ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΣΕ ΟΙΚΟΝΟΜΙΚΑ ΔΕΔΟΜΕΝΑ

ΕΠΙΜΕΛΕΙΑ: ΑΝΑΣΤΑΣΟΠΟΛΟΥ ΝΙΚΟΛΕΤΑ

ΣΚΟΥΡΑΣ ΙΩΑΝΝΗΣ-ΔΗΜΗΤΡΗΣ

ΕΠΙΒΛΕΠΟΥΣΑ: ΚΑΛΑΠΟΔΗ ΑΛΕΚΑ
ΚΑΘΗΓΗΤΡΙΑ

` **ΠΑΤΡΑ-2013**

ΠΡΟΛΟΓΟΣ

Το θέμα αφορά όπως αναφέρει και ο τίτλος του τι είδους συμπεράσματα προκύπτουν από την χρήση της στατιστικής σε οικονομικά δεδομένα. Καθώς το εύρος της στατιστικής επιστήμης είναι εξαιρετικά μεγάλο η παρούσα έρευνα έχει επικεντρωθεί στο να παρουσιάσει την χρήση των διαστημάτων εμπιστοσύνης και τον έλεγχο υποθέσεων έτσι ώστε να εισαγάγει ομαλά και αποτελεσματικά τον αναγνώστη στην χρήση της στατιστικής για εξαγωγή συμπερασμάτων σε πρακτικά οικονομικά προβλήματα. Επικεντρωθήκαμε ιδιαίτερα σε διαστήματα εμπιστοσύνης για την μέση τιμή σε μικρά και μεγάλα δείγματα, στην διαφορά μέσων τιμών δύο πληθυσμών, στο λόγο των διασπορών και στις αναλογίες εντός ενός πληθυσμού. Αντίστοιχα στον έλεγχο υποθέσεων ερευνήσαμε θέματα στις προαναφερθείσες περιοχές αλλά όχι από την οπτική της εύρεσης ενός διαστήματος εμπιστοσύνης. Ερευνήσαμε αυτές τις περιοχές μέσω της αποδοχής ή όχι της μηδενικής υπόθεσης που κάθε φορά θέταμε ανάλογα με το πρόβλημα που έπρεπε να διαχειριστούμε. Επιπλέον για να είναι πιο εύκολη η κατανόηση των διαστημάτων εμπιστοσύνης και του ελέγχου υποθέσεων αναφερθήκαμε εκτενώς στις στατιστικές κατανομές που χρησιμοποιούνται για την κατασκευή των παραπάνω. Μέσω της επεξήγησης του τρόπου κατασκευής των στατιστικών αυτών εργαλείων που βασίζονται σε συγκεκριμένες στατιστικές κατανομές είναι πιο εύκολη η χρήση τους σε συγκεκριμένα οικονομικά προβλήματα. Στο τέλος παραθέσαμε οικονομικές εφαρμογές που δείχνουν την χρήση της στατιστικής συμπερασματολογίας σε πρακτικά προβλήματα.

ΠΕΡΙΛΗΨΗ

Στη συγκεκριμένη έρευνα ερευνούμε τα στατιστικά συμπεράσματα που μπορούν να προκύψουν σε οικονομικά δεδομένα. Πιο συγκεκριμένα στο πρώτο κεφάλαιο εξετάζουμε τις κατανομές δειγματοληψίας, δείγματα από κανονικούς πληθυσμούς και τις κατανομές $\bar{X}$, X^2, S^2 την F-κατανομή και την t-κατανομή. Στο δεύτερο κεφάλαιο εξετάζουμε τα διαστήματα εμπιστοσύνης για τον μέσο όρο μεγάλων και μικρών δειγμάτων, την διασπορά για μικρά και μεγάλα δείγματα, του ποσοστού, την διαφορά των μέσων όρων για μεγάλα και μικρά δείγματα, την διαφορά των αναλογιών, και τον λόγο των διαπορών κανονικών διασπορών. Στη συνέχεια στο τρίτο κεφάλαιο εξετάζουμε τον έλεγχο υποθέσεων για όλες τις παραπάνω περιπτώσεις. Στο τέταρτο και τελευταίο κεφάλαιο παρουσιάζουμε εφαρμογές των κεφαλαίων δύο και τρία συνήθως σε πρακτικά οικονομικά προβλήματα.

ΠΕΡΙΕΧΟΜΕΝΑ

Εισαγωγή

Η έρευνα που κάνουμε στην Στατιστική αφορά ομάδες ατόμων, αντικειμένων ή άλλων οντοτήτων. Οι ομάδες αυτές αποτελούν το επίκεντρο σε μια μελέτη και ονομάζονται πληθυσμός. Επειδή συχνά η μελέτη ολόκληρου του πληθυσμού δεν είναι πρακτικά δυνατή, καθώς το δείγμα μπορεί να είναι πολύ μεγάλο, τότε αντικείμενο της μελέτης γίνεται ένα υποσύνολο του πληθυσμού, που ονομάζεται δείγμα. Δηλαδή, το δείγμα είναι ένα υποσύνολο ενός πληθυσμού που έχει επιλεγεί κατάλληλα, ώστε να αντιπροσωπεύει τον πληθυσμό. Η διαδικασία επιλογής του δείγματος είναι πολύ σημαντική, καθώς καθορίζει την εγκυρότητα των αποτελεσμάτων. Αν το δείγμα δεν είναι αντιπροσωπευτικό υπάρχει μεγάλη πιθανότητα η μελέτη να οδηγήσει σε εσφαλμένα συμπεράσματα.

Για να είναι ένα δείγμα αντιπροσωπευτικό πρέπει να χρησιμοποιηθεί η διαδικασία της τυχαίας δειγματοληψίας. Σε αυτή τη διαδικασία πρέπει κάθε υποσύνολο να έχει την ίδια πιθανότητα να επιλεγεί. Όπως στον πληθυσμό, ένα δείγμα μπορεί να είναι από πολύ μικρό έως και πολύ μεγάλο. Όσο μεγαλύτερο είναι το δείγμα, τόσο πιο αντιπροσωπευτικό θα είναι καθώς αυξάνεται ο αριθμός των στοιχείων του δείγματος που επιλέγονται από τον πληθυσμό. Ωστόσο, όσο αντιπροσωπευτικό και να θεωρείται ένα δείγμα πάντα θα υπάρχει ένα ποσοστό λάθους ανάμεσα στο στατιστικό δείκτη (οι δείκτες που περιγράφουν ένα δείγμα π.χ δειγματικός μέσος $\bar{X}$) και την αντίστοιχη τιμή της παραμέτρου του πληθυσμού (δείκτες που περιγράφουν τον πληθυσμό π.χ μέσος όρος μ). Αυτό το ποσοστό λάθους λέγεται σφάλμα δειγματοληψίας. Το σφάλμα δειγματοληψίας τονίζει ότι ένας στατιστικός δείκτης αποτελεί μόνο μια εκτίμηση της αντίστοιχης παραμέτρου του πληθυσμού.

Μια μελέτη απαιτεί τη συλλογή ενός συνόλου δεδομένων. Τα δεδομένα που συλλέγονται μπορεί να είναι ποιοτικά ή ποσοτικά.

Ποιοτικά είναι αυτά που αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό και δεν μπορούμε να τα μετρήσουμε, όπως το φύλο, το επίπεδο μόρφωσης.

Ποσοτικά ονομάζονται αυτά των οποίων οι τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μια μονάδα μέτρησης. Αυτές ειδικότερα διακρίνονται σε:

- Συνεχή: μπορούν να πάρουν οποιαδήποτε τιμή σε ένα συνεχές διάστημα, όπως βάρος, ύψος, αντοχή

- Διακριτά: μπορούν να πάρουν μόνο ακέραιες τιμές, όπως αριθμός ελαττωματικών προϊόντων, ο αριθμός παιδιών σε μια οικογένεια

Η Επαγωγική Στατιστική είναι ο κλάδος που περιλαμβάνει ένα σύνολο μεθόδων για την εκτίμηση, πρόβλεψη ή απόφαση σχετικά με τον πληθυσμό με βάση ένα τυχαίο δείγμα. Μας επιτρέπει να καθορίσουμε με σχετική ακρίβεια την πιθανότητα σφάλματος που περιέχεται στα συμπεράσματα στα οποία εξάγονται μετά από στατιστική ανάλυση.

Οι μέθοδοι της Επαγωγικής Στατιστικής διακρίνονται σε:

A. Παραμετρικές: εδώ προϋποθέτουμε ότι είναι γνωστή η κατανομή του πληθυσμού (κανονική) και άγνωστες οι παράμετροι της κατανομής.

B. Μη Παραμετρικές: Δεν προϋποθέτουμε ότι ο πληθυσμός από τον οποίο προέρχεται το δείγμα είναι κανονικός, αλλά ούτε ότι έχει κάποια άλλη γνωστή κατανομή.

Η Επαγωγική Στατιστική περιλαμβάνει:

- Στατιστική Εκτίμηση: είναι η διαδικασία με την οποία «εκτιμούμε» τις τιμές με βάση τα δεδομένα του δείγματος, δηλαδή υπολογίζουμε ποια είναι η πιθανότερη τιμή των χαρακτηριστικών του πληθυσμού. Έχουμε δυο είδη στατιστικής εκτίμησης:

  i. Εκτίμηση σημείου

  ii. Εκτίμηση διαστήματος

- Έλεγχοι Στατιστικών Υποθέσεων: είναι η διαδικασία με την οποία ελέγχουμε την ορθότητα μιας υπόθεσης που κάναμε για κάποιο χαρακτηριστικό

ΚΕΦΑΛΑΙΟ 1

Κατανομές Δειγματοληψίας

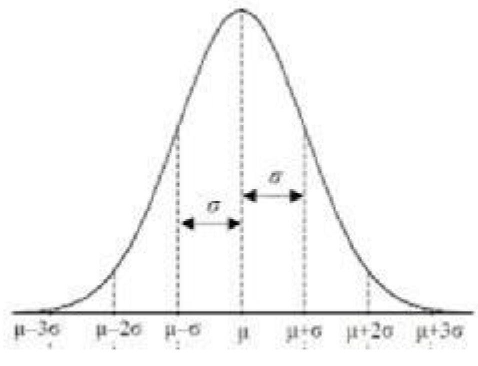
Στο κεφάλαιο αυτό παρουσιάζεται η κατανομή δειγματοληψίας.Με την κατανομή δειγματοληψίας μπορεί να προσδιοριστούν οι παράμετροι της κατανομής ενός πληθυσμού (για παράδειγμα μ, σ) από τις αντίστοιχες παραμέτρους των δειγμάτων που παίρνουμε από τον πληθυσμό. Δηλαδή να εκτιμήσουμε τα άγνωστα μ και σ από τα γνωστά $\overline{X}$ και S.

Αν από ένα πληθυσμό με Ν στοιχεία πάρουμε άπειρα ισοπληθή τυχαία δείγματα $\nu_1=\nu_2=\nu_3=\ldots=\nu$ και από κάθε δείγμα υπολογίσουμε μια ορισμένη παράμετρο (π. χ αριθμητικό μέσο, διακύμανση) τότε προκύπτει μια σειρά ομοειδών εκτιμήσεων των αντίστοιχων παραμέτρων του πληθυσμού. Οι εκτιμήσεις αυτές των δειγμάτων μπορούν να ταξινομηθούν σε κατανομές συχνοτήτων. Σε κάθε κατανομή, οι συχνότητες θα τείνουν να συγκεντρωθούν γύρω από μια κεντρική τιμή και θα κατανέμονται γύρω από αυτήν συμμετρικά ή ασύμμετρα. Όταν ο αριθμός των δειγμάτων θα είναι μεγάλος, τότε οι τιμές θα παρουσιάζουν μια συνέχεια (δηλ. η μια τιμή θα είναι πολύ κοντά στην άλλη) και η γραφική απεικόνιση της κατανομής θα τείνει να σχηματίσει μια ομαλή και συνεχή καμπύλη συχνοτήτων η οποία ονομάζεται κατανομή δειγματοληψίας της εξεταζόμενης παραμέτρου.

Δηλαδή, δειγματική κατανομή ονομάζεται η κατανομή συχνοτήτων ενός στατιστικού μεγέθους που μετράται σε όλα τα τυχαία δείγματα μεγέθους ν που μπορούμε να πάρουμε από τον πληθυσμό.

Πολύ συχνά χρησιμοποιούμε την κανονική κατανομή για διάφορους λόγους που θα δούμε παρακάτω ως προς το τι μας εξυπηρετεί. Αλλά ας δούμε τι είναι μια κανονική κατανομή. Η κανονική κατανομή είναι μια συνεχής κατανομή που χρησιμοποιήθηκε πρώτη φορά το1733 από το Γάλλο μαθηματικό Abraham de Moivre. Η εφαρμογή της όμως στην στατιστική άρχισε μετά τις εργασίες του Pierres Laplace και του μαθηματικού Karl Gauss. Τα χαρακτηριστικά γνωρίσματα της κανονικής κατανομής είναι:

Ο μαθηματικός της τύπος είναι $f(x)=\frac{1}{s\sqrt{2p}}e^{-\frac{1}{2}(\frac{x-m}{s})^2}$ όπου μ ο μέσος αριθμητικός της τυχαίας μεταβλητής Χ και σ η τυπική απόκλιση της. Η καμπύλη που παριστάνει τη συνάρτηση πυκνότητας πιθανότητας έχει την μορφή καμπάνας, είναι συμμετρική και έχει ασύμπτωτο τον άξονα των τετμημένων, δηλαδή τον άξονα των χ, είναι συμμετρική ως προς την ευθεία χ=μ που είναι ο μέσος όρος όλων των δυνατών τιμών της τυχαίας μεταβλητής Χ, η μέση τιμή η διάμεσος και η επικρατούσα τιμή συμπίπτουν λόγο συμμετρίας της καμπύλης κατανομής και το συνολικό εμβαδόν που περικλείεται από την καμπύλη και τον άξονα Χ είναι μοναδιαίο.



Για το Κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία από τις πηγές [3], [4].

1.1 Ο δειγματικός μέσος

Έστω δείγμα $\boldsymbol{X_1, X_2, \mathbf{K}, X_n}$ προερχόμενο από κάποιο πληθυσμό με μέση τιμή m και μεταβλητότητα s^2. Η μέση τιμή $\overline{X}$ και η $\boldsymbol{S}^2$ μεταβλητότητα του δείγματος υπολογίζονται ως εξής:

$$\overline{\boldsymbol{X}} = \frac{\sum \boldsymbol{X_i}}{\boldsymbol{n}}, \quad \boldsymbol{S}^2 = \frac{\sum \boldsymbol{X}_i^2 - \boldsymbol{n}\overline{\boldsymbol{X}}^2}{\boldsymbol{n} - 1}$$

Η τυποποιημένη μορφή μιας τυχαίας μεταβλητής Χ είναι μια άλλη μεταβλητή Ζ, η οποία συνδέεται με τη Χ με τη σχέση : $Z = \frac{X - m}{s}$

Όπου μ είναι ο μέσος και σ η τυπική απόκλιση της Χ.

Όταν η μεταβλητή Χ ακολουθεί την κανονική κατανομή, η τυποποιημένη μορφή της είναι η Ζ-κατανομή με μέσο μ=0 και διακύμανση $\sigma^2=1$

Αν η τυχαία μεταβλητή Χ είναι ο δειγματικός μέσος $\overline{X}$,τότε ο τυποποιημένος μέσος είναι:

$$z = \frac{X - m}{\frac{s}{\sqrt{n}}}$$

Έστω ότι έχουμε έναν πληθυσμό Ν στοιχείων με μέσο μ και διακύμανση σ^2 . Αν από αυτόν τον πληθυσμό πάρουμε μεγάλο αριθμό ισοπληθών δειγμάτων ($n_1=n_2=\ldots=n$), υπολογίσουμε τους μέσους αυτών των δειγμάτων ($\overline{X}_1=\overline{X}_2=\ldots=\overline{X}_n$) και βάση αυτών κατασκευάσουμε μια κατανομή συχνοτήτων, η κατανομή αυτή θα έχει τις εξής ιδιότητες:

- Αν ο πληθυσμός ακολουθεί κανονική κατανομή με γνωστή μέση τιμή μ και διακύμανση σ^2 , τότε η δειγματική κατανομή του μέσου $\overline{X}$ για δείγματα μεγέθους n, ακολουθεί κανονική κατανομή με μέσο μ= $m_{\bar{x}}$ και τυπική απόκλιση $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ τόσο περισσότερο όσο το ν αυξάνει.

- Αν ο πληθυσμός έχει άγνωστη κατανομή, αλλά το μέγεθος του δείγματος είναι μεγάλο ν > 30 τότε ισχύει το ίδιο με το παραπάνω .

Η τυπική απόκλιση $s_{\bar{x}}$ της δειγματικής κατανομής του μέσου $\overline{X}$ ονομάζεται *τυπικό σφάλμα του μέσου* $\overline{X}$ και φανερώνει πόσο καλά ο δειγματικός μέσος εκτιμά το μέσο μ του πληθυσμού. Όσο μικρότερη είναι η τυπική απόκλιση $s_{\bar{x}}$ τόσο πιο αξιόπιστη θα είναι η εκτίμηση του μέσου του πληθυσμού. Επίσης, από τον τύπο του τυπικού σφάλματος

$s_{\bar{x}} = \frac{s}{\sqrt{n}}$ φαίνεται:

- Όσο μικρότερη είναι η τυπική απόκλιση σ του πληθυσμού, τόσο μικρότερο είναι και το τυπικό σφάλμα.

§ Όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο μικρότερο είναι το τυπικό σφάλμα.

Για παράδειγμα αν το ύψος των Ελλήνων ακολουθεί την κανονική κατανομή με μ=170 εκ. και σ^2=100 τ.εκ τότε ο δειγματικός μέσος $\bar{X}$ του δείγματος 25 ατόμων ακολουθεί την κανονική κατανομή με μέση τιμή 170 και διασπορά 100/25=4.

1.2 Η κατανομή X^2

Έστω, ότι υπάρχουν $X_1, X_2 \ldots X_n$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση ίση με 1, τότε το άθροισμα τετραγώνων των τυχαίων μεταβλητών που αναφέρονται παραπάνω ακολουθούν την κατανομή X^2_n με n βαθμούς ελευθερίας.

Αν τώρα S^2 είναι η δειγματική διασπορά τυχαίου δείγματος μεγέθους n από μια κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 τότε η τυχαία μεταβλητή $\dfrac{(n-1)S^2}{s^2}$ ακολουθεί την κατανομή X^2 με n-1 βαθμούς ελευθερίας.

<u>Παράδειγμα</u>: Το πάχος ενός τμήματος ημιαγωγού το οποίο μας ενδιαφέρει απόλυτα η καλή λειτουργία του συστήματος. Η μέθοδος κατασκευής του θεωρείται ότι ελέγχεται όταν η πραγματική διακύμανση των παχών των κατασκευαζόμενων μερών έχει μια τυπική απόκλιση όχι μεγαλύτερη από σ=0,60 χιλιοστά του εκατοστού. Παίρνουμε κατά περιοδικά διαστήματα δείγματα μεγέθους n=20 και θεωρούνται εκτός ελέγχου αν η πιθανότητα το s^2 να πάρει τιμή μεγαλύτερη ή ίση από την παρατηρούμενη δειγματική τιμή είναι το πολύ 0,01. Σύμφωνα με τα παραπάνω, η μέθοδος κρίνεται εκτός ελέγχου, αν $\dfrac{(n-1)S^2}{s^2}$ με n=20, σ=0,60 υπερβεί το $X^2_{0,01,\,19}$=36,191 δηλαδή $\dfrac{(n-1)S^2}{s^2}$ >36,191=>s^2>36.191*0.6*0.6/19=0.68572=>s > 0.8281

1.3 Η F-κατανομή

Έστω ότι υπάρχουν X_1, X_2 δύο ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή X^2 με v_1 και v_2 βαθμούς ελευθερίας αντίστοιχα, τότε τη τυχαία μεταβλητή

$$Y = \frac{X_1 / n_1}{X_2 / n_2}$$

ακολουθεί την F κατανομή με v_1 και v_2 βαθμούς ελευθερίας.

