξάρτητα*** λέγονται δυο ενδεχόμενα Α και Β αν η πραγματοποίηση του ενός δεν επηρεάζει την πραγματοποίηση του άλλου.

2.3 Κλασικός ορισμός πιθανότητας

Σε ένα πείραμα με ***ισοπίθανα*** απλά ενδεχόμενα, ορίζουμε ως **πιθανότητα ενός ενδεχομένου Α** και συμβολίζουμε με **P(A)** το πηλίκο των ευνοϊκών περιπτώσεων του Α προς το πλήθος των συνολικών περιπτώσεων του πειράματος

$$\boxed{P(A) = \frac{\text{πλήθος ευνοικών περιπτώσεων}}{\text{πλήθος συνολικών περιπτώσεων}}}$$

• **$0 \le P(A) \le 1$** • **$P(\Omega) = 1$** • **$P(\varnothing) = 0$**

2.4 Αξιωματικός ορισμός πιθανότητας

Έστω $\Omega = \{\omega_1, \omega_2, \ldots, \omega_\nu\}$ ο δειγματοχώρος ενός πειράματος τύχης με πεπερασμένο πλήθος στοιχείων. Σε κάθε στοιχείο ω_i, i= 1,..,ν του δειγματοχώρου αντιστοιχίζουμε έναν πραγματικό αριθμό $P(\omega_i)$ με τις εξής ιδιότητες

• $0 \le P(\omega_i) \le 1$

• $P(\omega_1) + P(\omega_2) + \ldots + P(\omega_\nu) = 1$

Αν $A = \{ \alpha_1, \alpha_2, \ldots, \alpha_\kappa \}$ ένα ενδεχόμενο (υποσύνολο του δειγματοχώρου), τότε $P(A) = P(\alpha_1) + P(\sigma_2) + \ldots + P(\alpha_K)$.

2.5 Τύποι πιθανοτήτων

• $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

• $P(A \cup B) = P(A) + P(B)$ αν Α και Β δυο ασυμβίβαστα μεταξύ τους ενδεχόμενα

• $P(A \cup B \cup \Gamma) = P(A) + P(B) + P(\Gamma) - P(A \cap B) - P(A \cap \Gamma) - P(B \cap \Gamma) + P(A \cap B \cap \Gamma)$

• $P(\overline{A}) = 1 - P(A)$

• Αν $A \subseteq B$, τότε $P(A) \le P(B)$

• $P(A - B) = P(A) - P(A \cap B)$

• $P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B})$

• $P(\overline{A} \cup \overline{B}) = P(\overline{A \cap B})$

2.6. Ασκήσεις

2.6.1. Δίνονται $P(\overline{A}) = 0{,}3$, $P(B) = 0.4$ και $P(A \cap \overline{B}) = 0{,}5$. Να βρεθούν:
α) $P(A)$, β) $P(A \cap B)$ και γ) $P(A \cup B)$

2.6.2. Μια επιχείρηση πουλάει ανταλλακτικά σε συσκευασία των 5 τεμαχίων. Έστω ότι για κάθε κουτί η πιθανότητα να βρεθεί ένα ανταλλακτικό ελαττωματικό είναι ίση με 0.32, η πιθανότητα να βρεθούν δυο ανταλλακτικά ελαττωματικά είναι ίση με 0.08, η πιθανότητα να βρεθούν τρία ανταλλακτικά ελαττωματικά είναι ίση με 0.02, η πιθανότητα να βρεθούν τέσσερα ανταλλακτικά ελαττωματικά είναι ίση με 0.01 και η πιθανότητα να βρεθούν πέντε ανταλλακτικά ελαττωματικά είναι ίση με 0. Να βρεθεί η πιθανότητα να μην βρεθεί κανένα ανταλλακτικό ελαττωματικό.

2.6.3. Έστω Α, Β, Γ τα ενδεχόμενα μια οικογένεια να έχει τηλεόραση, βίντεο και DVD writer, αντίστοιχα. Αν ισχύουν οι σχέσεις $P(A \cap B \cap \Gamma) = 0.03$, $P(B \cap \Gamma) = 0.07$, $P(A \cap B)=0.02$, $P(A \cap \Gamma)=0.09$, $P(\Gamma)=0.13$, $P(A \cup B)=0.67$, $P(B)=0.4$, τότε να υπολογιστούν οι πιθανότητες α) να αγοράσει μια τουλάχιστον από τις 3 συσκευές, β) να αγοράσει μόνο DVD writer, γ) να αγοράσει DVD writer και μια τουλάχιστον ακόμα συσκευή.

Κεφάλαιο 3

ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ ΚΑΙ ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

3.1 ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ ΔΙΑΚΡΙΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Τυχαία μεταβλητή ονομάζεται η συνάρτηση **X:** $S \to R$ η οποία αντιστοιχίζει σε κάθε στοιχείο s ενός δειγματοχώρου S έναν και μοναδικό αριθμό **X(*s*)**. Το σύνολο των δυνατών τιμών της **X** ονομάζεται πεδίο τιμών της **X** και συμβολίζεται με R_X.

Μια τυχαία μεταβλητή ονομάζεται **απαριθμητή ή διακριτή τυχαία μεταβλητή** αν παίρνει πεπερασμένο ή άπειρο αριθμήσιμο πλήθος τιμών. Στην περίπτωση αυτή τα στοιχεία της X είναι αριθμημένα και μπορούν να καταγραφούν ως εξής: $X = \{x_1, x_2, \ldots, x_k, \ldots\}$. Δηλαδή, οι τιμές που μπορεί να πάρει η X είναι $x_1, x_2, \ldots, x_k, \ldots$ διατεταγμένες σε αύξουσα σειρά.

