ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

ΕΠΙΜΕΛΕΙΑ: ΝΙΚΟΛΕΤΟΣ ΛΑΜΠΡΟΣ
ΑΝΑΣΤΑΣΟΠΟΥΛΟΣ ΙΩΑΝΝΗΣ
ΣΟΦΟΓΙΑΝΝΗΣ ΒΑΣΙΛΕΙΟΣ

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΑΣ
ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

ΕΠΙΜΕΛΕΙΑ: ΝΙΚΟΛΕΤΟΣ ΛΑΜΠΡΟΣ
ΑΝΑΣΤΑΣΟΠΟΥΛΟΣ ΙΩΑΝΝΗΣ
ΣΟΦΟΓΙΑΝΝΗΣ ΒΑΣΙΛΕΙΟΣ

ΕΙΣΗΓΗΤΡΙΑ: ΜΠΟΥΜΠΟΥΛΗ ΑΘΑΝΑΣΙΑ

ΠΑΤΡΑ, **ΙΟΥΛΙΟΣ** 2009

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΤΙ ΕΙΝΑΙ ΣΤΑΤΙΣΤΙΚΗ

Στην καθομιλουμένη, στατιστική σημαίνει συστηματική απαρίθμηση και παρουσίαση αριθμητικών δεδομένων ή στοιχείων, τα οποία προέρχονται από πολλές παρατηρήσεις ή μετρήσεις. Οι παρατηρήσεις αυτές ή οι μετρήσεις αναφέρονται σε συγκεκριμένο αντικείμενο ή γεγονός. Ανάλογα με το αντικείμενο ή το γεγονός στο οποίο αναφέρονται τα αριθμητικά δεδομένα, η στατιστική παίρνει και ιδιαίτερη ονομασία. Έτσι όταν μιλάμε π.χ. για Γεωργική στατιστική, Στατιστική επιχειρήσεων ή Στατιστική εργατικού δυναμικού κ.τ.λ., εννοούμε αριθμητικά δεδομένα που αναφέρονται αντίστοιχα στη γεωργία, στις επιχειρήσεις ή στο εργατικό δυναμικό κ.τ.λ. Στην επιστημονική γλώσσα, η λέξη Στατιστική έχει ευρύτερη σημασία σημαίνει την επιστήμη που έχει ως αντικείμενο όχι μόνο τη συγκέντρωση και παρουσίαση, αλλά και τη μελέτη και ανάλυση των παρατηρήσεων ή μετρήσεων που αναφέρονται σε ένα συγκεκριμένο αντικείμενο ή γεγονός, οποιαδήποτε και αν είναι η φύση του. Έτσι, η στατιστική περιλαμβάνει τόσο τις μεθόδους συλλογής και επεξεργασίας στοιχείων, όσο και τις μεθόδους ανάλυσης και μελέτης τους, ανακαλύπτοντας έτσι σχέσεις που υπάρχουν ανάμεσα στα διάφορα φαινόμενα και διατυπώνοντας συμπεράσματα που είναι χρήσιμα για τη λήψη ορθών αποφάσεων. Μπορούμε λοιπόν να πούμε ότι:

Στατιστική είναι η επιστήμη που ασχολείται με τις επιστημονικές μεθόδους συλλογής, οργάνωσης, παρουσίασης και ανάλυσης των αριθμητικών εκείνων στοιχείων που αναφέρονται σε χαρακτηριστικές ιδιότητες διαφόρων οικονομικών, κοινωνικών, δημογραφικών, φυσικών φαινομένων κτλ. και έχει ως σκοπό τη συστηματική μελέτη αυτών των στοιχείων για την κατάληξη σε γενικά συμπεράσματα, που είναι χρήσιμα στη διαδικασία της λήψης ορθών αποφάσεων.

Αναλύοντας τον ορισμό αυτό της στατιστικής, παρατηρούμε ότι τα βασικά στάδια που ακολουθούμε για τη μελέτη των ιδιοτήτων των διαφόρων μονάδων μιας πολυπληθούς ομάδας είναι τα εξής:

α) Η συγκέντρωση των στατιστικών στοιχείων που είναι αναγκαία για τη μελέτη του προβλήματος που θέλουμε να ερευνήσουμε

β) Η μεθοδική επεξεργασία και παρουσίαση των στατιστικών στοιχείων σε μορφή αριθμητικών πινάκων και γραφικών παραστάσεων

γ) Η ανάλυση των στοιχείων αυτών και η εξαγωγή χρήσιμων συμπερασμάτων για να ληφθούν σωστές αποφάσεις

- Ιστορία της Στατιστικής

Η λέξη στατιστική προέρχεται από τη λατινική λέξη status(που σημαίνει κράτος) και δηλώνει αρχικά συλλογή στοιχείων για τις κρατικές ανάγκες(έκταση, παραγωγή, πληθυσμό). Έχει εξακριβωθεί ότι η πρώτη απογραφή πληθυσμού έγινε στην Κίνα από τον αυτοκράτορα Υάο το έτος 2238 π.Χ., ενώ στους Ρωμαίους η πρώτη απογραφή πληθυσμού έγινε επί Ρωμύλου (753-715 π.Χ.) και η τελευταία από τον αυτοκράτορα Βεσπασιανό το 73 μ.Χ. Στην Αγγλία η πρώτη καθολική απογραφή του πληθυσμού και του πλούτου γενικά έγινε το 1085 από τον Γουλιέλμο τον κατακτητή. Το 1853 γράφεται από τον Fr. Sansovino το πρώτο βιβλίο στατιστικού περιεχομένου και λίγο αργότερα εισάγεται από τον Konring (1606-1681) η στατιστική στην ανώτερη παιδεία.

Την ίδια εποχή εμφανίζεται το ενδιαφέρον για τις ασφάλειες ζωής και ο περίφημος Άγγλος αστρονόμος Halley , χρησιμοποιώντας τα ληξιαρχικά βιβλία γεννήσεων και θανάτων της πόλεως Breslaou, παρουσιάζει τον πρώτο πίνακα θνησιμότητας. Το ρεύμα αυτό των δημοσιογραφικών μελετών επεκτείνεται και στη Γερμανία, όπου ο πάστορας Sissmilch (1707-1767) συγκεντρώνει στοιχεία από τα ληξιαρχικά βιβλία των εφημερίων της Πρωσίας και καταλήγει το 1741 στο συμπέρασμα ότι το ποσοστό γέννησης των αγοριών είναι 51% και των κοριτσιών 49%, ενώ τα δύο φύλα έχουν ίσα ποσοστά κατά την εποχή του γάμου. Για το συγγραφέα το φαινόμενο αυτό δεν είναι τυχαίο γεγονός, αλλά νόμος θείας προέλευσης που αποσκοπεί στη διαιώνιση του είδους. Μέχρι την εποχή αυτή, η στατιστική έχει περιγραφικό χαρακτήρα και ασχολείται κυρίως με θέματα Δημογραφίας.

Η στατιστική θα ξεφύγει από τον περιγραφικό χαρακτήρα της με την ανάπτυξη ενός νέου κλάδου, του λογισμού των πιθανοτήτων, ο οποίος προήλθε από τη μελέτη των τυχερών παιγνιδιών (χαρακτηριστική μάλιστα είναι η αλληλογραφία ανάμεσα στους Γάλλους μαθηματικούς Pascal και ο ιππότης De Mere για τα παιγνίδια του κύβου). Από τους θεμελιωτές του Λογισμού των Πιθανοτήτων αναφέρουμε τον Bernoulli, ο οποίος στο βιβλίο του «Η τέχνη των προβλέψεων» διατυπώνει τον περίφημο νόμο των μεγάλων αριθμών και το Γάλλο μαθηματικό Laplace, στον οποίο οφείλεται η εφαρμογή του Λογισμού των Πιθανοτήτων στη σπουδή των φυσικών φαινομένων με πολυσύνθετες αιτίες. Στη νέα αυτή περίοδο της στατιστικής, ο Βέλγος αστρονόμος Quetelet επεκτείνει την εφαρμογή της στατιστικής στη σπουδή των φυσικών, διανοητικών και ηθικών ιδιοτήτων του ανθρώπου και παίρνει την πρωτοβουλία για τη σύγκληση του πρώτου Διεθνούς Συνεδρίου Στατιστικής που έγινε στις Βρυξέλλες το 1853, ενώ αργότερα ο F. Galton εφαρμόζει τη στατιστική στη Βιολογία και ειδικότερα στα προβλήματα της κληρονομικότητας. Η προσπάθεια του Galton συνεχίστηκε από τον Άγγλο μαθηματικό Pearson, στον οποίο οφείλεται κατά πολύ η σημερινή ανάπτυξη και θέση της Στατιστικής.

- Χρησιμότητα και πεδία εφαρμογής της Στατιστικής

Μια απλή αρίθμηση των εφαρμογών της που είναι βασικά εφαρμοσμένη επιστήμη, δείχνει ότι αυτή χρησιμοποιείται σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας. Η Στατιστική είναι απαραίτητη στη Διοίκηση γενικά, όπου η λήψη των ορθών αποφάσεων έχει μεγάλη σημασία για την πρόοδο ενός κράτους, ενός οργανισμού, μιας βιομηχανίας ή μιας επιχείρησης. Γι' αυτό και δεν υπάρχει σήμερα στις σύγχρονες επιχειρήσεις κανένας τομέας που να μην χρησιμοποιεί τις στατιστικές μεθόδους στη λήψη επιχειρηματικών αποφάσεων.

Μεγάλη σημασία έχει η εφαρμογή της στατιστικής στη Δημογραφία, όπου η μελέτη της γαμηλιότητας, της γεννητικότητας, της θνησιμότητας, της μετανάστευσης κ.λ.π, απαιτεί μακροχρόνιες στατιστικές παρατηρήσεις και επίπονες αναλύσεις. Επίσης η στατιστική εφαρμόζεται σήμερα στην Ιατρική, Φυσική, Γενετική, Αστρονομία, Βιολογία, Μετεωρολογία, Γεωργία, Βιομηχανία, στη μελέτη του φυσικού περιβάλλοντος, στη μελέτη των ανθρωπίνων ιδεών και προθέσεων, στη θεωρία των αποφάσεων, στον έλεγχο ποιότητας των προϊόντων κ.λ.π. Τέλος, η στατιστική βρίσκει πολύ μεγάλη εφαρμογή και στον οικονομικό τομέα, όπου η παρακολούθηση του γενικού επιπέδου των τιμών, του εθνικού εισοδήματος, της νομισματικής ισοτιμίας και των οικονομικών διακυμάνσεων, της απασχόλησης, της παραγωγικότητας, της κατάρτισης δεικτών οικονομικής δραστηριότητας των εθνικών πόρων και της εθνικής δαπάνης, είναι αντικείμενα στατιστικής επεξεργασίας. Η χρησιμότητα της στατιστικής φαίνεται και από το γεγονός ότι η Στατιστική διδάσκεται σήμερα σχεδόν σε όλες τις Ανώτατες και Ανώτερες Σχολές της χώρας μας.

- Το κράτος και η Στατιστική

Ένα καλά οργανωμένο κράτος οφείλει να γνωρίζει κάθε στιγμή τον πληθυσμό της χώρας, την κατανομή του πληθυσμού κατά φύλο, ηλικία, επάγγελμα, κ.λ.π, καθώς και την κίνηση και πιθανή εξέλιξή του. Πρέπει επίσης να παρακολουθεί τόσο τα οικονομικά φαινόμενα της χώρας (παραγωγή, εισαγωγές και εξαγωγές, κίνηση και εμπορία των αγαθών κ.λ.π) όσο και τα διοικητικά και κοινωνικά φαινόμενα της χώρας (διοίκηση, εργασία, δημόσια υγεία, πρόνοια, κοινωνικές ασφαλίσεις, εκπαίδευση, δικαιοσύνη, στέγαση, κατάρτιση τιμάριθμου κόστους ζωής κ.λ.π).

Για το σκοπό αυτό, κάθε κράτος έχει μια στατιστική υπηρεσία, η οποία συγκεντρώνει τα απαραίτητα στοιχεία και παρακολουθεί την εξέλιξη των παραπάνω φαινομένων. Μια τέτοια υπηρεσία πρέπει να είναι καλά οργανωμένη και να διαθέτει ένα πλούσιο κεντρικό αρχείο στατιστικών στοιχείων, από το οποίο θα αντλεί χρήσιμες πληροφορίες κάθε διοικητικός παράγοντας του κράτους και κάθε ερευνητής.

- Ορισμός και βασικές έννοιες της στατιστικής

Η λέξη στατιστική έχει τρεις διαφορετικές έννοιες:

1) Χρησιμοποιείται για να υποδηλώσει το αντικείμενο της στατιστικής επιστήμης δηλ. μεθόδους και τεχνικές που ασχολούνται με τη συλλογή, ταξινόμηση, παρουσίαση, ανάλυση και ερμηνεία αριθμητικών δεδομένων τα οποία είναι χρήσιμα για τη λήψη ορθών αποφάσεων.

2) Χρησιμοποιείται για να υποδηλώσει σειρά αριθμητικών δεδομένων: πληθυσμό, παραγωγή, εισόδημα, γεννήσεις, γάμοι, θάνατοι, κ.λ.π. Λέγοντας π.χ.

«δημογραφικές στατιστικές», εννοούμε τα αριθμητικά δεδομένα που αναφέρονται στον πληθυσμό, στις γεννήσεις, στους θανάτους, «γεωργικές στατιστικές», τα στατιστικά δεδομένα παραγωγής γεωργικών και κτηνοτροφικών προϊόντων, καλλιεργούμενων εκτάσεων, κ.λ.π.

3) Η λέξη «στατιστική» (statistic), τέλος είναι συνώνυμη με τη λέξη ισόμετρος δηλ. μια σταθερή ποσότητα, η οποία συνοψίζει μια ορισμένη ιδιότητα ενός συνόλου αριθμητικών δεδομένων. Π.χ. ο μέσος μισθός 1000 υπαλλήλων είναι μια στατιστική και εκφράζει ποιος ήταν ο μισθός κάθε υπαλλήλου αν όλοι αμείβονταν με τον ίδιο μισθό. Η στατιστική, με την πρώτη έννοια μπορεί να ορισθεί ως εξής:

Στατιστική είναι η επιστήμη, η οποία ασχολείται με τη συλλογή, επεξεργασία και παρουσίαση, ανάλυση και ερμηνεία αριθμητικών δεδομένων τα οποία είναι χρήσιμα για τη λήψη ορθών αποφάσεων.

Από τον παραπάνω ορισμό προκύπτει ότι κάθε στατιστική έρευνα ακολουθεί τέσσερα βασικά στάδια:

1. Συλλογή. Η συλλογή του στατιστικού υλικού γίνεται με διάφορες πρακτικές και μεθόδους που αναπτύσσονται πιο κάτω. Λέγοντας «συλλογή στατιστικού υλικού» εννοούμε τη συγκέντρωση αριθμητικών δεδομένων τα οποία θέλουμε να διερευνήσουμε.

2. Επεξεργασία και παρουσίαση. Τα αριθμητικά δεδομένα που θα συγκεντρώσουμε πρέπει να τα ταξινομήσουμε κατά συστηματικό τρόπο, ώστε να είναι εύκολη η διερεύνησή τους. Η παρουσίαση του στατιστικού υλικού ισούται με ταξινόμηση των αριθμητικών δεδομένων σε πίνακες και απεικόνιση αυτών με στατιστικά διαγράμματα.

3. Ανάλυση. Λέγοντας ανάλυση του στατιστικού υλικού εννοούμε το λογισμό των διαφόρων στατιστικών παραμέτρων που είναι αναγκαίες για τη μελέτη της συμπεριφοράς των κοινωνικών, δημογραφικών, οικονομικών φαινομένων που διερευνούμε. Η ανάλυση του στατιστικού υλικού γίνεται με διάφορους τρόπους της στατιστικής αναλύσεως.

4. Ερμηνεία. Τα στατιστικά μέτρα που θα προκύψουν από τη στατιστική ανάλυση θα ερμηνευθούν ανάλογα με το σκοπό της διεξαγόμενης έρευνας. Τα στατιστικά μέτρα αποτελούν χρησιμότατο στατιστικό εργαλείο για τη λήψη ορθών αποφάσεων, για τον προγραμματισμό και για την άσκηση οικονομικοκοινωνικής πολιτικής από την ηγεσία του Δημοσίου και Ιδιωτικού τομέα της Διοικήσεως.

· Στατιστικός πληθυσμός

Για να κατανοηθεί η έννοια του στατιστικού πληθυσμού πρέπει να έχουμε υπόψη μας ότι κάθε φαινόμενο έχει ένα πλήθος ιδιοτήτων ή χαρακτηριστικών. Π.χ. ο άνθρωπος έχει ανάστημα, βάρος, φύλο, οικογενειακή κατάσταση, κ.λ.π. Τα αναστήματα των σπουδαστών αποτελούν έναν πληθυσμό αναστημάτων. Οι βαθμοί όλων των σπουδαστών αποτελούν έναν πληθυσμό βαθμών. Η παραγωγή ενός γεωργικού ή βιομηχανικού προϊόντος, για μια μεγάλη χρονική περίοδο, είναι ένας στατιστικός πληθυσμός. Επομένως, πληθυσμός είναι το σύνολο των μετρήσεων ή απαριθμήσεων οι οποίες αναφέρονται σε ένα πλήθος έμψυχων όντων ή άψυχων αντικειμένων που έχουν ένα ή περισσότερα κοινά χαρακτηριστικά.

Πρέπει να έχουμε υπόψη μας ότι, αν και ο στατιστικός πληθυσμός περιλαμβάνει φυσικά αντικείμενα, εντούτοις, τα αντικείμενα αυτά δεν αποτελούν και τον αντικειμενικό σκοπό της στατιστικής αλλά μόνο τα χαρακτηριστικά τους, δηλ. μόνον οι ιδιότητες των αντικειμένων αποτελούν το πεδίο εφαρμογής της στατιστικής.

Οι στατιστικοί πληθυσμοί διακρίνονται σε άπειρους και πεπερασμένους. Άπειρος ονομάζεται ένας πληθυσμός όταν το πλήθος των στοιχείων του είναι πολύ μεγάλος αριθμός, ώστε να μπορεί να θεωρηθεί ως άπειρος. Π.χ. οι παραγόμενες μονάδες ενός βιομηχανικού προϊόντος, για μια μεγάλη χρονική περίοδο, σχηματίζουν έναν άπειρο πληθυσμό. Πεπερασμένοι είναι εκείνοι οι πληθυσμοί που το πλήθος των μονάδων τους είναι απολύτως καθορισμένο. Π.χ. ο πληθυσμός της Ελλάδος είναι πεπερασμένος, ενώ ο πληθυσμός της γης, πρακτικά, θεωρείται άπειρος. Οι πεπερασμένοι πληθυσμοί διακρίνονται σε πολυπληθείς, όταν το πλήθος των στοιχείων τους είναι πολύ μεγάλο και σε ολιγοπληθείς, όταν αποτελούνται από μικρό αριθμό στοιχείων. Π.χ. ο αριθμός των σπουδαστών, οι οποίοι εισάγονται κάθε χρόνο στα Ανώτατα Εκπαιδευτικά Ιδρύματα είναι ένας ολιγοπληθής πληθυσμός, ενώ ο αριθμός των υποψηφίων θεωρείται πολυπληθής πληθυσμός.

Για τη διερεύνηση ενός στατιστικού πληθυσμού πρέπει να είναι διαθέσιμα όλα τα στοιχεία του πληθυσμού. Αυτό όμως είναι πολύ δύσκολο και δαπανηρό έργο και σε μερικές περιπτώσεις τελείως αδύνατο, γιατί αν π.χ. μια βιομηχανία που παράγει ηλεκτρικούς λαμπτήρες, θελήσει να υπολογίσει τη μέση διάρκεια ζωής των παραγόμενων λαμπτήρων της πρέπει να «κάψει» όλους τους λαμπτήρες της για να μετρήσει τις ώρες ζωής των λαμπτήρων. Γι’ αυτό το λόγο, τα στοιχεία (μονάδες) του πληθυσμού δεν είναι συνήθως διαθέσιμα. Οι διάφορες στατιστικές έρευνες διεξάγονται, κατά κανόνα, με τη μέθοδο της Δειγματοληψίας (Sampling). Αν π.χ. ένας πληθυσμός αποτελείται από 1.000.000 νοικοκυριά και μια βιομηχανία θελήσει να διαπιστώσει το ποσοστό των νοικοκυριών που αγοράζουν τα προϊόντα της βιομηχανίας, είναι οικονομικώς ασύμφορη η έρευνα γιατί το κόστος της έρευνας θα είναι τεράστιο σε σχέση με το όφελος. Αν λοιπόν η βιομηχανία θελήσει να ρωτήσει 1.000.000 νοικοκυριά, επιλέγει, κατά κάποιον τρόπο ένα αντιπροσωπευτικό Δείγμα (Sample) από το σύνολο των νοικοκυριών και υπολογίζει ένα ποσοστό, το οποίο θα ισχύει για το σύνολο των νοικοκυριών αν το δείγμα είναι αντιπροσωπευτικό του συνόλου των νοικοκυριών. Η τεχνική της Δειγματοληψίας αποτελεί σήμερα ένα εξειδικευμένο κλάδο, τον οποίο χρησιμοποιούν όλοι οι επιστημονικοί κλάδοι για την εξαγωγή συμπερασμάτων σχετικών με τον τομέα των ερευνών τους.

- Στατιστικές Μεταβλητές

Είπαμε πιο πάνω ότι τα στοιχεία ενός στατιστικού πληθυσμού αποτελούνται από μετρήσεις ή απαριθμήσεις χαρακτηριστικών ή ιδιοτήτων διάφορων φαινομένων. Τα χαρακτηριστικά ή οι ιδιότητες των στοιχείων ενός πληθυσμού ονομάζονται μεταβλητές (Variables) και συμβολίζονται με τα γράμματα Ζ,... Οι αριθμητικές εκφράσεις που παίρνουν οι μεταβλητές, ονομάζονται τιμές των μεταβλητών.

Οι μεταβλητές διακρίνονται σε δύο κατηγορίες:

1. Σε εκείνες που επιδέχονται ποσοτική μέτρηση:

Οι τιμές τους εκφράζονται με αριθμούς και λέγονται ποσοτικές μεταβλητές. Τέτοιες μεταβλητές είναι π.χ. το βάρος και το ανάστημα των ατόμων, το εισόδημα, η βιομηχανική παραγωγή, ο αριθμός των ελαττωματικών προϊόντων σε μια μαζική παραγωγή, κ.ά.

2. Σε εκείνες που δεν μπορούμε να τις μετρήσουμε, αλλά τις εκφράζουμε με λέξεις και ονομάζονται ποιοτικές μεταβλητές. Τέτοιες μεταβλητές είναι π.χ. το φύλο, η φυλή, η οικογενειακή κατάσταση (άγαμος, έγγαμος, κ.λ.π).

Οι ποσοτικές μεταβλητές διακρίνονται σε συνεχείς και ασυνεχείς.

Συνεχείς ονομάζονται οι μεταβλητές εκείνες που μπορούν να πάρουν θεωρητικά οποιαδήποτε τιμή μεταξύ δύο πραγματικών αριθμών. Π.χ. το ανάστημα και το βάρος των ανθρώπων, η θερμοκρασία, το εισόδημα, οι εισπράξεις ενός καταστήματος, οι καταθέσεις στα ταμιευτήρια των τραπεζών, τα ημερομίσθια, κ.λ.π είναι συνεχείς μεταβλητές.

Ασυνεχείς λέγονται οι μεταβλητές εκείνες που παίρνουν μόνο ακέραιες τιμές. Π.χ. ο αριθμός των παιδιών σε μια οικογένεια, ο αριθμός των ελαττωματικών μονάδων ενός βιομηχανικού προϊόντος, ο αριθμός των τυπογραφικών λαθών ενός βιβλίου, κ.λ.π είναι ασυνεχείς μεταβλητές.

- Ο ρόλος της στατιστικής στις επιχειρήσεις

Η εφαρμογή της στατιστικής στον επιχειρηματικό τομέα αφορά κυρίως τις βιομηχανικές και εμπορικές επιχειρήσεις. Η εφαρμογή στατιστικών μεθόδων για τη λύση προβλημάτων του επιχειρηματικού τομέα έγινε πρώτα στις Η.Π.Α. στις αρχές του 20^{ου} αιώνα, ως συνέπεια της συγκεντρώσεως των δραστηριοτήτων σε μεγάλες επιχειρήσεις, με πολυάριθμο προσωπικό και υποκαταστήματα σε όλη την έκταση της απέραντης αυτής χώρας.

Είναι φανερό ότι για τον επιχειρηματία, ο οποίος γνωρίζει προσωπικά τους κυριότερους συνεργάτες του και όλο το μηχανισμό της επιχειρήσεώς του, η στατιστική δεν παρέχει πρόσθετη ωφέλεια. Αλλά για το γενικό διευθυντή μιας μεγάλης επιχείρησης η οποία απασχολεί εκατοντάδες ή και χιλιάδες πρόσωπα σε πολυάριθμες εγκαταστάσεις που βρίσκονται συνήθως μακριά από το κέντρο της επιχείρησης, οι στατιστικές πληροφορίες για την ικανότητα των εργοταξίων, των γραφείων, των αποθηκών, για τις αγορές και πωλήσεις, για την παραγωγή και τον έλεγχο καλής ποιότητας των παραγόμενων προϊόντων, κ.λ.π. η στατιστική είναι χρησιμότατο πληροφοριακό εργαλείο για τη λήψη ορθών επιχειρηματικών αποφάσεων και την άσκηση οικονομικής, κοινωνικής, τιμολογιακής κ.λ.π πολιτικής. Μπορούμε να πούμε, ότι η τεράστια ανάπτυξη των βιομηχανικών και εμπορικών επιχειρήσεων στις Η.Π.Α., κατά τα τελευταία χρόνια οφείλεται κυρίως στην καλύτερη οργάνωση της παραγωγής και διανομής των εμπορευμάτων με στατιστικές μεθόδους.

Ο ρόλος του στατιστικού μιας μεγάλης επιχείρησης είναι να συγκρίνει το στατιστικό υλικό, να το παρουσιάζει σε πίνακες και διαγράμματα και να υπολογίζει τους απαραίτητους στατιστικούς δείκτες. Όλα αυτά τα στατιστικά στοιχεία πρέπει να είναι στη διάθεση του γενικού διευθυντή, ο οποίος θα τα χρησιμοποιήσει για τη λήψη ορθών αποφάσεων για την επιχείρηση.

Οι κυριότερες στατιστικές δραστηριότητες μιας μεγάλης επιχείρησης είναι οι ακόλουθες:

1. Το τμήμα Οικονομικών ή Στατιστικών Μελετών στο οποίο ο προϊστάμενος οικονομολόγος ή στατιστικός αναλύει γενικές επιχειρηματικές τάσεις και προβλέψεις των επιχειρηματικών δραστηριοτήτων για τις τιμές των προϊόντων και άλλους οικονομικούς παράγοντες. Συντονίζει τη στατιστική εργασία των τμημάτων της επιχείρησης και εκδίδει συνοπτικές εκθέσεις των λειτουργιών της επιχείρησης, τις οποίες υποβάλλει στους διευθύνοντες της επιχείρησης.
2. Το τμήμα Marketing ασχολείται με την έρευνα των προτιμήσεων των καταναλωτών και τη σφυγμομέτρηση της κοινής γνώμης. Οι επιχειρήσεις επιδιώκουν την παραγωγή και διάθεση νέων προϊόντων, διενεργούν δειγματοληπτικές έρευνες για να εξακριβώσουν τις ανάγκες και τις προτιμήσεις των καταναλωτών. Η διερεύνηση της αγοράς για την

εξακρίβωση των αντιδράσεων του καταναλωτικού κοινού, σχετικά με την ποιότητα, την τιμή, κ.λ.π του προϊόντος, μπορεί να αποτρέψει τις επιχειρήσεις από επικίνδυνους πειραματισμούς. Το τμήμα Marketing αποτελεί σήμερα απαραίτητη λειτουργία των σύγχρονων επιχειρήσεων. Σκοπός του Marketing είναι η έρευνα και η πολιτική κατακτήσεως της αγοράς.