Εάν για παράδειγμα έχουμε δύο δείγματα από κανονικές κατανομές μεγέθους n_1 και n_2 αντίστοιχα τότε οι τυχαίες μεταβλητές $\dfrac{(n_1-1)S_1^2}{s_1^2}$ και $\dfrac{(n_2-1)S_2^2}{s_2^2}$ είναι ανεξάρτητες και ακολουθούν σύμφωνα με τα προηγούμενα την κατανομή X^2 με n_1-1 και n_2-1 βαθμούς ελευθερίας αντίστοιχα. Τότε η τυχαία μεταβλητή:

$$F=\frac{\frac{\frac{(n_1-1)S_1^2}{s_1^2}}{n_1-1}}{\frac{\frac{(n_2-1)S_2^2}{s_2^2}}{n_2-1}}=\frac{\frac{S_1^2}{s_1^2}}{\frac{S_2^2}{s_2^2}}=\frac{S_1^2}{S_2^2}\cdot\frac{s_2^2}{s_1^2}$$

Ακολουθεί την F κατανομή με n_1-1 και n_2-1 βαθμούς ελευθερίας. Στη περίπτωση που οι πληθυσμιακές διασπορές είναι ίσες, τότε αυτό σημαίνει ότι ο λόγος $\frac{S_1^2}{S_2^2}$ ακολουθεί την F κατανομή με n_1-1 και n_2-1 βαθμούς ελευθερίας. Λόγω αυτού, η κατανομή F είναι γνωστή και σαν κατανομή του λόγου των διακυμάνσεων.

1.4 Η t-κατανομή

Η κατανομή t είναι παράγωγος κατανομή της κανονικής συνάρτησης και την ορίζουμε ως εξής:
Αν η τυχαία μεταβλητή Z ακολουθεί την κανονική κατανομή με μέση τιμή 0 και διακύμανση 1 και η Y, μια άλλη μεταβλητή ανεξάρτητη της X, που ακολουθεί την κατανομή X^2 με ν=n-1 βαθμούς ελευθερίας τότε η τυχαία συνεχής μεταβλητή:

$$t_n=\frac{\mathrm{Z}}{\sqrt{\mathrm{Y}/n}}$$

ακολουθεί την κατανομή t με ν βαθμούς ελευθερίας και έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(t_n)=\frac{\Gamma[n+1/2]}{\sqrt{np}\,\Gamma(n/2)}\cdot\frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}},-\infty<t<\infty$$

όπου Γ η κατανομή Γάμμα.

Αν τώρα $\bar{X}$ και S^2 ο δειγματικός μέσος όρος και η δειγματική διακύμανση δείγματος μεγέθους n από μια κανονική κατανομή με μέση τιμή μ και διασπορά σ^2,τότε η τυχαία μεταβλητή $t=\frac{\bar{X}-m}{S/\sqrt{n}}$ ακολουθεί την t_{n-1} κατανομή.

Στη περίπτωση που $\bar{X}_1,\bar{X}_2$ οι μέσοι όροι και S_1^2,S_2^2 οι διακυμάνσεις δύο ανεξάρτητων τυχαίων δειγμάτων με μεγέθη n_1 και n_2 αντίστοιχα, από δύο πληθυσμούς με κανονικές κατανομές με μέση τιμή μ_1 και μ_2 αντίστοιχα και κοινή διασπορά σ^2 τότε η τυχαία μεταβλητή

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(m_1-m_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}$$

ακολουθεί την $t_{n_1+n_2-2}$ κατανομή.

ΚΕΦΑΛΑΙΟ 2

Εκτίμηση Σημείου & Διαστήματα Εμπιστοσύνης

Σκοπός της στατιστικής είναι η εξαγωγή συμπερασμάτων που αφορούν τον πληθυσμό ή το φαινόμενο που μελετάμε, με τη βοήθεια ενός τυχαίου δείγματος. Έτσι, αν υπάρχουν κάποιες παρατηρήσεις $x_1, x_2 .\mathbf{K}, x_n$ οι οποίες έχουν συγκεντρωθεί επαναλαμβάνοντας ένα πείραμα n φορές, με τη βοήθειά τους μπορούν να προσδιοριστούν οι άγνωστες παράμετροι της κατανομής του πληθυσμού. Δηλαδή αν θέλουμε να εκτιμήσουμε μια άγνωστη παράμετρο για τον πληθυσμό του παραπάνω δείγματος, για παράδειγμα την μέση τιμή τους τότε θα πρέπει να βρούμε την κατάλληλη στατιστική συνάρτηση μέσω της οποίας θα υπολογίσουμε-εκτιμήσουμε την άγνωστη παράμετρο για τον πληθυσμό Όταν υπάρχει μια μόνο άγνωστη παράμετρος, θα συμβολίζεται με $f(x,q)$ η συνάρτηση πιθανότητας του πληθυσμού. Έτσι λοιπόν καταλήγουμε στον παρακάτω ορισμό:

<u>Ορισμός:</u> Εκτιμήτρια συνάρτηση ή εκτιμητής μιας παραμέτρου θ, ονομάζεται εκείνη η στατιστική συνάρτηση Τ($X_1, X_2, \mathbf{K}, X_n$) που χρησιμοποιείται για την εκτίμηση της παραμέτρου θ.

Η τιμή της εκτιμήτριας συνάρτησης για ένα συγκεκριμένο τυχαίο δείγμα Τ($x_1, x_2, ..., x_n$) για την οποία θέλουμε να είναι όσο το δυνατόν πιο κοντά στο θ, λέγεται εκτίμηση της παραμέτρου θ και συμβολίζεται με $\hat{q}$.

Με την εκτίμηση παραμέτρων του πληθυσμού ασχολείται η εκτιμητική και προτείνει δύο ειδών εκτιμητές: εκτιμητές σε σημείο και εκτιμητές σε διάστημα. Μάλιστα υπάρχουν διάφορες μέθοδοι εκτίμησης παραμέτρων σε σημείο:

- η μέθοδος μέγιστης πιθανοφάνειας
- η μέθοδος των ροπών
- η μέθοδος ελάχιστων τετραγώνων

Βέβαια εδώ πρέπει να τονίσουμε ότι για μια συγκεκριμένη παράμετρο, μπορεί να προτείνονται περισσότεροι από ένας εκτιμητής. Σε αυτή την περίπτωση τα οποία είναι:

- Αμεροληψία
- Συνέπεια
- Αποτελεσματικότητα
- Επάρκεια

Για το Κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία από τις πηγές [5], [6].

2.1 Διαστήματα εμπιστοσύνης για τη μέση τιμή μ (μεγάλα δείγματα)

Με τον όρο διάστημα εμπιστοσύνης εννοούμε ένα διάστημα τιμών μέσα στο οποίο είναι πολύ πιθανό να κυμαίνεται η εκτιμώμενη παράμετρος του πληθυσμού. Το διάστημα αυτό τιμών είναι το ζητούμενο αυτής της ενότητας και των επόμενων που ακολουθούν σε αυτό το κεφάλαιο. Ακόμα προτού προχωρήσουμε στα διαστήματα εμπιστοσύνης πρέπει να τονίσουμε ότι με την παράμετρο α (συνήθως 1% ή 5%) συμβολίζουμε την πιθανότητα η παράμετρος του πληθυσμού για την οποία δημιουργούμε το διάστημα εμπιστοσύνης να μην περιέχεται τελικά εντός αυτού του διαστήματος εμπιστοσύνης αλλά εκτός αυτού.

Στη περίπτωση που ο πληθυσμός ακολουθεί την κανονική κατανομή και η μεταβλητότητα είναι γνωστή τότε το 100(1-α)% διάστημα εμπιστοσύνης για την μέση τιμή του πληθυσμού είναι το παρακάτω:

$$\overline{X} - z_{\frac{a}{2}} \frac{s}{\sqrt{n}} \leq m \leq \overline{X} + z_{\frac{a}{2}} \frac{s}{\sqrt{n}}.$$

Παράδειγμα

Σε μια ομάδα 36 ασθενών μετριέται το επίπεδο της χοληστερίνης τους. Ο μέσος όρος των τιμών της χοληστερίνης είναι 211 ενώ η τυπική απόκλιση είναι ίση με 23. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για το μέσο του πληθυσμού.

Έχουμε: $\overline{x}$ =211 σ=23 ν=36

<u>(βαθμός εμπιστοσύνης)</u>: γ=1-α → 95%=1-α → α= 0,05 άρα $z_{0,025} = 1.96$

$$\overline{x} \pm z\frac{s}{\sqrt{n}} = 211 \pm 1.96\frac{2}{\sqrt{36}} = 211 \pm 7.51 = \mathbf{203.49 \leq \mu \leq 218.51}$$

Παράδειγμα

Ένα ζυθοποιείο ξέρει ότι η ποσότητα μπύρας σε ένα κουτί ακολουθεί μια κανονική κατανομή με τυπική απόκλιση 0,2 ουγκιές. Παίρνοντας ένα τυχαίο δείγμα 25 κουτιών κατασκευάσθηκε ένα διάστημα εμπιστοσύνης για το μέσο του πληθυσμού ως εξής: $11{,}98 < m < 12{,}12$. Ποιο είναι το επίπεδο εμπιστοσύνης του παραπάνω διαστήματος;

Έχουμε κανονικό πληθυσμό και γνωστή διακύμανση, το (1-α)% ΔΕ για το μέσο δίνεται από,

$\overline{X} - z_{a/2}\frac{0{,}2}{\sqrt{25}} < m < \overline{X} + z_{a/2}\frac{0{,}2}{\sqrt{25}}$. Το εύρος του διαστήματος είναι $2\left(z_{a/2}\frac{0{,}2}{\sqrt{25}}\right)$,

Λύνοντας, $12{,}12 - 11{,}98 = 2\left(z_{a/2}\frac{0{,}2}{\sqrt{25}}\right)$, έχουμε $z_{a/2} = \frac{0{,}14}{2}\frac{5}{0{,}2} = 1{,}75$. Χρησιμοποιούμε το πίνακα της N(0,1) και το επίπεδο εμπιστοσύνης είναι 91,98%($1 - 2(1 - \Phi(1{,}75))$)

2.2 Διαστήματα εμπιστοσύνης για τη μέση τιμή μ (μικρά δείγματα)

Στη περίπτωση που ο πληθυσμός ακολουθεί την κανονική κατανομή και το δείγμα είναι μικρό τότε το 100(1-α)% διάστημα εμπιστοσύνης για την μέση τιμή του πληθυσμού είναι το παρακάτω

$$\overline{X} - t_{\frac{a}{2},n-1}\frac{S}{\sqrt{n}} \leq m \leq \overline{X} + t_{\frac{a}{2},n-1}\frac{S}{\sqrt{n}}.$$

Παράδειγμα

Στον επόμενο πίνακα φαίνεται ο μέσος χρόνος αναμονής πελατών σε ουρά μιας τράπεζας όπου έγιναν 6 επαναλήψεις με τα εξής αποτελέσματα:

Επανάληψη	Μέσος χρόνος αναμονής (λεπτά)
1	14,1
2	15,2
3	11,9
4	13,4
5	16
6	14,3

Να υπολογιστεί 95% διάστημα εμπιστοσύνης για το μέσο χρόνο αναμονής.

Υπολογίζουμε τη μέση τιμή και τη μεταβλητότητα του δείγματος.

$$\overline{X} = \frac{\sum X_i}{6} = 14{,}15$$

$$S^2 = \frac{\sum X_i^2 - 6\overline{X}^2}{5} = 2{,}035$$

Τα όρια του 95% διαστήματος εμπιστοσύνης είναι τα

$$\overline{X} \pm t_{0,025,5}\frac{S}{\sqrt{6}} = 14{,}15 \pm 2{,}57\sqrt{\frac{2{,}035}{6}} = 14{,}15 \pm 1{,}5.$$

2.3 Διάστημα εμπιστοσύνης για τη διασπορά ενός πληθυσμού (μικρά και μεγάλα δείγματα)

Για πληθυσμούς που ακολουθούν κανονική κατανομή, ένα $100(1-a)$**%** διάστημα εμπιστοσύνης για την μεταβλητότητα του πληθυσμού δίνεται από τον τύπο

$$\frac{(n-1)S^2}{c^2_{\frac{a}{2},n-1}} \le s^2 \le \frac{(n-1)S^2}{c^2_{1-\frac{a}{2},n-1}}.$$

Ο παραπάνω τύπος ισχύει όταν δεν είναι γνωστή η μέση τιμή του πληθυσμού. Στην περίπτωση που είναι γνωστή η μέση τιμή του πληθυσμού τότε ένα $100(1-a)$**%** διάστημα εμπιστοσύνης για την μεταβλητότητα του πληθυσμού δίνεται από τον τύπο

$$\frac{(n-1)S^2}{c^2_{\frac{a}{2},n}} \le s^2 \le \frac{(n-1)S^2}{c^2_{1-\frac{a}{2},n}}.$$

Παράδειγμα

Έστω ότι από έναν κανονικό πληθυσμό παίρνουμε δείγμα 4 μονάδων όπου οι παρατηρήσεις είναι: 50, 60, 48, 74. Αν ο μέσος του πληθυσμού είναι μ=50, ζητείται να εκτιμηθεί το 95% διάστημα εμπιστοσύνης της διακύμανσης.

Το διάστημα της διακύμανσης δίνεται από,

$$\frac{(n-1)S^2}{c^2_{\frac{a}{2},n}} \le s^2 \le \frac{(n-1)S^2}{c^2_{1-\frac{a}{2},n}}$$

$$\frac{\sum (x_i-m)^2}{c^2_{n,a/2}} < s^2 < \frac{\sum (x_i-m)^2}{c^2_{n,1-a/2}}$$

$$39,5 < s^2 < 916,7$$

όπου $a=0,05$ $\quad c^2_{4,a/2}=11,14$ $\quad c^2_{4,1-a/2}=0,48$. Να σημειωθεί ότι οι βαθμοί ελευθερίας είναι 4 και όχι 3 επειδή χρησιμοποιούμε μ και όχι $\overline{X}$ στον τύπο της διακύμανσης του δείγματος.

2.4 Διαστήματα εμπιστοσύνης για την διαφορά των μέσων τιμών δύο πληθυσμών (μεγάλα δείγματα)

Για μεγάλα δείγματα, υποθέτοντας πληθυσμούς με ίσες μεταβλητότητες, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για την διαφορά των μέσων τιμών δύο πληθυσμών δίνεται από τον τύπο

$$\overline{X}_1 - \overline{X}_2 - z_{1-\frac{a}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le m_1 - m_2 \le \overline{X}_1 - \overline{X}_2 + z_{1-\frac{a}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Για πληθυσμούς που ακολουθούν κανονική κατανομή και οι μεταβλητότητες είναι γνωστές το διάστημα εμπιστοσύνης είναι

$$\overline{X}_1 - \overline{X}_2 - z_{1-\frac{a}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le m_1 - m_2 \le \overline{X}_1 - \overline{X}_2 + z_{1-\frac{a}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Παράδειγμα

Ακολουθεί ένα σύνολο εσόδων από πωλήσεις (σε χιλιάδες φύλλα) μιας εφημερίδας σε δύο γειτονικές πόλεις για μια περίοδο λίγων ημερών:

Πόλη Α: 25,13, 14, 19, 23, 30, 35, 29, 28, 17, 17, 16, 13, 18, 20

Πόλη Β: 10,12, 15, 13, 7, 6, 11, 5, 9, 14, 15, 18, 17, 16, 12, 12, 10, 11, 13, 14

Να κατασκευάσετε ένα 99% διάστημα εμπιστοσύνης της διαφοράς των μέσων εάν οι αρχικοί πληθυσμοί είναι κανονικοί και $s_A^2 = 40, s_B^2 = 14$.

Έχουμε κανονικοί πληθυσμοί, μικρά δείγματα, γνωστές διακυμάνσεις.

$\overline{X}_A = 21{,}13$

$n_A = 15$

$\overline{X}_B = 12$

$n_B = 20$

$\overline{X}_A - \overline{X}_B = 9{,}13$

$a = 0{,}01$

$z_{1-a/2} = 2{,}58$

Τότε,

$$\overline{X}_A - \overline{X}_B - z_{1-a/2}\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}} < m_A - m_B < \overline{X}_A - \overline{X}_B + z_{1-a/2}\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}$$

$$4,40 < m_A - m_B < 13,86$$

2.5 Διαστήματα εμπιστοσύνης για την διαφορά των μέσων τιμών δύο πληθυσμών (μικρά δείγματα)

Για μικρά δείγματα, υποθέτοντας πληθυσμούς με ίσες μεταβλητότητες, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για την διαφορά των μέσων τιμών δύο πληθυσμών δίνεται από τον τύπο

$$\overline{X}_1 - \overline{X}_2 - t_{\frac{a}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le m_1 - m_2 \le \overline{X}_1 - \overline{X}_2 + t_{\frac{a}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\ .$$

Όπου $S_p = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

Παράδειγμα

Το ποσοστό (%) μίας ουσίας σε ένα προϊόν που παράγεται από μονάδα παραγωγής Α είναι τυχαία μεταβλητή Χ1 με κατανομή που ακολουθεί την κανονική κατανομή με μέση τιμή μ1 και τυπική απόκλιση σ1. Όμοια για προϊόν που παράγεται από μονάδα παραγωγής Β είναι τυχαία μεταβλητή Χ2 με κατανομή που ακολουθεί την κανονική κατανομή με μέση τιμή μ2 και τυπική απόκλιση σ2.

A 8 9 14 16 10 11 13

B 7 12 6 13 9 8 10 7

Να κατασκευαστεί ένα 0.98 δ.ε. της διαφοράς μ1 - μ2.

Υπολογίζουμε πρώτα τα εξής:

$$\bar{X}_1 = 11.57, \bar{X}_2 = 9, S_1^2 = 8.286, S_2^2 = 6.286$$

$$\overline{X}_1 - \overline{X}_2 - t_{\frac{a}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le m_1 - m_2 \le \overline{X}_1 - \overline{X}_2 + t_{\frac{a}{2},n_1+n_2-2} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\overline{X}_1 - \overline{X}_2 - t_{0.01,13} S_p\sqrt{\frac{1}{7}+\frac{1}{8}} \le m_1 - m_2 \le \overline{X}_1 - \overline{X}_2 + t_{0.01,13} S_p\sqrt{\frac{1}{7}+\frac{1}{8}}$$

Καταλήγουμε τελικά στο διάστημα (-1,113, 6,253)

2.6 Διάστημα εμπιστοσύνης για τον λόγο των διασπορών δύο κανονικών πληθυσμών

Για πληθυσμούς που ακολουθούν κανονική κατανομή, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για τον λόγο των μεταβλητοτήτων τους δίνεται από τον τύπο

$$\frac{S_1^2 / S_2^2}{F_{1-\frac{a}{2};n_1-1,n_2-1}} \le \frac{s_1^2}{s_2^2} \le \frac{S_1^2 / S_2^2}{F_{\frac{a}{2};n_1-1,n_2-1}}.$$

2.7 Διάστημα εμπιστοσύνης για το ποσοστό κάποιας κατηγορίας στον πληθυσμό

Για μεγάλα δείγματα, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για το ποσοστό κάποιας κατηγορίας στον πληθυσμό δίνεται από τον τύπο

$$\hat{p} - z_{\frac{a}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{a}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Παράδειγμα

Παίρνουμε ένα τυχαίο δείγμα 745 ατόμων που είχαν αποκτήσει βίντεο κάμερα για τουλάχιστον 12 μήνες και λιγότερο από 24 μήνες . Από τα μέλη του δείγματος, 321 δήλωσαν ότι μετά από την αγορά της βίντεο κάμερας πηγαίνουν λιγότερο συχνά στο Κινηματογράφο. Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για την αναλογία στο πληθυσμό

Έχουμε μεγάλο δείγμα οπότε χρησιμοποιούμε την κανονική κατανομή.

$$\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0{,}395 < p < 0{,}467$$

Όπου

$$\hat{p} = \frac{321}{745} \approx 0{,}43$$

$$n = 745$$

$$a = 0{,}05$$

$$z_{a/2} = 1{,}96$$

[πηγή 3]

2.8 Διάστημα εμπιστοσύνης για τη διαφορά ποσοστών κάποιας κατηγορίας σε δύο πληθυσμούς

Για μεγάλα δείγματα, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για την διαφορά των ποσοστών κάποιας κατηγορίας σε δύο πληθυσμούς δίνεται από τον τύπο

$$\hat{p}_1 - \hat{p}_2 - z_{1-\frac{a}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_{1-\frac{a}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Παράδειγμα

Να κατασκευαστεί ένα 95% και ένα 90% διάστημα εμπιστοσύνης για την διαφορά αναλογιών των ατόμων δύο πληθυσμών που είναι ενημερωμένοι για την ύπαρξη ενός προϊόντος πριν και μετά μιας διαφημιστικής εκστρατείας

	Μέγεθος δείγματος	Πόσοι ήταν ενημερωμένοι
Πριν	150	68
Μετά	120	65

Η διαφορά εκτιμάται $\frac{68}{150}-\frac{65}{120}=0.45-0.54=-0.088$

Επομένως το διάστημα εμπιστοσύνης θα είναι

$$-0.088 \pm 1.96 * \sqrt{\frac{0.45*0.55}{150}+\frac{0.54*0.56}{120}}$$

$$(-0.03, 0.21)$$

Στο 90% διάστημα εμπιστοσύνης αντί για συντελεστή 1,96 θα χρησιμοποιήσουμε 1,645

$$(-0.01, 0.19)$$

[πηγή 4]

ΚΕΦΑΛΑΙΟ 3

Έλεγχοι Στατιστικών Υποθέσεων

Ο στατιστικός έλεγχος υποθέσεων (hypothesis testing) πρόκειται για μια συμπερασματική διαδικασία η οποία μα δίνει την δυνατότητα να λάβουμε απόφαση μεταξύ δύο εναλλακτικών αποφάσεων. Τις περισσότερες περιπτώσεις η υπόθεση αφορά που τις παραμέτρους της κατανομής (πληθυσμού) αλλά μπορεί επίσης να αναφέρεται στη φύση ή τη συναρτησιακή μορφή της κατανομής. Στη περίπτωση που η υπόθεση θέτει τη συναρτησιακή μορφή της κατανομής καθώς και τις τιμές των παραμέτρων, τότε ονομάζεται απλή υπόθεση(simple hypothesis). Διαφορετικά ονομάζεται σύνθετη υπόθεση(composite hypothesis).