Συνάρτηση μάζας πιθανότητας ή συνάρτηση πιθανότητας ή κατανομή πιθανοτήτων της X ονομάζεται η συνάρτηση P η οποία σε κάθε τιμή x_i της **X,** αντιστοιχεί την πιθανότητα $P(x_i) = P(X = x_i)$ και ικανοποιεί τις συνθήκες:

(1) $P(x_i) \geq 0$ **για κάθε** x_i

(2) $\sum_i P(x_i) = 1$

Θεώρημα

Αν X διακριτή τυχαία μεταβλητή, όπου x_i και x_j είναι δυο τιμές της X και $P(X = x)$ η συνάρτηση πιθανότητας της, τότε θα ισχύει:

$$\boxed{P(x_i \leq X \leq x_j) = P(X=x_i) + P(X=x_{i+1}) + \ldots + P(X=x_j)}$$

Μέση τιμή διακριτής τυχαίας μεταβλητής

$$\mu = E(X) = \sum_x x\, P(X = x)$$

Διακύμανση και τυπική απόκλιση διακριτής τυχαίας μεταβλητής

$$\sigma^2 = Var(X) = \sum_x x^2 f(x) - \mu^2 \quad \text{και} \quad \sigma = \sqrt{Var(X)}$$

π.χ. Ρίχνουμε 2 ζάρια. Να βρεθεί η κατανομή πιθανοτήτων της μεταβλητής X η οποία αντιστοιχεί σε κάθε στοιχείο του δειγματοχώρου του πειράματος στο άθροισμα των αποτελεσμάτων των δυο ζαριών.

Ο δειγματοχώρος του πειράματος είναι $S = \left\{ \begin{matrix} (1,1), & (1,2), & \ldots, & (1,6) \\ (2,1), & (2,2), & \ldots, & (2,6) \\ & \ldots\ldots & & \\ (6,1), & (6,2), & \ldots, & (6,6) \end{matrix} \right\}$.

Η μεταβλητή Χ αντιστοιχεί σε κάθε στοιχείο του δειγματοχώρου (α, β) τον αριθμό α + β, δηλ. x_i = **α + β** ή Χ((α, β) = α + β).

Άρα το σύνολο τιμών της Χ είναι R_X = {2, 3, 4, …. , 12}.

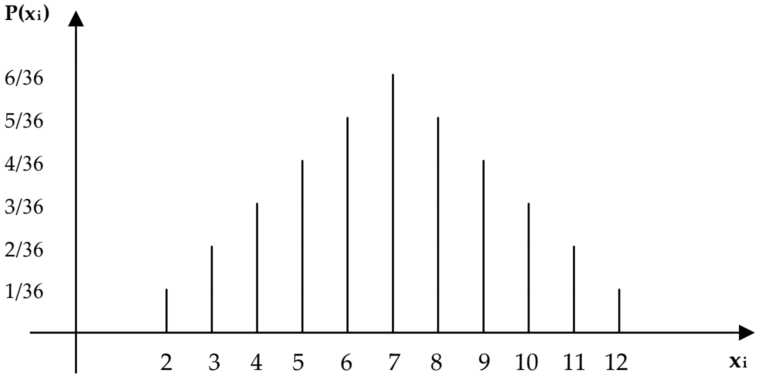
Η κατανομή πιθανοτήτων της Χ και η γραφική της παράσταση είναι:

x_i	2	3	4	5	6	7	8	9	10	11	12
P(x_i)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Είναι: x_i =2 : Χ((α, β) = 2) = {(1,1)} $\Rightarrow$ P(2) ≡ P(X = 2) = 1/36

x_i =3 : Χ((α, β) = 3) = {(1,2), (2,1)} $\Rightarrow$ P(3) ≡ P(X = 3) = 2/36

x_i =4 :Χ((α, β) = 4) = {(1,3), (2,2), (3,1)} $\Rightarrow$ P(4) ≡ P(X = 4) = 3/36, κ.τ.λ.



Αθροιστική συνάρτηση πιθανότητας ή αθροιστική κατανομή πιθανοτήτων ή συνάρτηση κατανομής F της διακριτής μεταβλητής Χ, είναι η συνάρτηση που αντιστοιχεί σε κάθε τιμή x_i της **Χ,** την αθροιστική πιθανότητα ***F*(x$_i$) = P(X ≤ x$_i$).** Ισχύει:

(1) 0 ≤ F(x_i) ≤ 1

(2) Αν **x_i** και **x_j** δυο τιμές της Χ με **$x_i \le x_j$** τότε **F(x_i) ≤ F(x_j)** δηλ. η συνάρτηση **F** είναι αύξουσα.

- $F(x_i) = P(X \le x_i) = \sum_{x \le x_i} P(x_i)$
- $P(x_i) = F(x_i) - F(x_{i-1})$

Π.χ. Ρίχνουμε ένα νόμισμα δυο φορές και θεωρούμε την μεταβλητή Χ που δείχνει το πλήθος των Κ που εμφανίστηκαν. Να βρεθούν η συνάρτηση μάζας πιθανότητας και η αθροιστική κατανομή πιθανοτήτων της Χ.

Ο δειγματοχώρος του πειράματος είναι **S** = {(Γ,Γ),(Κ,Γ),(Γ,Κ),(Κ,Κ)}. Η μεταβλητή Χ που δείχνει το πλήθος των Κ παίρνει τιμές x_i = 0, 1, 2. Είναι:

$P(0) = P(X = 0) = ¼$, $P(1) = P(X = 1) = ½$, $P(2) = P(X = 2) = ¼$

Η αθροιστική κατανομή πιθανοτήτων της Χ είναι: $F(x) = \begin{cases} 0 & x<0 \\ 1/4 & 0 \le x<1 \\ 3/4 & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$

3.2 ΒΑΣΙΚΕΣ ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΔΙΑΚΡΙΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

3.2.1 Διωνυμική κατανομή

Ένα τυχαίο πείραμα στο οποίο υπάρχουν δύο μόνο δυνατά αποτελέσματα λέγεται **δοκιμή Bernoulli**. Το ένα από τα δύο αποτελέσματα της δοκιμής Bernoulli λέγεται ***επιτυχία*** και το άλλο ***αποτυχία***. Για κάθε δοκιμή **Bernoulli** ορίζουμε μια τυχαία μεταβλητή X ως εξής: Αν το αποτέλεσμα ***s*** του πειράματος είναι ***επιτυχία*** τότε X(***s***)=1, αλλιώς X(***s***)=0. Η κατανομή πιθανοτήτων της X είναι:

$$P(x)=\begin{cases} p, & \alpha\nu\ x=1 \\ 1-p, & \alpha\nu\ x=0 \end{cases}$$ όπου *p* είναι η πιθανότητα επιτυχίας.