3. Το τμήμα Παραγωγής διενεργεί Στατιστικό Έλεγχο Ποιότητας παραγόμενων προϊόντων κατά μαζικό τρόπο. Είναι γεγονός, ότι η σύγχρονη τεχνική της μαζικής παραγωγής απαιτεί όπως κάθε μονάδα του παραγόμενου προϊόντος , ικανοποιεί ορισμένες «προδιαγραφές» , δηλαδή τα παραγόμενα προϊόντα πρέπει να έχουν ορισμένες διαστάσεις, χημικές συνθέσεις, κ.λ.π οι οποίες δεν πρέπει να αποκλίνουν (διαφέρουν) από το προκαθορισμένο παραγόμενο προϊόν. Με την εφαρμογή των μεθόδων του Στατιστικού Ελέγχου Ποιότητας, μπορούμε να εντοπίσουμε εγκαίρως τις υπάρχουσες αποκλίσεις του παραγόμενου προϊόντος από το πρότυπο προϊόν, οι οποίες δημιουργούνται κατά τη διάρκεια της παραγωγικής διαδικασίας, έτσι ώστε να ρυθμίσουμε τις συνθήκες της παραγωγής εγκαίρως, με αποτέλεσμα να περιορίσουμε τις αποκλίσεις σε λογικά επίπεδα και έτσι να ελαττώσουμε τον αριθμό ή το ποσοστό των ελαττωματικών προϊόντων, να μειώσουμε το κόστος παραγωγής, να αυξήσουμε την παραγωγικότητα, να διατηρήσουμε την ομοιομορφία της ποιότητας των παραγόμενων προϊόντων, να αυξήσουμε την κατανάλωση και τελικά να αυξήσουμε τα κέρδη της επιχείρησης.
4. Το τμήμα Οικονομικού Ελέγχου συνδυάζει στατιστικές και λογιστικές μεθόδους για την κατάρτιση του προϋπολογισμού του επόμενου οικονομικού έτους και περιλαμβάνει: πωλήσεις, πρώτες ύλες, εργατικό

δυναμικό, καθαρά κέρδη κ.ά. Ενδεχομένως τηρεί σύστημα πρότυπου κόστους (standard cost) για την άσκηση τιμολογιακής πολιτικής.

5. Το τμήμα Προσωπικού ασχολείται με την παρακολούθηση των εξής στοιχείων:

α) Δύναμη προσωπικού κατά κατηγορία εργαζομένων (εργάτες, υπάλληλοι, μαθητευόμενοι).

β) Σύνθεση προσωπικού κατά γένος, ειδικότητα, κ.λ.π.

γ) Αμοιβές προσωπικού (μισθοί, ημερομίσθια, ΙΚΑ, οικογενειακά επιδόματα, κ.λ.π.).

δ) Χρόνος εργασίας. Ο χρόνος εργασίας καθορίζει και το μέσο ωριαίο κόστος εργασίας για μισθούς – ημερομίσθια , κ.λ.π.

- Περιγραφική και Επαγωγική Στατιστική. Έννοια και περιεχόμενο

Η στατιστική ασχολείται με μεθόδους συλλογής, ταξινομήσεως και παρουσιάσεως, αναλύσεως και ερμηνείας αριθμητικών δεδομένων, τα οποία είναι στοιχεία χρήσιμα για τον προγραμματισμό και τη λήψη ορθών αποφάσεων. Σε μερικές περιπτώσεις, ο αντικειμενικός σκοπός της στατιστικής είναι κατά περιγραφικό τρόπο, διερεύνηση του στατιστικού υλικού, παρουσίαση στατιστικών δεδομένων με στατιστικούς πίνακες και διαγράμματα και υπολογισμός των αναγκαίων στατιστικών παραμέτρων – και η εξαγωγή συμπερασμάτων, τα οποία αναφέρονται μόνο σε ένα τμήμα του στατιστικού πληθυσμού, χωρίς να γενικεύονται τα συμπεράσματα για το σύνολο των στοιχείων του πληθυσμού. Στις περιπτώσεις αυτές λέμε ότι η διερεύνηση των στατιστικών δεδομένων γίνεται με τεχνικές και μεθόδους της Περιγραφικής Στατιστικής (Descriptive Statistics).

Σε άλλες περιπτώσεις, η στατιστική εφαρμόζει μεθόδους, βάσει των οποίων είναι δυνατόν, από τα δεδομένα ενός αντιπροσωπευτικού δείγματος να εξαχθούν, επαγωγικώς, γενικότερα συμπεράσματα, δηλ. συμπεράσματα που αφορούν το σύνολο των στοιχείων του πληθυσμού από τον οποίο έχει ληφθεί το δείγμα. Στις περιπτώσεις αυτές λέμε, ότι η διερεύνηση ενός στατιστικού πληθυσμού γίνεται με μεθόδους της Επαγωγικής Στατιστικής.

Οι κυριότερες μέθοδοι της Περιγραφικής Στατιστικής είναι οι ακόλουθες:

i. Η ταξινόμηση των στατιστικών δεδομένων σε Κατανομές Συχνοτήτων και η απεικόνισή τους με Στατιστικά Διαγράμματα.

ii. Ο υπολογισμός των βασικών στατιστικών παραμέτρων των διαφόρων κατανομών συχνοτήτων (π.χ. μέσων όρων, μέτρων διασποράς, ασυμμετρίας κ.λ.π)

iii. Οι διμεταβλητές ή πολυμεταβλητές κατανομές, στις οποίες οι μονάδες του πληθυσμού αναφέρονται σε δύο ή περισσότερες στατιστικές μεταβλητές βάσει των οποίων επιτυγχάνεται η διερεύνηση της αλληλεξαρτήσεως των συμπλεκομένων μεταβλητών και η τυχόν επίδραση των τιμών της άλλης μεταβλητής ή η επίδραση των τιμών πολλών μεταβλητών στη διαμόρφωση των τιμών μιας μεταβλητής. Π.χ. η διαμόρφωση του κόστους ενός βιομηχανικού προϊόντος δέχεται την επίδραση των μεταβλητών: πρώτες ύλες, μισθοί-ημερομίσθια, γενικά έξοδα, κ.ά. Ο υπολογισμός των διμεταβλητών ή πολυμεταβλητών μαθηματικών υποδειγμάτων, συντελεστών συσχετίσεως των συμπλεκόμενων μεταβλητών, κ.λ.π.

iv. Η Ανάλυση των Χρονολογικών σειρών, για τη διαπίστωση, περιγραφή και μελέτη της διαχρονικής εξελίξεως διαφόρων φαινομένων (οικονομικών, κοινωνικών, μετεωρολογικών, κ.α.).

- Το εθνικό στατιστικό σύστημα της Ελλάδος

Η συστηματική συλλογή, επεξεργασία, παρουσίαση και ανάλυση πάσης φύσεως στατιστικού υλικού σε μια χώρα, αποτελεί το στατιστικό σύστημα αυτής. Η συγκρότηση του στατιστικού συστήματος μιας χώρας εξαρτάται κυρίως από τη χρησιμότητα και τη ζήτηση στατιστικών δεδομένων από το κράτος, τους μεγάλους οργανισμούς, τις διάφορες επιχειρήσεις και τους ερευνητές. Προγραμματισμός οικονομικής και κοινωνικής αναπτύξεως είναι αδιανόητος, χωρίς την ύπαρξη των αναγκαίων στατιστικών δεδομένων, δηλαδή χωρίς την ύπαρξη ενός οργανωμένου στατιστικού συστήματος. Προγραμματισμός της δράσεως των διαφόρων επιχειρηματικών μονάδων είναι αδύνατος, χωρίς την παρουσία κάποιου στατιστικού συστήματος.

ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Η Επαγωγική Στατιστική ασχολείται με μεθόδους, οι οποίες γενικεύουν τα δειγματοληπτικά συμπεράσματα, δηλ. μελετώντας τις στατιστικές παραμέτρους ενός αντιπροσωπευτικού δείγματος, εξάγει συμπεράσματα για τις αντίστοιχες παραμέτρους του πληθυσμού, από τον οποίο έχει ληφθεί το δείγμα. Η Στατιστική Επαγωγή (Statistical Inference) περιλαμβάνει τρεις σπουδαίες στατιστικές μεθόδους: α) την Εκτίμηση Παραμέτρων του πληθυσμού, β) τον Έλεγχο Στατιστικών Υποθέσεων και γ) τη Θεωρία Λήψεως Αποφάσεων. Η Επαγωγική Στατιστική βασίζεται σε ένα σπουδαίο επιστημονικό κλάδο, ο οποίος ονομάζεται θεωρία Πιθανοτήτων.

Εμείς, στις επόμενες σελίδες θα χωρίσουμε την επαγωγική στατιστική σε δύο τομείς:

i) Την εκτίμηση σημείου και τα διαστήματα εμπιστοσύνης και

ii) Τον έλεγχο υποθέσεων.

Ο έλεγχος υποθέσεων θα αναλυθεί εκτενώς, μιας και από αυτόν θα εξάγουμε τα όποια συμπεράσματα.

ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Από τα στοιχεία ενός τυχαίου δείγματος μπορούμε να εκτιμήσουμε διάφορες παραμέτρους του πληθυσμού. Έστω ότι ο μέσος δείγματος είναι ($\bar{x}$)και είναι μια εκτίμηση του πληθυσμού (μ). Η διακύμανση ενός δείγματος (s^2) υπολογίζεται με n-1 παρατηρήσεις – παρέχει μια εκτίμηση διακυμάνσεως του πληθυσμού s^2.

Αν από Ν πληθυσμό, με Ν στοιχεία, πάρουμε ισοπληθή τυχαία δείγματα ($n_1 = n_2 = n_3 = ... = n$) και από κάθε δείγμα υπολογίσουμε ορισμένη παράμετρο (π.χ. μέσο αριθμητικό, διακύμανση, αναλογία) , τότε προκύπτει μια σειρά ομοειδών εκτιμήσεων των αντιστοίχων παραμέτρων του πληθυσμού, από τον οποίο προέρχονται τα δείγματα. 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 απ’ αυτή συμμετρικά ή ασυμμετρικά. Όταν ο αριθμός των δειγμάτων είναι αρκετά μεγάλος , τότε οι τιμές π.χ. των μέσων των δειγμάτων θα παρουσιάζουν μια συνέχεια (δηλ. η μια τιμή θα είναι πολύ κοντά στην άλλη) και , επομένως, η γραφική απεικόνιση της κατανομής θα τείνει να σχηματίσει μια συνεχή και ομαλή καμπύλη συχνοτήτων, η οποία ονομάζεται κατανομή δειγματοληψίας (Sampling Distribution) της εξεταζόμενης παραμέτρου. Οι κατανομές Δειγματοληψίας έχουν τεράστια σημασία στην στατιστική συμπερασματολογία, γιατί η ισχύ των στατιστικών συμπερασμάτων εξαρτάται από την γνώση την οποία έχουμε για τις κατανομές δειγματοληψίας των διαφόρων παραμέτρων των δειγμάτων.

- **Κατανομή δειγματοληψίας μέσου αριθμητικού**

Κεντρικό Οριακό Θεώρημα

Σε αυτό το σημείο θα πρέπει να αναφέρουμε και το κεντρικό οριακό θεώρημα, το οποίο είναι ένα από τα σπουδαιότερα θεωρήματα στη στατιστική. Το θεώρημα αναφέρεται στην οριακή κατανομή του τυποποιημένου μέσου, όταν το μέγεθος του δείγματος τείνει προς το άπειρο.

Η τυποποιημένη μορφή μιας τυχαίας μεταβλητής X είναι μια άλλη μεταβλητή z, η οποία συνδέεται με τη X με τη σχέση:

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

Όπου μ είναι ο μέσος, και σ η τυπική απόκλιση της X. Είναι εύκολο να δειχθεί ότι

E(z)=0

V(z)=1

Όταν η μεταβλητή X ακολουθεί την κανονική κατανομή, η τυποποιημένη μορφή της είναι γνωστή ως Z κατανομή. Δηλαδή, η κατανομή Z είναι η κανονική κατανομή με μέσο το μηδέν και διακύμανση τη μονάδα.

Αν η τυχαία μεταβλητή είναι ο μέσος $\bar{X}$,τότε:

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

είναι ο τυποποιημένος μέσος.

Έστω ότι έχουμε έναν πληθυσμό N στοιχείων με μέσο μ και διακύμανση s^2. Αν από αυτόν τον πληθυσμό πάρουμε μεγάλο αριθμό ισοπληθών δειγμάτων ($n_1 = n_2 = n_3 = ... = n$), υπολογίσουμε τους μέσους αυτών των δειγμάτων ($\bar{x}_1, \bar{x}_2, \bar{x}_3, ...$)

και βάσει αυτών κατασκευάσουμε μια κατανομή συχνοτήτων, τότε η κατανομή αυτή ονομάζεται κατανομή δειγματοληψίας του μέσου αριθμητικού και έχει τις εξής χαρακτηριστικές ιδιότητες:

Α) Η κατανομή θα είναι κανονική αν ο πληθυσμός είναι κανονικός και μάλιστα εξαρτάται από το μέγεθος των ισοπληθών δειγμάτων. Αν τώρα ο πληθυσμός, από τον οποίο παίρνουμε τα δείγματα, δεν είναι κανονικός, τότε η κατανομή δειγματοληψίας του μέσου προσεγγίζει την κανονική, όσο το μέγεθος του δείγματος αυξάνει (n>30).

Β) Ο μέσος της κατανομής δειγματοληψίας των μέσων ($=\overset{=}{x}$) ισούται με το μέσο του πληθυσμού από τον οποίο έχουν ληφθεί τα δείγματα. Δηλαδή είναι:

$$E(\overline{x})\overset{=}{x} = m$$

Όπου

$$m = \frac{\sum_{i=1}^{N} x_1}{N}$$

Γ) Η διακύμανση της κατανομής δειγματοληψίας του μέσου συνδέεται με την διακύμανση s^2 του πληθυσμού με σχέση:

$$s_{\overline{x}}^2 = \frac{s^2}{n}$$

Όπου

$$s^2 = \frac{\sum_{i=1}^{\mathrm{N}} (x_i - m)^2}{\mathrm{N}}$$

Δ) Η μέση απόκλιση τετραγώνου της κατανομής δειγματοληψίας του μέσου, ονομάζεται τυπικό σφάλμα εκτιμήσεως του μέσου και προσδιορίζεται από τη σχέση:

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

Από τη σχέση αυτή φαίνεται καθαρά ότι όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο μικρότερο θα είναι το τυπικό σφάλμα εκτιμήσεως του μέσου και, επομένως τόσο περισσότερο ακριβής θα είναι η εκτίμηση του δείγματος.

Η παραπάνω σχέση ισχύει όταν ο πληθυσμός από τον οποίο παίρνονται τα δείγματα είναι άπειρος ή πεπερασμένος αλλά η δειγματοληψία γίνεται με επανατοποθέτηση. Αν όμως ο πληθυσμός είναι πεπερασμένος με πλήθος στοιχείων τότε το τυπικό σφάλμα εκτιμήσεως του μέσου υπολογίζεται βάσει τύπου:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Στις πρακτικές όμως εφαρμογές, θεωρούμε ότι ελέγχουμε άπειρους πληθυσμούς και για αυτό χρησιμοποιούμε συνήθως τον τύπο

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

Είναι προφανές ότι, όσο μικρότερο είναι το $s_{\bar{x}}$, τόσο πιο αξιόπιστη είναι η εκτίμηση του μέσου πληθυσμού (μ), γιατί η διασπορά των μέσων $\bar{x}_i$ των δειγμάτων γύρω από το μέσο των μέσων $\bar{\bar{x}}$ των δειγμάτων θα είναι μικρή και επομένως οι μέσοι των δειγμάτων δεν θα διαφέρουν πολύ από το μέσο των μέσων των μέσων $\bar{\bar{x}}$ των δειγμάτων και επομένως από το μ του πληθυσμού, γιατί $\bar{\bar{x}}$ =μ.

Αν το δείγμα μας είναι πολύ μικρό τότε η βεβαιότητά μας μικραίνει και η s δεν είναι αξιόπιστη εκτιμήτρια της σ.

Έτσι λοιπόν ο λόγος $t = \frac{\overline{x}_i - m}{s_i}$ δεν ακολουθεί την κανονική κατανομή αλλά τη γνωστή ως κατανομή t (student) με n-1 βαθμούς ελευθερίας.

Ο Gosset υπέθεσε ότι ο πληθυσμός από τον οποίο παίρνουμε τα δείγματα είναι κανονικός, έχει γνωστή μέση τιμή μ και γνωστή τυπική απόκλιση. Οπότε οι τιμές για το $s_{\overline{x}}$ που θα χρησιμοποιήσουμε είναι οι z τιμές. Χρησιμοποιούμε το $s_{\overline{x}} = \frac{s}{\sqrt{n}}$, όπου S είναι η τυπική απόκλιση του δείγματος μεγέθους n και όχι το άγνωστο σ του πληθυσμού. Δηλαδή το τυπικό σφάλμα $s_{\overline{x}}$ (που υπολογίζεται από το δείγμα) χρησιμοποιήθηκε ως ένας εκτιμητής του $s_{\overline{x}}$, αφού το σ είναι άγνωστο.

Σε περίπτωση που η δειγματική κατανομή του $\overline{x}_i$ είναι κανονική αλλά δεν γνωρίζουμε την τυπική απόκλιση, τότε ο τύπος που θα αποδίδει την κατανομή πιθανότητας θα είναι ο εξής: $t = \frac{\overline{x} - m}{\frac{s}{\sqrt{n}}}$

Γενικά, όταν γνωρίζουμε το σ του πληθυσμού χρησιμοποιούμε τις τιμές z, ενώ όταν το σ είναι άγνωστο, χρησιμοποιούμε το στατιστικό t.

- **Κατανομή δειγματοληψίας για τη διακύμανση**

Στις περισσότερες των περιπτώσεων η κατανομή του πληθυσμού μπορεί να υποτεθεί κανονική.

Από έναν κανονικό πληθυσμό N(μ, σ^2), παίρνουμε ένα τυχαίο δείγμα X_1, X_2, X_3, …X_n, τότε και το άθροισμα:

$$\frac{\Sigma(X_i - \overline{X})^2}{s^2}$$

ακολουθεί την κανονική κατανομή x^2.

Γνωρίζουμε πως τα $\overline{x} = \frac{\Sigma x_i}{n}$ και $S^2 = \frac{\Sigma\left(x_i - x\right)^2}{n-1} \Rightarrow (n-1)S^2 = \Sigma\left(x_i - x\right)^2 \Rightarrow$

$$\Rightarrow \frac{(n-1)S^2}{s^2} = \frac{\Sigma(x_i - x)^2}{s^2} \Rightarrow \frac{(n-1)S^2}{s^2} = \frac{(x_1 - x)^2}{s^2} + \frac{(x_2 - x)^2}{s^2} + \ldots + \frac{(x_n - x)^2}{s^2}$$

Όπου s^2 η διακύμανση του πληθυσμού. Την τυχαία μεταβλητή $\frac{(n-1)S^2}{s^2}$ τη συμβολίζουμε με x^2 και έχει n-1 βαθμούς ελευθερίας.

- **Κατανομή δειγματοληψίας διαφοράς μέσων τιμών δύο δειγμάτων**

Η διαφορά των μέσων τιμών μιας μεταβλητής που μετριέται σε δύο πληθυσμούς μπορεί να μελετηθεί στη βάση της διαφοράς των μέσων τιμών τυχαίων και ανεξάρτητων δειγμάτων που παίρνουμε από τους πληθυσμούς.

Έστω ότι x η μεταβλητή που μας ενδιαφέρει να μελετηθεί σε δύο πληθυσμούς Α και Β. Τότε τη μεταβλητή αυτή τη συμβολίζουμε με X_A στον πληθυσμό Α και με X_B στον πληθυσμό Β. Παίρνουμε όλα τα δυνατά τυχόντα δείγματα μεγέθους n_A από τον πληθυσμό Α και υπολογίζουμε τη μέση τιμή $\overline{x}_A$ για καθένα από αυτά. Το ίδιο κάνουμε για τον πληθυσμό Β και έστω $\overline{x}_B$η μέση τιμή των δειγμάτων μεγέθους n_B.

Στη συνέχεια σχηματίζουμε όλες τις διαφορές $\overline{x}_A - \overline{x}_B$και κατασκευάζουμε την κατανομή συχνοτήτων και το διάγραμμα συχνοτήτων. Η κατανομή που προκύπτει ονομάζεται δειγματική κατανομή της διαφοράς των μέσων τιμών $\bar{x}_A - \bar{x}_B$.

Περιπτώσεις:

Α) Έστω ότι οι τυχαίες μεταβλητές X_A και X_B έχουν μέσες τιμές μ_A και μ_B και τυπικές αποκλίσεις σ_A και σ_B αντίστοιχα. Έστω ακόμη ότι τα X_A και X_B ακολουθούν την κανονική κατανομή (ήτοι $X_A \sim N(m_A, s_A^2)$ και $X_B \sim N(m_B, s_B^2)$.

Οπότε η δειγματική κατανομή της τυχαίας μεταβλητής $\bar{x}_A - \bar{x}_B$ είναι κανονική κατανομή με μέση τιμή $m_A - m_B$ και διακύμανση $\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$.

Δηλαδή, $\bar{x}_A - \bar{x}_B \sim$ N ($m_A - m_B$, $\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$).

Άρα η αναμενόμενη μέση τιμή της διαφοράς των μέσων είναι:

$$m_{\bar{x}_A - \bar{x}_B} = m_A - m_B$$

Και η διακύμανση των της διαφοράς μέσων τιμών $\bar{x}_A - \bar{x}_B$ είναι:

$$s^2_{\bar{x}_A - \bar{x}_B} = \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$$

Β) Αν οι τυχαίες μεταβλητές X_A και X_B ακολουθούν την κανονική κατανομή αλλά τα μεγέθη n_A και n_B είναι πολύ μεγάλα (δηλαδή μεγαλύτερα του 30), τότε πάλι η δειγματική κατανομή της διαφοράς των μέσων τιμών $\bar{x}_A$ και $\bar{x}_B$ δεν ακολουθεί την κανονική κατανομή, δηλαδή, $\bar{x}_A - \bar{x}_B \sim N(m_A - m_B, \frac{s_A^2}{n_A} \frac{s_B^2}{n_B})$ όταν $n_A, n_B \to +\infty$.

Επισήμανση: Τα δείγματα που επιλέγουμε από τους πληθυσμούς Α και Β πρέπει να είναι τυχαία και ανεξάρτητα μεταξύ τους.

Γ) Η δειγματική κατανομή της διασποράς των μέσων τιμών μπορεί να μετατραπεί σε τυποποιημένη κανονική κατανομή όταν γνωρίζουμε τις διασπορές s_{A}^{2} και s_{B}^{2}.

$$z=\frac{\left(\overline{x}_{A}-\overline{x}_{B}\right)-\left(m_{\mathrm{A}}-m_{\mathrm{B}}\right)}{\sqrt{\frac{s_{\mathrm{B}}^{2}}{n_{A}}+\frac{s_{\mathrm{B}}^{2}}{n_{B}}}} \sim \mathrm{N}(0,1)$$

Στην περίπτωση της t κατανομής όταν δεν γνωρίζουμε τις διασπορές s_{A}^{2} και s_{B}^{2}, χρησιμοποιούμε τις διασπορές s_{A}^{2} και s_{B}^{2}.

$$t=\frac{\left(\overline{x}_{A}-\overline{x}_{B}\right)-\left(m_{\mathrm{A}}-m_{\mathrm{B}}\right)}{\sqrt{\frac{s_{p}^{2}}{n_{A}}+\frac{s_{p}^{2}}{n_{B}}}}$$

Προσεγγίζει την t-κατανομή με $(n_A-1)+(n_B-1)=n_A+n_B-2$ βαθμούς ελευθερίας και

$$s_{p}^{2}=\frac{(n_{A}-1)s_{A}^{2}+(n_{B}-1)s_{B}^{2}}{n_{A}+n_{B}-2}.$$

- **Κατανομή δειγματοληψίας αναλογίας (ή ποσοστού)**

Υπάρχουν περιπτώσεις κατά τις οποίες ενδιαφερόμαστε να εκτιμήσουμε την αναλογία (το ποσοστό) των στοιχείων ενός πληθυσμού, τα οποία έχουν ένα ορισμένο χαρακτηριστικό (π.χ. το ποσοστό των νοικοκυριών που προτιμούν το προϊόν Α κ.λ.π.). Για να εκτιμήσουμε το άγνωστο ποσοστό του πληθυσμού (=P), παίρνουμε το πλήθος των μονάδων (=f) του δείγματος, οι οποίες έχουν αυτό το ποσοστό p=f/n. Προφανώς, το ποσοστό που δεν έχει το ερευνώμενο χαρακτηριστικό είναι q=1-p.

Αν από έναν πληθυσμό πάρουμε όλα τα δυνατά ισοπληθή δείγματα μεγέθους n, τότε θα σχηματιστεί η κατανομή,

$$n_1 = n_2 = n_3 = \ldots = n_k = n$$

$$p_1 = \tfrac{f_1}{n_1}, \quad p_2 = \tfrac{f_2}{n_2}, \; p_3 = \tfrac{f_3}{n_3}, \; \ldots, \quad p_k = \tfrac{f_k}{n_k}$$

η οποία ονομάζεται κατανομή δειγματοληψίας ποσοστού και έχει τις εξής χαρακτηριστικές ιδιότητες:

α) Η κατανομή δειγματοληψίας ακολουθεί την κανονική κατανομή αν το μέγεθος n του δείγματος είναι αρκετά μεγάλο ($n \geq 30$).