Όταν έχουμε τον έλεγχο μιας στατιστικής υποθέσεως είναι αναγκαίο να υπάρχει και η εναλλακτική υπόθεση (alternative hypothesis). Ο έλεγχος της υπόθεσης γίνεται με βάση κριτήρια που καταλήγουν στην αποδοχή ή στην απόρριψη της υποθέσεως. Φυσικά για να έχει έννοια ο έλεγχος θα πρέπει να υπάρχει μια εναλλακτική υπόθεση για την περίπτωση που η ελεγχόμενη υπόθεση δεν γίνεται αποδεκτή. Η εναλλακτική αυτή υπόθεση είναι αναγκαία καθώς μέσω αυτή καθορίζονται και τα κατάλληλα στατιστικά κριτήρια με τα όποια θα γίνει ο έλεγχος. Ο συμβολισμός της αρχικής υπόθεσης είναι H_o και ονομάζεται μηδενική υπόθεση (null hypothesis), και η άλλη με H_1 και ονομάζεται εναλλακτική υπόθεση (alternative hypothesis).

Ο στατιστικός έλεγχος μιας υπόθεσης βασίζεται στην εξής διαδικασία: στο πρώτο στάδιο γίνεται ο καθορισμός της μηδενικής και της εναλλακτικής υποθέσεως. Μετά στο επόμενο βήμα βρίσκουμε ένα κανόνα ή κριτήριο, όπου θα γίνει δεκτή ή θα απορριφθεί η μηδενική υπόθεση. Στην ουσία με βάση τις πληροφορίες από το δείγμα βλέπουμε αν ευνοούν την απόρριψη της μηδενικής υποθέσεως.

Στην πραγματικότητα θέτουμε την μηδενική υπόθεση (H_o) αυτή για την οποία αμφιβάλουμε, και έπειτα εξετάζουμε μέσω της επιλογής ενός τυχαίου δείγματος από τον πληθυσμό αν συμφωνεί υπέρ της απόρριψής της, έναντι της εναλλακτικής (H_1). Με βάση την παρατήρηση του τυχαίου δείγματος από τον πληθυσμό η μηδενική υπόθεση γίνεται δεχτή ή όχι. Πιο συγκεκριμένα και κάτω από την υπόθεση ότι η H_o είναι αληθής εξετάζουμε αν αυτό που παρατηρείται στο δείγμα είναι ακραίο. Με αυτό εννοούμε αν έχει πολύ μικρή πιθανότητα να συμβεί, τότε απορρίπτουμε την H_o. Στην αντίθετη περίπτωση που αυτό δεν είναι ακραίο κάτω από την υπόθεση πάλι ότι η μηδενική είναι αληθής τότε το δείγμα που πήραμε δε μας δίνει αρκετές ενδείξεις για την απόρριψη της H_o και «αποτυγχάνουμε να την απορρίψουμε».

Τα παραπάνω μπορούν να κατηγοριοποιηθούν στις παρακάτω περιπτώσεις στην περίπτωση που το δείγμα είναι ακραίο και απορρίψουμε την μηδενική υπόθεση:

(α) είτε η H_o δεν είναι αληθής, όποτε αποφασίσαμε σωστά,

(β) είτε η H_o είναι αληθής και το ακραίο τυχαία εμφανίστηκε οπότε συνέβη κάτι σπάνιο. Σε αυτή την περίπτωση αυτή έχουμε απορρίψει λανθασμένα την H_o και το ονομάζουμε **σφάλμα τύπου I (type I error)**. Ανάλογο σφάλμα υπάρχει και για την εναλλακτική υπόθεση, όταν είναι λάθος να μην απορρίψουμε την μηδενική υπόθεση, δηλαδή, η αποτυχία απόρριψης της μηδενικής, ενώ είναι αληθής η εναλλακτική, και αυτό το ονομάζουμε **σφάλμα τύπου II (type II error)**. Το «ρίσκο», επομένως, είναι διπλό, με πιθανότητα,

• **λανθασμένης απόρριψης** της μηδενικής υπόθεσης ,
P(σφάλμα τύπου I) = P(απόρριψη της H_o | αληθής η H_o) και

• **λανθασμένης μη απόρριψης** της H_o ,
P(σφάλμα τύπου II) = P(μη απόρριψη της H_o | αληθής η H_1).

Για να ξεκαθαρίσουμε τι εννοούμε όταν λέμε «αυτό που παρατηρείται στο δείγμα» και πώς κρίνουμε ότι αυτό που παρατηρείται στο δείγμα» είναι ή όχι «ακραίο» παραθέτουμε το παρακάτω παράδειγμα:

Ας υποθέσουμε ότι το όριο αντοχής ενός τύπου καλωδίων είναι τυχαία μεταβλητή X, με μέση τιμή μ = 1500Kgr και τυπική απόκλιση σ = 175Kgr. Το εργοστάσιο που κατασκευάζει αυτόν τον τύπο καλωδίων ισχυρίζεται ότι βελτίωσε τα υλικά που χρησιμοποιεί και πλέον το όριο αντοχής των καλωδίων έχει αυξηθεί.
Ελέγχοντας τον ισχυρισμό του εργοστασίου, ως μηδενική υπόθεση θέτουμε την
H_o: μ =1500 Kgr,

Αυτό που μπορούμε να πούμε αρχικά είναι ότι η H_o δηλώνει ότι στον πληθυσμό η κατάσταση δεν υπέστη αλλαγές ή αλλιώς ότι η ανεξάρτητη μεταβλητή δεν επιδρά στην εξαρτημένη μεταβλητή για τον πληθυσμό (στο παράδειγμά μας, ότι η βελτίωση των υλικών δεν επιδρά στο όριο αντοχής των καλωδίων). Μια δεύτερη σε σειρά αρχή για τον καθορισμό της H_o, είναι ο εξής: Ως μηδενική υπόθεση θέτουμε την υπόθεση της οποίας η λανθασμένη απόρριψη έχει τους περισσότερους κινδύνους.

Δηλαδή, αυτή που απαιτεί μεγαλύτερη προστασία από σφάλμα τύπου I. Ως εναλλακτική θέτουμε την H_1: μ>1500 Kgr, δηλαδή, η H_1 δηλώνει ότι η βελτίωση των υλικών επηρεάζει, και αυξάνει το όριο αντοχής των καλωδίων.

Γενικά, η H_1 δηλώνει ότι στον πληθυσμό η κατάσταση υπέστη αλλαγές ή αλλιώς, ότι η ανεξάρτητη μεταβλητή επιδρά στην εξαρτημένη μεταβλητή για τον πληθυσμό. Ο έλεγχος που μόλις διατυπώσαμε, είναι ένας μονόπλευρος και ειδικότερα
δεξιόπλευρος έλεγχος. Γενικότερα, οι έλεγχοι,

1) $H_o : \mu \leq \mu_o$ 2) $H_o : \mu \geq \mu_o$
$H_1 : \mu < \mu_o$ $H_1 : \mu > \mu_o$

ονομάζονται μονόπλευροι (one-tailed) έλεγχοι (δεξιόπλευρος και αριστερόπλευρος αντίστοιχα) και ο έλεγχος:

3) $H_o : \mu=\mu_o$
$H_1 : \mu\neq \mu_o$

ονομάζεται αμφίπλευρος (two-tailed).

Εδώ να σημειώσουμε ότι οι δύο υποθέσεις που διατυπώνουμε σε έναν έλεγχο πρέπει να είναι το ένα άρνηση του άλλου. Ακόμα αναφέρουμε ότι και οι δύο υποθέσεις αναφέρονται στον πληθυσμό γι'αυτό δηλώνονται με όρους παραμέτρων του πληθυσμού.

Ως προς το παράδειγμα παραπάνω εφόσον διατυπώσουμε την υπόθεση ότι η άγνωστη μέση τιμή του πληθυσμού των ορίων αντοχής των καλωδίων μετά τη βελτίωση των υλικών είναι 1500Kgr (H_o: μ=1500 Kgr), παίρνουμε ένα τυχαίο δείγμα καλωδίων από το σύνολο της παραγωγής του εργοστασίου και μετράμε το όριο αντοχής κάθε καλωδίου του δείγματος. Έστω ότι ένα τυχαίο δείγμα n X_1 , X_2 ,...X_{50} μεγέθους n=50, μας έδωσε τις μετρήσεις x_1 , x_2 ,..., x_{50} με την μέση τιμή ίση με 1550 Kgr.

Για το Κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία από τις πηγές [7], [8].

3.1 Έλεγχος για την μέση τιμή μ (μεγάλα δείγματα)

Έστω ότι πήραμε ένα τυχαίο δείγμα X1, X2, X3….,Xn από έναν πληθυσμό που ακολουθεί την κανονική κατανομή ή οποιαδήποτε κατανομή, με την προϋπόθεση ότι το μέγεθος του δείγματος είναι αρκετά μεγάλο (n>30), με αντικειμενικό σκοπό να ελέγξουμε μια στατιστική υπόθεση που αναφέρεται στον άγνωστο μέσο αριθμητικό του πληθυσμού και επιθυμούμε να ελέγξουμε αν είναι η υπόθεση στατιστικά αποδεκτή.

Οι έλεγχοι υποθέσεων σχετικά με την μέση τιμή m του πληθυσμού είναι της μορφής:

1) $\begin{matrix} \boldsymbol{H}_0 : m = m_0 \\ \boldsymbol{H}_1 : m > m_0 \end{matrix}$,

Αυτό σημαίνει ότι κάνουμε την υπόθεση για τους λόγους που πιστεύουμε ότι η άγνωστη παράμετρος μ είναι ίση με μια γνωστή τιμή μ_o.

2) $\begin{matrix} \boldsymbol{H}_0 : m = m_0 \\ \boldsymbol{H}_1 : m < m_0 \end{matrix}$.,

3) $\begin{matrix} \boldsymbol{H}_0 : m = m_0 \\ \boldsymbol{H}_1 : m \neq m_0 \end{matrix}$

***Για τον έλεγχο της παραμέτρου μ διακρίνουμε δυο περιπτώσεις*:**

- Η διακύμανση είναι γνωστή και ο πληθυσμός ακολουθεί κανονική κατανομή το δείγμα έχει μέγεθος n>30 ανεξάρτητα από την μορφή της κατανομής.

Γνωρίζουμε ότι αν το τυχαίο δείγμα προέρχεται από κανονικό πληθυσμό ή αν είναι αρκετά μεγάλο (n>30), τότε η κατανομή του εκτιμητή $\bar{x}$ είναι κανονική και επομένως η μεταβλητή:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

κατανέμεται σαν τυπική κανονική κατανομή N(0,1). Η συνάρτηση Z εξαρτάται μόνο από την τιμή $\bar{x}$ και χρησιμοποιείται ως κριτήριο για την αποδοχή ή την απόρριψη της υπόθεσης H_o.

Περίπτωση $1^η$:

Έλεγχος της υπόθεσης

H_o: $\mu=\mu_o$ VS H_1 :$\mu> \mu_o$ σε επίπεδο σημαντικότητας α.

Στην περίπτωση αυτή ο έλεγχος ονομάζεται μονόπλευρος προς τα πάνω και το επίπεδο σημαντικότητας τοποθετείται στο δεξιό άκρο της κανονικής καμπύλης. Στην περίπτωση που ελέγχουμε την παραπάνω μονόπλευρη υπόθεση, μόνο πολύ μεγάλες τιμές της Z λόγω μεγάλης τιμής της $\bar{x}$ οδηγούν στην απόρριψη της H_o και την αποδοχή της H_1.

Η υπόθεση H_o απορρίπτεται, αν ισχύει η ανισότητα

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} > Z_a$$

Την τιμή z_a την παίρνουμε από τον πίνακα κανονικής κατανομής.

Έστω ότι η μεταβλητότητα του πληθυσμού είναι άγνωστη. Για μεγάλα δείγματα, η υπόθεση $\boldsymbol{H}_0$ απορρίπτεται χάριν της $\boldsymbol{H}_1$ όταν αντίστοιχα ισχύουν

$$|\overline{\boldsymbol{X}} - m_0| > \boldsymbol{z}_{1-\frac{a}{2}} \frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}}, \quad \overline{\boldsymbol{X}} - m_0 > \boldsymbol{z}_{1-a} \frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}}, \quad \overline{\boldsymbol{X}} - m_0 < -\boldsymbol{z}_{1-a} \frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}},$$

όπου a το επίπεδο σημαντικότητας

Παράδειγμα

Μέχρι το έτος 1987 στις εξετάσεις στο μάθημα της Στατιστικής στη σχολή Ικάρων έχει διαπιστωθεί, ότι η μέση βαθμολογία των σπουδαστών ήταν 87 μονάδες με τυπική απόκλιση σ=9 μονάδες. Στις εξετάσεις της περιόδου Ιανουαρίου του τρέχοντας έτους από τυχαίο δείγμα 36 σπουδαστών διαπιστώνεται μέση βαθμολογία 90 μονάδων. Με την προϋπόθεση ότι η τυπική απόκλιση σ=9 ισχύει, μπορούμε να ισχυριστούμε ότι η μέση βαθμολογία δεν έχει αυξηθεί; Επίπεδο σημαντικότητας (α=5%)

Η υπόθεση θα είναι:
H_o: μ=87 εναντι H_1: μ>87

Επειδή σ είναι γνωστή και έχουμε μεγάλο δείγμα θα χρησιμοποιήσουμε την κανονική κατανομή.

Στατιστικό στοιχείο: z = $\dfrac{\overline{X} - m_0}{s / \sqrt{n}} = \dfrac{90-87}{9/6} = 2$

Από τον πίνακα της κανονικής κατανομής $z_a = z_{5\%} = 1{,}645$

Επειδή 2>1,645 απορρίπτεται ότι η βαθμολογία έχει παραμένει σταθερή. Άρα αυξήθηκε και επομένως απορρίπτεται η H_0.
[πηγή 5]

Περίπτωση 2η
Έλεγχος της υπόθεσης
H_o: $\mu=\mu_o$ VS H_1 :$\mu< \mu_o$ σε επίπεδο σημαντικότητας α.

Στην περίπτωση αυτή ο έλεγχος ονομάζεται μονόπλευρος προς τα κάτω και το επίπεδο σημαντικότητας τοποθετείται στο αριστερό άκρο της κανονικής καμπύλης.
Η υπόθεση H_o απορρίπτεται, αν ισχύει η ανισότητα

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_a$$

και γίνεται δεκτή αν $z > -z_a$
Την τιμή $-z_α$ την παίρνουμε από τον πίνακα κανονικής κατανομής.

Παράδειγμα

Από ένα δείγμα 35 καταστημάτων υπολογίστηκε ο μέσος όρος των πωλήσεων σε αυτά τα καταστήματα. Ο μέσος όρος πωλήσεων είναι 84.32 (σε χιλιάδες) με τυπική απόκλιση ίση με 7 (σε χιλιάδες).
Εξετάστε αν ο μέσος όρος των πωλήσεων είναι κάτω των 92 (σε χιλιάδες) με επίπεδο εμπιστοσύνης 99%.
Έχουμε έναν μονόπλευρο έλεγχο μέσου με άγνωστη διασπορά σε μικρό δείγμα (n=12):
Ho: $\mu = \mu_0$
H_1: $\mu < \mu_0$, όπου $\mu_0 = 92$.
Εξετάζουμε αν

$$\overline{X} - m_0 < -z_{1-a} \frac{S}{\sqrt{n}}$$

$$84,32 - 92 > -z_{0,01} \frac{7}{\sqrt{35}}$$

$$-7,68 > -2,33 \frac{7}{\sqrt{35}}$$

$$-5,58 > -2,756$$

Προφανώς η τελευταία ανισότητα δεν αληθεύει και επομένως δεν απορρίπτουμε την μηδενική υπόθεση.

Περίπτωση 3η

Έλεγχος της υπόθεσης
H_o: $\mu=\mu_o$ VS H_1 :$\mu\neq\mu_o$ σε επίπεδο σημαντικότητας α.

Στην περίπτωση αυτή ο έλεγχος ονομάζεται δίπλευρος και το επίπεδο σημαντικότητας α κατανέμεται ίσα, στο δεξιό και στο αριστερό άκρο της καμπύλης της κανονικής κατανομής. Η υπόθεση H_o απορρίπτεται, αν ισχύει η ανισότητα

$$Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_{a/2} \quad \text{ή} \quad Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_{a/2}$$

και γίνεται δεκτή αν $-z_{α/2} < z < z_{a/2}$

3.2 Έλεγχος για την μέση τιμή μ (μικρά δείγματα)

Η διακύμανση του πληθυσμού είναι άγνωστη και το μέγεθος του δείγματος είναι n<30 Η υπόθεση H_o των τριών περιπτώσεων που αφορούν το μέσο αριθμητικό ελέγχεται από το κριτήριο :

$$t=\frac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}}$$

Η μεταβλητή t, ακολουθεί την κατανομή με ν=n-1 βαθμούς ελευθερίας όπου :

$$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$$ η διακύμανση του δείγματος.

Στη συνέχεια συγκρίνουμε την τιμή t με μια τιμή που παίρνουμε από τον πίνακα κατανομής σε επίπεδο σημαντικότητας α και ν=n-1 βαθμούς ελευθερίας. Αναλυτικότερα :

1)η υπόθεση H_o στην περίπτωση Ho: $\mu=\mu_o$ VS H_1: $\mu>\mu_o$ γίνεται δεκτή αν $t<t_{\nu,\alpha}$

2)η υπόθεση H_o στην περίπτωση Ho: $\mu=\mu_o$ VS H_1: $\mu<\mu_o$ γίνεται δεκτή αν $-t_{\nu,\alpha}<t$

3)η υπόθεση H_o στην περίπτωση Ho: $\mu=\mu_o$ VS H_1: $\mu\neq\mu_o$ γίνεται δεκτή αν
$-t_{\nu,\alpha/2}< t < t_{\nu,\alpha/2}$

Για πληθυσμούς που ακολουθούν κανονική κατανομή, η υπόθεση $\boldsymbol{H}_0$ απορρίπτεται χάριν της $\boldsymbol{H}_1$ όταν ισχύουν

$$|\,\overline{\boldsymbol{X}}-m_0\,|>\boldsymbol{t}_{\frac{a}{2},\boldsymbol{n}-1}\frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}},\quad \overline{\boldsymbol{X}}-m_0>\boldsymbol{t}_{a,\boldsymbol{n}-1}\frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}},\quad \overline{\boldsymbol{X}}-m_0<-\boldsymbol{t}_{a,\boldsymbol{n}-1}\frac{\boldsymbol{S}}{\sqrt{\boldsymbol{n}}}.$$

Παράδειγμα

Παρακάτω έχουμε τις μετρήσεις για το μαγγάνιο με την νέα μέθοδο προσδιορισμού του σημείου στίξης μετάλλων:

1267, 1262, 1267, 1263, 1258, 1263, 1268.