Έστω μια **δοκιμή Bernoulli** με πιθανότητα επιτυχίας *p*, επαναλαμβάνεται *n* φορές, έτσι ώστε το αποτέλεσμα κάθε δοκιμής να είναι ανεξάρτητο από κάθε άλλη δοκιμή. Ορίζουμε την τυχαία μεταβλητή X που εκφράζει τον αριθμό των επιτυχιών σε *n* ανεξάρτητες επαναλήψεις του πειράματος (άρα X = 0, 1, 2,.., *n*). Η **κατανομή πιθανοτήτων** ή **συνάρτηση πιθανότητας** της *X* είναι

$$\boxed{P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}}$$ όπου $$\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

Η διωνυμική κατανομή προσδιορίζεται πλήρως από τις παραμέτρους *n* και *p*. Γράφουμε X ~ B(*n, p*) και διαβάζουμε «η X ακολουθεί τη διωνυμική κατανομή με παραμέτρους *n* και *p*».

- Η **μέση τιμή της διωνυμικής κατανομής** είναι **E(*X*) = *n p.***
- Η **διασπορά της διωνυμικής κατανομής** είναι **Var(*X*) = *n p (1-p).***

Π.χ. Το 25% των μαθητών ενός σχολείου δήλωσαν ότι δεν θα πάνε στην σχολική εκδρομή. Επιλέγουμε τυχαία 15 μαθητές. Να βρεθεί:
α) η πιθανότητα οι 3 από τους 15 μαθητές να μην πάνε στην εκδρομή.
β) η πιθανότητα τουλάχιστον 2 από τους 15 μαθητές να μην πάνε στην εκδρομή.
γ) η μέση τιμή και η διακύμανση των μαθητών.
<u>Λύση</u>
Έστω η μεταβλητή X = αριθμός μαθητών που δεν θα πάνε εκδρομή, η οποία ακολουθεί τη διωνυμική κατανομή με n = 15 και p = 0,25.

α) P(X=3) = $\binom{15}{3}0.25^3(1-0,25)^{15-3}$ **=0,2246**

β) P(X ≥ 2) = 1- P(X<2) = 1 – [P(X=0) +P(X=1)] =

= 1 - [$\binom{15}{0}0.25^0(1-0,25)^{15-0}$ **+** $\binom{15}{1}0.25^1(1-0,25)^{15-1}$] =

= 1 – 0,2138 = **0,7862**

γ) **E(X)** = *n p* = 15 0,25 = **3,75** και **Var(x)** = *n p (1-p)* = 15 0,25 (1-0,25) = **2,8125**

3.2.2 Κατανομή Poisson

Η **κατανομή Poisson** είναι η κατανομή των σπάνιων γεγονότων και μας ενδιαφέρει όταν ψάχνουμε τον αριθμό και την πιθανότητα των «συμβάντων» στην μονάδα μέτρησης (π.χ. τυπογραφικά λάθη ανά σελίδα, τηλεφωνικές κλίσεις ανά ώρα κ.τ.λ.). Η τυχαία μεταβλητή που ακολουθεί **κατανομή Poisson** και εκφράζει το πλήθος των «συμβάντων» ανά μονάδα μέτρησης έχει συνάρτηση πιθανότητας

$$\boxed{P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}}$$

όπου x = 0,1,2…. και λ >0 (παράμετρος).

- Η **μέση τιμή της κατανομής Poisson** είναι **E(*X*) = λ.**
- Η **διασπορά της κατανομής Poisson** είναι **Var(*X*) = λ.**

Π.χ. Για μια μεταφορική εταιρεία γνωρίζουμε ότι περίπου 12 φορτηγά ανά 5 ημέρες χρειάζονται επισκευή. Αν η εταιρεία έχει 2 εφεδρικά φορτηγά, να βρεθεί η πιθανότητα μια οποιαδήποτε ημέρα

α) να μην χρειαστεί η εταιρία κανένα εφεδρικό φορτηγό

β) ο αριθμός των εφεδρικών φορτηγών να μην είναι επαρκής.

γ) Να βρεθεί η μέση τιμή και η τυπική απόκλιση του αριθμού των φορτηγών που παθαίνουν βλάβη κάθε μέρα.

Έστω η μεταβλητή X = αριθμός των φορτηγών που παθαίνουν βλάβη κάθε μέρα. Η X ακολουθεί κατανομή Poisson με λ = 12/5 =2,4 (12 βλάβες ανά 5 ημέρες).

α) P(X = 0) = $e^{-2,4}\frac{2,4^0}{0!}$ = **0,0907**

β) P(X > 2) = 1 – P(X ≤ 2) = 1 – [P(X=0)+P(X=1) +P(X=2)]=

$$= 1 - [\,e^{-2,4}\frac{2,4^0}{0!}+e^{-2,4}\frac{2,4^1}{1!}+e^{-2,4}\frac{2,4^2}{2!}\,] =$$

= 1 – (0,0907+0,21777+0,2613) = **0,4303**

γ) Μέση τιμή E(X) = λ = 2,4

Διασπορά Var(X) = λ = 2,4 $\Rightarrow$ Τυπική απόκλιση σ = $\sqrt{Var(X)}=\sqrt{2,4}$ = **1,55**

3.3 ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ ΣΥΝΕΧΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Μια τυχαία μεταβλητή ονομάζεται ***συνεχής,*** αν το πεδίο τιμών της $\boldsymbol{R_X}$ είναι ένα διάστημα ή μια συλλογή διαστημάτων του ***R***. Αν η X είναι μια συνεχής τυχαία μεταβλητή, η πιθανότητα να πάρει η X μια ορισμένη τιμή είναι γενικά μηδέν. Συνεπώς δεν μπορούμε να ορίσουμε μια συνάρτηση πιθανότητας με τον ίδιο τρόπο, όπως μια διακριτή μεταβλητή. Για να ορίσουμε την κατανομή πιθανότητας για μια συνεχή τυχαία μεταβλητή, παρατηρούμε ότι η πιθανότητα να βρίσκεται η X μεταξύ δύο διαφορετικών τιμών έχει νόημα.

Συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση πυκνότητας της συνεχούς τυχαίας μεταβλητής X ονομάζεται η συνάρτηση f(x) για την οποία ισχύουν οι σχέσεις

(1) $\mathbf{f(x) \geq 0}$ για κάθε $x \in R_x$

(2) $\int_{R_x} f(x)dx = 1$

(3) $\mathbf{P(x_1 \leq X \leq x_2)} = \int_{x_1}^{x_2} f(x)dx$

Παρατηρήσεις

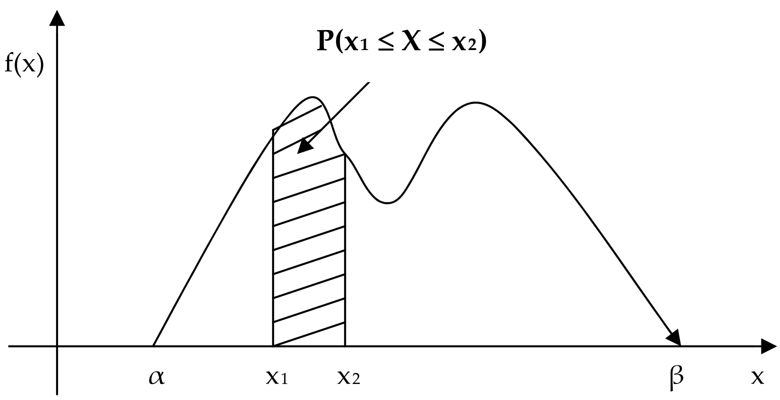
• Το εμβαδόν κάτω από την καμπύλη της συνάρτησης πυκνότητας πιθανότητας f(x) ισούται με 1.

• $\mathbf{P(x_1 \leq X \leq x_2)}$ είναι το εμβαδόν κάτω από την καμπύλη της συνάρτησης f(x) μεταξύ των σημείων $\mathbf{x_1}$ και $\mathbf{x_2}$.

Επίσης ισχύουν τα εξής:

(1) Αν $\mathbf{x_1}$ και $\mathbf{x_2}$ δυο τιμές της X με $\mathbf{x_1 < x_2}$ τότε $\mathbf{F(x_1) \leq F(x_2)}$**.**

(2) $\mathbf{P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)}$



Αθροιστική συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση κατανομής της συνεχούς τυχαίας μεταβλητής X ονομάζεται η συνάρτηση F(x) για την οποία ισχύει: $\mathbf{F(x) = P(X \leq x)} = \int_{-\infty}^{x} f(t)dt$

3.4 ΒΑΣΙΚΕΣ ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΝΕΧΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Αν μια συνεχής τυχαία μεταβλητή έχει ***συνάρτηση πυκνότητας πιθανότητας*** την

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

όπου **E(X) = μ** και **Var(X) = σ²**.

• Η κανονική κατανομή προσδιορίζεται πλήρως από τις παραμέτρους μ και σ². Γράφουμε Χ ~ Ν(μ, σ²) και διαβάζουμε «η Χ ακολουθεί τη κανονική κατανομή με παραμέτρους μ και σ²».

Αν η τυχαία μεταβλητή Χ ακολουθεί την κανονική κατανομή Ν(μ, σ²), τότε η μεταβλητή $Z = \frac{X-\mu}{\sigma}$ ακολουθεί την τυπική κανονική κατανομή Ν(0,1).

• Η κατανομή της Ζ ονομάζεται ***τυποποιημένη*** ή ***τυπική κανονική κατανομή***, με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

όπου **μ= 0** και **σ² =1**

• Η **συνάρτηση αθροιστικής κατανομής** της τυπικής κανονικής κατανομής Ν(0,1) συμβολίζεται με **Φ(x) = P(Z ≤ z)** και έχει τις εξής ιδιότητες:

$$\Phi(-\alpha) = 1 - \Phi(\alpha)$$

$$P(\alpha \leq Z \leq \beta) = \Phi(\beta) - \Phi(\alpha)$$

• Αν **F_X(x)** η **συνάρτηση αθροιστικής κατανομής** της Χ~(μ, σ²) και **F_Z(z)=Φ(z)** η **συνάρτηση αθροιστικής κατανομής** της Ζ~(0, 1) τότε

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right)$$

• Οι πιθανότητες Φ(z) = P(Z ≤ z) είναι γνωστές και δίνονται σε πίνακα.

3.5 ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ

Το θεώρημα αυτό αποδεικνύει ότι η κατανομή των μέσων τυχαίων δειγμάτων με μέγεθος σχετικά μεγάλο, πρακτικά μεγαλύτερο ή ίσο του τριάντα, δεν εξαρτάται από την κατανομή του αρχικού πληθυσμού αλλά ακολουθεί, κάτω από ορισμένες προϋποθέσεις, οι οποίες συνήθως πληρούνται, την κανονική κατανομή. Όταν το μέγεθος του δείγματος είναι μικρότερο του τριάντα, για να ακολουθεί η κατανομή δειγματοληψίας των μέσων την κανονική κατανομή, θα πρέπει και ο πληθυσμός από τον οποίο προέρχεται το δείγμα να ακολουθεί την κανονική κατανομή.

Αν οι τυχαίες μεταβλητές $X_1, X_2,\dots,X_\nu$ έχουν την ίδια κατανομή πιθανοτήτων, τότε ονομάζονται **ισόνομες τυχαίες μεταβλητές**.

Κεντρικό οριακό θεώρημα
Αν $X_1, X_2,\dots,X_\nu$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέση τιμή μ και πεπερασμένη διακύμανση σ^2, τότε όσο αυξάνει το ν, η κατανομή της τυχαίας μεταβλητής

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{\nu}}$$

προσεγγίζει την τυπική κανονική κατανομή.

3.6 Ασκήσεις

3.6.1. Έστω κάποιος αγοράζει 10 μετοχές προς 50 ευρώ τη μια. Πληροφορείται ότι σε έναν χρόνο:

- οι μετοχές δεν θα αξίζουν τίποτα, με πιθανότητα 0,1
- οι μετοχές θα αξίζουν όσο ήταν η αρχική τους τιμή με πιθανότητα 0,5
- οι μετοχές θα αξίζουν διπλάσια από την αρχική τους τιμή με πιθανότητα 0,4.

Αν X το κέρδος από την αγορά 10 μετοχών, να βρεθούν
(α) η συνάρτηση πιθανότητας της X
(β) η αθροιστική συνάρτηση πιθανότητας της X
(γ) η μέση τιμή της X
(δ) η διακύμανση και η τυπική απόκλιση της X

3.6.2. Έστω ότι για κάθε παιδί που γεννιέται η πιθανότητα να είναι κορίτσι είναι 0.5. Να υπολογιστεί η πιθανότητα σε μια οικογένεια με 10 παιδιά:
(α) να είναι όλα κορίτσια
(β) να είναι κορίτσια τουλάχιστον τα 5
(γ) να είναι το πολύ 8 κορίτσια.