β) Το μέσο ποσοστό της κατανομής δειγματοληψίας ισούται με το μέσο ποσοστό πληθυσμού. Δηλαδή:

$$E(p) = m_p = p$$

γ) Το τυπικό σφάλμα εκτιμήσεως της αναλογίας (ποσοστού) είναι:

$$s_{\,p} = \sqrt{\tfrac{P(1-P)}{n}}$$

Στην πράξη, εργαζόμαστε συνήθως με ένα δείγμα από το οποίο υπολογίζουμε το p και το q=1-p, οπότε το τυπικό σφάλμα εκτιμήσεως υπολογίζεται με τον τύπο:

$$s_p = \sqrt{\tfrac{p \cdot q}{n}}$$

Η παραπάνω σχέση ισχύει όταν ο πληθυσμός από τον οποίο παίρνουμε τα δείγματα είναι άπειρος. Αν ο πληθυσμός είναι πεπερασμένος ή οι μονάδες του δείγματος παίρνονται χωρίς επανατοποθέτηση, τότε εμφανίζεται ο τύπος:

$$s_p = \sqrt{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}}$$

- **Κατανομή δειγματοληψίας διαφοράς δύο αναλογίων (ποσοστών)**

Έστω ότι έχουμε δύο πληθυσμούς Π_1 και Π_2 και η αναλογία (ποσοστό) ενός χαρακτηριστικού στον Π_1 είναι P_1 και στον Π_2 είναι P_2. Αν τώρα από τους πληθυσμούς αυτούς πάρουμε όλα τα δυνατά ζεύγη δειγμάτων n_1, n_2, από κάθε ζεύγος υπολογίζουμε τα ποσοστά p_1,p_2 και στη συνέχεια τη διαφορά τους p_1-p_2, τότε θα προκύψουν τόσες διαφορές p_1-p_2 όσα και τα ζεύγη των δειγμάτων. Αν τις διαφορές p_1-p_2 τις ταξινομήσουμε σε μια κατανομή συχνοτήτων, τότε η κατανομή αυτή ονομάζεται κατανομή δειγματοληψίας διαφοράς δύο ποσοστών (αναλογιών) και έχει τα εξής βασικά χαρακτηριστικά:

1) Αν τα δείγματα n_1, n_2 είναι πολυπληθή $(n_1,\ n_2 \geq 30)$, τότε η κατανομή δειγματοληψίας της διαφοράς p_1,p_2 θα είναι κανονική ή θα προσεγγίζει την κανονική κατανομή όσο το μέγεθος των δειγμάτων συνεχώς αυξάνει.
2) Η μέση τιμή των διαφορών p_1-p_2 συμπίπτει με τη διαφορά p_1-p_2 των ποσοστών στους πληθυσμούς.

3) Για τη διακύμανση και το τυπικό σφάλμα εκτιμήσεως της παραπάνω κατανομής δειγματοληψίας ισχύουν τα ακόλουθα:

α) Αν οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα είναι διαφορετικοί, τότε η διακύμανση είναι:

$$S^2_{p_1-p_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Όπου $q_1=1-p_1$ και $q_2=1-p_2$

και το τυπικό σφάλμα εκτιμήσεως θα είναι:

$$S_{p_1-p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

β) Αν τώρα τα δείγματα προέρχονται από τον ίδιο πληθυσμό ή οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα είναι όμοιοι, τότε θα έχουν κοινή αναλογία η οποία εκτιμάται από τις αναλογίες p_1 και p_2 δειγμάτων με βάση τον τύπο:

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

και $q=1-p$.

Η διακύμανση είναι της κατανομής δειγματοληψίας της p_1-p_2 θα είναι:

$$s^2_{p_1-p_2} = \frac{pq}{n_1} + \frac{pq}{n_2}$$

και το τυπικό σφάλμα εκτιμήσεως:

$$s_{p_1-p_2} = \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}} = \sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}$$

- **Κατανομή δειγματοληψίας του λόγου των διασπορών (F κατανομή)**

Σε πολλές πρακτικές εφαρμογές προκύπτει η ανάγκη σύγκρισης των διακυμάνσεων δύο πληθυσμών. Έτσι π.χ. αν ένα μηχάνημα αξιολογείται με βάση την ομοιογένεια του παραγόμενου προϊόντος, τότε θα συγκρίνουμε δύο μηχανήματα, συγκρίνοντας τις διακυμάνσεις ενός χαρακτηριστικού ποιότητας του προϊόντος. Εξάλλου είδαμε ότι σε ορισμένες περιπτώσεις για να κάνουμε τον έλεγχο για την διαφορά δύο μέσων τιμών πληθυσμού θα πρέπει να ισχύει η υπόθεση της ισότητας των δύο διακυμάνσεων.

Αν s_1^2 είναι η διακύμανση τυχαίου δείγματος n_1 στοιχείων από έναν κανονικό πληθυσμό με διακύμανση s_1^2 και s_2^2 η διακύμανση τυχαίου δείγματος n_2 στοιχείων από έναν κανονικό πληθυσμό με διακύμανση s_2^2 και τα δύο δείγματα είναι ανεξάρτητα τότε, η τυχαία μεταβλητή

$$F = \frac{S_1^2 / s_1^2}{S_2^2 / s_2^2}$$

ακολουθεί την κατανομή F με $n_1 = n_1 - 1$ και $v_2 = n_2 - 1$ βαθμούς ελευθερίας.

ΕΚΤΙΜΗΣΗ ΣΗΜΕΙΟΥ

Αν θεωρήσουμε ότι έχουμε μια τυχαία μεταβλητή Χ, η οποία χαρακτηρίζεται από μια παράμετρο που παριστάνουμε με y. Η παράμετρος y είναι είτε μέσος είτε διακύμανση κτλ. Για να κάνουμε εκτίμηση αυτής της άγνωστης παραμέτρου y, συνδυάζουμε τις πληροφορίες που έχουμε ως γνωστές, με τις πληροφορίες που μας δίνει ένα τυχαίο δείγμα (x_1, x_2, x_3, $\ldots, x_n$). Το πρόβλημα είναι να διαλέξουμε μια συνάρτηση των παρατηρήσεων του δείγματος έτσι που η τιμή της για ένα δεδομένο δείγμα να είναι μια αποδεκτή εκτίμηση της παραμέτρου του πληθυσμού.

Μια συνάρτηση των παρατηρήσεων του δείγματος, που δίνει εκτιμήσεις της παραμέτρου του πληθυσμού y, ονομάζεται εκτιμητής και παριστάνεται:

$$\hat{y} = \hat{y}(x_1, x_2, \ldots x_n)$$

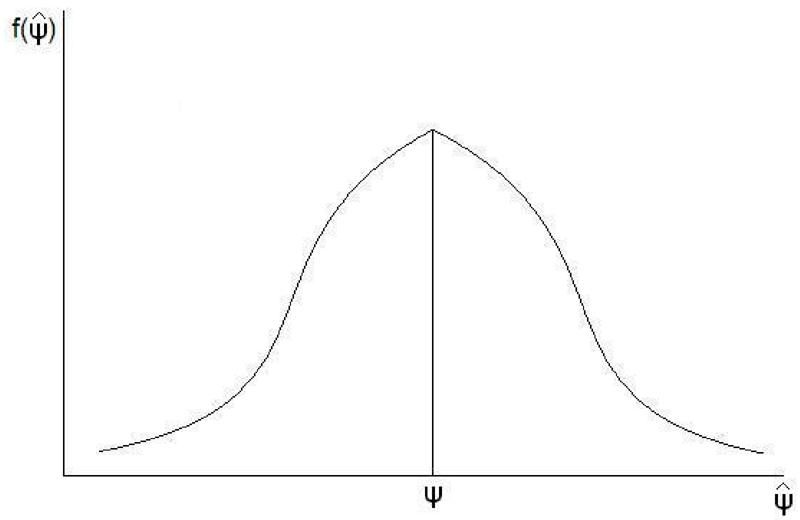
Η τιμή που παίρνει ο εκτιμητής $\hat{y}$ για ένα δεδομένο δείγμα ονομάζεται εκτίμηση της παραμέτρου *y*. Η παραπάνω συνάρτηση μπορεί να πάρει οποιαδήποτε τιμή , αλλά δεν μπορεί να περιλαμβάνει άγνωστες παραμέτρους του πληθυσμού.

Οι εκτιμητές είναι τυχαίες μεταβλητές αφού είναι συναρτήσεις των παρατηρήσεων του δείγματος, οι οποίες είναι τυχαίες μεταβλητές. Οι τιμές που παίρνει ένας εκτιμητής $\hat{y}$ είναι τιμές στατιστικών δείγματος και θα διαφέρουν από δείγμα σε δείγμα. Επομένως, ένας εκτιμητής $\hat{y}$ δεν μπορεί να δίνει πάντοτε την αληθινή τιμή της παραμέτρου του πληθυσμού. Η διαφορά μεταξύ της τιμής του εκτιμητή για ένα δεδομένο δείγμα και της αληθινής τιμής της παραμέτρου στον πληθυσμό ονομάζεται σφάλμα δειγματοληψίας. Η διαφορά μεταξύ της προσδοκώμενης τιμής, Ε$\hat{y}$, και της αληθινής τιμής της παραμέτρου ψ στον πληθυσμό ονομάζεται σφάλμα μεροληψίας.

Γενικά ισχύει ότι οι εκτιμήτριες συναρτήσεις, επειδή είναι συναρτήσεις τυχαίων μεταβλητών, είναι και αυτές τυχαίες μεταβλητές, ενώ οι εκτιμήσεις είναι σταθερές ποσότητες.

Αν εξετάσουμε όλα τα παραπάνω θα παρατηρήσουμε ότι αν πάρουμε άπειρα δείγματα μεγέθους n θα έχουμε άπειρες τιμές για το $\hat{y}$ μιας τυχαίας μεταβλητής, εφόσον ο εκτιμητής $\hat{y}$ είναι τυχαία μεταβλητή.

Η κατανομή δειγματοληψίας του εκτιμητή $\hat{y}$, παριστάνεται στο παρακάτω σχεδιάγραμμα.



ΙΔΙΟΤΗΤΕΣ ΕΚΤΙΜΗΤΩΝ

Οι ιδιότητες των εκτιμητών μικρών δειγμάτων είναι οι εξής:

α) *Αμεροληψία*

Ένας εκτιμητής $\hat{q}$ μιας παραμέτρου του πληθυσμού θ είναι αμερόληπτος όταν η προσδοκώμενη τιμή του είναι ίση με την παράμετρο, δηλαδή,

$E\hat{q} = \theta$

Ο μέσος του δείγματος είναι αμερόληπτος εκτιμητής του μέσου του πληθυσμού. Επίσης, η διακύμανση του δείγματος:

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

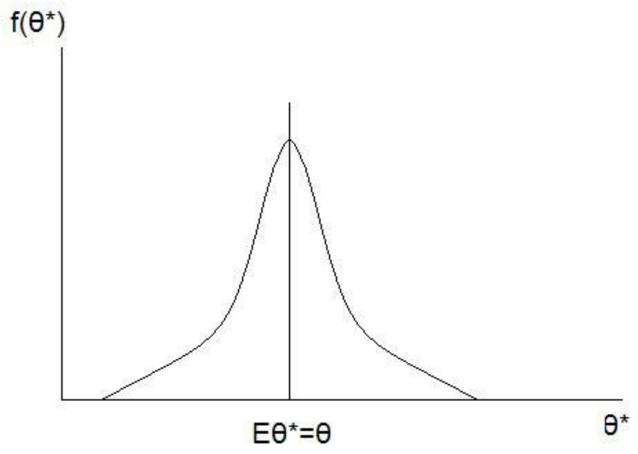
Είναι αμερόληπτος εκτιμητής της διακύμανσης στον πληθυσμό. Αντιθέτως, η τυπική απόκλιση (s) δεν είναι αμερόληπτος εκτιμητής της τυπικής απόκλισης στον πληθυσμό.

Η ιδιότητα της αμεροληψίας μπορεί να μην είναι αρκετή, γιατί μπορεί να υπάρχουν περισσότεροι από ένας αμερόληπτοι εκτιμητές μιας δεδομένης παραμέτρου του πληθυσμού. Επιπλέον η αμεροληψία ενός εκτιμητή δεν μας συμπεραίνει κάτι για τη διασπορά της κατανομής του. Ένας αμερόληπτος εκτιμητής με μεγάλη διακύμανση δεν είναι επιθυμητός, γιατί οι εκτιμήσεις του, τις περισσότερες φορές, θα διαφέρουν πολύ από την αληθινή τιμή της παραμέτρου. Αλλά και ένας εκτιμητής που έχει μικρή ή μηδέν διακύμανση δεν είναι επίσης επιθυμητός αν είναι μεροληπτικός. Μπορούμε εύκολα να βρούμε έναν εκτιμητή με μηδέν διακύμανση, έστω μια οποιαδήποτε σταθερά, πχ. $\hat{q} = 5$. Ένας τέτοιος εκτιμητής δεν μπορεί να είναι χρήσιμος, γιατί αγνοεί εντελώς τις πληροφορίες από το δείγμα. Επομένως, μεταξύ των αμερόληπτων εκτιμητών, ο εκτιμητής με τη μικρότερη διακύμανση είναι περισσότερο επιθυμητός.

β) *Αποτελεσματικότητα*

Ένας εκτιμητής $\hat{q}$ είναι αποτελεσματικός εκτιμητής της παραμέτρου θ αν ο i) $\hat{q}$ είναι αμερόληπτος και ii) $V(\hat{q})<V(\theta^*)$, όπου θ* είναι οποιοσδήποτε άλλος αμερόληπτος εκτιμητής της θ.

Μερικοί συγγραφείς χρησιμοποιούν το κριτήριο του ελάχιστου μέσου του τετραγώνου του σφάλματος για να ορίσουν την ιδιότητα της αποτελεσματικότητας. Δηλαδή, ένας εκτιμητής είναι αποτελεσματικός αν ελαχιστοποιεί τον μέσο του τετραγώνου του σφάλματος. Άρα, μπορεί να είναι αρκετά μικρή ώστε να αντισταθμίζει το σφάλμα μεροληψίας. Παρακάτω φαίνεται ότι ο εκτιμητής θ* είναι μεροληπτικός αλλά έχει μικρότερη διακύμανση από τον εκτιμητή $\hat{q}$ που είναι αμερόληπτος.



γ) *Συνέπεια*

Μια εκτιμήτρια συνάρτηση Τ μιας παραμέτρου θ λέγεται ότι είναι συνεπής, εάν, καθώς αυξάνει το μέγεθος του δείγματος, η κάθε φορά προκύπτουσα τιμή της εκτιμήτριας συνάρτησης τείνει προς την τιμή της παραμέτρου θ του πληθυσμού.

Με άλλα λόγια, αν μια εκτιμήτρια συνάρτηση Τ συγκλίνει στοχαστικά προς την παράμετρο θ με πιθανότητα 1, όταν το μέγεθος του δείγματος τείνει στο άπειρο, τότε η εκτιμήτρια συνάρτηση Τ είναι συνεπής εκτιμητής της παραμέτρου θ, δηλαδή :

$P\{|T-q|<e\}\to 1$ για ε>0

ή

$P\{|T-q|>e\}\to 0$ για $n\to\infty$

Πιο συγκεκριμένα, έστω ότι έχουμε τις εξής παρατηρήσεις $x_1, x_2, x_3, ..., x_n$ ενός τυχαίου δείγματος $X_1, X_2, X_3, ...X_n$, που πήραμε από ένα πληθυσμό που έχει συνάρτηση πιθανότητας$\hat{q}$ $f(x,q)$ και εκτιμήτρια συνάρτηση $\mathrm{T}=\mathrm{T}(X_1, X_2, X_3, ...X_n)$ της παραμέτρου θ. Αν πάρουμε τις παραπάνω παρατηρήσεις του δείγματος $x_1, x_2, x_3, ..., x_n$, τότε η εκτίμηση της παραμέτρου θ είναι $\mathrm{T}=\mathrm{T}(X_1, X_2, X_3, ...X_n)$, όπου n είναι το μέγεθος του δείγματος. Όταν αυξάνει το μέγεθος του δείγματος n, η εκτίμηση της πλησιάζει την τιμή της παραμέτρου θ, δηλαδή όταν το n τείνει στο άπειρο, η ακολουθία των εκτιμήσεων $\hat{\mathrm{T}}$ συγκλίνει στο θ.

δ) *Επάρκεια*

Ένας εκτιμητής $\hat{q}$ είναι επαρκής εκτιμητής της παραμέτρου θ όταν χρησιμοποιεί όλες τις πληροφορίες σχετικά με την παράμετρο θ, που περιέχονται στο δείγμα. Δηλαδή αν πάρουμε για παράδειγμα το μέσο του δείγματος, τότε θα δούμε ότι είναι επαρκής εκτιμητής του μέσου πληθυσμού, γιατί από το δείγμα δεν μπορούμε να πάρουμε άλλες πληροφορίες που να μας βοηθήσουν να μορφώσουμε καλύτερη γνώμη για τον μέσο του πληθυσμού, ενώ η διάμεσος που προκύπτει από το δείγμα δεν είναι επαρκής εκτιμήτρια, αφού δεν χρησιμοποιεί τις τιμές των παρατηρήσεων, αλλά μόνο τη θέση τους. Η ιδιότητα της ικανότητας από μόνη της δεν είναι τόσο σπουδαία ή επιθυμητή. Η σημασία της έγκειται στο ότι αποτελεί αναγκαία συνθήκη για την ιδιότητα της αποτελεσματικότητας.

ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

Όπως έχουμε πει προηγουμένως, η εξαγωγή γενικότερων συμπερασμάτων για τη δομή ενός στατιστικού πληθυσμού, με βάση ένα τυχαίο δείγμα, αποτελεί το κεντρικό πρόβλημα της στατιστικής και ονομάζεται στατιστική επαγωγή. Ένα πρόβλημα της στατιστικής επαγωγής είναι η εκτίμηση παραμέτρων του πληθυσμού κατά την οποία από τα στοιχεία ενός τυχαίου δείγματος βρίσκουμε τις άγνωστες παραμέτρους του πληθυσμού. Επειδή ο ακριβής προσδιορισμός της άγνωστης παραμέτρου του πληθυσμού από την αντίστοιχη παράμετρο ενός δείγματος είναι δύσκολος, γι αυτό προσδιορίζουμε συνήθως δύο σταθμούς, δύο όρια, εντός των οποίων αναμένεται ότι θα περιλαμβάνεται η πραγματική, αλλά άγνωστη, παράμετρος του πληθυσμού με προκαθορισμένη κάθε φορά πιθανότητα (συνήθως 95%).

Για την καλύτερη κατανόηση και εμπέδωση ενός τέτοιου προβλήματος, παραθέτουμε το ακόλουθο παράδειγμα:

Έστω ότι θέλουμε να εκτιμήσουμε μια παράμετρο θ ενός πληθυσμού. Για την εκτίμηση της βρίσκουμε με τη βοήθεια ενός τυχαίου δείγματος και με προκαθορισμένη πιθανότητα 1-α δύο δειγματικές συναρτήσεις ή πιο απλά δύο αριθμούς θ_1,θ_2, μεταξύ των οποίων περιμένουμε να βρίσκεται η παράμετρος θ του πληθυσμού με πιθανότητα 1-α δηλαδή:

$$P\{q_1 \le q \le q_2\} = 1 - a$$

Το διάστημα (θ_1, θ_2) ονομάζεται διάστημα εμπιστοσύνης για την παράμετρο θ του πληθυσμού, οι αριθμοί θ_1 και θ_2 λέγονται όρια του διαστήματος εμπιστοσύνης, το όριο θ_1 λέγεται κατώτερο και θ_2 λέγεται ανώτερο.

Σημ.: Πληροφορίες για όλα τα επίπεδα σημαντικότητας και για όλα τα είδη των κατανομών παρέχονται από την παρακάτω ιστοσελίδα:
http://instruct1.cit.cornell.edu/Courses/econ321/public_html/stattables/sttable.html

- **Διάστημα εμπιστοσύνης μέσου αριθμητικού**

Όπως προείπαμε, η κατανομή δειγματοληψίας του μέσου αριθμητικού (η οποία έχει προκύψει από ισοπληθή δείγματα, τα οποία προέρχονται από κανονικό πληθυσμό με γνωστή διακύμανση σ^2), ακολουθεί κανονική κατανομή με μέσο μ και διακύμανση σ^2/n. Η κατανομή δειγματοληψίας του μέσου αριθμητικού που έχει προκύψει από πολυπληθή δείγματα($n \geq 30$) (ανεξάρτητα αν ο πληθυσμός από τον οποίο προέρχονται τα δείγματα είναι κανονικός ή όχι, με γνωστή διακύμανση σ^2), ακολουθεί προσεγγιστικά την κανονική κατανομή (βάσει του κεντρικού οριακού θεωρήματος). Επομένως, κατά την εκτίμηση διαστημάτων εμπιστοσύνης των διαφόρων παραμέτρων του πληθυσμού με πολυπληθή δείγματα, τις κριτικές τιμές θα τις παίρνουμε από τους πίνακες της τυποποιημένης κανονικής καμπύλης, ενώ όταν το δείγμα είναι ολιγοπληθές (n<30) και η διακύμανση σ^2 άγνωστη, τότε η κατανομή δειγματοληψίας του μέσου αριθμητικού δεν ακολουθεί την κανονική κατανομή, αλλά την κατανομή t-student.

Ο τύπος λοιπόν για το διάστημα εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού με γνωστή διακύμανση είναι ο εξής:

$$\bar{x} \pm z_{a/2} \frac{s}{\sqrt{n}}$$

Και αντίστοιχα, το διάστημα εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού αλλά με άγνωστη τη διακύμανση είναι:

$$\overline{x} \pm t_{n-1,a/2} \frac{s}{\sqrt{n}}$$

Το διάστημα εμπιστοσύνης για το μέσο του πληθυσμού εκτιμάται με βάση τη διπλή ανισότητα:

$$\overline{x} - Z_{a/2} \cdot \sigma_{\overline{x}} < \mu < \overline{x} + Z_{a/2} \cdot \sigma_{\overline{x}}$$

- **Διάστημα εμπιστοσύνης για τη διακύμανση**

Το διάστημα εμπιστοσύνης για τη διακύμανση, με γνωστό το μέσο, μ, μπορούμε να το δούμε ως εξής:

$$\left[\frac{\sum_{i=1}^{n}(X_i - m)^2}{c^2_{n,a/2}}, \frac{\sum_{i=1}^{n}(X_i - m)^2}{c^2_{n,1-a/2}} \right]$$

Εάν ο μέσος του δείγματος είναι άγνωστος τότε θα ισχύει ότι s^2 η διακύμανση τυχαίου δείγματος μεγέθους n από μια κανονική τυχαία μεταβλητή με διακύμανση s^2, θα είναι η τυχαία μεταβλητή

$$x^2 = \frac{(n-1)s^2}{s^2}$$

Ακολουθεί την κατανομή x^2 με ν=n-1 βαθμούς ελευθερίας. Συνεπώς το 100(1-α)% διάστημα εμπιστοσύνης για τη διακύμανση μπορεί να υπολογιστεί χρησιμοποιώντας τον πίνακα τιμών της x^2 κατανομής.

Έχουμε:

$$1-a = P(x^2_{n,1-a/2} \le \frac{(n-1)s^2}{s^{\,2}} \le x^2_{n,a/2}) \Rightarrow$$

$$1-a = P(\frac{(n-1)s^2}{x^2_{n,a/2}} \le s^{\,2} \le \frac{(n-1)s^2}{x^2_{n,1-a/2}})$$

Επομένως, αν σε ορισμένο δείγμα n στοιχείων υπολογίσουμε τη διακύμανση s^2, τότε εκτιμούμε ότι το διάστημα που θα περιέχει την διακύμανση πληθυσμού σ^2 με επίπεδο εμπιστοσύνης 1-α είναι το:

$$\left[\frac{(n-1)s^2}{x^2_{n,a/2}}, \frac{(n-1)s^2}{x^2_{n,1-a/2}}\right]$$

- **Διάστημα εμπιστοσύνης διαφοράς δύο μέσων αριθμητικών**

Η διαφορά των μέσων δειγμάτων$(\bar{x}_1 - \bar{x}_2)$ ακολουθεί την κανονική κατανομή με μέσο τη διαφορά των μέσων $(m_1 - m_2)$των πληθυσμών και διακύμανση ίση με:

$$s^2_{\bar{x}_1 - \bar{x}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad \text{και} \qquad s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

αν οι διακυμάνσεις s_1^2 και s_2^2 είναι γνωστές. Αναφερόμαστε σε μεγάλα δείγματα.

Διαφορετικά, αν οι διακυμάνσεις είναι άγνωστες, αρκεί τα n_1 και n_2 να είναι τυχαία και πολυπληθή. Σε αυτή την περίπτωση ισούται με:

$$s^2_{\bar{x}_1 - \bar{x}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad \text{και} \qquad s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Συνεπώς, το διάστημα εμπιστοσύνης των μέσων δύο πληθυσμών με γνωστές διασπορές είναι το εξής:

$$\left[\bar{x}_1 - \bar{x}_2 - z_{a/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{a/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

Αντίστοιχα με άγνωστες διασπορές:

$$\left[\bar{x}_1 - \bar{x}_2 - z_{a/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{a/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

Και το διάστημα εμπιστοσύνης με άγνωστες διασπορές και ίσες και δείγματα μικρά:

$$\left[\bar{x}_1 - \bar{x}_2 - t_{v,a/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{v,a/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

Όπου $v = n_1 + n_2 - 2$

Και το βρίσκουμε με τον τύπο:

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

Και τέλος, έχουμε την περίπτωση του διαστήματος εμπιστοσύνης με διασπορές άγνωστες και άνισες και τα δείγματα μικρά:

$$\left[\bar{x}_1 - \bar{x}_2 - t_{v,a/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{v,a/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

$$\text{Όπου} \quad \nu = \frac{1}{\dfrac{\left(\dfrac{s_1^2/n_1}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}\right)^2}{n_1 - 1} + \dfrac{\left[1 - \left(\dfrac{s_1^2/n_1}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}\right)\right]^2}{n_2 - 1}} \quad \text{για } s_1^2 > s_2^2$$

- **Διάστημα εμπιστοσύνης αναλογίας (ή ποσοστού)**

Πολλές φορές είναι αναγκαίο να εκτιμήσουμε ένα διάστημα εμπιστοσύνης για το άγνωστο ποσοστό P ενός πληθυσμού. Για να εκτιμήσουμε ένα τέτοιο ποσοστό παίρνουμε ένα τυχαίο δείγμα n και προσδιορίζουμε τις μονάδες του δείγματος (=f), οι οποίες έχουν ένα ορισμένο χαρακτηριστικό και προσδιορίζουμε αυτό το ποσοστό p=f/n. Το ποσοστό των μονάδων του δείγματος που δεν έχει το χαρακτηριστικό είναι q=1-p. Το τυπικό σφάλμα εκτιμήσεως της κατανομής δειγματοληψίας του ποσοστού P γνωρίζουμε ότι είναι $s_p = \sqrt{pq/n}$.

Το διάστημα εμπιστοσύνης εντός του οποίου αναμένεται να βρίσκεται το πραγματικό ποσοστό P του πληθυσμού καθορίζεται από τη διπλή ανισότητα:

$$\boldsymbol{p - Z_{a/2} S_p < P < p + Z_{a/2} S_p}$$

- **Διάστημα εμπιστοσύνης για τη διαφορά των αναλογιών (ή των ποσοστών)**

$$\hat{p}_1 - \hat{p}_2 - z_{a/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{a/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- **Διάστημα εμπιστοσύνης του λόγου των διασπορών δύο πληθυσμών**

Θεωρούμε δύο πληθυσμούς που ακολουθούν κανονική κατανομή με διακυμάνσεις, s_1^2 και s_2^2, και παίρνουμε τυχαία δείγματα μεγέθους n_1 και n_2 και διακυμάνσεις s_1^2 και s_2^2.