Να εξεταστεί αν η νέα μέθοδος σφάλει με ε.σ. 0.05, δεδομένου ότι το σ.τ. του μαγγανίου είναι 1260°C **[πηγή 6]**

Έχουμε έναν αμφίπλευρο έλεγχο μέσου με άγνωστη διασπορά σε μικρό δείγμα (n=7):
Ho: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$, όπου $\mu_0 = 1260$.

Εύκολα υπολογίζουμε το δειγματικό μέσο των παρατηρήσεων μας $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=1264$ καθώς και τη δειγματική διασπορά $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=12.67$ Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή (περιοχή απόρριψης της H_0) είναι:

$$K: |t| = \frac{\left|\bar{X} - m_o\right|}{S/\sqrt{n}} > t_{a/2,n-1,} a = 0.05$$

Έτσι |t| = 2.98, ενώ από τους πίνακες της κατανομής t προκύπτει ότι $t_{a/2,n-1} = t_{0.025,6} = 2.447$, οπότε απορρίπτουμε την H_o, δηλαδή η νέα μέθοδος σφάλει.

Παράδειγμα

Από ένα δείγμα 12 καταστημάτων υπολογίστηκε ο μέσος όρος των πωλήσεων σε αυτά τα καταστήματα, που είναι 84.42 (σε χιλιάδες) με δειγματική τυπική απόκλιση ίση με 7.18 (σε χιλιάδες).

Εξετάστε αν ο μέσος όρος των πωλήσεων είναι άνω των 90 (σε χιλιάδες) με επίπεδο εμπιστοσύνης 99%.

Έχουμε έναν μονόπλευρο έλεγχο μέσου με άγνωστη διασπορά σε μικρό δείγμα (n=12):

Ho: $\mu = \mu_0$

H_1: $\mu > \mu_0$, όπου $\mu_0 = 90$.

Εξετάζουμε αν

$$\bar{X} - m_0 > t_{a,n-1}\frac{S}{\sqrt{n}}$$

$$84,42 - 90 > t_{0,01,12-1}\frac{7,18}{\sqrt{12}}$$

$$-5,58 > 2,718\frac{7,18}{\sqrt{12}}$$

$$-5,58 > 5,633$$

Προφανώς η τελευταία ανισότητα δεν αληθεύει και επομένως δεν απορρίπτουμε την μηδενική υπόθεση.

3.3 Έλεγχος σημαντικότητας για διαφορά δύο μέσων τιμών (μεγάλα δείγματα)

Η σύγκριση των μέσων τιμών m_1, m_2 δύο πληθυσμών γίνεται μέσω ελέγχων υποθέσεων της μορφής

1) $\begin{matrix} \boldsymbol{H}_0 : m_1 = m_2 \\ \boldsymbol{H}_1 : m_1 > m_2 \end{matrix}$

2) H_o: $\mu_1 = \mu_2$
H_1: $\mu_1 < \mu_2$

3) $\begin{matrix} \boldsymbol{H}_0 : m_1 = m_2 \\ \boldsymbol{H}_1 : m_1 \neq m_2 \end{matrix}$

Για μεγάλα δείγματα, υποθέτοντας πληθυσμούς με ίσες μεταβλητότητες, η υπόθεση $\boldsymbol{H}_0$ απορρίπτεται χάριν της $\boldsymbol{H}_1$ όταν αντίστοιχα ισχύουν

$$|\bar{X}_1-\bar{X}_2|> z_{1-\frac{a}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \quad \bar{X}_1-\bar{X}_2 > z_{1-a} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

Παράδειγμα

Τα παρακάτω δεδομένα αφορούν τα φορτία θραύσης (σε tn/cm^2) συνθετικών νημάτων δύο τύπων:

Τύπος Ι	1.2	0.3	0.8	0.5	0.4	0.9	1.0		
Τύπος ΙΙ	1.4	1.5	1.1	1.0	0.8	1.7	0.9	0.7	0.6

Υποθέτοντας ισότητα διασπορών να εξεταστεί αν οι δύο τύποι νημάτων έχουν την ίδια μέση αντοχή σε ε.σ. 0.05
Έχουμε έναν αμφίπλευρο έλεγχο ισότητας μέσων με άγνωστες αλλά κοινές διασπορές:
Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$, με $\sigma_1 = \sigma_2 = \sigma$.

Εύκολα προκύπτει ότι $\bar{X}_1 = 0.73, S_1^2 = 0.11, n_1 = 7$ & $\bar{X}_2 = 1,08, S_2^2 = 0.14, n_2 = 9$
Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K: |t| = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{S_p / \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{a/2,v,} \quad S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{v}, v = n_1+n_2-2, a = 0.05$$

Έτσι |t| = 1.91, ενώ από τους πίνακες της κατανομής t προκύπτει ότι $t_{a/2,n-1} = t_{0.025,14} = 2.145$, οπότε αποδεχόμαστε την H_o, δηλαδή οι δύο τύποι νημάτων έχουν την ίδια μέση αντοχή.
[πηγή 7]

3.4 Έλεγχος σημαντικότητας για διαφορά δύο μέσων τιμών (μικρά δείγματα)

Η υπόθεση H_o των τριών συνδυασμών ελέγχεται με το κριτήριο

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$S^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Όπου
και n_1+n_2-2 οι βαθμοί ελευθερίας
Επομένως:

1)η υπόθεση H_o του πρώτου συνδυασμού γίνεται δεκτή αν $t<t_{v,\alpha}$

2)η υπόθεση H_o του δεύτερου συνδυασμού γίνεται δεκτή αν: $-t_{v,\alpha}<t$

3)η υπόθεση H_o του τρίτου συνδυασμού γίνεται δεκτή αν : $-t_{v,\alpha/2} < t < t_{v,\alpha/2}$

Για πληθυσμούς που ακολουθούν κανονική κατανομή, η υπόθεση $\boldsymbol{H}_0$ απορρίπτεται χάριν της $\boldsymbol{H}_1$ όταν ισχύουν

$$|\overline{X}_1 - \overline{X}_2| > t_{\frac{a}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad \overline{X}_1 - \overline{X}_2 > t_{a, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Παράδειγμα

Από δύο κανονικούς πληθυσμούς πήραμε δυο δείγματα Α και Β, τα οποία μας έδωσαν τις παρακάτω τιμές

X_i: 0,114 0,127 0,143 0,132
Y_i: 0,131 0,107 0,104 0,111 0,108 0,110

Αν υποθέσουμε ότι $s_1^2 = s_2^2$ να ελεγχθεί η υπόθεση , δηλαδή και τα δυο δείγματα προέρχονται από δύο πληθυσμούς με τον ίδιο μέσο ως προς την εναλλακτική υπόθεση $m_1 \neq m_2$.

$$\overline{X} = 0{,}129$$
$$\overline{Y} = 0{,}1118$$
$$n_1 = 4$$
$$n_2 = 6$$
$$S_1^2 = \frac{\sum (X_i - \overline{X})^2}{n_1 - 1} = 0{,}000220$$
$$S_2^2 = \frac{\sum (Y_i - \overline{Y})^2}{n_2 - 1} = 0{,}000094$$
$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0{,}0001412$$
$$t = \frac{0{,}129 - 0{,}1118}{\sqrt{0{,}0001412\left(\frac{1}{4} + \frac{1}{6}\right)}} = 2{,}24$$

Οι βαθμοί ελευθερίας είναι $v=n_1+n_2-2=8$, από τους πίνακες $t_{8,5\%} = 2{,}306$. Επειδή 2,24<2,306 αποδεχόμαστε την H_o σε επίπεδο 5%.
[πηγή 8]

Παράδειγμα

Σε μια έρευνα για τη διάρκεια του 2ου σταδίου του τοκετού συγκρίθηκαν τα ιστορικά τοκετού 10 παιδιών, που πέθαναν ξαφνικά (ομάδα Α), με τα ιστορικά 10 παιδιών που αποτέλεσαν τη συγκριτική ομάδα (ομάδα Β), ως προς τη διάρκεια του 2ου σταδίου του τοκετού (δεν απαιτείται το n1 να είναι ίσο με n2).
Οι μετρήσεις ήταν:

Διάρκεια σε λεπτά (έχουμε λογαριθμίσει)

Ομάδα Α: 1.477 1.398 1.204 1.255 1.176 1.114 1.000
1.398 1.176 1.000

Ομάδα Β: 1.114 1.301 1.699 1.230 1.653 1.544 1.000
1.447 1.279 1.398

οπότε, $\mu_1 = 1.22$, $SD_1 = 0.164$ και $\mu_2 = 1.37$, $SD_2 = 0.226$

$SE_1 = 0{,}052$ και $SE_2 = 0{,}071$

Έχουμε: t-value= $\dfrac{m_1 - m_2}{\sqrt{SE_1^2 + SE_2^2}} = \dfrac{1.22 - 1.37}{\sqrt{0.052^2 + 0.071^2}} =$ =-1,71

Οι βαθμοί ελευθερίας είναι 10+10-2=18, οπότε από τον Πίνακα προκύπτει ότι η θεωρητική τιμή του επιπέδου P=5% είναι t=2,10. Επειδή η τιμή που υπολογίσαμε (κατά απόλυτη τιμή) είναι μικρότερη της θεωρητικής τιμής, η διαφορά των 2 μέσων είναι στατιστικά μη σημαντική στο επίπεδο 5%.

3.5 Έλεγχος σημαντικότητας για την διασπορά σ^2

Οι έλεγχοι υποθέσεων σχετικά με την μεταβλητότητα s^2 του πληθυσμού είναι της μορφής:

1) $\begin{matrix} \boldsymbol{H}_0 : s^2 = s_0^2 \\ \boldsymbol{H}_1 : s^2 > s_0^2 \end{matrix}$,

2) $\begin{matrix} \boldsymbol{H}_0 : s^2 = s_0^2 \\ \boldsymbol{H}_1 : s^2 < s_0^2 \end{matrix}$

3). $\begin{matrix} \boldsymbol{H}_0 : s^2 = s_0^2 \\ \boldsymbol{H}_1 : s^2 \neq s_0^2 \end{matrix}$

Η υπόθεση H_o των παραπάνω περιπτώσεων ελέγχεται με κριτήριο:

$$x^2 = \frac{\sum (x_i - \mu)^2}{\sigma_0^2}$$

αν ο μέσος μ είναι γνωστός,
εάν ο μέσος μ είναι άγνωστος χρησιμοποιούμε το κριτήριο:

$$x^2 = \frac{\sum (x_i - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2}$$

Αν από έναν κανονικό πληθυσμό με μέσο μ και διακύμανση σ^2 πάρουμε όλα τα δυνατά δείγματα του ιδίου μεγέθους n, τότε η ποσότητα $(n-1)\, S^2 / \sigma_o^2$ ακολουθεί την κατανομή x^2, με ν=n-1 βαθμούς ελευθερίας

1. αν ο έλεγχος αναφέρεται στην πρώτη περίπτωση, τότε συγκρίνουμε την τιμή x^2 με την τιμή $x^2_{\nu,\alpha}$

Για πληθυσμούς που ακολουθούν κανονική κατανομή, η υπόθεση $\boldsymbol{H}_0$ απορρίπτεται χάριν της $\boldsymbol{H}_1$ όταν αντίστοιχα ισχύουν

$$\frac{(\boldsymbol{n}-1)\boldsymbol{S}^2}{s_0^2} > c^2_{1-\frac{a}{2},\boldsymbol{n}-1} \text{ ή } \frac{(\boldsymbol{n}-1)\boldsymbol{S}^2}{s_0^2} < c^2_{\frac{a}{2},\boldsymbol{n}-1}, \quad \frac{(\boldsymbol{n}-1)\boldsymbol{S}^2}{s_0^2} > c^2_{1-a,\boldsymbol{n}-1}, \quad \frac{(\boldsymbol{n}-1)\boldsymbol{S}^2}{s_0^2} < c^2_{a,\boldsymbol{n}-1}.$$

Παράδειγμα

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές 30, 40, 28, 54. Ζητείται να ελεγχθεί η υπόθεση

H_0: σ^2 =30
H_1: σ^2 >30
σε επίπεδο σημαντικότητας 5%.

Στατιστικό στοιχείο: $x^2 = \frac{(n-1)S^2}{s_0^2} = \frac{3(141,3)}{30} = 14,13$ όπου $S^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{424}{3} = 141,3$

Κριτική τιμή: οι βαθμοί ελευθερίας είναι ν=n-1=4-1=3, α=5%, οπότε $c^2_{n,a} = c^2_{3,5\%} = 7,81$

Επειδή 14,13>7,81 απορρίπτεται η H_o, αλλιώς αν 14,13 <7,81 η υπόθεση H_o γίνεται δεκτή.
[πηγή 9]

2. αν ο έλεγχος αναφέρεται στην δεύτερη περίπτωση H_o γίνεται δεκτή αν $x^2 > x^2_{\nu,1-\alpha}$

3. αν ο έλεγχος αναφέρεται στην τρίτη περίπτωση η υπόθεση γίνεται δεκτή αν $x^2_{\nu,1-\alpha/2} < x^2 < x^2_{\nu,\alpha/2}$

3.6 Σύγκριση των διασπορών δύο πληθυσμών

Η σύγκριση των μεταβλητοτήτων s_1^2, s_2^2 δύο πληθυσμών γίνεται με την $F=s^2_1/s^2_2$ η οποία ακολουθεί την κατανομή F με n_1-1 και n_2-1 βαθμούς εμπιστοσύνης όπου n_1 to μέγεθος του δείγματος με διακύμανση s^2_1 και n_2 με διακύμανση s^2_2 .

Περιπτώσεις:

1) $\begin{matrix} H_0 : s_1^2 = s_2^2 \\ H_1 : s_1^2 \neq s_2^2 \end{matrix}$, αμφίπλευρος έλεγχος

Απορρίπτουμε την H_o αν : $F< F_{n1-1,n2-1,a/2}$ ή $F>F_{n1-1,n2-1,1-a/2}$ και αποδεχόμαστε την H_o αν $F_{n1-1,n2-1,a/2} < F< F_{n1-1,n2-1,1-a/2}$

2) $\begin{matrix} H_0 : s_1^2 = s_2^2 \\ H_1 : s_1^2 > s_2^2 \end{matrix}$, μονόπλευρος έλεγχος

Απορρίπτουμε την H_o αν : $F>F_{n1-1,n2-1,1-a}$ και αποδεχόμαστε την H_o αν $F< F_{n1-1,n2-1,1-α}$

3) H_o : $σ^2_1=σ^2_2$

H_1: $σ^2_1 < σ^2_2$

Μονόπλευρος έλεγχος

Απορρίπτουμε την H_o :αν $F< F_{n1-1,n2-1,α}$ και αποδεχόμαστε την H_o αν $F>F_{n1-1,n2-1,a}$

Για πληθυσμούς που ακολουθούν κανονική κατανομή, η υπόθεση H_0 απορρίπτεται χάριν της H_1 όταν αντίστοιχα ισχύουν

$$\frac{S_1^2}{S_2^2} > F_{1-\frac{a}{2};n_1-1,n_2-1} \quad ή \quad \frac{S_2^2}{S_1^2} < F_{1-\frac{a}{2};n_2-1,n_1-1}, \quad \frac{S_1^2}{S_2^2} > F_{1-a;n_1-1,n_2-1}.$$

Παράδειγμα

Δυο αναλυτές Α και Β έκαναν 6 μικροαναλυτικούς προσδιορισμούς της περιεκτικότητας σε άνθρακα ενός χημικού προϊόντος και πήραν τις ακόλουθες τιμές:

Αναλυτής Α	59.09	59.17	59.27	59.13	59.10	59.14
Αναλυτής Β	59.06	59.40	59.00	59.12	59.01	59.25

Να ελεγχθεί η υπόθεση Ho: $σ_1 = σ_2$ με εναλλακτική την H_1: $σ_1 \neq σ_2$ (α=0.02).

Εύκολα προκύπτει ότι $S_A^2 = 4.28*10^{-3}, n_A = 6$ $\& S_B^2 = 24.59*10^{-3}, n_B = 6$

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K : F = \frac{S_A^2}{S_B^2} > F_{n_A-1,n_B-1,a/2} \quad h' \quad F = \frac{S_A^2}{S_B^2} < F_{n_A-1,n_B-1,1-a/2}$$

Αλλά F = 5.75 και από τους πίνακες της κατανομής F έχουμε ότι $F_{n_A-1,n_B-1,a/2} = F_{5,5,0,01} = 10,97$, ενώ $F_{n_A-1,n_B-1,1-a/2} = \frac{1}{F_{n_A-1,n_B-1,a/2}} = \frac{1}{F_{5,5,0,01}} = \frac{1}{10,97} = 0,09$, οπότε αποδεχόμαστε την H_0.

3.7 Έλεγχος σημαντικότητας για το ποσοστό

Έστω ότι έχουμε n επαναλαμβανόμενα πειράματα στα οποία έχουμε ως αποτέλεσμα αποτυχία ή επιτυχία. Ονομάζουμε το πλήθος των επιτυχιών x και p την πιθανότητα επιτυχίας σε n επαναλήψεις (n>30). Το ποσοστό των επιτυχιών είναι $\hat{p} = \frac{x}{n}$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Οι έλεγχοι υποθέσεων σχετικά με το ποσοστό p κάποιας κατηγορίας στον πληθυσμό είναι της μορφής:

1) $\begin{matrix} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{matrix}$, αμφίπλευρος έλεγχος

Απορρίπτουμε την H_0 αν: $z=-z_{a/2}$ και αποδεχόμαστε την H_o αν $-z_{a/2} < z < z_{a/2}$

2) $\begin{matrix} H_0 : p = p_0 \\ H_1 : p > p_0 \end{matrix}$, μονόπλευρος έλεγχος

Απορρίπτουμε την H_0 αν: $z > z_a$ και αποδεχόμαστε την H_o αν $z < z_a$

3) $\begin{matrix} H_0 : p = p_0 \\ H_1 : p < p_0 \end{matrix}$, μονόπλευρος έλεγχος

Απορρίπτουμε την H_0 αν: $z < -z_a$ και αποδεχόμαστε την H_o αν $z > -z_a$

Για μεγάλα δείγματα, η υπόθεση H_0 απορρίπτεται χάριν της H_1 όταν αντίστοιχα ισχύουν

$$|\hat{p} - p_0| > z_{1-\frac{a}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}, \quad \hat{p} - p_0 > z_{1-a} \sqrt{\frac{p_0(1-p_0)}{n}}, \quad \hat{p} - p_0 < -z_{1-a} \sqrt{\frac{p_0(1-p_0)}{n}}.$$

Προϋπόθεση είναι το μικρότερο υποσύνολο των δεδομένων να έχει τουλάχιστον 10 παρατηρήσεις.

Παράδειγμα

Κατασκευαστής ισχυρίζεται ότι το πολύ 2% των προϊόντων του είναι ελαττωματικά. Σε τυχαίο δείγμα μεγέθους n=900 βρέθηκαν κ=27 ελαττωματικά. Με ε.σ. α=0.05 να ελεγχθεί ο ισχυρισμός του κατασκευαστή με εναλλακτική υπόθεση την H_1: p>0.02.