3.6.3. Έστω μια παραγγελία ηλεκτρικών συσκευών περιλαμβάνει 10% ελαττωματικούς. Εάν επιλέξουμε τυχαία 4 συσκευές, να βρεθεί η πιθανότητα
(α) ακριβώς μια συσκευή να είναι ελαττωματική

(β) τουλάχιστον μια συσκευή να είναι ελαττωματική, στο δείγμα των 4 συσκευών
(γ) Η μέση τιμή και η διακύμανση των συσκευών που είναι ελαττωματικές.

3.6.4. Σε μια ποσότητα μήλων, το 15% είναι χαλασμένα. Παίρνουμε τυχαία 5 μήλα. Να βρεθεί η πιθανότητα:
(α) ακριβώς ένα μήλο είναι χαλασμένο
(β) κανένα μήλο δεν είναι χαλασμένο
(γ) τουλάχιστον ένα μήλο είναι χαλασμένο
(γ) Να βρεθεί η μέση τιμή και η διακύμανση του αριθμού των χαλασμένων μήλων.

3.6.5. Έστω η πιθανότητα να κάνει κάποιος ένα λάθος στην δακτυλογράφηση μιας σελίδας είναι 0.01. Ποια η πιθανότητα σε ένα κείμενο 350 σελίδων
(α) να υπάρχει μια σελίδα με τουλάχιστον ένα σφάλμα
(β) το πολύ 4 σελίδες με τουλάχιστον ένα σφάλμα σε κάθε μια σελίδα

3.6.6. Οι πελάτες ενός καταστήματος φτάνουν στο ταμείο με ρυθμό 8 πελατών /ώρα. Αν ο αριθμός των αφίξεων των πελατών ακολουθεί κατανομή Poisson, τότε να βρεθεί για μια δεδομένη ώρα, η πιθανότητα
(α) να φτάνουν ακριβώς 8 πελάτες
(β) να φτάνουν το πολύ 3 πελάτες
(γ) να φτάνουν τουλάχιστον 2 πελάτες

3.6.7. Ο αριθμός X των πλοίων που φτάνουν σε ένα λιμάνι σε μια ημέρα ακολουθεί κατανομή Poisson με $\lambda=3$ πλοία. Αν στο λιμάνι δεν μπορούν να εξυπηρετηθούν πάνω από 5 πλοία την ημέρα, ποια η πιθανότητα σε μια ημέρα να υπάρξουν πλοία που δεν θα εξυπηρετηθούν;

3.6.8. Αν η τυχαία μεταβλητή Z ακολουθεί την τυπική κανονική κατανομή, να υπολογισθούν οι πιθανότητες
α) $P(Z \leq 1{,}25)$
β) $P(Z \leq -1{,}25)$
γ) $P(0{,}3 \leq Z \leq 1{,}3)$
δ) $P(Z \geq 2{,}1)$

3.6.9. Αν $X \sim N(200,100)$ να υπολογισθούν οι πιθανότητες
α) $P(190 < X < 210)$
β) $P(180{,}4 < X \leq 219{,}6)$

3.6.10. Αν $X \sim N(4,9)$ να υπολογισθεί η πιθανότητα $P(-2 \leq X \leq 5)$.

Κεφάλαιο 4

ΕΚΤΙΜΗΤΙΚΗ

4.1 Εισαγωγή

Στατιστική συμπερασματολογία είναι η μεθοδολογία με την οποία συγκεντρώνονται συμπεράσματα για έναν στατιστικό πληθυσμό βάση αποτελεσμάτων που λαμβάνονται από ένα τυχαίο δείγμα.

Ένας **εκτιμητής σε σημείο** είναι μια συνάρτηση των δεδομένων του τυχαίου δείγματος. Ο **εκτιμητής σε σημείο** είναι μια τιμή ενός στατιστικού που χρησιμοποιείται ως η καλύτερη προσέγγιση στην αντίστοιχη παράμετρο του πληθυσμού την οποία εκτιμά ο εκτιμητής.
π.χ. ο δειγματικός μέσος $\overline{x}$ είναι ο εκτιμητής σε σημείου του πληθυσμιακού μέσου **μ** ενός πληθυσμού.

Ο **εκτιμητής σε διάστημα ή διάστημα εμπιστοσύνης** είναι ένα διάστημα για το οποίο μπορεί να καθοριστεί με διάφορους βαθμούς εμπιστοσύνης ότι περιέχει την παράμετρο του πληθυσμού που εκτιμάται.

Συντελεστής ή βαθμός εμπιστοσύνης ονομάζεται η πιθανότητα με την οποία η εκτιμώμενη παράμετρος περιέχεται στο διάστημα εμπιστοσύνης. Αν ο συντελεστής εμπιστοσύνης είναι **1-α** τότε το διάστημα εμπιστοσύνης λέγεται **100(1-α)% διάστημα εμπιστοσύνης.**

4.2. ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΓΙΑ ΤΗ ΜΕΣΗ ΤΙΜΗ ΕΝΟΣ ΠΛΗΘΥΣΜΟΥ

Α. Αν είναι **γνωστή η διασπορά σ²** ενός πληθυσμού τότε το **100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ** του είναι

$$\boxed{(\overline{x} - z_{1-a/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-a/2}\frac{\sigma}{\sqrt{n}})}$$

π.χ. Το καθαρό βάρος Χ μιας συσκευασίας βούτυρο ακολουθεί την κανονική κατανομή με τυπική απόκλιση σ = 5 gr. Σε τυχαίο δείγμα 25 πακέτων υπολογίσαμε το μέσο καθαρό βάρος $\overline{x}$=243 gr. Να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 0.99 για το μέσο καθαρό βάρος στο σύνολο των συσκευασιών.