Αν έχουμε γνωστούς τους μέσους

$$\frac{1}{F_{n_1,n_2,a/2}}\frac{\frac{\Sigma(x_1-m_1)^2}{n_1}}{\frac{\Sigma(y_1-m_2)^2}{n_2}}<\frac{s_1^2}{s_2^2}<\frac{1}{F_{n_1,n_2,1-a/2}}\frac{\frac{\Sigma(x_1-m_1)^2}{n_1}}{\frac{\Sigma(y_1-m_2)^2}{n_2}}$$

Αν έχουμε άγνωστους τους μέσους

$$\frac{1}{F_{n_1,n_2,a/2}}\frac{\frac{\Sigma(x_1-\overline{x})^2}{n_1-1}}{\frac{\Sigma(y_1-\overline{y})^2}{n_2-1}}<\frac{s_1^2}{s_2^2}<\frac{1}{F_{n_1,n_2,1-a/2}}\frac{\frac{\Sigma(x_1-\overline{x})^2}{n_1-1}}{\frac{\Sigma(y_1-\overline{y})^2}{n_2-1}}$$

Από το οποίο συνεπάγεται ότι

$$\frac{1}{F_{n_1,n_2,a/2}}\frac{s_1^2}{s_2^2}<\frac{s_1^2}{s_2^2}<\frac{1}{F_{n_1,n_2,1-a/2}}\frac{s_1^2}{s_2^2}$$

ΕΛΕΓΧΟΙ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

- **Γενικά για τους ελέγχους στατιστικών υποθέσεων**

Ο έλεγχος υποθέσεων αναφέρεται στη διαδικασία ή στον τρόπο με τον οποίο καταλήγουμε στην αποδοχή ή στην απόρριψη μιας στατιστικής υποθέσεως. Με τον όρο στατιστική υπόθεση εννοούμε μια υπόθεση σχετικά με την κατανομή μιας τυχαίας μεταβλητής. Η υπόθεση συνήθως αφορά τις παραμέτρους της κατανομής (πληθυσμού), αλλά μπορεί επίσης να αναφέρεται στη φύση ή τη συναρτησιακή μορφή της κατανομής.

Αν η στατιστική υπόθεση καθορίζει ή μάλλον εξειδικεύει τελείως τη συναρτησιακή μορφή της κατανομής καθώς και τις τιμές των παραμέτρων, τότε ονομάζεται απλή υπόθεση (simple hypothesis). Στην αντίθετη περίπτωση, ονομάζεται σύνθετη υπόθεση (composite hypothesis).

Για τον έλεγχο μιας στατιστικής υποθέσεως θα πρέπει να υπάρχει και η εναλλακτική υπόθεση (alternative hypothesis). Με άλλα λόγια, εφόσον ο έλεγχος μιας υποθέσεως καταλήγει, βάσει κριτηρίων στην αποδοχή ή απόρριψη της υποθέσεως, για να έχει έννοια ο έλεγχος θα πρέπει να υπάρχει μια εναλλακτική υπόθεση για την περίπτωση που η ελεγχόμενη υπόθεση δεν γίνεται αποδεκτή. Επιπλέον, η εναλλακτική υπόθεση είναι αναγκαία και για την εξεύρεση των κατάλληλων στατιστικών κριτηρίων με τα οποία θα γίνει ο έλεγχος. Θα παριστάνουμε την εναλλακτική υπόθεση με H_1 και την ελεγχόμενη υπόθεση η οποία στη στατιστική αποκαλείται μηδενική υπόθεση (null hypothesis) με H_0.

Όπως η μηδενική υπόθεση έτσι και η εναλλακτική μπορεί να είναι απλή ή σύνθετη. Ας σημειωθεί ότι στην πράξη σπάνια έχουμε απλή μηδενική και απλή εναλλακτική υπόθεση. Συνήθως η μηδενική ή η εναλλακτική , ή και οι δύο είναι σύνθετες.

Ο καθορισμός της μηδενικής και της εναλλακτικής υποθέσεως αποτελεί το πρώτο στάδιο στη διαδικασία ελέγχου μιας στατιστικής υποθέσεως. Το επόμενο βήμα είναι η εξεύρεση ενός κανόνα ή κριτηρίου, σύμφωνα με το οποίο θα γίνει δεκτή ή θα απορριφθεί η μηδενική υπόθεση. Δηλαδή σύμφωνα με το οποίο θα αποφασισθεί αν οι πληροφορίες από το δείγμα ευνοούν την απόρριψη της μηδενικής υποθέσεως.

Στον έλεγχο των στατιστικών υποθέσεων κάνουμε μια υπόθεση, ότι η παράμετρος του πληθυσμού έχει μια ορισμένη δοσμένη τιμή και προσπαθούμε έχοντας ένα τυχαίο δείγμα το οποίο παίρνουμε από το σχετικό πληθυσμό να ελέγξουμε αν η υπόθεση που μας δίνεται είναι σωστή ή πρέπει να την απορρίψουμε, γιατί δεν συμβιβάζεται με τα δεδομένα ενός δείγματος. Σε αυτήν την περίπτωση εξετάζουμε το γεγονός κατά πόσο η διαφορά μεταξύ μιας παραμέτρου του δείγματος και της υποθετικής παραμέτρου του πληθυσμού είναι στατιστικά ασήμαντη, δηλαδή είναι διαφορά που μπορεί να αποδοθεί στις τυχαίες διακυμάνσεις της δειγματοληψίας ή είναι διαφορά στατιστικά σημαντική, δηλαδή πραγματική που δεν μπορεί να αποδοθεί μόνο στις διακυμάνσεις της τυχαίας δειγματοληψίας , αλλά σε σημαντικά αίτια και ακολούθως την υπόθεση αυτή πρέπει να την απορρίψουμε.

Αν μια διαφορά είναι τόσο μεγάλη, ώστε να μπορεί να συμβεί λόγω της δειγματοληψίας μόνο α φορές στις εκατό, τότε λέμε ότι η διαφορά είναι σημαντική σε επίπεδο σημαντικότητας α%.

Την διαφορά ανάμεσα στην τιμή της παραμέτρου του πληθυσμού, που την υποθέτουμε γνωστή και την τιμή του δείγματος, την ελέγχουμε με διάφορα κριτήρια. Αν δεν υπάρχει μεγάλη διαφορά σύμφωνα με το κριτήριο που χρησιμοποιούμε, τότε την υπόθεση αυτή την αποδεχόμαστε, διαφορετικά την απορρίπτουμε.

Την περισσότερο συνηθισμένη στατιστική υπόθεση την αποκαλούμε υπόθεση μηδενική και τη συμβολίζουμε με H_0. Στην υπόθεση αυτή, από την αρχή θεωρείται γνωστό, ότι η διαπιστωμένη διαφορά μεταξύ της παραμέτρου του δείγματος και του πληθυσμού είναι στατιστικά ασήμαντη, δηλαδή αν δεν υπήρχαν τα σφάλματα της δειγματοληψίας, τότε οι δύο παράμετροι θα είχαν την ίδια τιμή και επομένως η διαφορά τους θα ήταν μηδέν. Επίσης μια άλλη συνηθισμένη διατύπωση της μηδενικής υπόθεσης είναι ότι η άγνωστη παράμετρος θ του πληθυσμού είναι ίση με μια υποθετική τιμή θ_0 και συμβολίζεται ως εξής:

$H_0 : \theta = \theta_0$

Η ονομασία μηδενική υπόθεση οφείλεται στο γεγονός ότι συνήθως η μηδενική υπόθεση δηλώνει ότι η διαφορά μεταξύ θ και θ_0 είναι μηδενική.

Σε κάθε μηδενική υπόθεση αντιστοιχεί και μια άλλη υπόθεση, η οποία ονομάζεται εναλλακτική υπόθεση και συμβολίζεται με H_1. Η απόρριψη της υπόθεσης H_0 συνεπάγεται την αποδοχή της H_1 και αντίστροφα.

Στην υπόθεση H_0, συνήθως, θέτουμε την πρόταση ότι η τιμή της παραμέτρου θ ισούται με μια ορισμένη τιμή θ_0, δηλαδή:

$H_0 : \theta = \theta_0$

Στη συνέχεια θέτουμε την εναλλακτική υπόθεση H_1 που μπορεί να έχει τις εξής μορφές:

α) $H_1 : \theta \neq \theta_0$, που σημαίνει ότι η τιμή της παραμέτρου του πληθυσμού είναι διαφορετική από τη γνωστή τιμή θ_0, είτε μικρότερη, είτε μεγαλύτερη και ο έλεγχος αυτός λέγεται δίπλευρος.

β) $H_1 : \theta < \theta_0$ ή $H_1 : \theta > \theta_0$. Ο έλεγχος στην περίπτωση αυτή ονομάζεται μονόπλευρος.

Είναι δυνατόν η στατιστική υπόθεση H_0 να έχει και τη μορφή:

1) $\theta = \theta_0$

2) $\theta \geq \theta_0$ και

3) $\theta \leq \theta_0$

Οι εναλλακτικές υποθέσεις H_1 που αντιστοιχούν αντίστοιχα στις τρεις παραπάνω μηδενικές υποθέσεις είναι:

1) $\theta \neq \theta_0$,

2) $\theta < \theta_0$ και

3) $\theta > \theta_0$

Η εκάστοτε μορφή της κατάλληλης μηδενικής H_0 και της εναλλακτικής υπόθεσης H_1 εξαρτάται από τα δεδομένα του προβλήματος. Η μαθηματική διατύπωση του ελέγχου υποθέσεων έχει ως εξής:

Ας υποθέσουμε ότι έχουμε έναν πληθυσμό που έχει συνάρτηση πυκνότητας πιθανότητας f (x, θ).

Η άγνωστη παράμετρος θ του πληθυσμού, έστω ότι έχει πεδίο ορισμού το θ. Το σύνολο των δυνατών τιμών της παραμέτρου θ, ονομάζεται συνήθως παραμετρικός χώρος. Π.χ. ο παραμετρικός χώρος του μέσου μ ενός κανονικού πληθυσμού είναι το διάστημα $(-\infty, \infty)$, ενώ ο παραμετρικός χώρος της παραμέτρου p μιας διωνυμικής κατανομής θα είναι (0,1). Από τον παραπάνω πληθυσμό $\chi \sim f(\chi, \theta)$ παίρνουμε ένα τυχαίο δείγμα $X_1, X_2, \ldots\ldots, X_n$. Από το δείγμα αυτό παίρνουμε μια εκτίμηση $\hat{q}$ της παραμέτρου θ, δηλαδή:

$$C2 = m_0 + t_{n,a}\frac{S}{\sqrt{n}}\quad \hat{q} = \theta(\chi1,\chi2,\ldots,\chi n)$$

Στη συνέχεια, με τη βοήθεια της τιμής $\hat{q}$ χρειάζεται να αποφασίσουμε αν η παράμετρος θ του πληθυσμού ανήκει στο υποσύνολο θ_0 ή στο υποσύνολο θ_1, όπου θ_0 και θ_1 είναι δύο υποσύνολα του θ (παραμετρικού χώρου) και είναι ασυμβίβαστα μεταξύ τους, δηλαδή:

$$\theta_0 \cap \theta_1 = \emptyset \text{ και } \theta_0 + \theta_1 = \theta$$

που σημαίνει ότι η πραγματοποίηση του ενός αποκλείει την πραγματοποίηση του άλλου.

Για να αποφασίσουμε κατά πόσο η παράμετρος του πληθυσμού θ ανήκει στο υποσύνολο θ_0 ή στο υποσύνολο θ_1, κάνουμε μια υπόθεση και με τη βοήθεια ενός κριτηρίου συμπεραίνουμε αν πρέπει να γίνει δεκτή η υπόθεση αυτή ή να απορριφθεί.

Την υπόθεση που κάνουμε, ότι η παράμετρος θ ανήκει στο υποσύνολο θ_0, την καλούμε υπόθεση μηδέν και, όπως είπαμε, τη συμβολίζουμε με H_0, ενώ την αντίθετη προς αυτή υπόθεση τη λέμε εναλλακτική υπόθεση και τη συμβολίζουμε με H_1.

Επομένως μια στατιστική υπόθεση μπορούμε να τη συμβολίζουμε ως εξής:

$H_0 : \theta \in \theta_0$

$H_1 : \theta \in \theta_1$

Αν το υποσύνολο θ_0 ή θ_1 αποτελείται από ένα μόνο σημείο, τότε την υπόθεση H_0 ή H_1 τη λέμε απλή, διαφορετικά την ονομάζουμε σύνθετη. Για παράδειγμα:

$H_0 : \theta = \theta_0$ (απλή)

$H_1 : \theta > \theta_0$ (σύνθετη)

$H_0 : \theta \geq \theta_0$ (σύνθετη)

$H_1 : \theta < \theta_0$ (σύνθετη)

Από την παραπάνω ανάλυση παρατηρούμε ότι ο έλεγχος μιας στατιστικής υπόθεσης είναι μια διαδικασία που διαιρεί το δειγματικό χώρο Ω (όπου δειγματικός χώρος είναι το σύνολο των δυνατών αποτελεσμάτων μιας δειγματοληψίας) σε δύο ξένα μεταξύ τους υποσύνολα Α και Κ και με τρόπο ώστε, αν συμβαίνει η εκτίμηση $\hat{q}$ που προκύπτει από το δείγμα να ανήκει στο Κ ($\hat{q}$ Є Α), απορρίπτουμε την υπόθεση H_0, ενώ αν το $\hat{q}$ ανήκει στο Α($\hat{q}$ Є Α), δεχόμαστε την υπόθεση H_0.

Το υποσύνολο Κ το λέμε περιοχή απόρριψης ή κρίσιμη περιοχή, ενώ το υποσύνολο Α το λέμε περιοχή αποδοχής

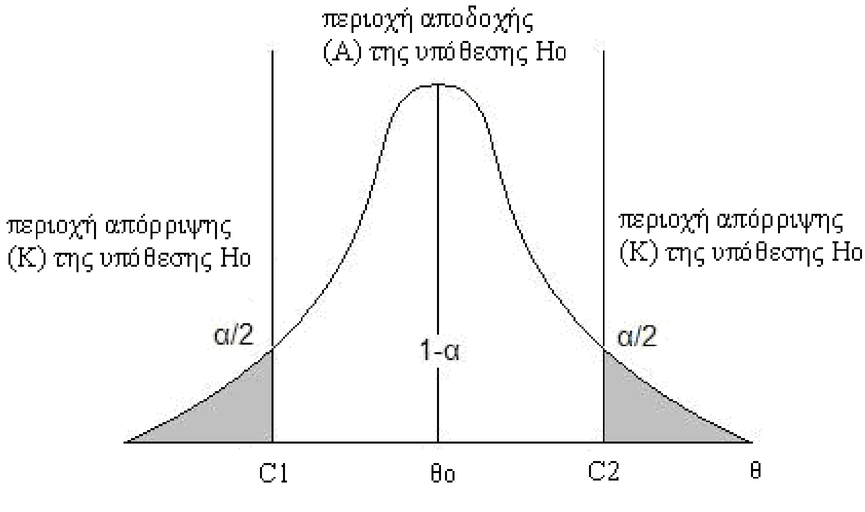
Προκειμένου να ελέγξουμε την υπόθεση $H_0 : \theta = \theta_0$ έναντι της H_1: $\theta \neq \theta_0$, παίρνουμε ένα συγκεκριμένο δείγμα χ1,χ2,χ3,….χn από τον πληθυσμό που έχει παράμετρο θ. Με τη βοήθεια των παραπάνω παρατηρήσεων του δείγματος υπολογίζουμε μια εκτίμηση $\hat{q}$ της παραμέτρου θ. Αν η $\hat{q}$ που προέκυψε από το δείγμα βρεθεί μεταξύ του διαστήματος που ορίζουν οι ακραίες τιμές C1 (κατώτερο όριο) και C2 (ανώτερο όριο) της περιοχής αποδοχής , δηλαδή $C1 \leq \hat{q} \leq C2$, δεχόμαστε την υπόθεση, ενώ αν $\hat{q}$<C1 ή $\hat{q}$>C2, απορρίπτουμε την υπόθεση.
Η περιοχή που ορίζεται από τις σχέσεις $\hat{q}$<C1 και $\hat{q}$>C2 είναι η κρίσιμη περιοχή.

Οι κριτικές τιμές C1 και C2, αποτελούν το κατώτερο και το ανώτερο όριο του διαστήματος αποδοχής μιας υπόθεσης, καθορίζονται με βάση μια ορισμένη από την αρχή πιθανότητα, που ονομάζεται επίπεδο σημαντικότητας α και ισχύει η σχέση:

$P\{\hat{q}<C1\} + P\{\hat{q}>C2\}=\alpha$

Ανεξάρτητα όμως από το επίπεδο σημαντικότητας α η αποδοχή μιας υπόθεσης ή η απόρριψή της στηρίζεται στο μέγεθος της διαφοράς $\hat{q}-\theta_0$, Το κριτήριο όμως αυτό δεν μας κατοχυρώνει πάντοτε. Γιατί είναι δυνατόν, ενώ η εκτίμηση $\hat{q}$ που προέκυψε από το δείγμα να απέχει πολύ από την τιμή θ_0, στην πραγματικότητα η τιμή θ του πληθυσμού να είναι ίση με θ_0. Επίσης, μπορεί να συμβεί η $\hat{q}$ να βρίσκεται κοντά στη θ_0 αλλά η παράμετρος θ του πληθυσμού να είναι διαφορετική από την τιμή θ_0 που ελέγχουμε. Στις περιπτώσεις αυτές ο έλεγχος στατιστικών υποθέσεων οδηγεί σε εσφαλμένα συμπεράσματα.

Επομένως όταν αποδεχόμαστε ή απορρίπτουμε μια μηδενική υπόθεση H_0, με τη βοήθεια ενός τυχαίου δείγματος, είναι δυνατόν να κάνουμε δύο ειδών σφάλματα. Με άλλα λόγια, αν το δείγμα πέσει στην περιοχή του Α, δεχόμαστε την υπόθεση H_0 (βλ. παρακάτω διάγραμμα 1.1)

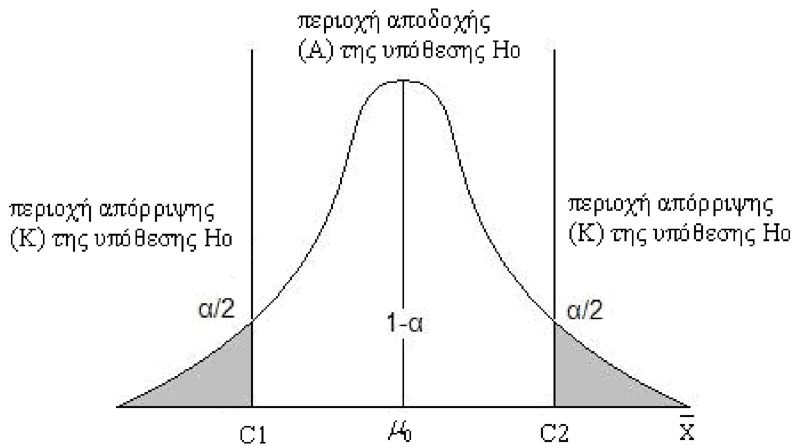


Διάγραμμα 1.1

Αν π.χ. πάρουμε σαν εκτιμητή του μέσου αριθμητικού μ το μέσο του δείγματος $\bar{x}$, τότε απορρίπτουμε την H_0, όταν:

$$\bar{x} \le C1 = \mu_0 - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad ή \quad \bar{x} \ge C2 = \mu_0 + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

και αποδεχόμενοι την H_0, όταν $C1 < \bar{x} < C2$, έχουμε επιτύχει, όπως φαίνεται στο διάγραμμα 1.2 , το διαμερισμό του συνόλου των τιμών $\bar{x}$ στην περιοχή απόρριψης και στην περιοχή αποδοχής.

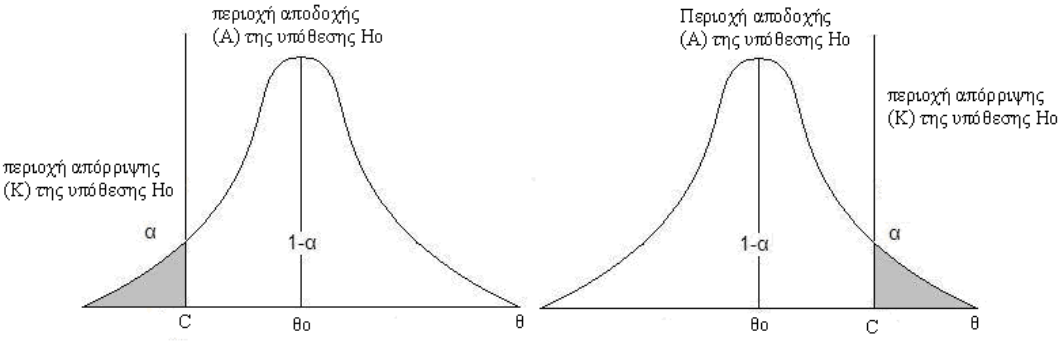


Διάγραμμα 1.2

Εάν ο έλεγχος αναφέρεται σε μονόπλευρο έλεγχο της μορφής:

$$\begin{matrix} Ho: \theta=\theta o \\ H1: \theta<\theta o \end{matrix} \quad και \quad \begin{matrix} Ho: \theta=\theta o \\ H1: \theta>\theta o \end{matrix}$$

θα έχουμε τις περιοχές αποδοχής και απόρριψης του διαγράμματος 1.3



Διάγραμμα 1.3

- **Είδη σφαλμάτων**

Κατά τον έλεγχο μιας στατιστικής υπόθεσης είναι δυνατόν να γίνουν, όπως αναφέραμε, δύο ειδών σφάλματα.

α) Σφάλμα πρώτου είδους

Ονομάζεται το σφάλμα που κάνουμε όταν απορρίπτουμε τη μηδενική υπόθεση H_0, ενώ η υπόθεση αυτή είναι αληθινή. Η πιθανότητα να κάνουμε σφάλμα πρώτου είδους συμβολίζεται ως εξής:

α=P{απορρίπτουμε την H_0, ενώ η H_0 είναι ορθή}

Η πιθανότητα α να απορρίψουμε εσφαλμένα μια υπόθεση H_0 ονομάζεται επίπεδο σημαντικότητας.

β) Σφάλμα δεύτερου είδους

Ονομάζεται το σφάλμα που κάνουμε, όταν δεχόμαστε τη μηδενική υπόθεση H_0 ενώ αυτή δεν είναι αληθινή. Η πιθανότητα να κάνουμε σφάλμα δεύτερου είδους συμβολίζεται ως εξής:

β=P{αποδοχή H_0 ενώ η H_0 είναι εσφαλμένη}

ή

β=P{αποδοχή H_0 ενώ η H_1 είναι ορθή}

Δηλαδή η πιθανότητα β είναι ο κίνδυνος να κάνουμε αποδεκτή την εσφαλμένη υπόθεση H_0. Και στις δύο παραπάνω περιπτώσεις θα πάρουμε λανθασμένη απόφαση.

Εκτός από τις παραπάνω πιθανότητες α και β μπορούμε να υπολογίσουμε και την πιθανότητα να απορρίψουμε τη μηδενική υπόθεση H_0, όταν αυτή είναι πράγματι εσφαλμένη. Η πιθανότητα αυτή ονομάζεται δύναμη ή ισχύ του στατιστικού ελέγχου.

Η δύναμη του στατιστικού ελέγχου συμβολίζεται ως εξής:

γ=1-β=P{Απορρίπτουμε την H_0, γιατί η H_1 είναι ορθή}=

=P{απόρριψη της H_0 / H_0 εσφαλμένη}

Όσο πιο μεγάλη είναι η τιμή της ισχύος του ελέγχου, τόσο πιο σωστή θα είναι η απόφασή μας.

Στον έλεγχο στατιστικών υποθέσεων επιδιώκεται η σχετική ελαχιστοποίηση του α και η σχετική μεγιστοποίηση της δύναμης γ.

- **Έλεγχος υποθέσεων του μέσου αριθμητικού**

Έστω ότι πήραμε ένα τυχαίο δείγμα X1,X2,….,Xn από έναν πληθυσμό που ακολουθεί την κανονική κατανομή ή οποιαδήποτε κατανομή, με την προϋπόθεση όμως 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 διαπιστώσουμε και κατά πόσο αυτή η υπόθεση είναι στατιστικά αποδεκτή.

Οι συνηθισμένες υποθέσεις για τον έλεγχο του μέσου αριθμητικού μ του πληθυσμού είναι οι παρακάτω:

1) H_0: $\mu=\mu_0$, έναντι της εναλλακτικής υπόθεσης ότι H1: $\mu>\mu_0$. Αυτό σημαίνει ότι κάνουμε την υπόθεση πως έχουμε λόγους να πιστεύουμε ότι η άγνωστη παράμετρος μ του πληθυσμού είναι ίση με μια γνωστή τιμή μ_0. Η ισχύς της παραπάνω υπόθεσης ελέγχεται σε σχέση με την υπόθεση ότι ο μέσος αυξήθηκε, δηλαδή H1: $\mu>\mu_0$.

2) H_0: $\mu=\mu_0$, έναντι H1: $\mu<\mu_0$

3) H_0: $\mu=\mu_0$, έναντι H1: $\mu\neq\mu_0$

Για τον έλεγχο της παραμέτρου μ διακρίνουμε δύο περιπτώσεις:

- **Η διακύμανση του πληθυσμού είναι γνωστή και ο πληθυσμός ακολουθεί την κανονική κατανομή, ή το δείγμα έχει μέγεθος n>30 ανεξάρτητα από τη μορφή της κατανομής, ώστε να μπορούμε να εφαρμόσουμε το κεντρικό οριακό θεώρημα.**

Γνωρίζουμε ότι, αν το τυχαίο δείγμα προέχεται από κανονικό πληθυσμό ή αν αυτό 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_0.

Περίπτωση 1η :

Έλεγχος της υπόθεσης:

H_0: $\mu=\mu_0$

έναντι της εναλλακτικής:

H_1: $\mu>\mu_0$

σε επίπεδο σημαντικότητας α.

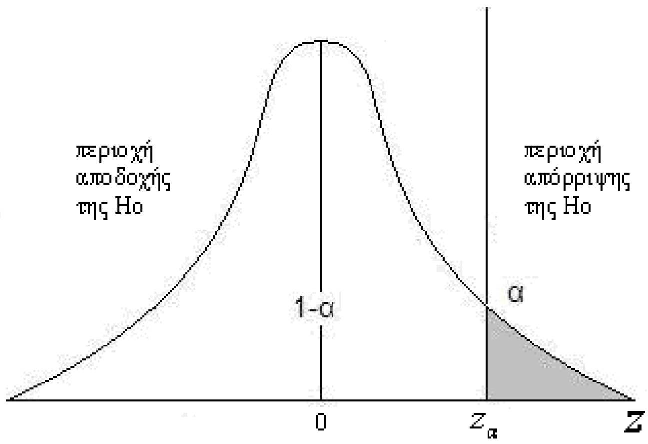
Στην περίπτωση αυτή ο έλεγχος ονομάζεται μονόπλευρος προς τα πάνω και το επίπεδο σημαντικότητας α τοποθετείται στο δεξιό άκρο της κανονικής καμπύλης.

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_0 και την αποδοχή της H_1.