Έχουμε τον παρακάτω έλεγχο:
H_0: p = 0.02 = p_0
H_1: p > 0.02 = $p_{0.}$

Από την εκφώνηση προκύπτει ότι $\hat{p} = \frac{27}{900} = 0.03$ και n=900

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K: Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} > z_a, a = 0.05$$

Έτσι Z=2.14 και από τους πίνακες της κανονικής κατανομής προκύπτει ότι Z_α=1.64, οπότε απορρίπτουμε την H_0 και τον ισχυρισμό του κατασκευαστή.

Παράδειγμα

Οι δημοτικές αρχές της πόλης XYZ θα αναγείρουν πολιτιστικό κέντρο σε συγκεκριμένη περιοχή της πόλης. Το δημοτικό συμβούλιο επιθυμεί να εκτιμήσει στην αποδοχή που τυγχάνει αυτή η προσπάθεια του από τους δημότες της πόλης. Διεξήγαγε έρευνα κατά την οποία 1500 δημότες ερωτήθηκαν κατά πόσο εγκρίνουν ή όχι την πρωτοβουλία αυτή. Εννιακόσιοι πενήντα (950) απάντησαν ότι εγκρίνουν την ανέγερση του πολιτιστικού κέντρου.

Να εξεταστεί αν άνω του 50% των δημοτών εγκρίνει το νέο πολιτιστικό κέντρο του δήμου με επίπεδο σημαντικότητας 5%.

Έχουμε τον παρακάτω έλεγχο:
H_0: p = 0.50 = p_0
H_1: p > 0.50 = $p_{0.}$

Από την εκφώνηση προκύπτει ότι $\hat{p} = \frac{950}{1500} = 63.33\%$ και n=1500

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K: Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} > z_a, a = 0.05$$

Έτσι Z=10,713 και από τους πίνακες της κανονικής κατανομής προκύπτει ότι Z_α=1.64, οπότε απορρίπτουμε την H_0.

3.8 Σύγκριση ποσοστών

Τις περισσότερες φορές θέλουμε να εξετάσουμε τη στατιστική σημαντικότητα της διαφοράς δυο ποσοστών τα οποία έχουν προκύψει από δυο πολυπληθή δείγματα. Η σύγκριση των ποσοστών $\boldsymbol{p_1, p_2}$ που αντιστοιχούν σε δυο μεγάλα δείγματα n_1, n_2 κάποιας κατηγορίας σε δύο πληθυσμούς γίνεται μέσω ελέγχων υποθέσεων της μορφής :

1) $\begin{matrix} \boldsymbol{H_0 : p_1 = p_2} \\ \boldsymbol{H_1 : p_1 \neq p_2} \end{matrix}$, αμφίπλευρος έλεγχος

Απορρίπτουμε την H_0 αν: $z<-z_{a/2}$ ή $z > z_{a/2}$ και αποδεχόμαστε την H_o :αν $–z_{a/2} < z < z_{a/2}$

2) $\begin{matrix} \boldsymbol{H_0 : p_1 = p_2} \\ \boldsymbol{H_1 : p_1 > p_2} \end{matrix}$, μονόπλευρος έλεγχος

Απορρίπτουμε την Ho: $z > z_a$ και αποδεχόμαστε την Ho αν $z<z_a$

3) H_o: $p_1 = p_2$
H_1: $p_1 < p_2$
Μονόπλευρος έλεγχος

Απορρίπτουμε την H_o αν $z < -z_a$ και αποδεχόμαστε την H_o αν $z > -z_a$

Ως γνωστόν οι έλεγχοι υποθέσεων βασίζονται σε κατανομές δειγματοληψίας όταν ισχύει η $\boldsymbol{H_0}$. Στην συγκεκριμένη περίπτωση η ισχύς της $\boldsymbol{H_0}$ σημαίνει $\boldsymbol{p_1 = p_2 = p}$, άρα ως εκτίμηση της άγνωστης παραμέτρου $\boldsymbol{p}$ χρησιμοποιούμε το ποσοστό $\boldsymbol{\hat{p}}$ της κατηγορίας συνολικά στα δύο δείγματα.
Επομένως, για μεγάλα δείγματα, η υπόθεση $\boldsymbol{H_0}$ απορρίπτεται χάριν της $\boldsymbol{H_1}$ όταν αντίστοιχα ισχύουν

$$|\hat{p}_1 - \hat{p}_2| > z_{1-\frac{a}{2}} \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \quad \hat{p}_1 - \hat{p}_2 > z_{1-a} \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

όπου $\hat{p} = \dfrac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$.

Παράδειγμα

Από την παραγωγή δύο μηχανών τυχαία δείγματα μεγέθους n_1=450 και n_2=400 έδωσαν αριθμό ελαττωματικών προϊόντων 17 και 12 αντίστοιχα. Να εξεταστεί αν υπάρχει διαφορά στα ποσοστά παραγωγής ελαττωματικών προϊόντων από τις δύο μηχανές.

Έχουμε τον παρακάτω αμφίπλευρο έλεγχο:
H_0: $p_1 = p_2$
H_1: $p_1 \neq p_{2.}$

Από την εκφώνηση προκύπτει ότι $\hat{p}_1 = \frac{17}{450} = 0.038, \hat{p}_2 = \frac{12}{400} = 0.03$

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K : |Z| = \frac{\left|\hat{p}_1 - \hat{p}_2\right|}{\sqrt{p(1-p) * (\frac{1}{n_1} + \frac{1}{n_2})}} > z_{a/2}$$

με $p = \frac{17+12}{450+400} = 0.034, a = 0.05$.

Έτσι |Z|=0.64 και από τους πίνακες της κανονικής κατανομής προκύπτει ότι $Z_{α/2}$=1.96, οπότε αποδεχόμαστε την H_0 για την μη ύπαρξη διαφοράς στα ποσοστά παραγωγής ελαττωματικών προϊόντων από τις δύο μηχανές.

Παράδειγμα

Σ’ ένα παιχνίδι baseball, ένας παίκτης έκανε συνολικά 233 βολές, απ’ τις οποίες οι 84 ήταν επιτυχημένες. Ένας άλλος, πέτυχε στις 103 από τις 350 βολές. Ο πρώτος παίκτης ισχυρίζεται ότι είναι καλύτερος. Είναι αλήθεια αυτό; $(\alpha = 5\%)$

Έχουμε τον παρακάτω μονόπλευρο έλεγχο:

H_0: $p_1 = p_2$

H_1: $p_1 > p_2$.

Από την εκφώνηση προκύπτει ότι $\hat{p}_1 = \frac{84}{233} = 0.3605, \hat{p}_2 = \frac{103}{350} = 0.2943$

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K : |Z| = \frac{\left|\hat{p}_1 - \hat{p}_2\right|}{\sqrt{p(1-p) * (\frac{1}{n_1} + \frac{1}{n_2})}} > z_{a}$$

με $p = \frac{84+103}{233+350} = 0.3207, a = 0.05$.

Έτσι |Z|=1.68 και από τους πίνακες της κανονικής κατανομής προκύπτει ότι $Z_α$=1.65, οπότε απορρίπτουμε την H_0 για την μη ύπαρξη διαφοράς.

Κεφάλαιο 4

Εφαρμογές

Στη παρούσα ενότητα παρουσιάζουμε εφαρμογές της εκτιμητικής, των διαστημάτων εμπιστοσύνης και του έλεγχου υποθέσεων. Έχουμε λοιπόν:

Για το Κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία από τις πηγές [9-15].

Εφαρμογή 1

Έστω δύο δείγματα με 5 και 4 παρατηρήσεις το καθένα. Η μέση τιμή του πρώτου είναι 1,8 και η διασπορά του είναι 0,5, ενώ η μέση τιμή του δεύτερου είναι 1,4 και η διασπορά του 0,24. Να βρεθεί το διάστημα εμπιστοσύνης των δύο μέσων τιμών και έπειτα να ελεγχθεί αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο μέσων τιμών. Επιπλέον να ελεγχτεί αν ο πρώτος μέσος όρος είναι κατά 50% μεγαλύτερος ή μικρότερος του δεύτερου. Ζητάμε ένα διάστημα εμπιστοσύνης με επίπεδο σημαντικότητας α=5% για την διαφορά των μέσων.

Χρησιμοποιούμε την κατανομή Student διότι έχουμε μικρά δείγματα από δύο διαφορετικούς πληθυσμούς και επομένως θα έχουμε ως διασπορά της διαφοράς των μέσων τιμών $\bar{T} = \bar{Y} - \bar{X}$ την $s_{XY}^2 = \dfrac{(v_x - 1)\cdot s_c^2 + (v_y - 1)\cdot s_y^2}{v_x + v_y - 2}$. Επομένως το διάστημα εμπιστοσύνης για την διαφορά των μέσων τιμών απόδοσης των δύο χαρτοφυλακίων.

$$\bar{Y} - \bar{X} - t_{v_x+v_y-2,a/2} \frac{s_{XY}}{\sqrt{\frac{1}{n_c} + \frac{1}{n_y}}} < m_y - m_c < \bar{Y} - \bar{X} + t_{v_x+v_y-2,a/2} \frac{s_{XY}}{\sqrt{\frac{1}{n_c} + \frac{1}{n_y}}}$$

Κάνουμε αντικατάσταση και έχουμε ότι

$$s_{XY}^2 = \frac{(5-1)\cdot 0.5 + (4-1)\cdot 0.24}{5+4-2}$$

$$s_{XY}^2 = \frac{2+0.72}{7}$$

$$s_{XY}^2 = \frac{2.72}{7}$$

$$s_{XY}^2 = 0.105$$

$$s_{XY} = 0.324$$

Έχουμε τελικά

$$1,8 - 1,4 - t_{7,0.05/2} 0.324\sqrt{\frac{1}{5} + \frac{1}{4}} < m_y - m_c < 1,8 - 1,4 + t_{7,0.05/2} 0.324\sqrt{\frac{1}{5} + \frac{1}{4}}9$$

$$1,8 - 1,4 - 2.26\cdot 0.217 < m_y - m_c < 1,8 - 1,4 + 2.26\cdot 0.217$$

$$0.4 - 0,491 < m_y - m_c < 0.4 + 0,491$$

$$-0.091 < m_y - m_c < .891$$

Άρα το διάστημα εμπιστοσύνης είναι το (-0,091, 0,891)

Η στατιστική που χρησιμοποιούμε για τον έλεγχο της υπόθεσης

H_o:μ_y-μ_χ=0
H_1: Διαφορετικά
Είναι

είναι η $T = \dfrac{\bar{Y} - \bar{X}}{s_{XY}(\frac{1}{n_c} + \frac{1}{n_y})}$ =0,4/0,324 = 1,234

όμως η $t_{7,0.05/2}$=2,26 >1,234 πράγμα που σημαίνει ότι οι μέσες τιμές των δύο χαρτοφυλακίων δεν διαφέρουν στατιστικά
Αυτό που θέλουμε να ελέγξουμε στο ερώτημα αυτό είναι η παρακάτω υπόθεση

H_o: 1,5 μ_y-μ_χ=0
H_1: Διαφορετικά
Για να γίνει αυτός ο έλεγχος παίρνουμε την στατιστική

$T = \dfrac{1.5\bar{Y} - \bar{X}}{s_{XY}(\frac{1}{n_c} + \frac{1}{n_y})}$ =1,3/0,324 = 4,01

όμως η $t_{7,0.05/2}$=2,26 <4,01 πράγμα που σημαίνει ότι η μέση τιμή του χαρτοφυλακίου X δεν μπορεί να είναι ίση με 1,5 φορά την μέση τιμή του χαρτοφυλακίου Y. Απορρίπτουμε την H_o.

Σχόλια: από το διάστημα εμπιστοσύνης στην αρχή είδαμε ότι η διαφορά των μέσων αποδόσεων των χαρτοφυλακίων έχει ένα διάστημα εμπιστοσύνης (-0,091, 0,891). Επομένως στο δεύτερο ερώτημα που κάναμε τον έλεγχο για τον η διαφορά των μέσων αποδόσεων διαφέρει γνωρίζαμε εκ των προτέρων ότι θα δεχτούμε την μηδενική υπόθεση για το ότι οι μέσες τιμές δεν διαφέρουν αφού η διαφορά τους πέφτει μέσα στο διάστημα εμπιστοσύνης.

Εφαρμογή 2

Έστω δύο δείγματα Y και X με διασπορές 6 και 19 αντίστοιχα. Το πλήθος των δειγμάτων είναι 40 και 25. Να ελεγχτεί αν η τυπική απόκλιση του Y δείγματος είναι ίση ή μεγαλύτερη του X
Ο έλεγχος τον οποίο εξετάζουμε είναι ο παρακάτω

H_0: $\sigma_y = \sigma_X$
H_1: $\sigma_y < \sigma_X$

Υπολογίζουμε την στατιστική F= $\dfrac{s_Y^2}{s_X^2}$ =6/19 = 0,316

Από τον πίνακα για την κατανομή F έχουμε $F_{40,25,0,05}$ =1,87
Επειδή 0.316<1,87 έχουμε ότι η μηδενική υπόθεση γίνεται αποδεκτή.

Εφαρμογή 3

Σε μια έρευνα που διεξήχθη ζητήθηκε από 172 εργαζόμενους να απαντήσουν αν θεωρεί τον αρχικό μισθό σε μια δουλειά ασήμαντο ή όχι. Ο μέσος όρος των απαντήσεων από τους εργαζομένους ήταν 3,31, η τυπική απόκλιση 0,7 ενώ της κλίμακας που χρησιμοποιήθηκε ο μέσος όρος ήταν 3.
Ο έλεγχος τον οποίο εξετάζουμε στην παρούσα άσκηση είναι ο παρακάτω

H_0: Μέσος όρος απαντήσεων =3
H_1: Μέσος όρος απαντήσεων >3

Υπολογίζουμε την στατιστική συνάρτηση Ζ= (3,31-3)/(0.7/ $\sqrt{172}$)=0.31/0,05337=5,808

Όμως $Z_{0.01}$ =2,57. Επομένως 2,57<5,808 το οποίο σημαίνει ότι απορρίπτω την μηδενική υπόθεση ότι ο μέσος όρος των απαντήσεων είναι μικρότερος του 3. Άρα δεχόμαστε την εναλλακτική.
Το συμπέρασμα του ελέγχου είναι ότι αν θεωρήσουμε το 3 ως μια ενδιάμεση κατάσταση (ούτε σημαντικός, ούτε ασήμαντος) παρατηρούμε ότι το μεγαλύτερο μέρος του δείγματος θεωρεί, τείνει στο να θεωρεί τον αρχικό μισθό του σε μια δουλειά ως πολύ σημαντικό και όχι ως ασήμαντο.

Εφαρμογή 4

Σε έρευνα πελατών ενός καταστήματος σε δείγμα 62 ατόμων ο μέσος των εβδομαδιαίων αγορών είναι 112,3 και η τυπική απόκλιση 22,8. Να δημιουργηθεί διάστημα εμπιστοσύνης 95% και 99%. Να εξηγηθεί η σημασία των διαστημάτων αυτών

Το διάστημα εμπιστοσύνης που θέλουμε είναι της μορφής

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

a=0,01 => $z_{a/2} = 2.58$

a=0.05=> $z_{a/2} = 1.96$

$$112.3 - z_{a/2}\frac{22.8}{\sqrt{62}} < m < 112.3 + z_{a/2}\frac{22.8}{\sqrt{62}}$$

a=0,01 => (104,8, 119.75)
a=0.05=> (106,6, 118)

Παρατηρούμε ότι το διάστημα εμπιστοσύνης για α=0,01 να είναι μεγαλύτερο του διαστήματος εμπιστοσύνης για α=0,05. Αυτό είναι όντως λογικό διότι όσο μειώνεται το επίπεδο σημαντικότητας τόσο αυξάνει η τιμή του z.

Εφαρμογή 5

Σε έρευνα 531 εμπειρογνωμόνων 284 χρησιμοποιούν Microsoft και 236 websoft. Να δημιουργηθεί διάστημα εμπιστοσύνης 99% για το ποσοστό που χρησιμοποιούν Microsoft και για αυτού που χρησιμοποιούν websoft.

Το επίπεδο σημαντικότητας είναι α=0,01=> $z_{a/2} = 2.58$
Το ποσοστό για Microsoft είναι ρ=284/531=0.535

$$p - z_{a/2}\sqrt{p(1-p)/N} < P < p + z_{a/2}\sqrt{p(1-p)/N}$$
$$0.535 - 2.58\sqrt{0.535*0.465/531} < P < 0.535 + 2.58\sqrt{0.535*0.465/531}$$
$$(0.479, 0.591)$$

Το ποσοστό για websoft είναι ρ=206/531=0.388

$$0.535 - 2.58\sqrt{0.388*0.612/531} < P < 0.535 + 2.58\sqrt{0.388*0.612/531}$$
$$(0.034, 0.442)$$

Εφαρμογή 6

Η εταιρεία σιδηροδρόμων λειτουργεί τηλεφωνικό κέντρο πληροφοριών για δρομολόγια και τιμές εισιτηρίων. Ένα τυχαίο δείγμα 150 τηλεφωνημάτων καταγράφηκε και υπολογίσθηκε ο μέσος χρόνος αναμονής σε 95 δευτερόλεπτα με τυπική απόκλιση 13 δευτερόλεπτα.

i) Να υπολογιστεί ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνος αναμονής των τηλεφωνικών κλήσεων.
ii) Ο υπεύθυνος του τηλεφωνικού κέντρου ισχυρίζεται ότι ο μέσος χρόνος αναμονής είναι το πολύ 90 δευτερόλεπτα. Υπάρχει στατιστικά σημαντική διαφορά μεταξύ του ισχυρισμού και των στοιχείων του δείγματος σε επίπεδο 5%;
iii) Θα είχαμε διαφορετικό αποτέλεσμα εάν το επίπεδο σημαντικότητας ήταν 1%;

i) Το διάστημα εμπιστοσύνης που ψάχνουμε είναι της μορφής (για μεγάλα δείγματα)

$$\bar{Y} - Z_{a/2}\frac{s}{\sqrt{n}} < m < \bar{Y} + Z_{a/2}\frac{s}{\sqrt{n}}$$

Επομένως στην περίπτωση μας θα έχουμε

$$95 - Z_{0,025}\frac{13}{\sqrt{150}} < m < 95 + Z_{0,025}\frac{13}{\sqrt{150}}$$
$$95 - 2,08 < m < 95 + 2,08$$

92,92 < μ < 97,08

ii) Θέλουμε να διεξάγουμε τον παρακάτω έλεγχο (μονόπλευρος)

Ho: μ = 90
H1: μ <90

Παίρνουμε την τυχαία μεταβλητή $z = \dfrac{\bar{Y} - m_o}{\dfrac{s}{\sqrt{n}}} = \dfrac{95-90}{\dfrac{13}{\sqrt{150}}} = \dfrac{5}{1.061} = 4.71$

Απορρίπτουμε την **Ho** αν $Z>Z_\alpha$. Επομένως επειδή 4,71 > 1,64 θα απορρίψουμε την μηδενική υπόθεση. Εδώ $Z_\alpha = Z_{0.05}$.

iii) Για $Z_\alpha = Z_{0.01}=2{,}33$ έχουμε το ίδιο αποτέλεσμα

Εφαρμογή 7

Η περιεκτικότητα σε λιπαρά κάποιου προϊόντος μετρήθηκε από τον παραγωγό και τον αγοραστή του προϊόντος. Ο παραγωγός έκανε 80 ελέγχους και βρήκε τη μέση περιεκτικότητα σε λιπαρά ίση με 13,108 και με τυπική απόκλιση 0,014. Ο αγοραστής έκανε 65 ελέγχους και βρήκε τη μέση περιεκτικότητα σε λιπαρά ίση με 13,101 και με τυπική απόκλιση 0,016. Να σχολιαστεί το αποτέλεσμα.

Θα διεξάγουμε τον παρακάτω έλεγχο για να δούμε αν τα αποτελέσματα των δύο ελέγχων διαφέρουν ή όχι.