1 - α = 0.99 $\Leftrightarrow$ $\alpha = 0.01$ $\Leftrightarrow$ **1- α/2 = 0.995** $\rightarrow$ **Φ(** $z_{1-a/2}$ **) = 1- α/2**

Φ($z_{1-a/2}$ **) = 0.995** $\rightarrow$ $z_{1-a/2}$ **= 2.58**

Άρα το **100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ** είναι:

$$(\overline{x} - z_{1-a/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-a/2}\frac{\sigma}{\sqrt{n}}) \Leftrightarrow (243 - 2.58\frac{5}{\sqrt{25}}, 243 + 2.58\frac{5}{\sqrt{25}}) \Leftrightarrow (240.42, 245.58)$$

Β. Αν δεν είναι **γνωστή η διασπορά σ²** ενός πληθυσμού και το δείγμα είναι μεγάλο τότε το **100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ** του είναι

$$\boxed{(\bar{x} - z_{1-a/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-a/2}\frac{s}{\sqrt{n}})}$$

όπου **s** είναι η **δειγματική τυπική απόκλιση**.

π.χ. Σε φορολογικό έλεγχο που έγινε σε 255 καταστήματα υπολογίστηκε το μέσο ποσό φοροκλοπής σε 540 ευρώ, με τυπική απόκλιση 110 ευρώ. Να εκτιμηθεί διάστημα εμπιστοσύνης με πιθανότητα 99% για το μέσο ποσό φοροκλοπής στο σύνολο των καταστημάτων.

1 - α = 0.99 $\Leftrightarrow$ α = 0.01 $\Leftrightarrow$ **1- α/2 = 0.995** $\rightarrow$ **Φ(**$z_{1-a/2}$**) = 1- α/2**

Φ($z_{1-a/2}$**) = 0.995** $\rightarrow$ $z_{1-a/2}$ **= 2.58**

Άρα το **100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ** είναι:

$$\bar{x} - z_{1-a/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{1-a/2}\frac{s}{\sqrt{n}} \Leftrightarrow$$

$$540 - 2.58\frac{110}{\sqrt{255}} < \mu < 540 + 2.58\frac{110}{\sqrt{255}} \Leftrightarrow$$

$$522.228 < \mu < 557.772$$

4.3. Ασκήσεις

4.3.1. Σε τυχαίο δείγμα 9 συσκευασιών γάλατος υπολογίσαμε το μέσο βάρος $\bar{x}$=96 gr. Αν είναι γνωστό ότι το βάρος X μιας συσκευασίας γάλατος ακολουθεί κανονική κατανομή με τυπική απόκλιση σ = 4.5 gr, να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 95% για το μέσο βάρος σε όλες τις συσκευασίες.

4.3.2. Σε τυχαίο δείγμα 35 μπαταριών υπολογίσαμε τη μέση διάρκεια ζωής 217 ώρες. Αν η τυπική απόκλιση του δείγματος είναι 47 ώρες, να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 95% για τη μέση διάρκεια ζωής στο σύνολο των μπαταριών.

Κεφάλαιο 5
ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

5.1. Εισαγωγή

Υπόθεση ονομάζεται η απόφαση που παίρνουμε για τα θέματα σχετικά με τους πληθυσμούς, βασιζόμενοι στις πληροφορίες που παίρνουμε από τα δείγματα των πληθυσμών. **Έλεγχος υποθέσεων ή έλεγχος σημαντικότητας ή κανόνες αποφάσεων** ή στατιστικός έλεγχος ονομάζεται η διαδικασία που χρησιμοποιείται ώστε να αποφασίσουμε αν θα δεχτούμε ή θα απορρίψουμε τις υποθέσεις που έχουμε κάνει. Μ**ηδενική υπόθεση H_0** ορίζεται ως η υπόθεση που κάνουμε αρχικά με σκοπό να την απορρίψουμε. **Εναλλακτική υπόθεση H_1** ορίζεται ως η ασυμβίβαστη υπόθεση σε σχέση με την μηδενική υπόθεση. Η απόφαση αν θα γίνει δεκτή ή αν θα απορριφθεί η μηδενική υπόθεση **H_0** στηρίζεται σε ένα στατιστικό που ονομάζεται **στατιστικό του τεστ**, το οποίο υπολογίζεται από τα δεδομένα του δείγματος. **Απορριπτική περιοχή R** της **H_0** ονομάζεται η περιοχή στα σημεία της οποίας η **H_0** απορρίπτεται.

Τα στοιχειά ενός στατιστικού ελέγχου είναι τα εξής:

- Ορίζεται η μηδενική υπόθεση **H_0**
- Ορίζεται η εναλλακτική υπόθεση **H_1**
- Ορίζεται το στατιστικό του ελέγχου από τα δεδομένα του δείγματος
- Ορίζεται η απορριπτική περιοχή **R** της υπόθεσης **H_0**
- Εξάγονται τα συμπεράσματα

Μονόπλευρος έλεγχος ορίζεται ως ο έλεγχος που κάνουμε όταν θέλουμε να δούμε αν η υπόθεση ότι η διαδικασία που επιλέξαμε είναι καλύτερη από μία άλλη είναι ευσταθής. Δηλαδή αν $H_0 : \theta = \theta_0$, τότε $H_1 : \theta > \theta_0$ ή $H_1 : \theta < \theta_0$. **Δίπλευρος έλεγχος** ορίζεται ως ο έλεγχος που γίνεται στις τιμές των άκρων της κατανομής που μελετάμε. Δηλαδή αν $H_0 : \theta = \theta_0$, τότε $H_1 : \theta \neq \theta_0$.

Σφάλμα τύπου Ι ονομάζεται η απόρριψη της υπόθεση **H_0** ενώ είναι σωστή. Η πιθανότητα αυτού του σφάλματος συμβολίζεται με α και ονομάζεται **επίπεδο σημαντικότητας ή σημαντικότητα ενός ελέγχου**. **Σφάλμα τύπου ΙΙ** ονομάζεται η αποδοχή της υπόθεση **H_0** ενώ είναι λάθος. Η πιθανότητα αυτού του σφάλματος συμβολίζεται με β.

Ο πληθυσμός βάση κάποιου κριτηρίου χωρίζεται σε δυο περιοχές: στην περιοχή αποδοχής της H_0 και στην περιοχή απόρριψης της που είναι η μικρότερη. Κατόπιν γίνεται ο έλεγχος βάση του κριτηρίου που ορίστηκε και αποφασίζεται που ανήκει το δείγμα. Εάν ανήκει στην περιοχή αποδοχής, τότε η H_0 γίνεται δεκτή. Θεωρούμε ότι ο εκτιμητής της παραμέτρου που εξετάζουμε στον έλεγχο ακολουθεί κανονική κατανομή.