Η υπόθεση H_0 απορρίπτεται, αν ισχύει η ανισότητα:

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

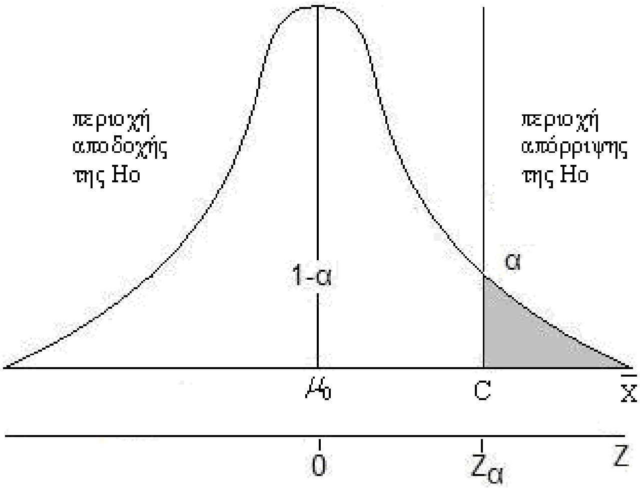
Την τιμή Z_α την παίρνουμε από τον πίνακα της κανονικής κατανομής.



Διάγραμμα 1.4

Άλλο κριτήριο αποδοχής ή απόρριψης της υπόθεσης H_0, ισοδύναμο με το προηγούμενο, είναι να βρούμε μια ακραία τιμή C που χωρίζει την περιοχή αποδοχής από την περιοχή απόρριψης. Την κριτική αυτή τιμή C τη βρίσκουμε από τη σχέση $P\{\bar{x} \geq C\} = \alpha$, από την οποία αποδεικνύεται ότι είναι:

$$C = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_0 + Z_\alpha \sigma_{\bar{x}}$$



Διάγραμμα 1.5

Αν ο μέσος αριθμητικός $\bar{x}$ του δείγματος είναι $\bar{x} > \mu_0 + Z_\alpha \sigma_{\bar{x}}$ απορρίπτουμε την υπόθεση. Δηλαδή αν ο μέσος αριθμητικός πάρει τιμή από $-\infty$ μέχρι C, δεχόμαστε την υπόθεση H_0, ενώ αν $\bar{x}>C$ απορρίπτουμε την υπόθεση.

Παραδείγματα:

1η περίπτωση

Έχει διαπιστωθεί έπειτα από παρατηρήσεις πολλών ετών, ότι μέχρι το έτος 1987 στις εξετάσεις στο μάθημα της στατιστικής σε μια σχολή, η μέση βαθμολογία των σπουδαστών ήταν 87 μονάδες, με τυπική απόκλιση σ=9 μονάδες. Στις εξετάσεις της περιόδου Ιανουαρίου του τρέχοντος 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; (α=5%)

H_0: μ=87

έναντι

H_1: μ>87

Επειδή η διακύμανση του πληθυσμού(σ^2) είναι γνωστή, θα χρησιμοποιήσουμε το κριτήριο:

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

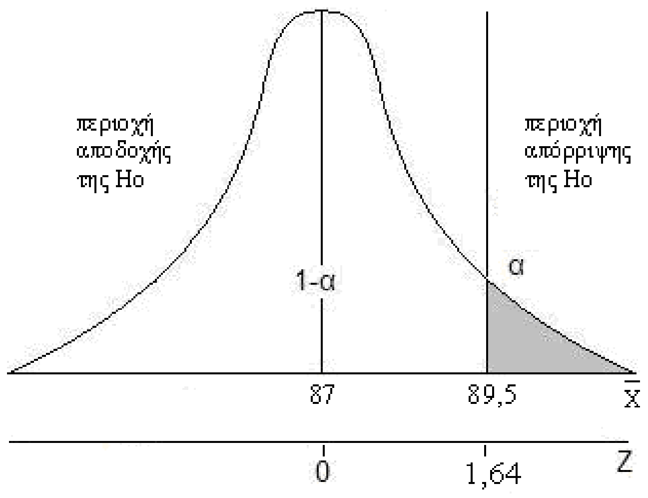
και θα έχουμε:

$$z = \frac{90-87}{\frac{9}{\sqrt{36}}} = \frac{3}{\frac{9}{6}} = \frac{3}{1,5} = 2$$

Από τους πίνακες της κανονικής κατανομής βρίσκουμε $z_\alpha = z_{5\%} = 1,64$

Επειδή $z > z_\alpha$, απορρίπτεται η υπόθεση ότι η βαθμολογία έχει παραμείνει σταθερή. Άρα αυξήθηκε και επομένως, η H_0 απορρίπτεται.

$$C = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}} = 87+1,645\,\frac{9}{6} = 89,5$$



$\bar{X} = 90$ και $C = 89,5$

Επειδή $\bar{x}$>C απορρίπτουμε την υπόθεση H_0

Περίπτωση 2η:

Έλεγχος της υπόθεσης:

H_0: $\mu=\mu_0$

έναντι της:

H_1: $\mu<\mu_0$

σε επίπεδο σημαντικότητας α.

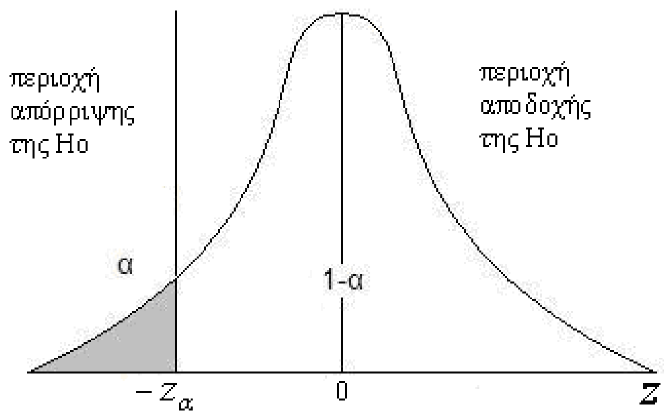
Κατά την περίπτωση αυτή ο έλεγχος ονομάζεται μονόπλευρος προς τα κάτω και το επίπεδο σημαντικότητας α βρίσκεται στο αριστερό άκρο της κανονικής καμπύλης.

Η παραπάνω υπόθεση H_0 απορρίπτεται, αν ισχύει η ανισότητα:

$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_\alpha$$

και γίνεται δεκτή αν $Z > -Z_\alpha$.

Την τιμή $-Z_\alpha$ την παίρνουμε από τον πίνακα της κανονικής κατανομής .



Διάγραμμα 1.6

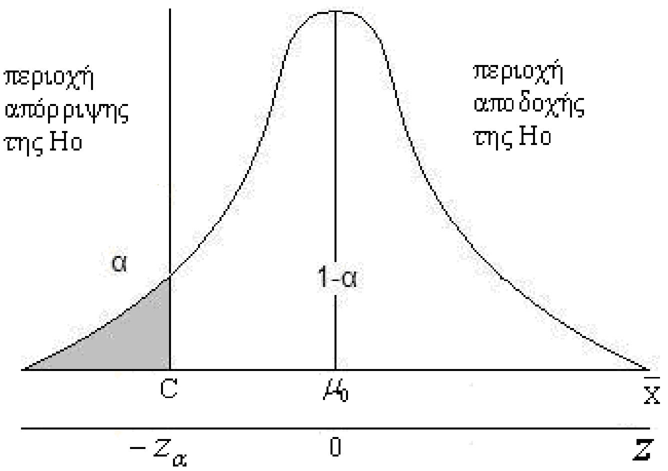
Άλλο κριτήριο αποδοχής ή απόρριψης της υπόθεσης H_0 στην παραπάνω περίπτωση, ισοδύναμο με το προηγούμενο, είναι η εύρεση μιας ακραίας τιμής C που χωρίζει την περιοχή αποδοχής από την περιοχή απόρριψης. Την κριτική αυτή τιμή C τη βρίσκουμε από τη σχέση $P\{\bar{x}<C\}=\alpha$, από την οποία έχουμε:

$$C=\mu_0-Z_\alpha \frac{\sigma}{\sqrt{n}}=\mu_0-Z_\alpha \sigma_{\bar{x}}$$

Αν ο μέσος αριθμητικός πάρει τιμή:

$$\bar{x}<C=\mu_0-Z_\alpha \sigma_{\bar{x}}$$

απορρίπτουμε την υπόθεση H_0, ενώ αν $\bar{x}>C$ δεχόμαστε την υπόθεση H_0.



Διάγραμμα 1.7

Ο παρακάτω πίνακας μας δίνει τα κυριότερα επίπεδα σημαντικότητας που χρησιμοποιούμε στην πράξη και τις αντίστοιχες κριτικές τιμές της μεταβλητής Ζ για δίπλευρους και μονόπλευρους ελέγχους.

Επίπεδο σημαντικότητας α	0,01	0,05	0,10
Κριτικές τιμές της Ζ για μονόπλευρο έλεγχο	-2,33 +2,33	-1,645 +1,645	-1,28 +1,28
Κριτικές τιμές της Ζ για δίπλευρο έλεγχο	-2,58 +2,58	-1,96 +1,96	-1,645 +1,645

Παράδειγμα:

Μια βιομηχανία ηλεκτρικών λαμπτήρων επιθυμεί να προσδιορίσει το μέσο όρο ζωής των παραγόμενων λαμπτήρων ορισμένου τύπου. Η διάρκεια ζωής των λαμπτήρων ακολουθεί την κανονική κατανομή. Οι υπεύθυνοι παίρνουν ένα τυχαίο δείγμα 100 λαμπτήρων και βρίσκουν $\bar{x}$=1570 ώρες. Να ελεγχθεί σε επίπεδο σημαντικότητας α=5% η υπόθεση ότι η μέση διάρκεια ζωής των λαμπτήρων είναι μικρότερη από 1600 ώρες αν γνωρίζουμε ότι σ=120 ώρες (α=5%).

H_0: μ=1600

έναντι

H_1: μ<1600

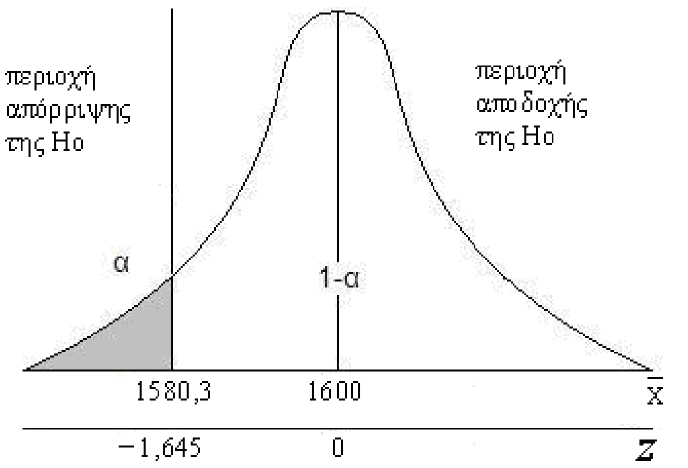
$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1570-1600}{\frac{120}{\sqrt{100}}} = \frac{-30}{12} = -2,5$$

Για α=5% έχουμε $-z_\alpha = -1,64$

Βλέπουμε ότι $Z < -Z_\alpha$ αφού -2,5<-1,64, άρα απορρίπτουμε την υπόθεση H_0.

$$C = \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} = 1600 - 1,64\, \frac{120}{\sqrt{100}} = 1580,3$$

Επειδή $\bar{x}$<C αφού 1570<1580,3 η υπόθεση H_0 απορρίπτεται και γίνεται δεκτή η υπόθεση H1:μ<1600, δηλαδή ότι η μέση διάρκεια ζωής των λαμπτήρων είναι μικρότερη από 1600 ώρες.



Περίπτωση 3η :

Έλεγχος της υπόθεσης:

H_0: $\mu=\mu_0$

έναντι της:

H_1: $\mu\neq\mu_0$

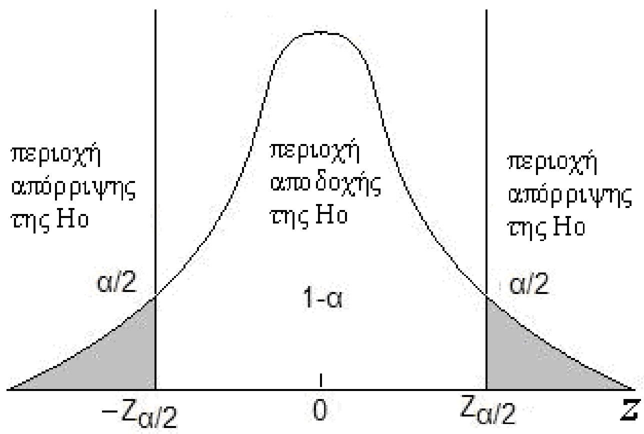
σε επίπεδο σημαντικότητας α.

Σε αυτήν την περίπτωση ο έλεγχος ονομάζεται δίπλευρος και το επίπεδο σημαντικότητας α ισοκατανέμεται στο δεξιό και αριστερό άκρο της καμπύλης της τυπικής κανονικής κατανομής.

Η περιοχή απόρριψης για το επίπεδο σημαντικότητας α ορίζεται από τις σχέσεις:

$$\boldsymbol{z} = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{\alpha/2} \quad \text{ή} \quad \boldsymbol{z} = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{\alpha/2}$$

και η υπόθεση γίνεται δεκτή αν ισχύει: $-z_{\alpha/2} < \boldsymbol{z} < z_{\alpha/2}$



Διάγραμμα 1.8

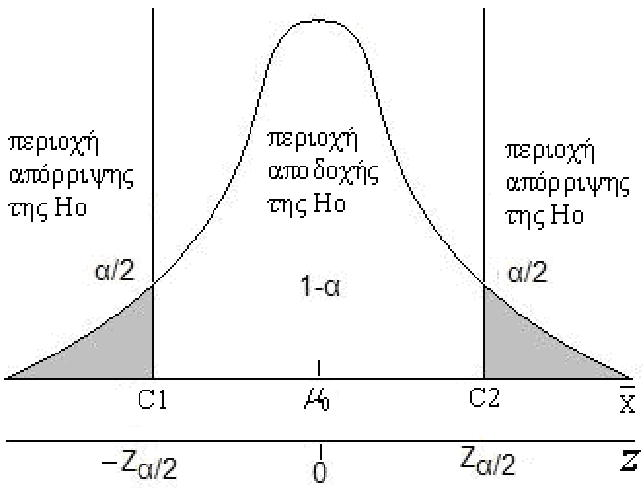
Άλλο κριτήριο αποδοχής ισοδύναμο με το προηγούμενο είναι η εύρεση δύο ακραίων τιμών C1 και C2 που ορίζουν την περιοχή αποδοχής και απόρριψης. Για το σκοπό αυτό χρησιμοποιούμε τη σχέση:

$P\{C1 < \bar{x} < C2\} = 1-\alpha$

από την οποία τελικά βρίσκουμε:

$$C1 = \mu_0 - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{και} \quad C2 = \mu_0 + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Αν ο μέσος $\bar{x}$ περιλαμβάνεται στο διάστημα (C1,C2), η υπόθεση γίνεται δεκτή, ενώ, αν η τιμή του $\bar{x}$ βρίσκεται εκτός του διαστήματος, η υπόθεση απορρίπτεται και δεχόμαστε την εναλλακτική υπόθεση.



Διάγραμμα 1.9

Παράδειγμα:

Το τμήμα πωλήσεων μιας εταιρείας θέλει να ελέγξει αν ο μέσος όρος πωλήσεων του προϊόντος της είναι 3500 κομμάτια την εβδομάδα ή περισσότερα. Σε 36 τυχαίες εβδομάδες διαπιστώθηκε ότι ο μέσος όρος πωλήσεων του προϊόντος ήταν 4000 κομμάτια εβδομαδιαίως με τυπική απόκλιση σ = 300 κομμάτια. Να ελεγχθεί η παραπάνω υπόθεση σε επίπεδο σημαντικότητας α=5%.

H_0: μ=μο=4000

έναντι

H_1: μ≠4000

$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{3500\text{-}4000}{\frac{300}{\sqrt{36}}} = \frac{-500}{50} = -10$$

Οι κριτικές τιμές για δίπλευρο έλεγχο και επίπεδο σημαντικότητας α=5% είναι:

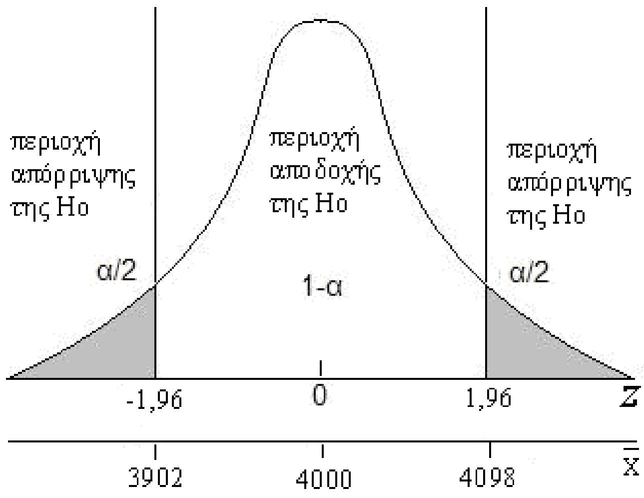
$$-Z_{\alpha/2} = -1,96 \quad \text{και} \quad Z_{\alpha/2} = 1,96$$

Βλέπουμε ότι $z < -Z_{\alpha/2}$ αφού -10<-1,96, άρα απορρίπτουμε την υπόθεση H_0 και γίνεται δεκτή η υπόθεση H1:μ≠4000

$$C1 = \mu_0 - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 4000 - 1,96\,\frac{300}{\sqrt{36}} = 4000 - \frac{588}{6} = 4000\text{-}98 = 3902$$

$$C2 = \mu_0 + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 4000 + 1,96\,\frac{300}{\sqrt{36}} = 4000 + \frac{588}{6} = 4000+98=4098$$

Επειδή $\bar{x}$=3500<C1=3902 απορρίπτουμε την υπόθεση γιατί η $\bar{x}$ είναι μικρότερη από την τιμή C1 που είναι το κατώτατο όριο αποδοχής.


περιοχή
απόρριψης
της Ho
περιοχή
αποδοχής
της Ho
περιοχή
απόρριψης
της Ho
α/2
1-α
α/2
-1,96
0
1,96
Z
3902
4000
4098
X̄

- **Η διακύμανση του πληθυσμού είναι άγνωστη και το μέγεθος του δείγματος είναι n<30**

Τότε η υπόθεση H_0 των τριών περιπτώσεων που αφορούν το μέσο αριθμητικό, ελέγχεται με το κριτήριο:

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

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

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

η διακύμανση του δείγματος.

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

(1) Η υπόθεση H_0 στην περίπτωση:

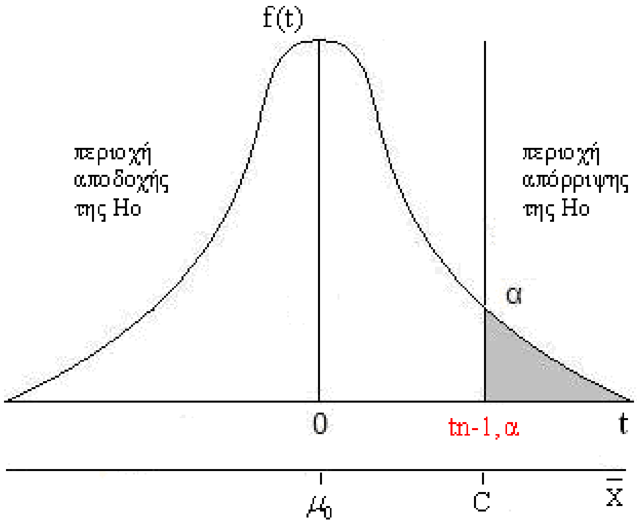
H_0: $\mu=\mu_0$

H_1: $\mu>\mu_0$

γίνεται δεκτή αν:

$t < t_{\nu,\alpha}$

ή αν $\bar{X} < C = \mu_0 + t_{\nu,\alpha} \frac{s}{\sqrt{n}}$



Διάγραμμα 1.10

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές: 42, 52, 40, και 74. Να ελεγχθεί η υπόθεση:

$$H_o: \mu=50$$

vs

$$H_1: \mu>50$$

σε επίπεδο σημαντικότητας α=5%.

	xi	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
	42	-10	100
	52	0	0
	40	-12	144
	74	22	484
Σύνολο	208		728

$$\bar{x} = \frac{\sum xi}{N} = \frac{208}{4} = 52$$

$$s^2 = \frac{\sum (xi - \bar{x})^2}{n-1} = \frac{728}{3} = 242{,}66 \quad \text{και} \quad S = \sqrt{242{,}66} = 15{,}57$$

όπου ν=n-1=4-1=3 οι βαθμοί ελευθερίας.

Για τον έλεγχο της υπόθεσης H_o:μ=50 έναντι της υπόθεσης H_1:μ>50, επειδή το μέγεθος του δείγματος είναι μικρό και η διακύμανση άγνωστη, θα χρησιμοποιήσουμε το εξής κριτήριο:

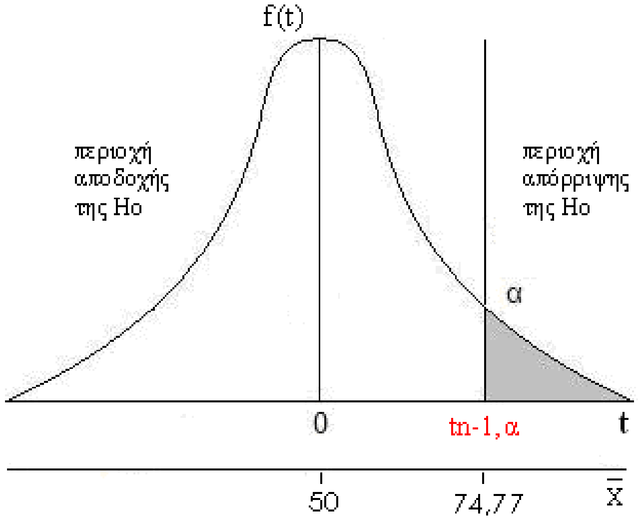
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{52-50}{\frac{15{,}57}{\sqrt{4}}} = \frac{2}{7{,}78} = 0{,}257$$

$t_{\nu,\alpha} = t_{3,0.05} = 3{,}182$

Βλέπουμε ότι $t < t_{\nu,\alpha}$ αφού 0,257<3,182, άρα η υπόθεση γίνεται δεκτή.

$$C = \mu_0 + t_{\nu,\alpha}\frac{s}{\sqrt{n}} = 50 + 3{,}182\frac{15{,}57}{\sqrt{4}} = 50 + 24{,}77 = 74{,}77$$

Επειδή $\bar{x} = 52 < C = 74{,}77$ η υπόθεση γίνεται δεκτή.



Διάγραμμα 1.10

(2) Η υπόθεση H_0 στην περίπτωση:

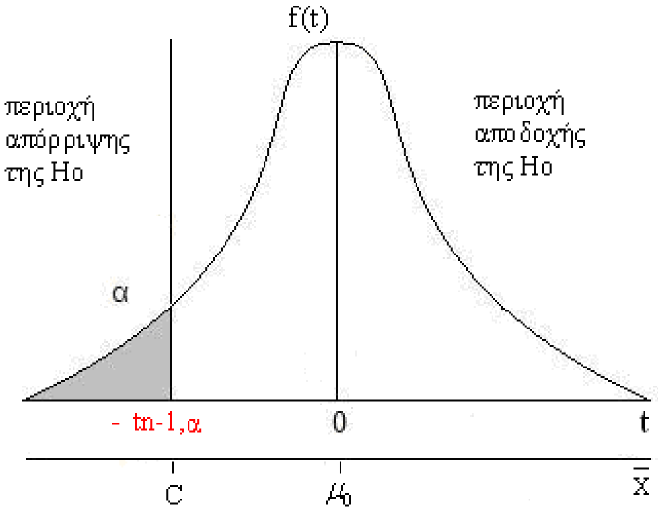
H_0: $\mu=\mu_0$

H_1: $\mu<\mu_0$

γίνεται δεκτή αν:

$-t_{\nu,\alpha} < t$

ή αν $\overline{X} > C = \mu_0 - t_{\nu,\alpha}\dfrac{s}{\sqrt{n}}$



Διάγραμμα 1.11

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές 4, 3 ,7, και 8 . Να ελεγχθεί η υπόθεση

$$H_0: \mu=10$$

vs

$$H_1: \mu<10$$

σε επίπεδο σημαντικότητας α=5%.

$$\bar{x} = \frac{\sum xi}{N} = \frac{22}{4} = 5,5$$

$$s^2 = \frac{\sum (xi - \bar{x})^2}{n-1} = \frac{(4-5,5)^2+(3-5,5)^2+(7-5,5)^2+(8-5,5)^2}{3} = \frac{2,25+6,25+2,25+6,25}{3} = \frac{17}{3} = 5,67$$

$$\text{και} \quad S = \sqrt{5,67} = 2,38$$

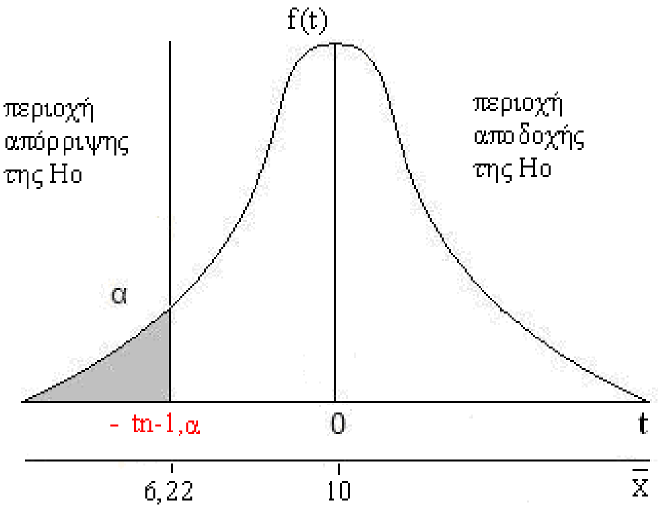
$$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{5,5-10}{\frac{2,38}{\sqrt{4}}} = -3,78$$

$- t_{\nu,\alpha} = - t_{3,0.05} = -3,182$

Βλέπουμε ότι $t < - t_{\nu,\alpha}$ αφού -3,78<-3,182, άρα η υπόθεση H_0 απορρίπτεται.

$$C = \mu_0 - t_{\nu,\alpha}\frac{s}{\sqrt{n}} = 10-3,182\frac{2,38}{\sqrt{4}} = 10 - 3,78 = 6,22$$

Επειδή $\bar{x} = 5,5 < C = 6,22$ η υπόθεση H_0 απορρίπτεται.