Ho: $\mu_1 = \mu_2$
H1: $\mu_1 \neq \mu_2$

Παίρνουμε την τυχαία μεταβλητή

$$z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{13.108 - 13.101}{\sqrt{\frac{0.014^2}{80} + \frac{0.016^2}{65}}} = \frac{0.007}{\sqrt{0.00000245 + 0.000004}} = \frac{0.007}{0.00252} = 2.769$$

Απορρίπτουμε την **Ho** αν $|Z| > Z_{\alpha/2}$. Επομένως επειδή 2,769 > 1,96 θα απορρίψουμε την μηδενική υπόθεση. Εδώ $Z_{\alpha/2} = Z_{0.025}$. Άρα τα αποτελέσματα των ελέγχων διαφέρουν μεταξύ τους.

Εφαρμογή 8

Σε ένα τυχαίο δείγμα 18 φοιτητών από κάποια σχολή της Αθήνας υπολογίσθηκε ο μέσος δείκτης νοημοσύνης σε 109 με τυπική απόκλιση ίση με 9. Σε ένα δεύτερο τυχαίο δείγμα 15 Αθηναίων φοιτητών από άλλη σχολή υπολογίσθηκε ο μέσος δείκτης νοημοσύνης τους σε 113 με τυπική απόκλιση ίση με 10. Υπάρχει σημαντική διαφορά μεταξύ των μέσων δεικτών νοημοσύνης των φοιτητών των δύο σχολών όπως προκύπτει από τα δείγματα σε επίπεδο σημαντικότητας 1% και 5%;

Θα διεξάγουμε τον παρακάτω έλεγχο (μικρά δείγματα) για να δούμε αν τα αποτελέσματα των δύο δειγματοληψιών διαφέρουν ή όχι.

Ho: $\mu_1 = \mu_2$
H1: $\mu_1 \neq \mu_2$

Παίρνουμε την τυχαία μεταβλητή $z = \dfrac{\bar{Y}_1 - \bar{Y}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ όπου $s = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

$$s=\sqrt{\frac{(18-1)9^2+(15-1)10^2}{18+15-2}}=\sqrt{\frac{(17)81+(14)100}{31}}=9.464$$

$$z=\frac{\bar{Y}_1-\bar{Y}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{109-113}{9.464\sqrt{\frac{1}{18}+\frac{1}{15}}}=\frac{-4}{9.464\sqrt{\frac{1}{18}+\frac{1}{15}}}=-1.209$$

Απορρίπτουμε την **Ho** αν $|t|>t_{n_1+n_2-2,0.025}$. Επομένως επειδή 1.209 < 2.04 δεν θα απορρίψουμε την μηδενική υπόθεση. Εδώ $t_{n_1+n_2-2,0.025}=t_{31,0.025}=2.04$

Για $t_{n_1+n_2-2,0.005}=t_{31,0.005}=2.75$.

Και στις δύο περιπτώσεις δεχόμαστε την μηδενική υπόθεση ότι οι μέσες τιμές των δεικτών νοημοσύνης δεν διαφέρουν στατιστικά μεταξύ τους.

Εφαρμογή 9

Έστω από έναν κανονικό πληθυσμό του οποίου δεν γνωρίζουμε τη διακύμανση παίρνουμε δείγμα μεγέθους $n=14$. Από το δείγμα αυτό βρέθηκε να έχει δειγματικό μέσο ίσο με 52,52 και τυπική απόκλιση ίση με 3,37. Ζητείται το 95% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού.

Επειδή έχουμε αγνωστη πληθυσμιακή διακύμανση και $n=14<30$, χρησιμοποιούμε την κατανομή $t-Student$, με $n-1=13$ βαθμούς ελευθερίας, για να κατασκευάσουμε το 95% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού. Έτσι έχουμε

$$\bar{X}-t_{n-1,a/2}\frac{S}{\sqrt{n}}<m<\bar{X}+t_{n-1,a/2}\frac{S}{\sqrt{n}}$$

$$50{,}58<m<54{,}47$$

όπου $t_{n-1,a/2}=2{,}16$.

Εφαρμογή 10

Η κατανάλωση βενζίνης για ένα τυχαίο δείγμα 6 αυτοκίνητων δίνεται πιο κάτω:
18,6 18,4 19,2 20,8 19,4 20,5.
Να κατασκευάσετε ένα 90% διάστημα εμπιστοσύνης για την μέση κατανάλωση στο πληθυσμό κάτω από την υπόθεση ότι η κατανομή του πληθυσμού είναι κανονική.

Πρέπει να υπολογίσουμε το μέσο και την διακύμανση του δείγματος.

$$\bar{X}=\frac{1}{6}\sum X_i=19{,}48$$
$$S_X^2=\frac{1}{n-1}\left(\sum X_i-n\bar{X}^2\right)=1{,}12$$
$$S_X=1{,}06$$

Επειδή έχουμε κανονικό πληθυσμό με άγνωστη διακύμανση το Δ.Ε. δίνεται από,

$$\overline{X}-\frac{t_{n-1,a/2}S_X}{\sqrt{n}}<m<\overline{X}+\frac{t_{n-1,a/2}S_X}{\sqrt{n}}$$ (με $\overline{X}=19{,}48 \quad S_X=1{,}06 \quad n=6 \quad a=0{,}10 \quad t_{n-1,a/2}=2{,}015$)

$18{,}61<m<20{,}35$

Εφαρμογή 11

Εμβολιάζονται 200 άνθρωποι για το νέο ιό της γρίπης και οι 185 αποκτούν ανοσία. Ζητείται να ορίσουμε το ποσοστό επιτυχιών p 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_o και είναι ίσο με 2.58. Βρίσκουμε διάστημα εμπιστοσύνης:

$0.925 \pm 2.58 \text{ X } (0.925 \text{ X } 0.075/200)^{1/2} = 0.925 \pm 0.048 = 0.877$ έως 0.973.

Εφαρμογή 12

Το ACT σκορ ενός τυχαίου δείγματος 61 τελειόφοιτων λυκείου αναλύθηκε και βρέθηκε ότι η μέση τιμή είναι ίση με 25,1 και η τυπική απόκλιση ίση με 3,6. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης.

$$\overline{X}-z_{\frac{a}{2}}\frac{s}{\sqrt{n}}\leq m\leq \overline{X}+z_{\frac{a}{2}}\frac{s}{\sqrt{n}}$$

$$25.1-1.96\frac{3.6}{\sqrt{61}}\leq m\leq 25.1+1.96\frac{3.6}{\sqrt{61}}$$

$$24.197\leq m\leq 26.003$$

Εφαρμογή 13

Ένα τεστ IQ δόθηκε σε ένα τυχαίο δείγμα 101 παιδιά της Γ' λυκείου. Ο μέσος δειγματικός όρος ήταν 104,2 και η τυπική απόκλιση ίση με 12. Να βρεθεί ένα 99% διάστημα εμπιστοσύνης.

$$\overline{X}-z_{\frac{a}{2}}\frac{s}{\sqrt{n}}\leq m\leq \overline{X}+z_{\frac{a}{2}}\frac{s}{\sqrt{n}}$$

$$104.2-2.58\frac{12}{\sqrt{101}}\leq m\leq 104.2+2.58\frac{12}{\sqrt{101}}$$

$$101.117\leq m\leq 107.283$$

Εφαρμογή 14

Το μέσο ύψος ενός τυχαίου δείγματος 41 ενήλικων γυναικών είναι 66 inches με τυπική απόκλιση 1,9 inche. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού.

$$\overline{X} - z_{\frac{a}{2}} \frac{s}{\sqrt{n}} \le m \le \overline{X} + z_{\frac{a}{2}} \frac{s}{\sqrt{n}}$$

$$66 - 1.96 \frac{1.9}{\sqrt{41}} \le m \le 66 + 1.96 \frac{1.9}{\sqrt{41}}$$

$$65.419 \le m \le 66.581$$

Εφαρμογή 15

Το μέσο ύψος ενός τυχαίου δείγματος 21 ενήλικων ανδρών είναι 70 inches με τυπική απόκλιση 2.2 . Να βρεθεί ένα 99% διάστημα εμπιστοσύνης.

$$\bar{x} - t^* \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t^* \frac{s}{\sqrt{n}}$$

$$70 - 2.845 \frac{2.2}{\sqrt{21}} \le \mu \le 70 + 2.845 \frac{2.2}{\sqrt{21}}$$

$$68.63 \le \mu \le 71.37$$

Εφαρμογή 16

Έξη συνθετικά νήματα κόπηκαν στη μέση. Στο ένα τμήμα από κάθε ζεύγος εφαρμόστηκε μία ειδική χημική επεξεργασία για την αύξηση της αντοχής του, ενώ το άλλο αφέθηκε όπως είχε. Με βάση τα παρακάτω δεδομένα, όπου x_1 εκφράζει το δείκτη αντοχής του τμήματος με χημική επεξεργασία και x_2 το δείκτη αντοχής του τμήματος χωρίς χημική επεξεργασία, να εξεταστεί αν αυξάνει κατά 2.0 τουλάχιστον μονάδες ο δείκτης αντοχής των τμημάτων με χημική επεξεργασία (α=0.10).

x_1	15.2	13.4	14.6	15.1	13.1	15.3
x_2	12.7	10.8	12.8	12.9	11.0	13.0

Έχουμε έναν έλεγχο διαφοράς μέσων εξαρτημένων δειγμάτων.

x_1	15.2	13.4	14.6	15.1	13.1	15.3
x_2	12.7	10.8	12.8	12.9	11.0	13.0
Διαφορές D_i	2.5	2.6	1.8	2.2	2.1	2.3

Ho: $\mu_1 - \mu_2 = 2 = \mu_D$

H_1: $\mu_1 - \mu_2 > 2 = \mu_D$.

Εύκολα προκύπτει ότι $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = 2.25, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2 = 0.083, \quad n = 6$

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K : |t| = \frac{|\bar{D} - m_D|}{S_D / \sqrt{n}} > t_{a,n-1}, a = 0.10$$

Έτσι |t| = 2.13, ενώ από τους πίνακες της κατανομής t προκύπτει ότι $t_{a,n-1} = t_{0.10,5} = 1.476$, οπότε αποδεχόμαστε την H_0, δηλαδή πράγματι αυξάνει κατά 2.0 τουλάχιστον μονάδες ο δείκτης αντοχής των τμημάτων με χημική επεξεργασία.

Εφαρμογή 17

Από την παραγωγή τυχαίο δείγμα μεγέθους n=16 έδωσε δειγματική απόκλιση S = 5.5. Αν η μεγαλύτερη επιτρεπόμενη τυπική απόκλιση είναι σ_0 = 4 να εξεταστεί αν η παραπάνω υπέρβαση είναι στατιστικά σημαντική ή όχι (α=0.05).

Έχουμε έναν έλεγχο διασποράς της μορφής

Ho: $\sigma = 4 = \sigma_0$

H_1: $\sigma > 4 = \sigma_0$.

Από τους πίνακες προκύπτει ότι η κρίσιμη περιοχή είναι:

$$K : |x^2| = \frac{(n-1)S^2}{s_o^2} > X^2_{n-1,a}, S^2 = 5.5^2, n = 16, a = 0.05$$

Έτσι | x^2 | =28.36, ενώ από τους πίνακες της κατανομής χ^2 προκύπτει ότι $X^2_{n-1,a} = X^2_{15,0.05} = 25$, οπότε απορρίπτουμε την H_0 και άρα η υπέρβαση είναι στατιστικά σημαντική.

Εφαρμογή 18

Να υπολογιστούν τα διαστήματα εμπιστοσύνης για τις ακόλουθες περιπτώσεις:

a) n = 44 $\hat{p}$ =.51 99% C.I.

b) n = 300 $\hat{p}$ = .82 95% C.I.

c) n = 115 $\hat{p}$ = .48 90% C.I.

a) $z_{.005}$ = 2.575

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .51 \pm 2.575\sqrt{\frac{(.51)(.49)}{44}} = .51 \pm .194 = .316 \leq p \leq .704$$

b) $z_{.025}$ = 1.96

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .82 \pm 1.96\sqrt{\frac{(.82)(.18)}{300}} = .82 \pm .043 = .777 \le p \le .863$$

c) $z_{.05} = 1.645$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .48 \pm 1.645\sqrt{\frac{(.48)(.52)}{1150}} = .48 \pm .024 = .456 \le p \le .504$$

Εφαρμογή 19

Να υπολογιστούν τα διαστήματα εμπιστοσύνης για τις ακόλουθες περιπτώσεις:
a) $n = 116$ $x = 57$ 99% C.I.
b) $n = 60$ $x = 21$ 90% C.I.

a) $z_{.005} = 2.575$

$$\hat{p} = \frac{x}{n} = \frac{57}{116} = .49$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .49 \pm 2.575\sqrt{\frac{(.49)(.51)}{116}} = .49 \pm .12 = .37 \le p \le .61$$

d) $z_{.05} = 1.645$

$$\hat{p} = \frac{x}{n} = \frac{21}{60} = .35$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .35 \pm 1.645\sqrt{\frac{(.35)(.65)}{60}} = .35 \pm .10 = .25 \le p \le .45$$

Εφαρμογή 20

Να υπολογιστούν τα διαστήματα εμπιστοσύνης για τις ακόλουθες περιπτώσεις:

a) $n = 12$ $\bar{x} = 28.4$ $s^2 = 44.9$ 99% C.I. df = 12 – 1 = 11

$\chi^2_{.995,11} = 2.60321$ $\chi^2_{.005,11} = 26.7569$

$$\frac{(12-1)(44.9)}{26.7569} \le s^2 \le \frac{(12-1)(44.9)}{2.60321}$$

$18.46 \leq s^2 \leq 189.73$

b) $n = 7$ $\bar{x} = 4.37$ $s = 1.24$ $s^2 = 1.5376$ 95% C.I. df = 12 – 1 = 11

$\chi^2_{.975,6} = 1.237347$ $\chi^2_{.025,6} = 14.4494$

$$\frac{(7-1)(1.5376)}{14.4494} \leq s^2 \leq \frac{(7-1)(1.5376)}{1.237347}$$

$0.64 \leq s^2 \leq 7.46$

c) $n = 20$ $\bar{x} = 105$ $s = 32$ $s^2 = 1024$ 90% C.I. df = 20 – 1 = 19

$\chi^2_{.95,19} = 10.117$ $\chi^2_{.05,19} = 30.1435$

$$\frac{(20-1)(1024)}{30.1435} \leq s^2 \leq \frac{(20-1)(1024)}{10.117}$$

$645.45 \leq s^2 \leq 1923.10$

Εφαρμογή 21

Από ελέγχους που έγιναν σε 600 μικρομεσαίες ανώνυμες εταιρείες με έδρα την Αθήνα και άλλες τόσες με έδρα τη Θεσσαλονίκη διαπιστώθηκαν παραβάσεις της εργατικής νομοθεσίας σε 98 επιχειρήσεις της Αθήνας και 74 της Θεσσαλονίκης.

Α) Υπολογίστε κι ερμηνεύστε ένα 99% διάστημα εμπιστοσύνης για τη διαφορά των αναλογιών των μικρομεσαίων ανωνύμων εταιρειών Αθήνας και Θεσσαλονίκης που παραβιάζουν την εργατική νομοθεσία.

Β) Με κατάλληλο έλεγχο υπόθεσης εξετάστε, σε επίπεδο σημαντικότητας 0.01, αν οι αναλογίες των μικρομεσαίων ανωνύμων εταιρειών που παραβιάζουν την εργατική νομοθεσία διαφέρουν μεταξύ Αθήνας και Θεσσαλονίκης.

Α) Θέλουμε να βρούμε ένα διάστημα εμπιστοσύνης για τη διαφορά δύο αναλογιών. Το .δ.ε στην περίπτωση αυτή βρίσκεται από τον τύπο

$$\hat{p}_1 - \hat{p}_2 - z_{1-\frac{a}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{1-\frac{a}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Όπου

$n_1 = n_2 = 600$

$p_1 = 98/600 = 0.163$, $q_1 = 0.837$

$p_2 = 74/600 = 0.123$, $q_2 = 0.877$

$a = 1-0.99 = 0.01 \Leftrightarrow \alpha/2 = 0.005$ και $z\alpha/2 = z0.005 = 2.576$, Αντικαθιστώντας στον παραπάνω τύπο προκύπτει:

$$-0.012 < p_1 – p_2 < 0.092$$

B) Θα ακολουθήσουμε τα εξής βήματα:
Κάνουμε τον αμφίπλευρο έλεγχο
H0 : p1 – p2 = 0
H1 : p1 – p2 $\neq$ 0
Η συνάρτηση που θα χρησιμοποιηθεί είναι η εξής:

$$Z = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Η κρίσιμη περιοχή είναι η

K: z < - zα/2 ή z > zα/2

Εφόσον α = 0.01 βρίσκουμε $z_{\alpha/2} = z_{0,005} = 2.576$. Έχουμε $n_1 = n_2 = 600$, $p_1 = 98/600 = 0.163$, $q_1 = 0.837$, $p_2 = 74/600 = 0.123$, $q_2 = 0.877$
.Αντικαθιστώντας στις σχέσεις έχουμε z = 1.982
Επειδή $|z| = 1.982 < z_{0.005} = 2.576$ η παρατηρούμενη τιμή. βρίσκεται εκτός της κρίσιμης περιοχής. Επομένως αποδεχόμαστε την H_o. Τελικά δεν υπάρχουν σημαντικές ενδείξεις στο 0.01 επίπεδο σημαντικότητας ότι οι αναλογίες των μικρομεσαίων ανωνύμων εταιρειών Αθήνας και Θεσσαλονίκης που παραβιάζουν την εργατική νομοθεσία διαφέρουν.

Εφαρμογή 22

Μια εταιρεία πλακιδίων διαφημίζει ότι παραδίδει κάθε παραγγελία πλακιδίων που δέχεται εντός 15 ημερών από την ημέρα της αγοράς. Ένα δείγμα 49 πελατών της λαμβάνεται για να ελέγξει τον ισχυρισμό της. Ο μέσος χρόνος παράδοσης είναι 16,2 μέρες με τυπική απόκλιση 5,6 μέρες. Χρησιμοποιώντας επίπεδο σημαντικότητα 5% να ελεγχθεί ο ισχυρισμός της εταιρείας.

Θέλουμε να ελέγξουμε την παρακάτω υπόθεση

$$H_0 : \mu = 15$$
$$H_A : \mu \neq 15$$

Όπου το μ αντιστοιχεί στο χρόνο παράδοσης. Για τον έλεγχο της υπόθεσης χρησιμοποιούμε την στατιστική συνάρτηση

$$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

Ενώ η τυπική απόκλιση είναι ίση με 5,6, το πλήθος είναι 49. Η κρίσιμη περιοχή είναι K: $Z > Z_{\alpha/2}$ ή $Z < -Z_{\alpha/2}$,

Έχουμε λοιπόν ότι

$$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{16.2 - 15}{5.6 / \sqrt{49}} = 1.5$$

$Z_{\alpha/2}=1.96$ 1.5< 1.96 οπότε αποδεχόμαστε την H_0.

Εφαρμογή 23

Στη παρούσα εφαρμογή εξετάζουμε διάφορες μεταβλητές για το αν διαφέρουν ως προς τα δημογραφικά του δείγματος (δείγμα 200 ατόμων) όπως το φύλο, η ηλικία κ.α. Παρακάτω παρουσιάζουμε πρώτα τα περιγραφικά χαρακτηριστικά των ποσοτικών μεταβλητών και έπειτα συνεχίζουμε με τον έλεγχο t-test και την ανάλυση διασποράς.