5.2. ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗ ΜΕΣΗ ΤΙΜΗ ΠΛΗΘΥΣΜΟΥ

5.2.1. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 γνωστό

Αν $\bar{x}$ είναι η δειγματική μέση τιμή και σ η τυπική απόκλιση και μ_0 η επιθυμητή μέση τιμή, τότε υπολογίζουμε το στατιστικό:

$$z_\pi = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Μονόπλευρο τεστ

H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{z_\pi > z_{1-\alpha}\}$

ή

H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{z_\pi < -z_{1-\alpha}\}$

Δίπλευρο τεστ

H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{|z_\pi| > z_{1-\alpha/2}\}$

5.2.2. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 άγνωστο και $n \geq 30$

Ομοίως με πριν αλλά το σ αντικαθίσταται με το s (δειγματική τυπική απόκλιση).

5.2.3. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 άγνωστο και $n < 30$

Στην περίπτωση αυτή η μεταβλητή $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ ακολουθεί την κατανομή Student με $\nu = n-1$. Αν μ_0 είναι η επιθυμητή μέση τιμή, τότε τα κριτήρια απόφασης είναι:

Μονόπλευρο τεστ

H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{t_\pi > t_{\nu,\alpha}\}$ ή

H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{t_\pi < -t_{\nu,\alpha}\}$

Δίπλευρο τεστ

H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$ $\Rightarrow$ Απορριπτική περιοχή $R = \{|t_\pi| > t_{\nu,\alpha/2}\}$

5.3. Ασκήσεις

5.3.1. Το βάρος Χ των βρεφών που γεννήθηκαν το 2000 ακολουθεί την κανονική κατανομή Ν(3300,900). Σε τυχαίο δείγμα n=20 βρεφών το 2009 βρέθηκε ότι $\overline{x}$=3380. Να ελεγχθεί η υπόθεση ότι το μέσο βάρος των βρεφών δεν αυξήθηκε κατά την περίοδο 2000-2009, σε στάθμη σημαντικότητας α=0.05.

5.3.2. Ο αριθμός Χ των αμειβόμενων υπερωριών μιας επιχείρησης ακολουθεί κανονική κατανομή Ν(43,100). Η διοίκηση της επιχείρησης έδωσε εντολή να μειωθούν οι υπερωρίες. Σε τυχαίο δείγμα n=20 εργαζομένων βρέθηκε ότι ο μέσος όρος των υπερωριών είναι 32 ώρες. Να ελεγχθεί η υπόθεση ότι δεν μειώθηκε ο μέσος όρος των υπερωριών, σε στάθμη σημαντικότητας α=0.01.

5.3.3. Επιχείρηση σοκολατοποιΐας διαφημίζει ότι τα προϊόντα της έχουν μέση επικάλυψη σοκολάτας ίση με 10 γραμμάρια. Η επικάλυψη σοκολάτας Χ είναι γνωστό ότι ακολουθεί κανονική κατανομή με τυπική απόκλιση σ=1 γραμμάρια. Αν ο ιδιοκτήτης δέχεται να καταστραφεί η ημερήσια παραγωγή 1 στις 100 περιπτώσεις, και ο ποιοτικός έλεγχος μια ημέρα σε n = 25 προϊόντα έδειξε ότι $\overline{x}$=9.3 γραμμάρια, τότε θα αχρηστευτεί η ημερήσια παραγωγή;

5.3.4. Έστω το μέσο βάρος κάποιων προϊόντων δεν διαφέρει από 700 γραμμάρια σε επίπεδο σημαντικότητας α = 0.05. Αν σε τυχαίο δείγμα n=18 προϊόντων υπολογίσαμε μέσο βάρος $\overline{x}$= 692 γραμμάρια και τυπική απόκλιση s = 20 γραμμάρια, και η κατανομή του βάρους μπορεί να υποτεθεί κανονική, τότε να ελεγχθεί η παραπάνω υπόθεση.

Κεφάλαιο 6

ΕΠΑΝΑΛΗΠΤΙΚΕΣ ΑΣΚΗΣΕΙΣ

6.1. Δίνονται τα παρακάτω δεδομένα που παριστάνουν το πλήθος των χρωστούμενων μαθημάτων από μια ομάδα 25 φοιτητών του τμήματος :
1, 0, 0, 4, 5, 10, 7, 6, 2, 2, 3, 3, 4, 1, 0, 0, 1, 2, 2, 3, 3, 1, 1, 0, 0.
α) Ποιο είναι το ποσοστό των φοιτητών που χρωστούν δύο μαθήματα;
β) Βρείτε τη διάμεσο, το πρώτο και τρίτο τεταρτημόριο των παραπάνω παρατηρήσεων.
γ) Σχεδιάστε το θηκόγραμμα των παραπάνω παρατηρήσεων.
δ) Υπολογίστε το μέσο πλήθος των χρωστούμενων μαθημάτων και την τυπική απόκλιση αυτού.

6.2. Μια έρευνα έδειξε ότι το 35% των νέων αγοραστών αυτοκινήτων είναι γυναίκες. Επιλέγουμε τυχαία 5 νέους αγοραστές αυτοκινήτων. Να βρεθεί:
α) η πιθανότητα και οι 5 αγοραστές να είναι γυναίκες.
β) η πιθανότητα ακριβώς 3 από τους 5 αγοραστές να είναι γυναίκες
γ) το πολύ 2 αγοραστές να είναι γυναίκες
δ) η μέση τιμή και η διακύμανση του αριθμού των γυναικών που θα μπορούσαν να εκλεγούν.