Διάγραμμα 1.11

(3) Η υπόθεση H_0 στην περίπτωση:

H_0: $\mu=\mu_0$

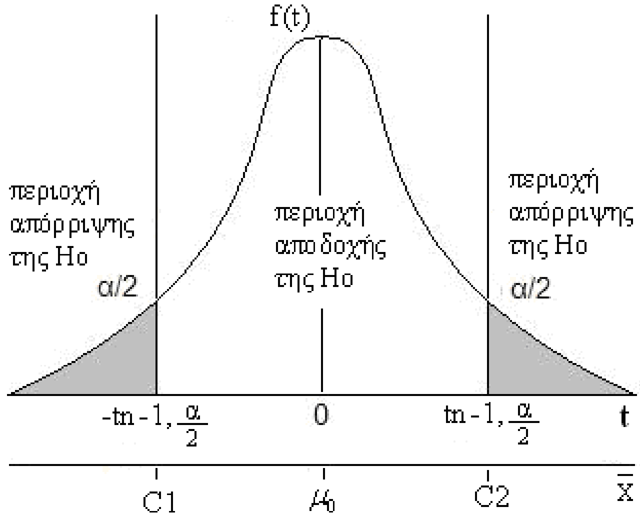
H_1: $\mu\neq\mu_0$

γίνεται δεκτή αν: $-t_{\nu,\frac{\alpha}{2}} < t < t_{\nu,\frac{\alpha}{2}}$

ή αν $C1 < \bar{x} < C2$

$$C1 = \mu_0 - t_{\nu,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$C2 = \mu_0 + t_{\nu,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$



Διάγραμμα 1.12

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα 9 παρατηρήσεων. Από το δείγμα προέκυψε $\bar{x}$=64 και S = 4. Να ελεγχθεί η υπόθεση Ho: μ=67 έναντι της H1: μ≠67 σε επίπεδο σημαντικότητας α=5%.

$$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{64-67}{\frac{4}{\sqrt{9}}} = \frac{-3}{1{,}33} = -2{,}26$$

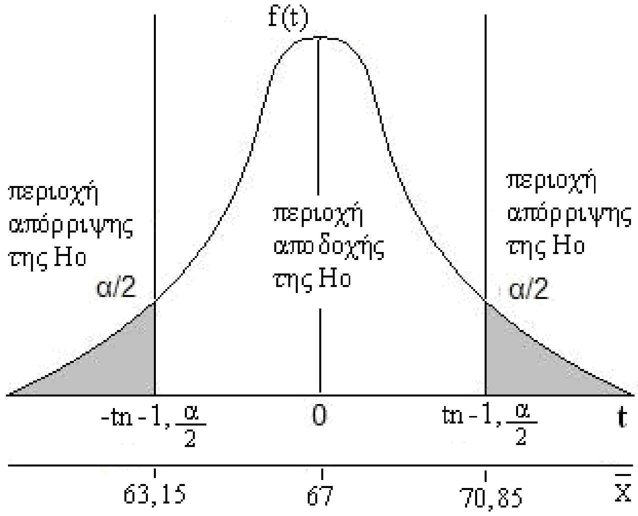
$$t_{\nu,\alpha/2} = t_{8,0.025} = 2{,}896$$

Βλέπουμε ότι $-t_{\nu,\alpha/2} < t < t_{\nu,\alpha/2}$ αφού -2,896<-2,26<2,896, άρα δεχόμαστε την υπόθεση H_0.

$$C1 = \mu_0 - t_{\nu,\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 67\text{-}2{,}896 \frac{4}{\sqrt{9}} = 67\text{-}3{,}85 = 63{,}15$$

$$C2 = \mu_0 + t_{\nu,\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 67 + 2{,}896 \frac{4}{\sqrt{9}} = 67 + 3{,}85 = 70{,}85$$

Επειδή $C1 < \bar{x} < C2$ αφού 63,15<64<70,85 αποδεχόμαστε την υπόθεση H_0.



Διάγραμμα 1.12

- **Έλεγχος υποθέσεων της διακύμανσης ενός κανονικού πληθυσμού**

Οι συνηθέστεροι συνδυασμοί για τον έλεγχο της διακύμανσης (σ^2) του πληθυσμού είναι οι εξής:

$$\begin{aligned}
&(1)\ H_0: \sigma^2 = \sigma_0^2 \\
&\quad\ \ H_1: \sigma^2 > \sigma_0^2 \\
&(2)\ H_0: \sigma^2 = \sigma_0^2 \\
&\quad\ \ H_1: \sigma^2 < \sigma_0^2 \\
&(3)\ H_0: \sigma^2 = \sigma_0^2 \\
&\quad\ \ H_1: \sigma^2 \neq \sigma_0^2
\end{aligned}$$

Η υπόθεση H_0 των παραπάνω περιπτώσεων ελέγχεται με το κριτήριο:

$$\chi^2 = \frac{\sum (xi - \mu)^2}{\sigma_0^2}$$

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

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

Αν από έναν κανονικό πληθυσμό με μέσο μ και διακύμανση σ^2 πάρουμε όλα τα δυνατά δείγματα του ίδιου μεγέθους n, τότε η ποσότητα $(n-1)\, s^2/\sigma_0^2$ ακολουθεί την κατανομή χ^2, με ν=n-1 βαθμούς ελευθερίας.

1) Αν ο έλεγχος αναφέρεται στην πρώτη περίπτωση, τότε συγκρίνουμε την τιμή χ^2 με την τιμή $\chi^2_{\nu,\alpha}$

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές: 30, 40, 28 και 54. Να ελεγχθεί η υπόθεση:

$$H_o: \sigma^2 = 30$$

$$vs$$

$$H_1: \sigma^2 > 30$$

σε επίπεδο σημαντικότητας α=5%.

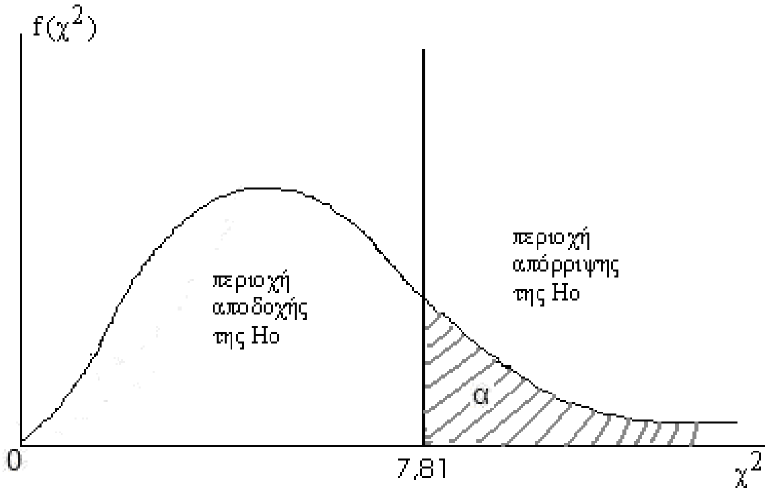
$$\bar{x} = \frac{\sum xi}{n} = \frac{152}{4} = 38$$

$$s^2 = \frac{\sum(xi-\bar{x})^2}{n-1} = \frac{(30-38)^2+(40-38)^2+(28-38)^2+(54-38)^2}{3} = \frac{64+4+100+256}{3} = 141{,}3$$

$$\chi^2 = \frac{\sum(xi-\bar{x})^2}{\sigma_0^2} = \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{3 \cdot 141{,}3}{30} = 14{,}13$$

$\chi^2_{v,a} = \chi^2_{3,0.05} = 7{,}81$

Επειδή $\chi^2 > \chi^2_{v,a}$ η υπόθεση απορρίπτεται.



Αν $\chi^2 < \chi^2_{v,\alpha}$, η υπόθεση H_0 γίνεται δεκτή.

2) Αν ο έλεγχος αναφέρεται στη δεύτερη περίπτωση, η H_0 γίνεται δεκτή, αν $\chi^2 > \chi^2_{v,1-\alpha}$

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα n=25 στο οποίο βρήκαμε $s^2 = 2$, σε επίπεδο σημαντικότητας α=5%. Να ελεγχθεί η υπόθεση:

$$H_o: \sigma^2 = 2{,}2$$

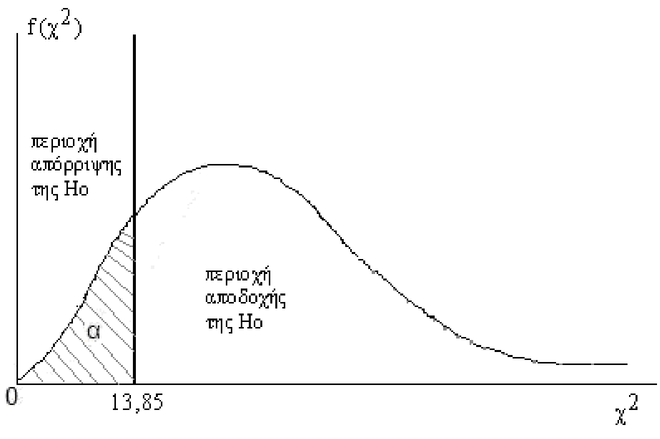
$$\text{vs}$$

$$H_1: \sigma^2 < 2{,}2$$

$$\chi^2 = \frac{\sum(xi - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)\, s^2}{\sigma_0^2} = \frac{24 \cdot 2}{2{,}2} = 21{,}82$$

$\chi^2_{v,1-\alpha} = \chi^2_{24,\,0.95} = 13{,}85$

Επειδή $\chi^2 > \chi^2_{v,1-\alpha}$ αφού 21,82 > 13,85 η υπόθεση H_o γίνεται δεκτή.



3) Αν ο έλεγχος αναφέρεται στην τρίτη περίπτωση, η υπόθεση γίνεται δεκτή, αν $\chi^2_{v,1-\alpha/2} < \chi^2 < \chi^2_{v,\alpha/2}$

Παράδειγμα:

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές: χi=51, 50, 52, 53, 50, 52, 54, 51, 52, 53, 50. Να ελεγχθεί η υπόθεση:

$$H_o: \sigma^2 = 4$$

$$\text{vs}$$

$$H_1: \sigma^2 \neq 4$$

σε επίπεδο σημαντικότητας α=5%.

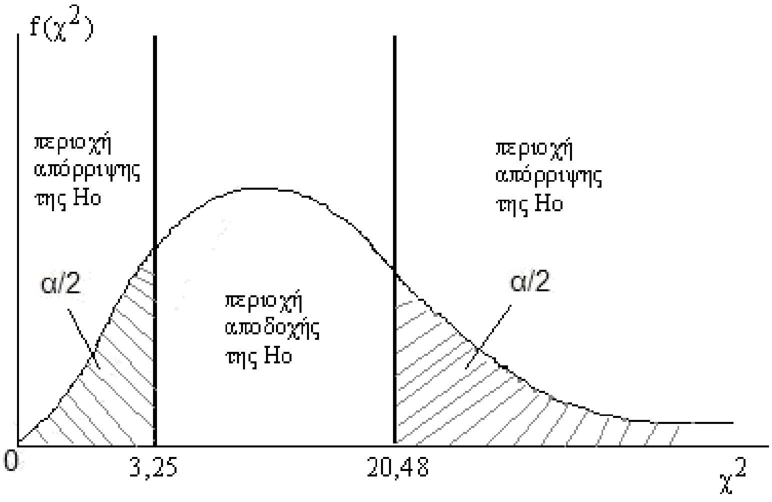
$$\bar{x} = \frac{\sum xi}{N} = \frac{568}{11} = 51{,}63$$

$$s^2 = \frac{\sum (xi - \bar{x})^2}{n-1} = \frac{18{,}47}{10} = 1{,}847$$

$$\chi^2 = \frac{\sum (xi - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)\, s^2}{\sigma_0^2} = \frac{10 \cdot 1{,}847}{4} = 4{,}61$$

$$\chi^2_{v,1-\alpha/2} = \chi^2_{10,0{,}975} = 3{,}25 \quad \text{και} \quad \chi^2_{v,\alpha/2} = \chi^2_{10,0.025} = 20{,}48$$

Επειδή $\chi^2_{v,1-\alpha/2} < \chi^2 < \chi^2_{v,\alpha/2}$ αφού 3,25 < 4,61< 20,48 η υπόθεση H_0 γίνεται δεκτή.


f(χ2)
περιοχή
απόρριψης
της Ho
περιοχή
απόρριψης
της Ho
α/2
περιοχή
αποδοχής
της Ho
α/2
0
3,25
20,48
χ2

- **Έλεγχος υποθέσεων της διαφοράς των μέσων δύο κανονικών πληθυσμών**

Αν από δύο κανονικούς πληθυσμούς που έχουν μέσους μ_1 και μ_2 και διακυμάνσεις σ_1^2 και σ_2^2, πάρουμε δύο τυχαία και ανεξάρτητα μεταξύ τους δείγματα μεγέθους n_1 και n_2 αντίστοιχα και υπολογίσουμε τους μέσους αριθμητικούς των δειγμάτων αυτών $\bar{x}_1$ και $\bar{x}_2$ αντίστοιχα, τότε μπορούμε να ελέγξουμε την υπόθεση H_0: μ_1- μ_2 με τους παρακάτω συνδυασμούς:

(1) Ho: $\mu1-\mu2=0$ ή Ho: $\mu1=\mu2$
H1: $\mu1-\mu2>0$ ή H1: $\mu1>\mu2$

(2) Ho: $\mu1-\mu2=0$ ή Ho: $\mu1=\mu2$
H1: $\mu1-\mu2<0$ ή H1: $\mu1<\mu2$

(3) Ho: $\mu1-\mu2=0$ ή Ho: $\mu1=\mu2$
H1: $\mu1-\mu2\neq0$ ή H1: $\mu1\neq\mu2$

Ανάλογα με το αν οι διακυμάνσεις του πληθυσμού σ_1^2 και σ_2^2 είναι γνωστές ή άγνωστες, διακρίνουμε τις εξής περιπτώσεις:

α) Οι διακυμάνσεις σ_1^2 και σ_2^2 είναι γνωστές

Εφόσον $\bar{x}_1$ και $\bar{x}_2$ είναι οι μέσοι αριθμητικοί των δύο τυχαίων και ανεξαρτήτων δειγμάτων, τότε η κατανομή δειγματοληψίας της διαφοράς $\bar{x}_1$-$\bar{x}_2$ έχει μέσο $\mu_1 - \mu_2$, διακύμανση $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ και τυπική απόκλιση $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ και επομένως η τυχαία μεταβλητή:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu 1 - \mu 2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

θα ακολουθεί την τυπική κατανομή έχοντας ως βασική υπόθεση:

$$Ho{:}\mu 1 = \mu 2 \quad ή \quad Ho{:}\ \mu 1 - \mu 2 = 0$$

και θα έχουμε:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(1) Η υπόθεση H_0 του πρώτου συνδυασμού γίνεται δεκτή αν:

$$Z < Z_\alpha$$

(2) Η υπόθεση H_0 του δεύτερου συνδυασμού γίνεται δεκτή αν:

$$-Z_\alpha < Z$$

(3) Η υπόθεση H_0 του τρίτου συνδυασμού γίνεται δεκτή αν:

$$-Z_{\alpha/2} < Z < Z_{\alpha/2}$$

Υποθέτουμε ότι από δύο κανονικούς πληθυσμούς $N_1(\mu_1,16)$ και $N_2(\mu_2,9)$ πήραμε τα παρακάτω δείγματα:

x_i: 28, 26, 18, 16

ψ_i: 30, 32, 26, 24

Να ελεγχθεί η υπόθεση ότι H_0: $\mu_1-\mu_2=0$ έναντι της εναλλακτικής H_1: $\mu_1<\mu_2$ σε επίπεδο σημαντικότητας α=5%.

Οι διακυμάνσεις των δύο πληθυσμών είναι γνωστές άρα θα χρησιμοποιήσουμε το εξής κριτήριο για τον έλεγχο της δοθείσας υπόθεσης:

$$\bar{x}=\frac{\sum xi}{n_1}=22$$

$$\bar{\psi}=\frac{\sum \psi i}{n_2}=28$$

$$z=\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{22-28}{\sqrt{\frac{9}{4}+\frac{16}{4}}}=-2,4$$

$-z_{0.05}=-1,645$

Αφού $z<-z_\alpha$, άρα η υπόθεση απορρίπτεται.

β) Οι διακυμάνσεις σ_1^2 και σ_2^2 είναι άγνωστες και άνισες και το μέγεθος του δείγματος μικρό

Στην περίπτωση αυτή η υπόθεση των τριών παραπάνω συνδυασμών ελέγχεται με το κριτήριο:

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

Ένας καταναλωτής έχει να επιλέξει ανάμεσα σε δύο τύπους λαμπτήρων. Ο δειγματικός έλεγχος από κάθε τύπο μεγέθους $n_1=20$ και $n_2=20$ έδωσε μέση διάρκεια ζωής του πρώτου τύπου $\bar{x}_1 = 1000$ ώρες και $\bar{x}_2 = 950$ ώρες του δεύτερου τύπου και επίσης τυπική απόκλιση $S_1=30$ ώρες και $S_2=40$ ώρες. Να ελεγχθεί η υπόθεση: H_0: $\mu_1=\mu_2$ έναντι της εναλλακτικής H_1: $\mu_1>\mu_2$ σε επίπεδο σημαντικότητας α=5%.

Επειδή δεν γνωρίζουμε τις διακυμάνσεις του πληθυσμού και το μέγεθος του δείγματος είναι μικρό n=20, θα χρησιμοποιήσουμε ως κριτήριο ελέγχου τη σχέση:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1000\text{-}950}{\sqrt{\frac{900+1600}{20}}} = 4{,}5$$

$t_{n1+n2-2,\,\alpha} = t_{38,0.05} = 2$

Επειδή $t>t_{v,\alpha}$ αφού 4,5>2 η υπόθεση απορρίπτεται.

γ) Οι διακυμάνσεις σ_1^2 και σ_2^2 είναι άγνωστες και ίσες και το μέγεθος του δείγματος μικρό

Στην περίπτωση αυτή η υπόθεση H_0 των τριών συνδυασμών ελέγχεται με το κριτήριο:

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

όπου:

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

και $n_1 + n_2 - 2$ οι βαθμοί ελευθερίας.

Η τιμή t και στη (β) και στη (γ) περίπτωση συγκρίνεται με την τιμή που παίρνουμε από τους πίνακες της κατανομής t Student με $\nu = n_1 + n_2 - 2$ βαθμούς ελευθερίας. Επομένως:

(1) Η υπόθεση H_0 του πρώτου συνδυασμού γίνεται δεκτή αν:

$$t < t_{\nu,\alpha}$$

(2) Η υπόθεση H_0 του δεύτερου συνδυασμού γίνεται δεκτή αν:

$$-t_{\nu,\alpha} < t$$

(3) Η υπόθεση H_0 του τρίτου συνδυασμού γίνεται δεκτή αν:

$$-t_{\nu,\frac{\alpha}{2}} < t < t_{\nu,\frac{\alpha}{2}}$$

Παράδειγμα

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

xi: 0,114, 0,127, 0,143, 0,132

ψi: 0,131, 0,107, 0,104, 0,111, 0,108, 0,110

Αν υποθέσουμε ότι $\sigma_1^2 = \sigma_2^2$ να ελεγχθεί η υπόθεση H_0: $\mu_1 = \mu_2$ έναντι της εναλλακτικής H_1: $\mu_1 \neq \mu_2$ σε επίπεδο σημαντικότητας α=5%.

$$\bar{x} = \frac{\sum xi}{n_1} = \frac{0,516}{4} = 0,129$$

$$\bar{\psi} = \frac{\sum \psi i}{n_2} = \frac{0,671}{6} = 0,1118$$

$$S_1^2 = \frac{\sum (xi - \bar{x})^2}{n_1 - 1} = 0,000220$$

$$S_2^2 = \frac{\sum (\psi i - \bar{\psi})^2}{n_2 - 1} = 0,000094$$

$$\boldsymbol{S}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{0,00066 + 0,00047}{8} = \frac{0,00113}{8} = 0,0001412$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\boldsymbol{S}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,129 - 0,1118}{\sqrt{\left(\frac{1}{4} + \frac{1}{6}\right) \cdot 0,0001412}} = \frac{0,0172}{\sqrt{0,00005874}} = \frac{0,0172}{0,0076} = 2,26$$

$t_{8,\,0.025} = 2,896$

Επειδή -2,896 < 2,26 < 2,896, δεχόμαστε την υπόθεση H_0.

- **Έλεγχος υποθέσεων της διαφοράς των μέσων δύο μη κανονικών πληθυσμών**

Έστω ότι έχουμε δύο ανεξάρτητα τυχαία δείγματα, τα οποία πήραμε από δύο μη κανονικούς πληθυσμούς. Αν το μέγεθος των δειγμάτων n_1 και n_2 είναι μεγάλο ($n_1, n_2 > 30$), τότε ο έλεγχος των στατιστικών υποθέσεων που αναφέρεται στη διαφορά $\mu_1 - \mu_2$ των δύο μέσων γίνεται όπως και στην περίπτωση που τα δείγματα προέρχονται από κανονικούς πληθυσμούς.

- **Έλεγχος Υποθέσεων για τη μέση τιμή των διαφορών ζευγών στατιστικών δεδομένων (εξαρτημένα δείγματα)**

Θεωρούμε ότι επιλέγουμε ένα τυχαίο δείγμα ενός πληθυσμού και μετράμε τις τιμές ενός χαρακτηριστικού αυτού του δείγματος πριν και μετά την εκτέλεση του πειράματος. Έτσι κάθε άτομο του δείγματος θα έχει ένα ζεύγος τιμών πριν το πείραμα και μετά το πείραμα. Στη συνέχεια υπολογίζουμε όλες τις διαφορές των ζευγών αυτών, όπου di,t=1,2,…,n οι διαφορές αυτές. Έπειτα μπορούμε να υπολογίσουμε τη μέση τιμή $\bar{d}$ των διαφορών.

$$\bar{d} = \frac{\Sigma di}{n}$$

και τη διακύμανση s_d^2

$$s_d^2 = \frac{\Sigma(di - \bar{d})^2}{n-1}$$

Αν δεχτούμε ότι οι δειγματικές διαφορές αποτελούν ένα τυχαίο δείγμα από έναν κανονικά κατανεμημένο πληθυσμό διαφορών d με μέση τιμή μ_d θα έχουμε:

(i) $Ho: \mu_d = 0$ vs $H1: \mu_d \neq 0$ Αμφίπλευρος Έλεγχος

(ii) $Ho: \mu_d = 0$ vs $H1: \mu_d > 0$ Μονόπλευρος Έλεγχος

(iii) $Ho: \mu_d = 0$ vs $H1: \mu_d < 0$ Μονόπλευρος Έλεγχος

Το κριτήριο που θα χρησιμοποιήσουμε είναι:

$$t = \frac{\bar{d} - do}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1}$$

(i) $Ho: \mu_d = 0$ vs $H1: \mu_d \neq 0$

Απορρίπτουμε την H_0 αν:

$t < -t_{n-1,\alpha/2}$ ή $t > t_{n-1,\alpha/2}$

Αποδεχόμαστε την H_0 αν:

$-t_{n-1,\alpha/2} < t < t_{n-1,\alpha/2}$

(ii) $Ho: \mu_d = 0$ vs $H1: \mu_d > 0$

Απορρίπτουμε την H_0 αν:

$t > t_{n-1,\alpha}$

Αποδεχόμαστε την H_0 αν:

$t < t_{n-1,\alpha}$

(iii) $Ho: \mu_d = 0$ vs $H1: \mu_d < 0$

Απορρίπτουμε την H_0 αν:

$t < -t_{n-1,\alpha}$

Αποδεχόμαστε την H_0 αν:

$t > -t_{n-1,\alpha}$

1η σημείωση:

Στην περίπτωση που οι διαφορές δεν είναι κανονικά κατανεμημένες αλλά ισχύει n>30, μπορούμε να χρησιμοποιήσουμε το Κ.Ο.Θ.

2η σημείωση

Δ.Ε. $\left[\bar{d} - t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}\right]$

Παραδείγματα:

Περιπτώσεις

1η περίπτωση:

Από τους πανελλήνιους διαγωνισμούς της ελληνικής μαθηματικής εταιρείας (Ε.Μ.Ε) επιλέχτηκαν τελικά έξι μαθητές Α, Β, Γ, Δ, Ε, ΣΤ που θα αποτελέσουν την ολυμπιακή ομάδα των μαθηματικών. Η Ε.Μ.Ε αποφάσισε να κάνει κάθε Σάββατο και για έξι μήνες μάθημα στους μαθητές αυτούς για να τους ετοιμάσει καλύτερα για την ολυμπιάδα. Πριν αρχίσουν τα μαθήματα οι έξι μαθητές έγραψαν ένα τέστ ώστε να εκτιμηθεί το επίπεδο γνώσεων με το οποίο ξεκινάνε. Μετά από έξι μήνες προετοιμασίας έγραψαν και πάλι ένα τέστ. Τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα:

Μαθητές	Πριν τη προετοιμασία Χ	Μετά τη προετοιμασία Ψ	d_i=Μετά-Πριν
Α	93	98	5
Β	90	94	4
Γ	95	96	1
Δ	92	91	-1
Ε	95	97	2
ΣΤ	91	97	6

Η σύγκριση των δεδομένων αυτών δείχνει ότι υπάρχει διαφορά στην επίδοση των μαθητών. Να γίνει έλεγχος σε επίπεδο σημαντικότητας α= 0,05

Έστω μ_d ο μέσος όρος των διαφορών Ψ-Χ (τιμές μετά την προετοιμασία – τιμές πριν την προετοιμασία).

$$Ho: \mu_d = 0 \quad vs \quad H1: \mu_d \neq 0$$

$$\bar{d} = \frac{\Sigma di}{n} = \frac{17}{6} = 2,83$$

$$s_d^2 = \frac{\Sigma(di - \bar{d})^2}{n-1} = \frac{(5-2,83)^2 + (4-2,83)^2 + (1-2,83)^2 + (-1-2,83)^2 + (2-2,83)^2 + (6-2,83)^2}{5} = \frac{4,7+1,36+3,34+14,66+0,68+10,04}{5} = \frac{34,78}{5} = 6,95$$

$$s_d = \sqrt{6,95} = 2,63$$

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{2,83}{\frac{2,63}{\sqrt{6}}} = \frac{2,83}{1,07} = 2,64$$

$t_{n-1,\alpha/2} = t_5{,}0.025 = 2{,}571$

$t > t_{n-1,\alpha/2}$ αφού 2,64 > 2,571 άρα απορρίπτουμε την H_0, δηλαδή υπάρχει σημαντική διαφορά στην επίδοση των μαθητών πριν και μετά την προετοιμασία.

2η περίπτωση:

Στον επόμενο πίνακα φαίνονται τα βάρη 8 ανθρώπων πριν σταματήσουν το κάπνισμα και πέντε εβδομάδες αφού σταματήσουν το κάπνισμα. Σε επίπεδο σημαντικότητας α=5% μπορεί να υποστηριχθεί η άποψη ότι το βάρος αυξάνεται όταν κάποιος σταματήσει το κάπνισμα;

Περίπτωση	**1η**	**2η**	**3η**	**4η**	**5η**	**6η**	**7η**	**8η**
Πριν Χ	74	88	75	58	65	64	60	66
Μετά Ψ	77	90	74	61	66	68	62	64
Διαφορά d=Ψ-Χ	3	2	-1	3	1	4	2	-2

$$Ho: \mu_d = 0 \text{ vs } H1: \mu_d > 0$$

$$\bar{d} = \frac{\Sigma di}{n} = \frac{12}{8} = 1,5$$

$$s_d^2 = \frac{\Sigma(di-\bar{d})^2}{n-1} = \frac{(3-1,5)^2+(2-1,5)^2+(-1-1,5)^2+(3-1,5)^2+(1-1,5)^2+(4-1,5)^2+(2-1,5)^2+(-2-1,5)^2}{7} = \frac{2,25+0,25+6,25+2,25+0,25+6,25+0,25+12,25}{7} = \frac{30}{7} = 4,28$$

$$S_d = \sqrt{4,28} = 2,07$$

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{1,5}{\frac{2,07}{\sqrt{8}}} = \frac{1,5}{0,73} = 2,05$$

$t_{n-1,\alpha} = t_{7,0.05} = 1,895$

$t > t_{n-1,\alpha}$ αφού 2,05 > 1,895, άρα απορρίπτουμε την H_0, δηλαδή με βάση τα δεδομένα υποστηρίζεται η αύξηση του βάρους μετά το σταμάτημα του τσιγάρου.