Descriptive Statistics

	N	Mean	Std. Deviation	Variance
power_distance	200	-1,8500	4,21501	17,766
individ	200	,1000	4,26391	18,181
mas_fem	199	-2,5879	3,99887	15,991
u_av	200	2,2900	3,20268	10,257
l_t_o	200	2,1950	3,78399	14,319
risk	200	1,8173	,31414	,099

Από τον παραπάνω πίνακα παρατηρούμε ότι η μέση τιμή του δείκτη της ανισότητας είναι ίσος με -1,85, της ατομικότητας ίσος με 0,1, της θηλυκότητας - ανδρισμού είναι ίσος με -2,58, της αποφυγής – αβεβαιότητας είναι ίσος με 2,29, του μακροπρόθεσμου – βραχυπρόθεσμου προσανατολισμού είναι ίσος με 2,19 και του ρίσκου είναι ίσος με 1,81. Επιπλέον πρέπει να παρατηρήσουμε ότι οι τιμές για τις τυπικές αποκλίσεις των έξι μεταβλητών παρουσιάζουν υψηλές τιμές σε σχέση με τις μέσες τιμές τους. Αυτό σημαίνει ότι υπάρχει μεγάλη μεταβλητότητα στα δεδομένα μας. Ακόμα κάποια συμπεράσματα που μπορούν να εξαχθούν είναι ότι το επίπεδο της ισότητας θεωρείται καλό με βάσει την τιμή του δείκτη, χαμηλό είναι επίσης το επίπεδο της ατομικότητας που σημαίνει ότι υπάρχει σε αναπτυγμένο επίπεδο το αίσθημα της συλλογικότητας, ο δείκτης θηλυκότητας – ανδρισμού είναι χαμηλός που σημαίνει ότι έχουμε περισσότερο θηλυκούς ρόλους μεταξύ των φύλων, το επίπεδο της αβεβαιότητας είναι και αυτό χαμηλό που σημαίνει ότι η κοινωνία αποδέχεται την αβεβαιότητα, μπορεί να την διαχειριστεί, το επίπεδο του μακροπρόθεσμου προσανατολισμού είναι χαμηλό που σημαίνει ότι η κοινωνία έχει βραχυπρόθεσμους στόχους κυρίως και τέλος το επίπεδο του επιχειρηματικού ρίσκου παρουσιάζεται υψηλό, άνω του μέσου όρου.

T – test έλεγχος

Στη παρούσα ενότητα πρόκειται να εξετάσουμε αν υπάρχουν διαφορές στις απόψεις μεταξύ των ανδρών και των γυναικών στους έξι παράγοντες που έχει το ερωτηματολόγιο μας. Για τον σκοπό αυτό χρησιμοποιούμε τον έλεγχο t - test.

Group Statistics

	Φύλο	N	Mean	Std. Deviation	Std. Error Mean
power_distance	Άντρας	107	-,9813	4,04918	,39145
	Γυναίκα	93	-2,8495	4,20131	,43566
individ	Άντρας	107	,5047	4,46448	,43160
	Γυναίκα	93	-,3656	3,99398	,41416
mas_fem	Άντρας	106	-1,3868	3,76341	,36553
	Γυναίκα	93	-3,9570	3,83325	,39749
u_av	Άντρας	107	1,8972	3,37005	,32580
	Γυναίκα	93	2,7419	2,95217	,30613
l_t_o	Άντρας	107	2,2897	3,85592	,37277
	Γυναίκα	93	2,0860	3,71734	,38547
risk	Άντρας	107	1,9044	,30328	,02932
	Γυναίκα	93	1,7171	,29752	,03085

Από τον παραπάνω πίνακα παρατηρούμε τις μέσες τιμές, τις τυπικές αποκλίσεις και το τυπικό σφάλμα της μέσης τιμής για τις έξι μεταβλητές. Οι μέσες τιμές ανά κατηγορία θα μας βοηθήσουν να εντοπίσουμε κατά ποιο τρόπο διαφέρει η κατηγορία με την άλλη στην περίπτωση που ο έλεγχος t – test φανερώσει στατιστικά σημαντικές διαφορές μεταξύ των κατηγοριών.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
power_distance	Equal variances assumed	,000	,989	3,198	198	,002	1,86815	,58417	,71616	3,02015
	Equal variances not assumed			3,190	191,936	,002	1,86815	,58569	,71295	3,02336
individ	Equal variances assumed	1,937	,166	1,444	198	,150	,87026	,60285	-,31857	2,05910
	Equal variances not assumed			1,455	197,827	,147	,87026	,59817	-,30933	2,04986
mas_fem	Equal variances assumed	,001	,969	4,765	197	,000	2,57020	,53936	1,50653	3,63386
	Equal variances not assumed			4,760	192,669	,000	2,57020	,54001	1,50510	3,63529
u_av	Equal variances assumed	1,685	,196	-1,872	198	,063	-,84474	,45121	-1,73453	,04506
	Equal variances not assumed			-1,890	197,986	,060	-,84474	,44705	-1,72633	,03686
l_t_o	Equal variances assumed	,220	,639	,379	198	,705	,20370	,53761	-,85648	1,26388
	Equal variances not assumed			,380	195,862	,704	,20370	,53623	-,85383	1,26122
risk	Equal variances assumed	,014	,904	4,394	198	,000	,18726	,04262	,10322	,27131
	Equal variances not assumed			4,400	195,099	,000	,18726	,04256	,10333	,27120

Από τον παραπάνω πίνακα παρατηρούμε ότι διαφέρουν οι απαντήσεις των ανδρών και γυναικών στο θέμα της ανισότητας, της θηλυκότητας και του ρίσκου διότι 0,02 <0,05, 0,00.. <0,05 και 0,00..<0,05 αντίστοιχα. Οι γυναίκες θεωρούν ότι υπάρχει υψηλότερο επίπεδο ισότητας, θηλυκότητας έναντι των ανδρών ενώ αναλαμβάνουν και μικρότερου επιπέδου επιχειρηματικό ρίσκο.

Οικονομική εκπαίδευση

Group Statistics

	Οικονομική Εκπαίδευση	N	Mean	Std. Deviation	Std. Error Mean
power_distance	Ναι	79	-2,0253	3,94830	,44422
	Όχι	121	-1,7355	4,39274	,39934
individ	Ναι	79	-,2405	4,24329	,47741
	Όχι	121	,3223	4,28022	,38911
mas_fem	Ναι	79	-2,0000	3,99679	,44967
	Όχι	120	-2,9750	3,96934	,36235
u_av	Ναι	79	2,4557	3,27303	,36825
	Όχι	121	2,1818	3,16491	,28772
l_t_o	Ναι	79	3,0886	3,61688	,40693
	Όχι	121	1,6116	3,79116	,34465
risk	Ναι	79	1,9951	,28016	,03152
	Όχι	121	1,7012	,27969	,02543

Από τον παραπάνω πίνακα παρατηρούμε τις μέσες τιμές, τις τυπικές αποκλίσεις και το τυπικό σφάλμα της μέσης τιμής για τις έξι μεταβλητές. Οι μέσες τιμές ανά κατηγορία θα μας βοηθήσουν να εντοπίσουμε κατά ποιο τρόπο διαφέρει η κατηγορία με την άλλη στην περίπτωση που ο έλεγχος t – test φανερώσει στατιστικά σημαντικές διαφορές μεταξύ των κατηγοριών.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
power_distance	Equal variances assumed	,889	,347	-,474	198	,636	-,28978	,61088	-1,49444	,91488
	Equal variances not assumed			-,485	179,017	,628	-,28978	,59733	-1,46849	,88893
individ	Equal variances assumed	,044	,834	-,912	198	,363	-,56282	,61702	-1,77960	,65396
	Equal variances not assumed			-,914	167,894	,362	-,56282	,61589	-1,77871	,65307
mas_fem	Equal variances assumed	,435	,510	1,691	197	,092	,97500	,57667	-,16225	2,11225
	Equal variances not assumed			1,688	166,239	,093	,97500	,57750	-,16518	2,11518
u_av	Equal variances assumed	,120	,729	,590	198	,556	,27388	,46402	-,64117	1,18893

	Equal variances not assumed			,586	162,853	,559	,27388	,46732	-,64891	1,19666
l_t_o	Equal variances assumed	,405	,525	2,742	198	,007	1,47704	,53859	,41493	2,53914
	Equal variances not assumed			2,770	172,383	,006	1,47704	,53327	,42446	2,52962
risk	Equal variances assumed	,016	,899	7,260	198	,000	,29392	,04048	,21409	,37376
	Equal variances not assumed			7,258	166,662	,000	,29392	,04050	,21397	,37388

Από τον παραπάνω πίνακα παρατηρούμε ότι διαφέρουν οι απαντήσεις των ατόμων με οικονομική εκπαίδευση και των ατόμων χωρίς οικονομική εκπαίδευση στο θέμα του μακροπρόσθεσμου – βραχυπρόθεσμου προσανατολισμού και του επιχειρηματικού ρίσκου διότι 0,07 <0,05 και 0,00..<0,05 αντίστοιχα. Τα άτομα με οικονομική εκπαίδευση έχουν μακροπρόθεσμο προσανατολισμό και αναλαμβάνουν μεγαλύτερο επιχειρηματικό ρίσκο σε σχέση με τα άτομα που δεν έχουν οικονομική εκπαίδευση.

Οικογενειακό εισόδημα

Group Statistics

	Οικογενειακό εισόδημα	N	Mean	Std. Deviation	Std. Error Mean
risk	< 30.000	88	1,7404	,25158	,02682
	50.000 - 100.000	24	2,0641	,40238	,08213

Από τον παραπάνω πίνακα παρατηρούμε τις μέσες τιμές, τις τυπικές αποκλίσεις και το τυπικό σφάλμα της μέσης τιμής για το επιχειρηματικό ρίσκο. Οι μέσες τιμές ανά κατηγορία θα μας βοηθήσουν να εντοπίσουμε κατά ποιο τρόπο διαφέρει η κατηγορία με την άλλη στην περίπτωση που ο έλεγχος t – test φανερώσει στατιστικά σημαντικές διαφορές μεταξύ των κατηγοριών.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
risk	Equal variances assumed	9,475	,003	-4,853	110	,000	-,32372	,06671	-,45592	-,19152
	Equal variances not assumed			-3,747	28,081	,001	-,32372	,08640	-,50068	-,14675

Από τον παραπάνω πίνακα παρατηρούμε ότι διαφέρουν οι απαντήσεις των ατόμων με οικογενειακό εισόδημα κάτω των 30.000 ευρώ και των ατόμων με οικογενειακό εισόδημα από 50.000 έως 100.000 ευρώ στο θέμα του επιχειρηματικού ρίσκου διότι 0,001<0,05 αντίστοιχα. Τα άτομα με οικογενειακό εισόδημα από 50.000 έως 100.000 ευρώ αναλαμβάνουν υψηλότερο επιχειρηματικό ρίσκο έναντι των ατόμων με οικογενειακό εισόδημα κάτω των 30.000 ευρώ.

Ηλικία

Group Statistics

	Ηλικία	N	Mean	Std. Deviation	Std. Error Mean
risk	30	6	1,7179	,17985	,07342
	50	4	2,0577	,51170	,25585

Από τον παραπάνω πίνακα παρατηρούμε τις μέσες τιμές, τις τυπικές αποκλίσεις και το τυπικό σφάλμα της μέσης τιμής για το επιχειρηματικό ρίσκο. Οι μέσες τιμές ανά κατηγορία θα μας βοηθήσουν να εντοπίσουμε κατά ποιο τρόπο διαφέρει η κατηγορία με την άλλη στην περίπτωση που ο έλεγχος t – test φανερώσει στατιστικά σημαντικές διαφορές μεταξύ των κατηγοριών.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
risk	Equal variances assumed	28,669	,001	-1,530	8	,165	-,33974	,22212	-,85194	,17246
	Equal variances not assumed			-1,276	3,500	,280	-,33974	,26618	-1,12230	,44281

Από τον παραπάνω πίνακα παρατηρούμε ότι δεν διαφέρουν οι απαντήσεις των ατόμων με ηλικία 30 ετών σε σχέση με άτομα που έχουν ηλικία 50 ετών σε σχέση με το επιχειρηματικό ρίσκο.

Σχέση με την χρηματαγορά

Group Statistics

	Σχέση με τη χρηματαγορά	N	Mean	Std. Deviation	Std. Error Mean
risk	Ναι	67	2,0321	,31749	,03879
	Όχι	133	1,7091	,25129	,02179

Από τον παραπάνω πίνακα παρατηρούμε τις μέσες τιμές, τις τυπικές αποκλίσεις και το τυπικό σφάλμα της μέσης τιμής για τις έξι μεταβλητές. Οι μέσες τιμές ανά κατηγορία θα μας βοηθήσουν να εντοπίσουμε κατά ποιο τρόπο διαφέρει η κατηγορία με την άλλη στην περίπτωση που ο έλεγχος t – test φανερώσει στατιστικά σημαντικές διαφορές μεταξύ των κατηγοριών.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
risk	Equal variances assumed	4,704	,031	7,838	198	,000	,32307	,04122	,24178	,40435
	Equal variances not assumed			7,262	108,813	,000	,32307	,04449	,23489	,41124

Από τον παραπάνω πίνακα παρατηρούμε ότι διαφέρουν οι απαντήσεις των ατόμων που έχουν σχέση με την χρηματαγορά και των ατόμων που δεν έχουν σχέση με την χρηματαγορά διότι 0,00<0,05 αντίστοιχα. Τα άτομα που έχουν σχέση με την χρηματαγορά αναλαμβάνουν υψηλότερο επιχειρηματικό ρίσκο έναντι των ατόμων που δεν έχουν σχέση με την χρηματαγορά.

Ανάλυση διασποράς

Με τον έλεγχο της ανάλυσης διασποράς εξετάζουμε αν τα επίπεδα της εκπαίδευσης διαφέρουν στις μέσες τιμές των απαντήσεων τους για την κάθε μεταβλητή μας.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
POWER_DI	**Between Groups**	88,394	3	29,465	1,807	,151
	Within Groups	1500,095	92	16,305		
	Total	1588,490	95			
INDIVID	**Between Groups**	66,142	3	22,047	1,224	,306
	Within Groups	1657,264	92	18,014		
	Total	1723,406	95			
MAS_FEM	**Between Groups**	270,579	3	90,193	5,834	,001
	Within Groups	1422,410	92	15,461		
	Total	1692,990	95			
U_AV	**Between Groups**	40,124	3	13,375	1,229	,304
	Within Groups	1001,532	92	10,886		
	Total	1041,656	95			
L_T_O	**Between Groups**	277,612	3	92,537	7,418	,000
	Within Groups	1147,722	92	12,475		
	Total	1425,333	95			
RISK	**Between Groups**	,694	3	,231	3,239	,026
	Within Groups	6,572	92	,071		
	Total	7,266	95			

Από τον παραπάνω πίνακα παρατηρούμε ότι διαφέρουν τα επίπεδα της εκπαίδευσης για την μεταβλητή που αφορά την θηλυκότητα, τον μακροπρόθεσμο προσανατολισμό και το επιχειρηματικό ρίσκο διότι 0,001<0,05, 0,00<0,05 και 0,026<0,05 αντίστοιχα. Για την ανισότητα, την αβεβαιότητα και την ατομικότητα το επίπεδο της εκπαίδευσης δεν παίζει κανέναν ρόλο. Στο παρακάτω πίνακα εξετάζουμε ποια επίπεδα εκπαίδευσης διαφέρουν μεταξύ τους για τις μεταβλητές της θηλυκότητας, του μακροπρόθεσμο προσανατολισμού και του επιχειρηματικού ρίσκου.

Multiple Comparisons Bonferroni							
			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
Dependent Variable	(I) Αν ναι σε ποιό επίπεδο	(J) Αν ναι σε ποιό επίπεδο				Lower Bound	Upper Bound
MAS_FEM	**Β' Βάθμια Εκπαίδευση**	**ΙΕΚ**	3,4444	1,75147	,313	-1,2784	8,1673
		ΑΕΙ - ΤΕΙ	-1,4756	1,08082	1,000	-4,3900	1,4389
		Μεταπτυχιακό	-3,2698	1,26300	,067	-6,6755	,1359
	ΙΕΚ	**Β' Βάθμια Εκπαίδευση**	-3,4444	1,75147	,313	-8,1673	1,2784
		ΑΕΙ - ΤΕΙ	-4,9200(*)	1,58680	,015	-9,1988	-,6412
		Μεταπτυχιακό	-6,7143(*)	1,71609	,001	-11,3417	-2,0869
	ΑΕΙ - ΤΕΙ	**Β' Βάθμια Εκπαίδευση**	1,4756	1,08082	1,000	-1,4389	4,3900
		ΙΕΚ	4,9200(*)	1,58680	,015	,6412	9,1988
		Μεταπτυχιακό	-1,7943	1,02248	,496	-4,5514	,9628
	Μεταπτυχιακό	**Β' Βάθμια Εκπαίδευση**	3,2698	1,26300	,067	-,1359	6,6755
		ΙΕΚ	6,7143(*)	1,71609	,001	2,0869	11,3417
		ΑΕΙ - ΤΕΙ	1,7943	1,02248	,496	-,9628	4,5514
L_T_O	**Β' Βάθμια Εκπαίδευση**	**ΙΕΚ**	-4,0317	1,57329	,072	-8,2741	,2106
		ΑΕΙ - ΤΕΙ	-3,4289(*)	,97086	,004	-6,0468	-,8110
		Μεταπτυχιακό	-5,2222(*)	1,13452	,000	-8,2814	-2,1630
	ΙΕΚ	**Β' Βάθμια Εκπαίδευση**	4,0317	1,57329	,072	-,2106	8,2741
		ΑΕΙ - ΤΕΙ	,6029	1,42537	1,000	-3,2407	4,4464
		Μεταπτυχιακό	-1,1905	1,54150	1,000	-5,3471	2,9662
	ΑΕΙ - ΤΕΙ	**Β' Βάθμια Εκπαίδευση**	3,4289(*)	,97086	,004	,8110	6,0468
		ΙΕΚ	-,6029	1,42537	1,000	-4,4464	3,2407
		Μεταπτυχιακό	-1,7933	,91846	,323	-4,2700	,6833
	Μεταπτυχιακό	**Β' Βάθμια Εκπαίδευση**	5,2222(*)	1,13452	,000	2,1630	8,2814
		ΙΕΚ	1,1905	1,54150	1,000	-2,9662	5,3471
		ΑΕΙ - ΤΕΙ	1,7933	,91846	,323	-,6833	4,2700
RISK	**Β' Βάθμια Εκπαίδευση**	**ΙΕΚ**	-,0702	,11905	1,000	-,3912	,2508
		ΑΕΙ - ΤΕΙ	-,1135	,07347	,755	-,3116	,0846
		Μεταπτυχιακό	-,2607(*)	,08585	,019	-,4922	-,0292
	ΙΕΚ	**Β' Βάθμια Εκπαίδευση**	,0702	,11905	1,000	-,2508	,3912
		ΑΕΙ - ΤΕΙ	-,0433	,10786	1,000	-,3341	,2475
		Μεταπτυχιακό	-,1905	,11665	,635	-,5050	,1241
	ΑΕΙ - ΤΕΙ	**Β' Βάθμια Εκπαίδευση**	,1135	,07347	,755	-,0846	,3116
		ΙΕΚ	,0433	,10786	1,000	-,2475	,3341
		Μεταπτυχιακό	-,1472	,06950	,221	-,3346	,0402
	Μεταπτυχιακό	**Β' Βάθμια Εκπαίδευση**	,2607(*)	,08585	,019	,0292	,4922
		ΙΕΚ	,1905	,11665	,635	-,1241	,5050
		ΑΕΙ - ΤΕΙ	,1472	,06950	,221	-,0402	,3346
* The mean difference is significant at the .05 level.							

Από τον παραπάνω πίνακα παρατηρούμε ότι για την θηλυκότητα παρατηρούμε όσους έχουν εκπαίδευση ΙΕΚ να εμφανίζουν υψηλότερα επίπεδα θηλυκότητα σε σχέση με όσους εκπαίδευση ΑΕΙ ή μεταπτυχιακού διότι 0,015<0,05 και 0,001<0,05 αντίστοιχα. Όσον αφορά τον μακροπρόθεσμο – βραχυπρόθεσμο προσανατολισμό παρατηρούμε όσους έχουν δευτεροβάθμια εκπαίδευση να εμφανίζουν υψηλότερα επίπεδα μακροπρόθεσμου προσανατολισμού σε σχέση με όσους έχουν εκπαίδευση ΑΕΙ ή μεταπτυχιακού διότι 0,004<0,05 και 0,000<0,05 αντίστοιχα. Ακόμα παρατηρούμε όσους έχουν δευτεροβάθμια εκπαίδευση να εμφανίζουν υψηλότερα επίπεδα επιχειρηματικού ρίσκου σε σχέση με όσους εκπαίδευση μεταπτυχιακού διότι 0,0019<0,05.