6.3. Ένα εργοστάσιο παράγει μπαταρίες συγκεκριμένου τύπου για χρήση σε συσκευές κινητής τηλεφωνίας τις οποίες συσκευάζει σε πακέτα των 120. Σύμφωνα με στοιχεία του εργοστασίου ο χρόνος ζωής μιας μπαταρίας του τύπου αυτού μπορεί να θεωρηθεί ότι ακολουθεί την κανονική κατανομή με μέση τιμή μ=40 ώρες και τυπική απόκλιση σ=5 ώρες. Ακόμη σύμφωνα με τις προδιαγραφές που έχει θέσει η Ευρωπαϊκή Ένωση αν μία μπαταρία του συγκεκριμένου τύπου έχει χρόνο ζωής μικρότερο από 30 ώρες θεωρείται ελαττωματική. Με βάση τα στοιχεία αυτά:
α) Αν επιλεγεί τυχαία ένα πακέτο μπαταριών να υπολογισθεί ο αριθμός των μπαταριών των οποίων ο χρόνος ζωής θα ξεπερνά τις 45 ώρες καθώς και ο αριθμός των μπαταριών των οποίων ο χρόνος ζωής θα είναι μεταξύ 30 και 50 ώρες.
β) Αν επιλεγεί τυχαία ένα πακέτο μπαταριών να υπολογισθεί ο αριθμός των ελαττωματικών μπαταριών που περιέχει.
γ) Αν επιλεγούν τυχαία 4 μπαταρίες από ένα τυχαία επιλεγμένο πακέτο να υπολογισθεί η πιθανότητα δύο τουλάχιστον να είναι ελαττωματικές;

6.4. Ο χρόνος εκτύπωσης των αρχείων που αποστέλλονται από τους υπαλλήλους μιας εταιρείας στον κεντρικό εκτυπωτή ακολουθεί κανονική κατανομή με μέση τιμή 2 min και διακύμανση 6,25 min^2.
α) Να υπολογισθεί η πιθανότητα το επόμενο αρχείο που θα σταλεί για εκτύπωση να χρειαστεί χρόνο:
i) Μεγαλύτερο από 2,3 λεπτά.
ii) Μεταξύ 1,35 και 3 λεπτών.

β) Να υπολογισθεί η πιθανότητα μεταξύ 3 αρχείων που στέλνονται ανεξάρτητα για εκτύπωση το ένα ακριβώς να εκτυπωθεί σε χρόνο μικρότερο του 1,5 min.

6.5. Υποθέτουμε ότι ένα περιφερειακό κέντρο υπολογιστών θέλει να υπολογίσει την απόδοση της μνήμης του συστήματος με βάση τον μέσο χρόνο μεταξύ αποτυχιών εκτέλεσης μιας λειτουργίας του δίσκου της μνήμης. Για να υπολογίσει αυτήν την τιμή, το κέντρο ανέγραψε το χρόνο αποτυχιών για ένα τυχαίο δείγμα 45 αποτυχιών. Αν η δειγματική μέση τιμή είναι 1762 ώρες και η τυπική απόκλιση 215 ώρες:
α) να εκτιμήσετε την πραγματική μέση τιμή του χρόνου μεταξύ αποτυχιών με ένα 90% διάστημα εμπιστοσύνης.
β) Εάν ο δίσκος μνήμης του συστήματος τρέχει σωστά, η πραγματική μέση τιμή ξεπερνά τις 1700 ώρες. Βάση του διαστήματος εμπιστοσύνης που εκτιμήσατε στο ερώτημα α, το μπορείτε να συμπεράνετε για το δίσκο μνήμης.

6.6. Το IQ ενός ενήλικα μετρημένο σύμφωνα με ορισμένο τεστ ακολουθεί την κανονική κατανομή με μέση τιμή 100 και διασπορά 48,36. Ποια είναι η πιθανότητα ένας ενήλικας να έχει IQ πάνω από 107?

6.7. Σοκολατοποιία διαφημίζει ότι ορισμένο προϊόν της έχει μέση επικάλυψη σοκολάτας ίση με 10gr. Σε τυχαίο δείγμα 16 προϊόντων μετρήσαμε την επικάλυψη Χ και πήραμε τα ακόλουθα αποτελέσματα:
8.4, 8.5, 8.5, 8.7, 8.7, 8.8, 8.9, 9.0, 9.0, 9.2, 9.4, 9.5, 9.8, 9.9, 10.0, 10.5
Να βρεθεί το διάστημα εμπιστοσύνης με πιθανότητα 99% για τη μέση επικάλυψη σοκολάτας, αν η Χ ακολουθεί κανονική κατανομή.

6.8. Ο σύλλογος καταναλωτών για να εξακριβώσει αν τα κουτιά των 100 γραμμαρίων καφέ περιέχουν πραγματικά 100 γραμμάρια , πήρε 37 κουτιά καφέ τυχαία και βρήκε μέσο περιεχόμενο σε καφέ 96 γραμμάρια με τυπική απόκλιση 1,8 γραμμάρια. Να ελεγχθεί η υπόθεση ότι τα κουτιά των 100 γραμμαρίων είναι ελλιποβαρή, σε στάθμη σημαντικότητας α=0,05.

6.9. Ο αριθμός Χ των υπερωριών των εργαζομένων μιας τράπεζας ακολουθεί κανονική κατανομή με μέση τιμή 18 και διασπορά 4. Σε ένα τυχαίο δείγμα από 35 εργαζόμενους βρέθηκε ότι ο μέσος όρος των υπερωριών που έκαναν είναι 20 ώρες. Να ελεγχθεί η υπόθεση ότι δεν αυξήθηκε ο μέσος όρος ωρών των υπερωριών, σε στάθμη σημαντικότητας α=0.01.

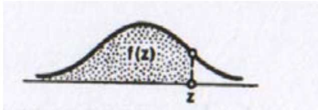
6.10. Ρωτήθηκαν 10 φοιτητές μιας σχολής που είναι στο πτυχίο για τον αριθμό Χ των μαθημάτων που χρωστάνε και έστω ότι οι απαντήσεις τους είναι οι εξής:

6, 2, 1, 9, 7, 3, 2, 1, 5, 1, 4, 6.

α) Να υπολογίσετε την δειγματική μέση τιμή και τυπική απόκλιση.
β) Αν γνωρίζετε ότι η μεταβλητή Χ ακολουθεί κανονική κατανομή, να υπολογίστε το 90% διάστημα εμπιστοσύνης για τη μέση τιμή του δείγματος.

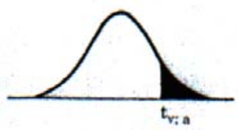
ΤΥΠΟΛΟΓΙΑ

Πίνακας 1. Αθροιστικές πιθανότητες της τυπικής κανονικής κατανομής



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Πίνακας 2. Τιμές $t_{v,a}$ της κατανομής Student ώστε $P(t_v > t_{v,a}) = a$.



β.ε.	a = .10	a = .05	a = .025	a = .010	a = .005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576