$3^{η}$ περίπτωση:

Η διοίκηση μιας ποδοσφαιρικής ομάδας σε συνεργασία με το ιατρικό επιτελείο της, θέλει να διαπιστώσει αν η κατάσταση του χλοοτάπητα του προπονητικού κέντρου έχει αντίκτυπο στους συχνούς τραυματισμούς των παιχτών. Για να διαπιστώσει αν αυτό ισχύει, αποφάσισε να χωρίσει τους 22 παίχτες σε 2 ομάδες, όπου η μία θα προπονείται σε παλιό προπονητικό κέντρο και η άλλη σε υπερσύγχρονο προπονητικό κέντρο με χλοοτάπητα σε άριστη κατάσταση. Έπειτα απ' τις προπονήσεις που έκαναν οι παίχτες, το ιατρικό επιτελείο διαπίστωσε ότι τα άτομα που προπονήθηκαν στο υπερσύγχρονο προπονητικό κέντρο, εμφάνισαν λιγότερους τραυματισμούς $\bar{d}$=-15,3 με s_d=23. Σε επίπεδο σημαντικότητας α=5%, να ελεγχθεί αν η καλή κατάσταση του χλοοτάπητα μειώνει τους τραυματισμούς των παιχτών.

$H_0: \mu_d = 0$ vs $H_1: \mu_d < 0$

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-15,3}{\frac{23}{\sqrt{22}}} = \frac{-15,3}{\frac{23}{4,69}} = \frac{-71,7}{23} = -3,11$$

-tn-1,α = -t21, 0.05 = -1,721

t < -tn-1,α αφού -3,11< -1,721, άρα απορρίπτουμε την H_o και συμπεραίνουμε ότι η καλή κατάσταση του χλοοτάπητα μειώνει τους τραυματισμούς των παιχτών.

- **Έλεγχος Υποθέσεων για το λόγο των διασπορών δύο κανονικών πληθυσμών**

Ο έλεγχος της υπόθεσης $H_0 : \sigma^2_1 = \sigma^2_2$ 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 το μέγεθος του δείγματος με διακύμανση s^2_1 και n_2 το μέγεθος του δείγματος με διακύμανση s^2_2.

Περιπτώσεις:

i) $H_0 : \frac{\sigma^2_1}{\sigma^2_2} = 1$ ή $H_0 : \sigma^2_1 = \sigma^2_2$

vs Αμφίπλευρος Έλεγχος

$H_1 : \sigma^2_1 \neq \sigma^2_2$

Απορρίπτουμε την H_0 αν:

$F < F_{n1-1,n2-1,\alpha/2}$ ή $F > F_{n1-1,n2-1,1-\alpha/2}$

και αποδεχόμαστε την H_0 αν:

$F_{n1-1,n2-1,\alpha/2} < F < F_{n1-1,n2-1,1-\alpha/2}$

ii) $H_o : \sigma^2{}_1 = \sigma^2{}_2$

vs Μονόπλευρος Έλεγχος

$H1: \frac{\sigma^2{}_1}{\sigma^2{}_2} > 1$ ή $H1: \sigma^2{}_1 > \sigma^2{}_2$

Απορρίπτουμε την H_0 αν:

$F > F_{n1-1,n2-1,1-\alpha}$

και αποδεχόμαστε την H_0 αν:

$F < F_{n1-1,n2-1,1-\alpha}$

iii) $H_o : \sigma^2{}_1 = \sigma^2{}_2$

vs Μονόπλευρος Έλεγχος

$H_1: \frac{\sigma^2{}_1}{\sigma^2{}_2} < 1$ ή $H_1: \sigma^2{}_1 < \sigma^2{}_2$

Απορρίπτουμε την H_0 αν:

$F < F_{n1-1,n2-1,\alpha}$

και αποδεχόμαστε την H_0 αν:

$F > F_{n1-1,n2-1,\alpha}$

<u>Σημείωση:</u> $F_{n_1-1,n_2-1,\frac{\alpha}{2}} = \dfrac{1}{F_{n_2-1,n_1-1,1-\frac{\alpha}{2}}}$

Παραδείγματα

1η περίπτωση

Ένας επενδυτής αγόρασε δύο μετοχές τις οποίες αριθμούμε με 1,2 πριν από 5 και 3 χρόνια αντίστοιχα. Οι αποδόσεις που είχαν οι δύο αυτές μετοχές εμφάνισαν διακυμάνσεις s^2_1=150 και s^2_2=430. Να συγκριθούν οι δύο διακυμάνσεις σε επίπεδο σημαντικότητας α=5%.

$$H_o : \sigma^2_1 = \sigma^2_2 \quad vs \quad H_1 : \sigma^2_1 \neq \sigma^2_2$$

$$F = \frac{s^2_1}{s^2_2} = \frac{150}{430} = 0,34$$

$$F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}} = F_{5-1, 3-1, 1\frac{-0,05}{2}} = F_{4,2,0.95} = 19,2$$

$$F_{n_1-1, n_2-1, \frac{\alpha}{2}} = \frac{1}{F_{n_2-1, n_1-1, 1-\frac{\alpha}{2}}} = \frac{1}{F_{3-1, 5-1, 1-\frac{0.05}{2}}} = \frac{1}{F_{2,4,0.975}} = \frac{1}{6,94} = 0,14$$

$$F_{n_1-1, n_2-1, \frac{\alpha}{2}} < F < F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}$$

αφού 0,14 < 0,34 < 19,2

άρα αποδεχόμαστε την H_0, δηλαδή συμπεραίνουμε ότι οι αποδόσεις των δύο μετοχών δεν εμφανίζουν σημαντική διαφορά ως προς τη διασπορά τους.

2η περίπτωση

Από δύο κανονικούς πληθυσμούς πήραμε τα εξής δείγματα:

x_i : 52,48,36,24,18 ψ_i : 44,32,28,16,8

Να ελεγχθεί σε επίπεδο σημαντικότητας α=5% η υπόθεση:

$H_o:\sigma^2_1=\sigma^2_2$ vs $H_1:\sigma^2_1>\sigma^2_2$

$$\overline{x}_i = \frac{\Sigma x_i}{n} = \frac{52+48+36+24+18}{5} = \frac{178}{5} = 35{,}6$$

$$\overline{\psi}_i = \frac{\Sigma \psi_i}{n} = \frac{44+32+28+16+8}{5} = \frac{128}{5} = 25{,}6$$

$$s^2_1 = \frac{\Sigma(x_i - \overline{x})^2}{n-1} = \frac{(52-35,6)^2+(48-35,6)^2+(36-35,6)^2+(24-35,6)^2+(18-35,6)^2}{5-1} = \frac{268,96+153,76+0,16+134,56+309,76}{4} = \frac{767,2}{4} = 191{,}8$$

$$s^2_2 = \frac{\Sigma(\psi i-\overline{\psi})^2}{n-1} = \frac{(44-25,6)^2+(32-25,6)^2+(28-25,6)^2+(16-25,6)^2+(8-25,6)^2}{5-1} = \frac{338,56+40,96+5,76+92,16+309,76}{4} = \frac{787,2}{4} = 196{,}8$$

$$F = \frac{s^2_1}{s^2_2} = \frac{191,8}{196,8} = 0{,}97$$

$F_{n1-1,n2-1,1-\alpha}=F_{5-1,5-1,1-0.05}=F_{4,4,0.95}=6{,}39$

$F < F_{n1-1,n2-1,1-\alpha}$

αφού 0,97<6,39 άρα αποδεχόμαστε την H_0.

3η περίπτωση

Ο αντιπρόσωπος μιας εταιρείας προσφέρει στον μάνατζερ μιας επιχείρησης τον ίδιο τύπο μηχανής με εκείνο που προτίθεται να αγοράσει από άλλη εταιρεία. Η τιμή της μηχανής M_1 που προσφέρει ο αντιπρόσωπος είναι μεγαλύτερη της M_2, επειδή όπως ισχυρίζεται η διασπορά της διαμέτρου των προϊόντων που κατασκευάζει η μηχανή του είναι μικρότερη της άλλης. Η επιχείρηση θέλει να ελέγξει την ορθότητα των λόγων του αντιπροσώπου. Ο αγοραστής παίρνει τυχαίο δείγμα 21 ανταλλακτικών από κάθε μηχανή. Οι σχετικοί υπολογισμοί έδωσαν s^2_1=0,00024 και s^2_2=0,00032. Ποια απόφαση πρέπει να πάρει ο υποψήφιος αγοραστής σε επίπεδο σημαντικότητας α=5%;

$H_o:\sigma^2_1=\sigma^2_2$ vs $H_1:\sigma^2_1<\sigma^2_2$

$$F=\frac{s^2_1}{s^2_2}=\frac{0,00024}{0,00032}=0,75$$

$$F_{n1-1,n2-1,\alpha}=\frac{1}{F_{n2-1,n1-1,1-\alpha}}=\frac{1}{F_{21-1,21-1,1-0.05}}=\frac{1}{F_{20,20,0.95}}=\frac{1}{2,12}=0,47$$

$F > F_{n1-1,n2-1,\alpha}$

αφού 0,75>0,47 άρα αποδέχομαι την H_0 δηλαδή ο αγοραστής δεν πείθεται από τα λόγια του αντιπροσώπου.

- **Έλεγχος υποθέσεων για το ποσοστό P του πληθυσμού**

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

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

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

Περιπτώσεις:

i) $H_0: p=p_0$ vs $H_1: p \neq p_0$ Αμφίπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z<-z_{\alpha/2}$ ή $z>z_{\alpha/2}$

και αποδεχόμαστε την H_0 αν:

$-z_{\alpha/2}<z<z_{\alpha/2}$

ii) H_0:p=p_0 vs H_1:p>p_0 Μονόπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z>z_\alpha$

και αποδεχόμαστε την H_0 αν:

$z<z_\alpha$

iii) H_0:p=p_0 vs H_1:p<p_0 Μονόπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z<-z_\alpha$

και αποδεχόμαστε την H_0 αν:

$z>-z_\alpha$

Παραδείγματα:

1η περίπτωση:

Ένας καθηγητής ισχυρίζεται ότι το 78% των φοιτητών του περνάει το μάθημα το οποίο διδάσκει. Για να θεμελιώσει τον ισχυρισμό του διενεργεί δειγματοληπτική έρευνα ανάμεσα στους φοιτητές του. Σε τυχαίο δείγμα 140 φοιτητών, οι 112 πέρασαν το μάθημα. α) Να εξεταστεί σε επίπεδο σημαντικότητας α=5% αν ευσταθεί ο ισχυρισμός του καθηγητή, β)Να βρεθούν τα όρια στα οποία περιλαμβάνεται το ποσοστό των φοιτητών που πέρασαν το μάθημα, σε όλους τους φοιτητές του καθηγητή που διδάσκει το συγκεκριμένο μάθημα με πιθανότητα 95%.

H_0:p=0,78 vs H_1:p≠0,78

$$\hat{p} = \frac{112}{140} = 0,8$$

$$z = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0,8-0,78}{\sqrt{\frac{0,78(1-0,78)}{140}}} = \frac{0,02}{\sqrt{\frac{0,1716}{140}}} = \frac{0,02}{\sqrt{0,00122}} = \frac{0,02}{0,035} = 0,57$$

$z_{\alpha/2} = z_{0.025} = \pm 1,96$

z=0,57

$-z_{\alpha/2} < z < z_{\alpha/2}$

αφού -1,96 < 0,57 < 1,96 άρα αποδεχόμαστε την H_0, δηλαδή ο ισχυρισμός του καθηγητή είναι βάσιμος.

Το 95% Δ.Ε. είναι:

$$\left[\hat{p}-z_{\alpha/2}\sqrt{\frac{p_o(1-p_o)}{n}}\ ,\ \hat{p}+z_{\alpha/2}\sqrt{\frac{p_o(1-p_o)}{n}}\right]=\left[0,8-(1,96 \cdot 0,035)\ ,\ 0,8+(1,96 \cdot 0,035)\right]=$$

$$=\left[0,8-0,068\ ,\ 0,8+0,068\right]=\left[0,732\ ,\ 0,868\right]\ \text{ή}\ \left[73,2\%\ ,\ 86,8\%\right]$$

2η περίπτωση:

Είναι γνωστό ότι 1 στους 10 καπνιστές προτιμά τη μάρκα Α. Μετά από μια διαφημιστική καμπάνια για τη μάρκα Α, ρωτήσαμε 200 καπνιστές και είδαμε ότι την προτιμούν οι 26. Να ελεγχθεί σε επίπεδο σημαντικότητας α=5% , αν αυξήθηκε το ποσοστό των καπνιστών που προτιμούν τη μάρκα Α.

$H_o{:}p=0,10$ vs $H_1{:}p>0,10$

$$p_o=\frac{1}{10}=0,1$$

$$\hat{p}=\frac{26}{200}=0,13$$

$$z=\frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}=\frac{0,13-0,10}{\sqrt{\frac{0,10(1-0,10)}{200}}}=\frac{0,03}{\sqrt{\frac{0,09}{200}}}=\frac{0,03}{0,021}=1,42$$

$z_{0.05}=1,64$

$z=1,42$

$z<z_\alpha$ αφού 1,42 < 1,64 άρα αποδεχόμαστε την H_0, δηλαδή δεν υπάρχει καμιά σημαντική βελτίωση από τη διαφημιστική καμπάνια.

3η περίπτωση:

Το προηγούμενο έτος, μια αυτοκινητοβιομηχανία κατέγραψε ποσοστό πωλήσεων 30% στην ελληνική αγορά. Παίρνουμε ένα τυχαίο δείγμα 100 υποψηφίων αγοραστών αυτοκινήτων και βλέπουμε ότι το 25% θα αγοράσουν αυτοκίνητο της συγκεκριμένης αυτοκινητοβιομηχανίας. Να ελεγχθεί η υπόθεση ότι η αυτοκινητοβιομηχανία αρχίζει να χάνει αγοραστικό κοινό, σε επίπεδο σημαντικότητας α=5%.

H_0:p=0,3 vs H_1:p<0,3

$$\hat{p} = 0{,}25$$

$$z = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0{,}25-0{,}3}{\sqrt{\frac{0{,}3(1-0{,}3)}{100}}} = \frac{-0{,}05}{\sqrt{\frac{0{,}21}{100}}} = \frac{-0{,}05}{\sqrt{0{,}0021}} = \frac{-0{,}05}{0{,}045} = -1{,}11$$

$-z_\alpha = -1{,}64$

$z = -1{,}11$

$z > -z_\alpha$ αφού -1,11>-1,64 άρα αποδέχομαι την H_0, δηλαδή το ποσοστό του αγοραστικού κοινού δεν αλλάζει για τη συγκεκριμένη αυτοκινητοβιομηχανία.

- **Έλεγχος υποθέσεων για τη διαφορά των ποσοστών p_1-p_2 δύο πληθυσμών**

Πολλές φορές θέλουμε να εξετάσουμε τη στατιστική σημαντικότητα της διαφοράς δύο ποσοστών τα οποία έχουν προκύψει από δύο πολυπληθή δείγματα. Θεωρούμε p_1 και p_2 τα ποσοστά τα οποία αντιστοιχούν σε δύο μεγάλα δείγματα n_1 και n_2 που προέρχονται από δύο πληθυσμούς με ποσοστά p_1 και p_2.

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{όπου} \quad \hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

Περιπτώσεις:

i) $H_0: p_1 = p_2$ vs $H_2: p_1 \neq p_2$ Αμφίπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z < -z_{\alpha/2}$ ή $z > z_{\alpha/2}$

και αποδεχόμαστε την H_0 αν:

$-z_{\alpha/2} < z < z_{\alpha/2}$

ii) $H_0: p_1 = p_2$ vs $H_1: p_1 > p_2$ Μονόπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z>z_\alpha$

και αποδεχόμαστε την H_0 αν:

$z<z_\alpha$

iii) H_0:$p_1=p_2$ vs H_1:$p_1<p_2$ Μονόπλευρος Έλεγχος

Απορρίπτουμε την H_0 αν:

$z<-z_\alpha$

και αποδεχόμαστε την H_0 αν:

$z>-z_\alpha$

Παραδείγματα:

1η περίπτωση

Θέλουμε να διενεργήσουμε δειγματοληπτική έρευνα στα δύο δημοφιλέστερα αθλήματα, το ποδόσφαιρο και το μπάσκετ, σχετικά με τα ποσοστά των φιλάθλων γυναικών. Στο ποδόσφαιρο, βάσει δείγματος 1500 φιλάθλων, οι 315 από αυτούς ήταν γυναίκες. Στο μπάσκετ, βάσει δείγματος 700 φιλάθλων, οι 126 από αυτούς ήταν γυναίκες. Ύστερα από την έρευνα αυτή, μπορούμε να ισχυριστούμε ότι το ποσοστό των φιλάθλων γυναικών είναι το ίδιο και στα δύο αθλήματα; (α=5%)

$H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$

$$p_1 = \frac{315}{1500} = 0,21$$

$$p_2 = \frac{126}{700} = 0,18$$

$$\hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{(0,21 \cdot 1500) + (0,18 \cdot 700)}{1500 + 700} = \frac{315 + 126}{2200} = \frac{441}{2200} = 0,20$$

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0,21 - 0,18}{\sqrt{0,20(1-0,20) \cdot (\frac{1}{1500} + \frac{1}{700})}} = \frac{0,03}{\sqrt{0,16 \cdot 0,00209}} = \frac{0,03}{\sqrt{0,0003344}} = \frac{0,03}{0,018} = 1,66$$

$z_{0,025} = \pm 1,96$

$-z\alpha/2 < z < z\alpha/2$

αφού $-1,96 < 1,66 < 1,96$, άρα αποδεχόμαστε την H_0, δηλαδή το ποσοστό των φιλάθλων γυναικών είναι το ίδιο και στα δύο αθλήματα.

2η περίπτωση

Σε 1000 γυναίκες που περνάνε μπροστά από ένα κατάστημα, οι 52 μπαίνουν μέσα, ενώ σε 1000 άνδρες οι 35 μπαίνουν μέσα. Μπορούμε να δεχθούμε σε επίπεδο σημαντικότητας α=5% ότι το ποσοστό των γυναικών που μπαίνουν μέσα σε ένα κατάστημα περνώντας μπροστά από αυτό, είναι σημαντικά μεγαλύτερο από το αντίστοιχο των ανδρών;

$H_0: p_1 = p_2$ vs $H_1: p_1 > p_2$

$$p_1 = \frac{52}{1000} = 0{,}052$$

$$p_2 = \frac{35}{1000} = 0{,}035$$

$$\hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{(0{,}052 \cdot 1000) + (0{,}035 \cdot 1000)}{1000 + 1000} = \frac{52+35}{2000} = \frac{87}{2000} = 0{,}0435$$

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}052 - 0{,}035}{\sqrt{0{,}0435 \cdot (1 - 0{,}0435) \cdot \left(\frac{1}{1000} + \frac{1}{1000}\right)}} = \frac{0{,}017}{0{,}00912} = 1{,}86$$

$z_{0,05} = 1{,}64$

$z > z_{\alpha}$

αφού 1,86>1,64 άρα απορρίπτουμε την H_0, δηλαδή δεχόμαστε ότι το ποσοστό των γυναικών είναι σημαντικά μεγαλύτερο από το αντίστοιχο των ανδρών.

3η περίπτωση

Σε μια πόλη της Ελλάδας, σε τυχαίο δείγμα 200 παιδιών, το 25% επιλέγει στον ελεύθερο χρόνο του να αθληθεί. Σε ένα χωριό της Ελλάδας, σε τυχαίο δείγμα 350 παιδιών, το 56% επιλέγει στον ελεύθερο χρόνο του να ασχοληθεί με τον αθλητισμό. Μπορεί να υποστηριχθεί η άποψη ότι στην πόλη το ποσοστό των παιδιών που αθλούνται είναι μικρότερο από αυτό στο χωριό; (α=5%)

$H_0: p_1 = p_2$ vs $H_1: p_1 < p_2$

$p_1 = 0,25$ και $p_2 = 0,56$

$$\hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{(0,25 \cdot 200) + (0,56 \cdot 350)}{200 + 350} = \frac{50 + 196}{550} = \frac{246}{550} = 0,4472$$

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,25 - 0,56}{\sqrt{0,4472 \cdot (1 - 0,4472) \cdot \left(\frac{1}{200} + \frac{1}{350}\right)}} = \frac{-0,31}{\sqrt{0,247 \cdot 0,00786}} = \frac{-0,31}{\sqrt{0,00194}} = \frac{-0,31}{0,044} = -7,045$$

$-z_{0,05} = -1,64$

$z < -z_\alpha$

αφού $-7,045 < -1,64$ άρα απορρίπτουμε την H_0, δηλαδή το ποσοστό των παιδιών στην πόλη που ασχολούνται με τον αθλητισμό είναι μικρότερο από το αντίστοιχο στο χωριό.

Εισαγωγικά εφαρμογής

Στο σημείο αυτό θα ασχοληθούμε με μια πρακτική εφαρμογή των όσων είπαμε μέχρι στιγμής σχετικά με τους ελέγχους. Για να βγάλουμε όσο το δυνατόν πιο ακριβή αποτελέσματα, πήραμε ως βάση πραγματικά δεδομένα. Αυτά είναι, τραπεζικές μετοχές στο Χρηματιστήριο Αθηνών.

Οι μετοχές που θα ασχοληθούμε είναι η μετοχή της Αγροτικής Τράπεζας και η μετοχή της Εμπορικής Τράπεζας. Θα εξετάσουμε της μεταβολές που είχαν οι δύο μετοχές κατά το διάστημα, από αρχή Μαρτίου έως και το τέλος του ίδιου μήνα. Οι μεταβλητές με τις οποίες θα εξάγουμε στοιχεία είναι χωρισμένες σε 7 κατηγορίες. Αυτές είναι, η τιμή κλεισίματος της μετοχής μια συγκεκριμένη ημέρα, η τιμή ανοίγματος, η υψηλή τιμή ημέρας, η χαμηλή τιμή ημέρας, ο όγκος των συναλλαγών, ο τζίρος των συναλλαγών καθώς και η μέση τιμή κατά τη διάρκεια της ημέρας.

Κατ' αρχάς θα πραγματοποιήσουμε έλεγχο κανονικότητας για κάθε μια μεταβλητή κάθε μετοχής. Κατόπιν θα κάνουμε ελέγχους για την σύγκριση των μέσων τιμών των αποδόσεων των ιδίων κατηγοριών μετοχών μεταξύ των δύο τραπεζών. Από τον έλεγχο κανονικότητας που πραγματοποιούμε πρωτίστως, θα κρίνουμε αν θα συνεχίσουμε στους ελέγχους υποθέσεων για τις συγκρίσεις των μέσων αποδόσεων των μετοχών των δύο τραπεζών.

Αναλυτικότερα θα τα δούμε παρακάτω, χρησιμοποιώντας τα δεδομένα των παρακάτω πινάκων. Τα δεδομένα τα έχουμε συλλέξει από ιστότοπο στο διαδίκτυο ο οποίος τα διαθέτει για κοινή χρήση.

- **Έλεγχος κανονικότητας**

H_0: Το δείγμα μας προέρχεται από κανονικά κατανεμημένο πληθυσμό.

Vs

H_1: Το δείγμα μας δεν προέρχεται από κανονικά κατανεμημένο πληθυσμό.

Σύμφωνα με τα παρακάτω δεδομένα θα πραγματοποιήσουμε έλεγχο κανονικότητας σε κάθε μια από τις περιπτώσεις ανάμεσα στα στοιχεία των δύο παρακάτω πινάκων.

ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ (ΚΟ) – Μάρτιος 2009							
Ημερ/νία	Κλείσιμο	Άνοιγμα	Υψηλ. Ημέρας	Χαμηλ. Ημέρας	Όγκος	Τζίρος	Μέση Τιμή
3/3/2009	0,98	0,99	0,99	0,96	375179	366,24	0,98
4/3/2009	1,02	1,00	1,04	0,97	487409	492,35	1,01
5/3/2009	1,00	1,02	1,05	0,99	311637	317,62	1,02
6/3/2009	1,00	1,00	1,06	0,99	482723	488,70	1,01
9/3/2009	1,00	1,01	1,01	0,99	317256	317,04	1,00
10/3/2009	1,08	1,01	1,09	1,00	433704	450,33	1,04
11/3/2009	1,09	1,10	1,14	1,06	446247	494,66	1,11
12/3/2009	1,05	1,09	1,09	1,04	263811	278,44	1,06
13/3/2009	1,03	1,09	1,10	1,03	422683	449,49	1,06
16/3/2009	1,05	1,03	1,07	1,03	396360	416,08	1,05
17/3/2009	1,03	1,06	1,07	1,02	398419	416,26	1,04
18/3/2009	1,04	1,05	1,06	1,03	243699	254,10	1,04
19/3/2009	1,07	1,05	1,07	1,04	1157605	1221,51	1,06
20/3/2009	1,06	1,07	1,07	1,04	676307	714,57	1,06
23/3/2009	1,10	1,08	1,12	1,07	1048476	1150,09	1,10
24/3/2009	1,10	1,13	1,15	1,09	971040	1086,28	1,12
26/3/2009	1,17	1,12	1,18	1,11	1057350	1212,64	1,15

27/3/2009	1,09	1,19	1,19	1,09	875520	1004,78	1,15
30/3/2009	1,06	1,05	1,08	1,04	424641	451,44	1,06
31/3/2009	1,11	1,08	1,11	1,07	282824	308,38	1,09

Πίνακας 1. Χαρακτηριστικά μετοχής «Αγροτικής Τράπεζας» για το μήνα Μάρτιο του 2009

ΕΜΠΟΡΙΚΗ ΤΡΑΠΕΖΑ (ΚΟ) – Μάρτιος 2009							
Ημερ/νία	Κλείσιμο	Άνοιγμα	Υψηλ. Ημέρας	Χαμηλ. Ημέρας	Όγκος	Τζίρος	Μέση Τιμή
3/3/2009	4,74	4,44	4,80	4,44	27128	125,78	4,64
4/3/2009	4,70	4,80	4,88	4,56	18216	86,56	4,75
5/3/2009	4,62	4,70	4,70	4,54	17372	80,28	4,62
6/3/2009	4,60	4,60	4,62	4,50	21967	100,48	4,57
9/3/2009	4,60	4,60	4,60	4,54	11256	51,60	4,58
10/3/2009	4,70	4,54	4,70	4,54	14165	65,90	4,65
11/3/2009	4,68	4,70	4,70	4,62	9274	43,42	4,68
12/3/2009	4,70	4,66	4,72	4,60	13673	63,57	4,65
13/3/2009	4,68	4,70	4,76	4,62	15235	71,26	4,68
16/3/2009	4,66	4,70	4,82	4,65	2614	12,22	4,68
17/3/2009	4,61	4,66	4,80	4,56	16476	75,93	4,61
18/3/2009	4,64	4,62	4,66	4,56	7496	34,63	4,62
19/3/2009	4,84	4,72	4,93	4,72	16668	79,87	4,79
20/3/2009	4,91	4,84	4,95	4,75	8649	41,74	4,83
23/3/2009	4,94	5,00	5,06	4,90	12817	64,49	5,03
24/3/2009	4,89	5,04	5,05	4,81	13574	67,16	4,95
26/3/2009	5,01	4,89	5,10	4,89	22632	113,62	5,02
27/3/2009	4,96	5,01	5,03	4,85	25316	125,93	4,97
30/3/2009	4,78	4,64	5,00	4,64	38114	182,76	4,80
31/3/2009	4,65	4,80	4,88	4,60	118535	553,43	4,67

Πίνακας 2. Χαρακτηριστικά μετοχής «Εμπορικής Τράπεζας» για το μήνα Μάρτιο του 2009

1. Έλεγχος κανονικότητας για το κλείσιμο των μετοχών.