Συμπεράσματα

Από την παραπάνω ανάλυση έχουμε τα παρακάτω αποτελέσματα:

Βρήκαμε ότι το επίπεδο της ισότητας θεωρείται καλό με βάσει την τιμή του δείκτη, χαμηλό είναι επίσης το επίπεδο της ατομικότητας που σημαίνει ότι υπάρχει σε αναπτυγμένο επίπεδο το αίσθημα της συλλογικότητας, ο δείκτης θηλυκότητας – ανδρισμού είναι χαμηλός που σημαίνει ότι έχουμε περισσότερο θηλυκούς ρόλους μεταξύ των φύλων, το επίπεδο της αβεβαιότητας είναι και αυτό χαμηλό που σημαίνει ότι η κοινωνία αποδέχεται την αβεβαιότητα, μπορεί να την διαχειριστεί, το επίπεδο του μακροπρόθεσμου προσανατολισμού είναι χαμηλό που σημαίνει ότι η κοινωνία έχει βραχυπρόθεσμους στόχους κυρίως και τέλος το επίπεδο του επιχειρηματικού ρίσκου παρουσιάζεται υψηλό, άνω του μέσου όρου.

Ακόμα βρήκαμε τις γυναίκες να θεωρούν ότι υπάρχει υψηλότερο επίπεδο ισότητας, θηλυκότητας έναντι των ανδρών ενώ αναλαμβάνουν και μικρότερου επιπέδου επιχειρηματικό ρίσκο. Τα άτομα με οικονομική εκπαίδευση έχουν μακροπρόθεσμο προσανατολισμό και αναλαμβάνουν μεγαλύτερο επιχειρηματικό ρίσκο σε σχέση με τα άτομα που δεν έχουν οικονομική εκπαίδευση. Επίσης τα άτομα με οικογενειακό εισόδημα από 50.000 έως 100.000 ευρώ αναλαμβάνουν υψηλότερο επιχειρηματικό ρίσκο έναντι των ατόμων με οικογενειακό εισόδημα κάτω των 30.000 ευρώ. Τα άτομα με ηλικία 30 ετών σε σχέση με άτομα που έχουν ηλικία 50 ετών δεν παρουσιάζουν διαφορές επιχειρηματικό ρίσκο που αναλαμβάνουν. Επίσης τα άτομα που έχουν σχέση με την χρηματαγορά αναλαμβάνουν υψηλότερο επιχειρηματικό ρίσκο έναντι των ατόμων που δεν έχουν σχέση με την χρηματαγορά.

Όσον αφορά την εκπαίδευση βρήκαμε ότι διαφέρουν τα επίπεδα της εκπαίδευσης για το θέμα της θηλυκότητας, του μακροπρόθεσμου προσανατολισμού και του επιχειρηματικού ρίσκου. Για την ανισότητα, την αβεβαιότητα και την ατομικότητα το επίπεδο της εκπαίδευσης δεν παίζει κανέναν ρόλο. Βρήκαμε ακόμα ότι για την θηλυκότητα όσους έχουν εκπαίδευση ΙΕΚ να εμφανίζουν υψηλότερα επίπεδα θηλυκότητας σε σχέση με όσους εκπαίδευση ΑΕΙ ή μεταπτυχιακού. Ακόμα για τον μακροπρόθεσμο – βραχυπρόθεσμο προσανατολισμό παρατηρούμε όσους έχουν δευτεροβάθμια εκπαίδευση να εμφανίζουν υψηλότερα επίπεδα μακροπρόθεσμου προσανατολισμού σε σχέση με όσους έχουν εκπαίδευση ΑΕΙ ή μεταπτυχιακού. Τέλος όσοι έχουν δευτεροβάθμια εκπαίδευση εμφανίζουν υψηλότερα επίπεδα επιχειρηματικού ρίσκου σε σχέση με όσους εκπαίδευση μεταπτυχιακού.

Εφαρμογή 24

Τα στελέχη του τηλεοπτικού σταθμού Ωμέγα θεωρούν ότι τα Δελτία Ειδήσεων ενός τηλεοπτικού σταθμού είναι ανταγωνιστικά εάν ο μέσος όρος τηλεθέασης σε μηνιαία βάση είναι 15%.
Στον έλεγχο που έγινε εκ μέρους των εκπρόσωπων του διαφημιστικών εταιρειών, για να καταλήξουν στο ποσοστό της διαφημιστικής δαπάνης που θα δώσουν στον τηλεοπτικό σταθμό, έθεσαν την εξής υπόθεση:

$$\left.\begin{array}{l} \boldsymbol{H_0: \mu = \mu_0 \ (\mu_0 = 15)} \\ \boldsymbol{H_1: \mu \neq \mu_0} \end{array}\right\} \qquad \text{ΙΣΧΥΕΙ ΑΥΤΗ Η ΥΠΟΘΕΣΗ?}$$

Πρέπει να εξεταστεί η περίπτωση του ελέγχου της υπόθεσης H_0 ότι ο μέσος του πληθυσμού έχει την τιμή μ_0. Δηλαδή θα ελεγχτεί η υπόθεση H_0: $\mu = \mu_0$ έναντι της εναλλακτικής H_1: μ μ_0. Ο έλεγχος της υπόθεσης διενεργείται με βάση την εκτίμηση του μέσου $\overline{X}$ από τα δεδομένα ενός τυχαίου δείγματος μεγέθους ***n***. Εάν η τυπική απόκλιση σ του πληθυσμού είναι γνωστή, τότε το κριτήριο ελέγχου είναι το:

$$Z = \frac{\overline{X - m}}{s / \sqrt{n}} =$$

Το οποίο σύμφωνα με το κεντρικό θεώρημα, ακολουθεί την κανονική κατανομή για μέγεθος δείγματος n>30 (Χαλικιάς, 2003).
Στην περίπτωση των τηλεθεάσεων των Δελτίων Ειδήσεων του σταθμού για την περίοδο 01/01/2009 – 31/03/2009 έχουμε:

- *Μέσος Όρος Τηλεθέασης Δελτίων Ειδήσεων* ***(Average) =*** $\overline{\mathrm{X}}$
- $\Sigma\,(X - \overline{\mathrm{X}})^2 = 588{,}08$
- $a = 0{,}05$
- $\sigma^2 = 3{,}21$
- $\sigma = \sqrt{3{,}23} = 1{,}79$
- $Z = +0{,}013$

Ανατρέχοντας στον πίνακα με Ζ στρογγυλοποιημένο +0,0 και με
α = 0,05 *η τιμή του Ζ είναι 0,51994 .*
Εφόσον Ζ = +0,51994 βλέπουμε ότι -1,96 < Ζ < +1,96, που σημαίνει ότι η τιμή του κριτηρίου Ζ βρίσκεται στην περιοχή αποδοχής της υπόθεσης μηδέν. *Επομένως το συμπέρασμα είναι ότι ο μέσος του δείγματος* $\overline{\mathrm{X}}$ *δεν διαφέρει στατιστικά σημαντικά από το μέσο* μ_0 *της υπόθεσης μηδέν και η υπόθεση μας γίνεται δεκτή.*
Αποδεικνύεται λοιπόν, ότι τα δελτία ειδήσεων του τηλεοπτικού σταθμού θεωρούνται ανταγωνιστικά εφόσον ο μέσος όρος είναι ίσος με 15.

Εφαρμογή 25

Κατά την ολοκλήρωση της συμφωνίας σχετικά με την διαφημιστική δαπάνη προς το κανάλι, από πλευράς των στελεχών του media shop τέθηκε ακόμη μία προϋπόθεση. Στα πλαίσια του λανσαρίσματος μιας νέας διαφημιστικής καμπάνιας ενός μεγάλου πολυεθνικού πελάτη, ζητήθηκε από τα στελέχη τουνα ερευνήσουν και να τους ενημερώσουν για το ερώτημα:

Κατά το Α΄ τρίμηνο του 2009, ποιο από τα δύο δελτία -μεσημβρινό ή κεντρικό- θεωρείται ανταγωνιστικότερο (με μέσο όρο μ>15);

Οπότε τίθεται η εξής υπόθεση:

$$\left.\begin{aligned} &\boldsymbol{H_0: \mu = \mu_0\ (\mu_0 = 15)} \\ &\boldsymbol{H_1: \mu > \mu_0} \end{aligned}\right\}$$

ΙΣΧΥΕΙ ΑΥΤΗ Η ΥΠΟΘΕΣΗ?

Για την εξαγωγή συμπεράσματος της υποθέσεως σχετικά με την σύγκριση του Μεσημβρινού και του Κεντρικού Δελτίου Ειδήσεων ακολουθείται κοινή στατιστική μεθοδολογία με εκείνη της προηγούμενης παραγράφου.

Οπότε για το μεσημβρινό δελτίο έχουμε:

- *Μέσος Όρος τηλεθέασης μεσημβρινού* = $\overline{X}$
- $\Sigma (X - \overline{X})^2 = 270{,}59$
- $a = 0{,}05$
- $\sigma^2 = 2{,}97$
- $\sigma = \sqrt{2{,}97} = 1{,}72$
- $Z = +1{,}33$

Ανατρέχοντας στον πίνακα με Ζ στρογγυλοποιημένο +1,3 (στρογγυλοποίηση στην πλησιέστερη τιμή προς τα κάτω) και με α = 0,05 *η τιμή του Ζ είναι* ***0,91149*** (βλ. Παράρτημα Εμβαδά Τυποποιημένης Κανονικής Κατανομής).

Εφόσον Ζ = +0,91149 συμπεραίνεται ότι Ζ < +1,645, που σημαίνει ότι η τιμή του κριτηρίου Ζ βρίσκεται στην περιοχή αποδοχής της υπόθεσης μηδέν. *Επομένως το συμπέρασμα είναι ότι ο μέσος του δείγματος* $\overline{X}$ *δεν διαφέρει στατιστικά σημαντικά από το μέσο* μ_0 *της υπόθεσης μηδέν* (Χαλικιάς 2003) *και η υπόθεση γίνεται δεκτή.*

Και για το Κεντρικό δελτίο έχουμε:

- *Μέσος Όρος τηλεθέασης κεντρικού* = $\overline{X}$
- $\Sigma (X - \overline{X})^2 = 286{,}49$
- $a = 0{,}05$
- $\sigma^2 = 3{,}14$
- $\sigma = \sqrt{3{,}14} = 1{,}77$
- $Z = +2{,}77$

Ανατρέχοντας στον πίνακα με Ζ στρογγυλοποιημένο +2,8 (στρογγυλοποίηση στην πλησιέστερη τιμή προς τα επάνω) και με

α = 0,05 *η τιμή του Ζ είναι* ***0,99781*** (βλ. Παράρτημα Εμβαδά Τυποποιημένης Κανονικής Κατανομής).

Εφόσον Ζ = +0,99781 βλέπουμε ότι Ζ < +1,645, που σημαίνει ότι η τιμή του κριτηρίου Ζ βρίσκεται στην περιοχή αποδοχής της υπόθεσης μηδέν. *Επομένως το συμπέρασμα είναι οτι ο* *μέσος του δείγματος* $\overline{X}$ *δεν διαφέρει στατιστικά σημαντικά από το μέσο* μ_0 *της υπόθεσης μηδέν και η υπόθεση γίνεται δεκτή.*

Επαληθεύουμε την σύγκριση των δύο μέσων όρων μέσω του στατιστικού προγράμματος SPSS

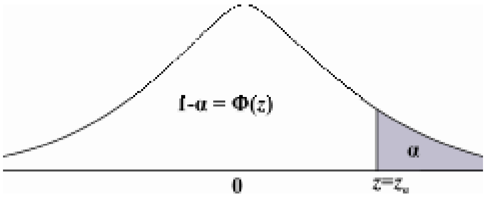
	Τηλεθέαση Μεσημβρινού	*Τηλεθέαση Κεντρικού*
Mean	…………..	…………..
Variance	3,00651392	3,1725885
Observations	91	91
Hypothesized Mean Difference	0	
df	180	
t Stat	-3,2164022	
P(T<=t) two-tail	0,00153919	
t Critical two-tail	1,97323078	

Στην υπόθεση αυτή, συμπεραίνεται ότι ο μέσος όρος του τριμήνου του μεσημβρινού δελτίου ισούται με τον μέσο όρο του κεντρικού
[πηγή 2]

5.Παράρτημα-Στατιστικοί Πίνακες

ΠΙΝΑΚΑΣ Α-ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Τιμές των πιθανοτήτων $\Phi(z) = P(Z \le z) = P(Z < z)$ της τυποποιημένης κανονικής κατανομής $N(0,1)$ για $z \ge 0$. Για $z < 0$ ισχύει $\Phi(z) = 1 - \Phi(-z)$.

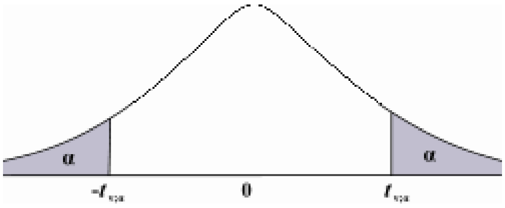


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
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	0.84850	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	0.92786	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
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	0.99897	0.99900

α	0.0005	0.001	0.005	0.01	0.025	0.05	0.10
z_α	3.29	3.09	2.576	2.326	1.960	1.645	1.282

ΠΙΝΑΚΑΣ Β: t-ΚΑΤΑΝΟΜΗ

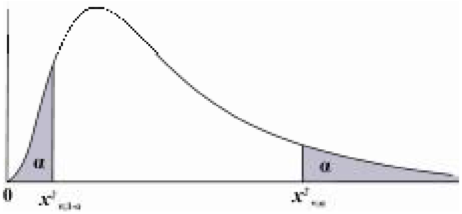
Τιμών $t_{\nu;a}$ της t_ν-κατανομής ώστε $P(T_\nu > t_{\nu;a}) = P(T_\nu \geq t_{\nu;a}) = a$.



ν	α = 0.10	α = 0.05	α = 0.025	α = 0.01	α = 0.005
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
¥	1.282	1.645	1.960	2.326	2.576

ΠΙΝΑΚΑΣ Γ:Χ^2-ΚΑΤΑΝΟΜΗ

Των τιμών $\chi^2_{v,1-a}$ της χ^2 κατανομής για τις οποίες $P(X < \chi^2_{v,1-a}) = P(X \leq \chi^2_{v,1-a}) = a$.



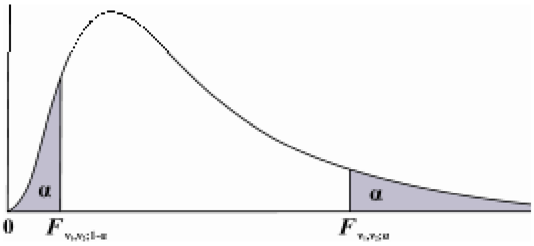
v	α = 0.005	α = 0.01	α = 0.025	α = 0.05	α = 0.10
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

ΠΙΝΑΚΑΣ Δ1: ΚΑΤΑΝΟΜΗ F

Τιμές $F_{v_1,v_2,a}$ της F κατανομής για τις οποίες $P(X > F_{v_1,v_2,a}) = P(X \geq F_{v_1,v_2,a}) = a$

$(a = 0.01)$.

Για τα α - κάτω ποσοστιαία σημεία $F_{v_1,v_2,1-a}$ ισχύει η σχέση $F_{v_1,v_2,1-a} = 1 / F_{v_2,v_1,a}$.



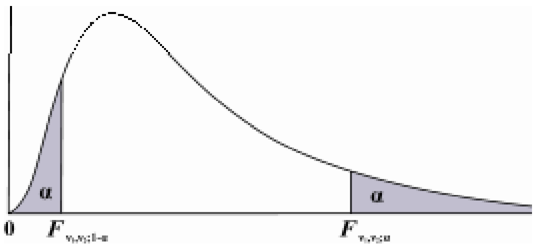
v_2 \ v_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
¥	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

ΠΙΝΑΚΑΣ Δ2: ΚΑΤΑΝΟΜΗ F (συνέχεια)

Τιμές $F_{\nu_1,\nu_2,a}$ της F κατανομής για τις οποίες $P(X > F_{\nu_1,\nu_2,a}) = P(X \geq F_{\nu_1,\nu_2,a}) = a$

$(a = 0.01)$.

Για τα α - κάτω ποσοστιαία σημεία $F_{\nu_1,\nu_2,1-a}$ ισχύει η σχέση $F_{\nu_1,\nu_2,1-a} = 1 \big/ F_{\nu_2,\nu_1,a}$.



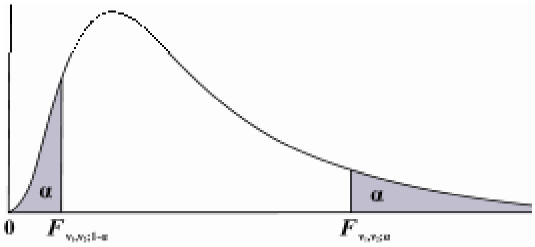
ν_1 / ν_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**¥**
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
¥	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ΠΙΝΑΚΑΣ Ε1: ΚΑΤΑΝΟΜΗ F (συνέχεια)

Τιμές $F_{v_1,v_2,a}$ της F κατανομής για τις οποίες $P(X > F_{v_1,v_2,a}) = P(X \geq F_{v_1,v_2,a}) = a$

(α=0.05).

Για τα a - κάτω ποσοστιαία σημεία $F_{v_1,v_2,1-a}$ ισχύει η σχέση $F_{v_1,v_2,1-a} = 1 \big/ F_{v_2,v_1,a}$.



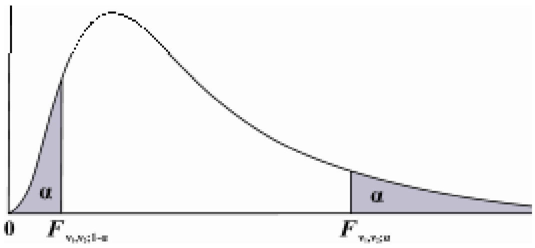
v_1 / v_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
¥	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

ΠΙΝΑΚΑΣ Ε2: ΚΑΤΑΝΟΜΗ F (συνέχεια)

Τιμές $F_{v_1,v_2,a}$ της F κατανομής για τις οποίες $P(X > F_{v_1,v_2,a}) = P(X \geq F_{v_1,v_2,a}) = a$

(α=0.05).

Για τα a - κάτω ποσοστιαία σημεία $F_{v_1,v_2,1-a}$ ισχύει η σχέση $F_{v_1,v_2,1-a} = 1 / F_{v_2,v_1,a}$.



v_2 \ v_1	10	12	15	20	24	30	40	60	120	¥
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
¥	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

6.Βιβλιογραφία

<u>Έντυπη</u>

1) Κουνιάς Σ, 'Εισαγωγή στην Στατιστική', Εκδόσεις Χριστοδουλίδη, Θεσσαλονίκη, 2000
2) Ι. Χαλικιάς, "Στατιστική – Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις", Εκδόσεις Rosili, Αθήνα 2003, σελ. 175
3) Αναστασιάδη Ι. Στοιχεία Στατιστικής Θεσσαλονίκη (1978)
4) Bhattscharyya C. Jonson R. Statistical Concepts and Methods Wiley (1977)
5) Φραγκάκη Χαρ. 'Στατιστική' Θεσσαλονίκη, 1985
6) Λαμπράκη Δ, 'Στατιστική', Αθήνα 1980
7) Αποστολοπούλου Θ, 'Στατιστική Επιχειρήσεως, Αθήνα, 1988
8) Μπένου Β, 'Εφαρμογές επαγωγικής Στατιστικής' Πειραιάς, 1984

<u>Ηλεκτρονική</u>

9) ftp://ftp.soc.uoc.gr/students/stao131/askhseis/askhseis.stao131.doc
10) *stat-athens.aueb.gr/~moustaki/stats1/set2.pdf*
11) ftp://ftp.soc.uoc.gr/students/stao131/.../luseis.askhsewn.stao131.doc
12) http://androulakis.bma.upatras.gr/mediawiki/index.php
13) www.semfe.gr/files/.../analysh_dedomenvn-b_ergasia_ekfvnhseis.pdf
14) ftp://ftp.soc.uoc.gr/students/stao131/.../luseis.askhsewn.stao131.doc
15) ftp://ftp.soc.uoc.gr/students/stao131/.../luseis.askhsewn.stao131.doc