Για το κλείσιμο της μετοχής της Αγροτική Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,088	20	,200*	,968	20	,706

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

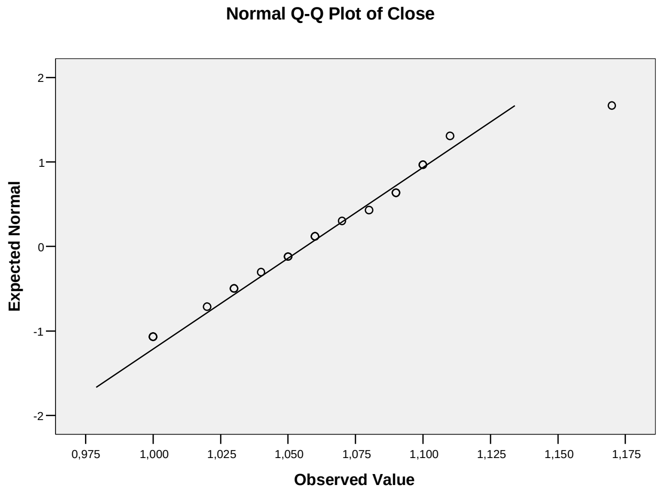
Kolmogorov – Smirnov:

Sig.$_1$=0,200 > 0.05

Shapiro-Wilk

Sig.$_1$=0,706 > 0.05

Άρα αποδεχόμαστε την H_0. Συνεπώς η μεταβλητή του κλεισίματος για την Αγροτική Τράπεζα είναι κανονικά κατανεμημένη.



Επίσης, για το κλείσιμο της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,235	20	,005	,880	20	,018

a. Lilliefors Significance Correction

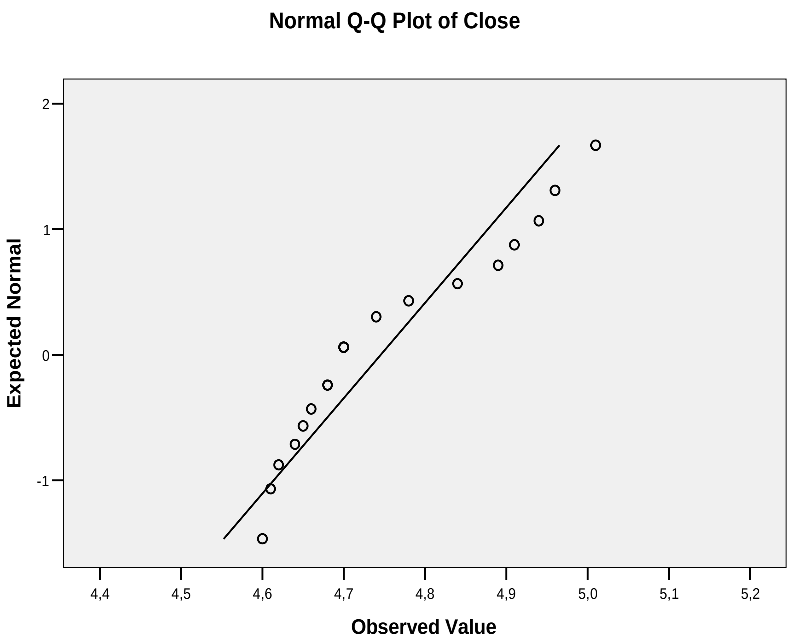
Kolmogorov – Smirnov:

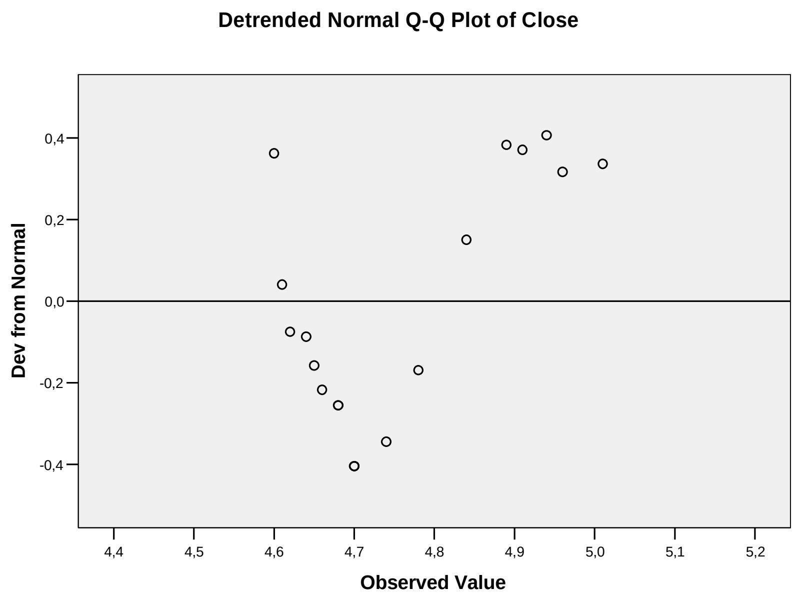
$Sig._2$=0,005 < 0.05

Shapiro-Wilk

$Sig._2$=0,018 < 0.05

Αντιστοίχως, έχουμε αποδοχή της H_1 οπότε για την Εμπορική Τράπεζα η μεταβλητή του κλεισίματος δεν είναι κανονικά κατανεμημένη.





2. Έλεγχος κανονικότητας για το άνοιγμα των μετοχών.

Για το άνοιγμα της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,091	20	,200*	,949	20	,359

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

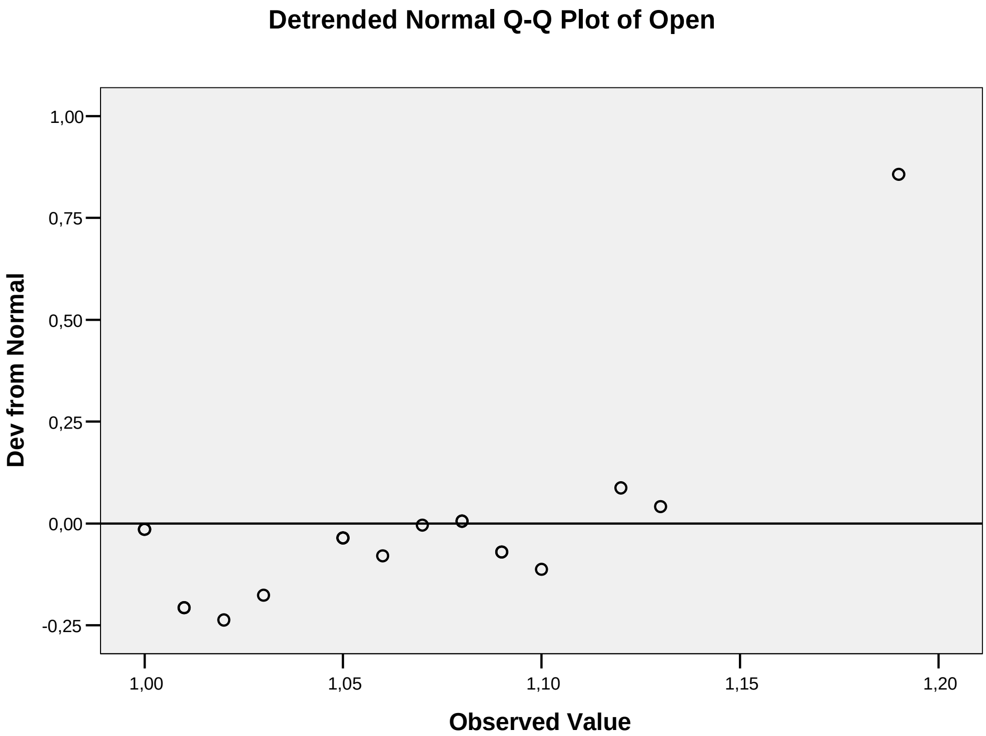
Kolmogorov – Smirnov:

$Sig._1$=0,200 > 0.05

Shapiro-Wilk

$Sig._1$=0,359 > 0.05

Άρα αποδεχόμαστε την H_0. Συνεπώς η μεταβλητή του ανοίγματος για την Αγροτική Τράπεζα είναι κανονικά κατανεμημένη.



Για το άνοιγμα της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,183	20	,080	,944	20	,286

a. Lilliefors Significance Correction

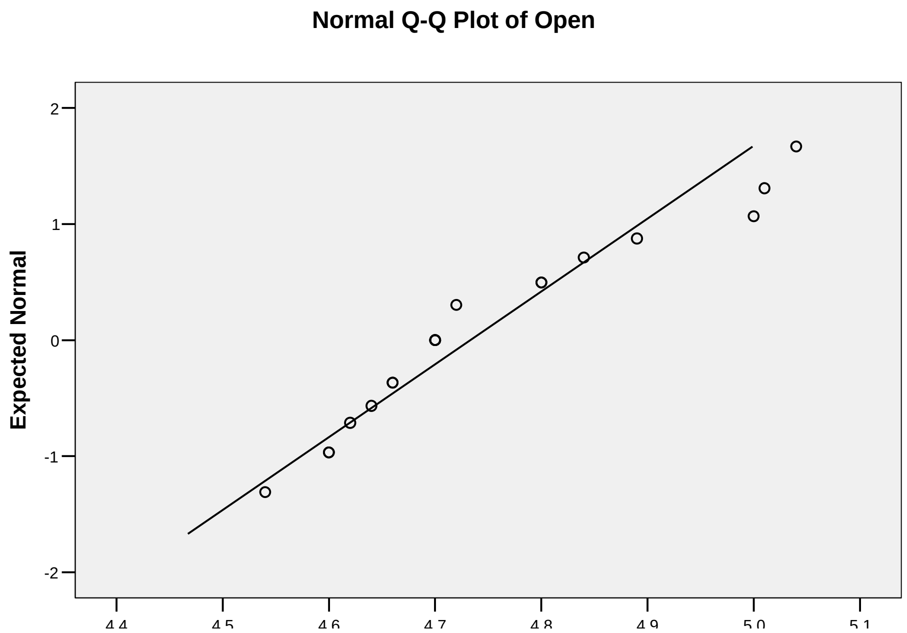
Kolmogorov – Smirnov:

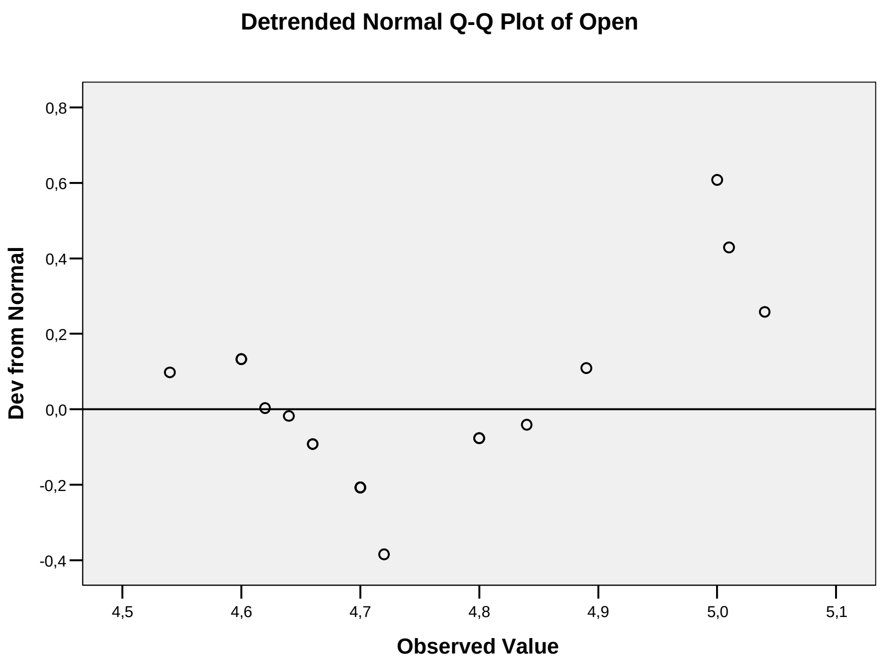
$Sig._1$=0,080 > 0.05

Shapiro-Wilk

$Sig._1$=0,286 > 0.05

Αντιστοίχως, έχουμε αποδοχή της H_0, οπότε και για την Εμπορική Τράπεζα η μεταβλητή του ανοίγματος είναι κανονικά κατανεμημένη.





3. Έλεγχος κανονικότητας για την υψηλή τιμή ημέρας των μετοχών.

Για την υψηλή τιμή ημέρας της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,131	20	,200*	,966	20	,669

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

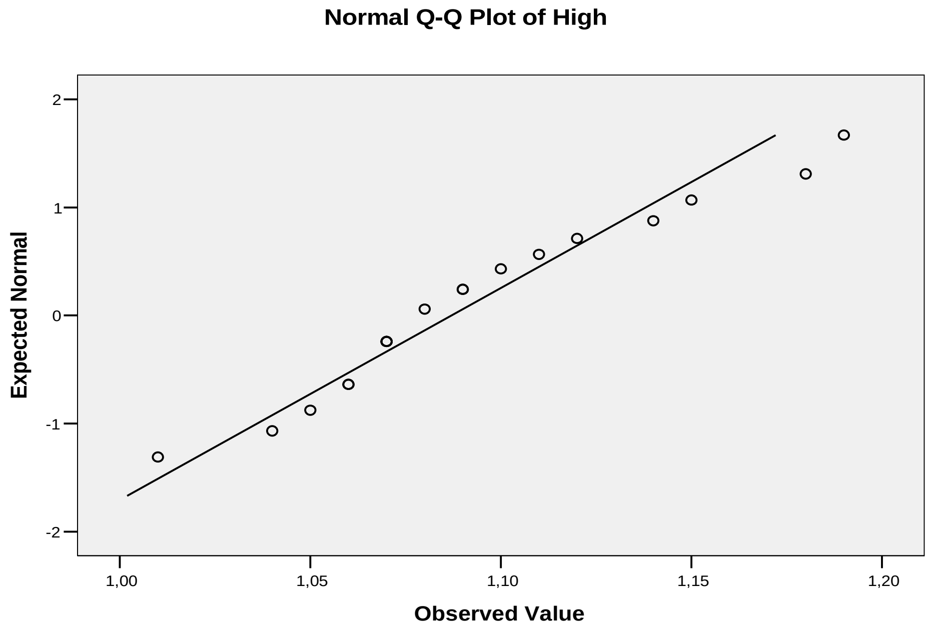
Kolmogorov – Smirnov:

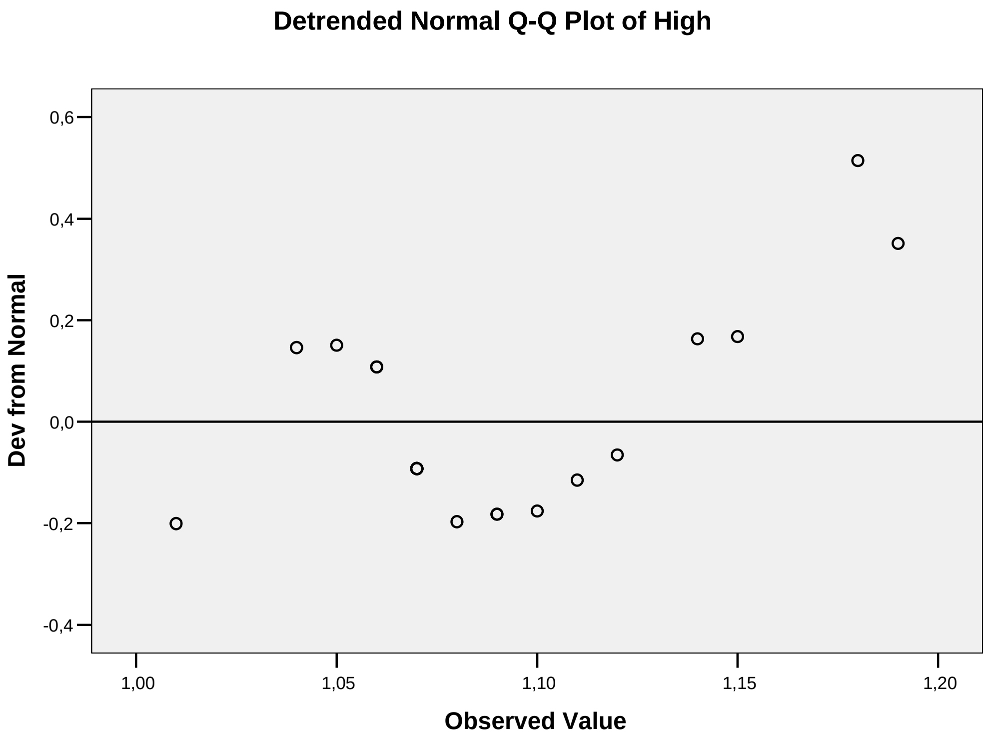
$Sig._1=0,200 > 0.05$

Shapiro-Wilk

$Sig._1=0,669 > 0.05$

Άρα αποδεχόμαστε την H_0. Συνεπώς η μεταβλητή της υψηλής τιμής ημέρας για την Αγροτική Τράπεζα είναι κανονικά κατανεμημένη.





Για την υψηλή τιμή ημέρας της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,124	20	,200*	,945	20	,301

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

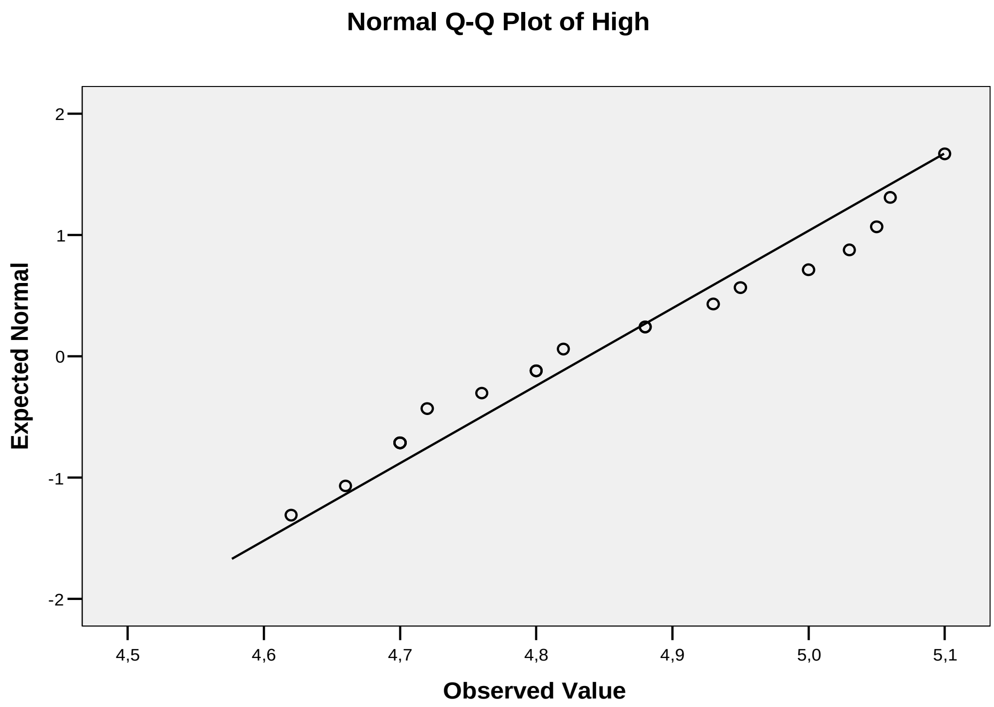
<u>Kolmogorov – Smirnov:</u>

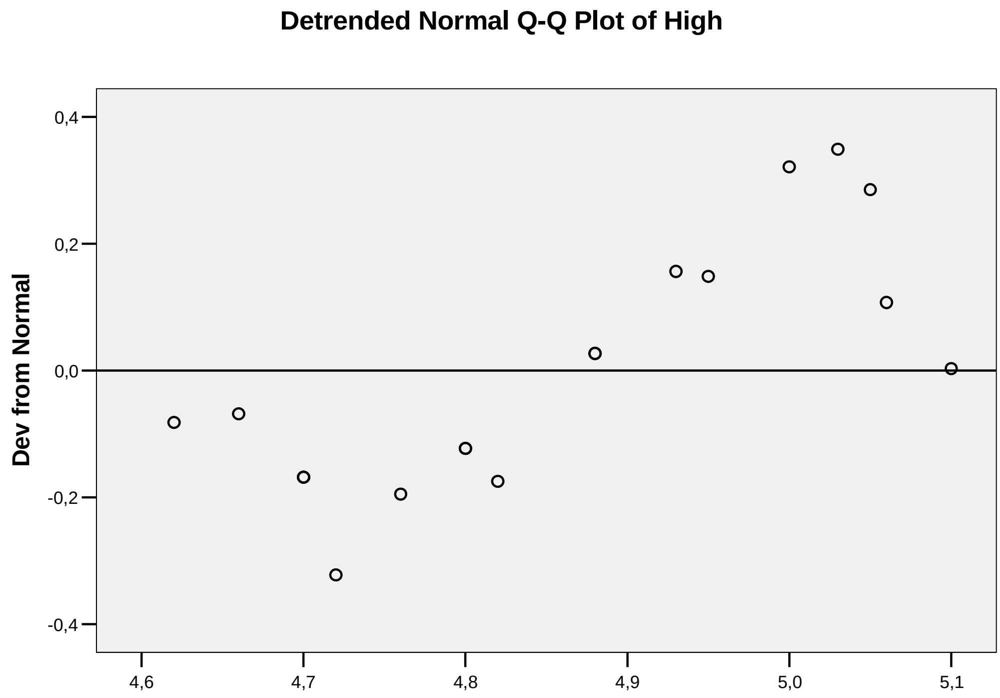
$Sig._1$=0,200 > 0.05

<u>Shapiro-Wilk</u>

$Sig._1$=0,301 > 0.05

Αντιστοίχως, έχουμε αποδοχή της H_0 οπότε και για την Εμπορική Τράπεζα η μεταβλητή της υψηλής τιμής ημέρας είναι κανονικά κατανεμημένη.


Normal Q-Q Plot of High
Expected Normal
Observed Value
2
1
0
-1
-2
4,5
4,6
4,7
4,8
4,9
5,0
5,1


Detrended Normal Q-Q Plot of High
Dev from Normal
0,4
0,2
0,0
-0,2
-0,4
4,6
4,7
4,8
4,9
5,0
5,1

4. Έλεγχος κανονικότητας για τη χαμηλή τιμή ημέρας των μετοχών.

Για την χαμηλή τιμή ημέρας της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,133	20	,200*	,969	20	,725

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

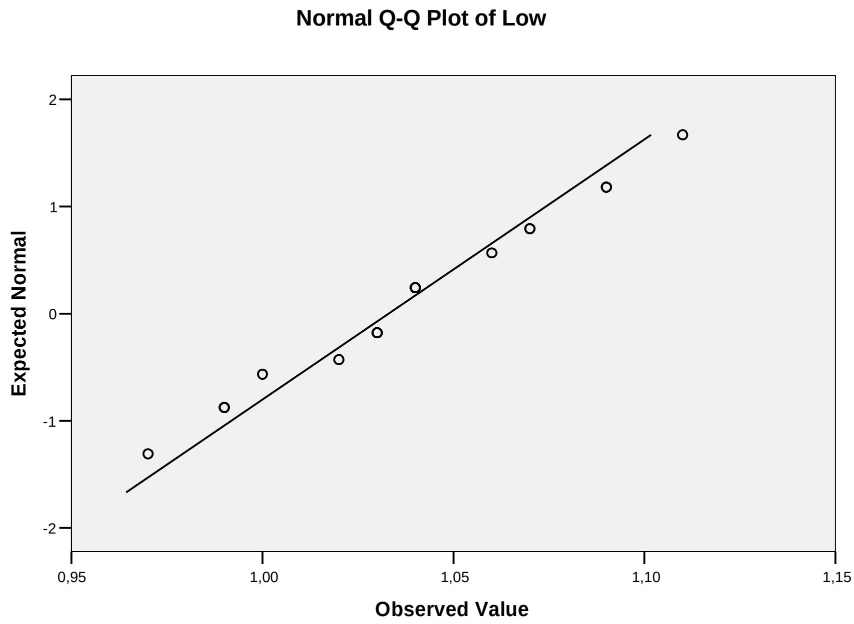
Kolmogorov – Smirnov:

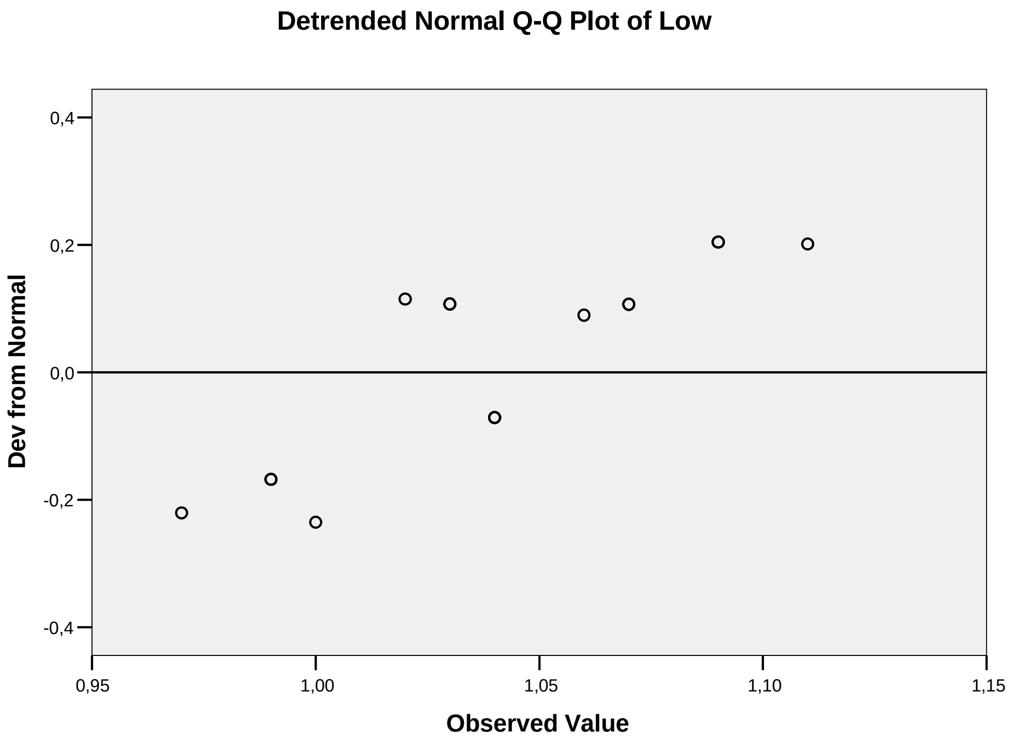
$Sig._1=0,200 > 0.05$

Shapiro-Wilk

$Sig._1=0,725 > 0.05$

Άρα αποδεχόμαστε την H_0. Συνεπώς η μεταβλητή της χαμηλής τιμής ημέρας για την Αγροτική Τράπεζα είναι κανονικά κατανεμημένη.





Για την χαμηλή τιμή ημέρας της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,183	20	,076	,910	20	,065

a. Lilliefors Significance Correction

Kolmogorov – Smirnov:

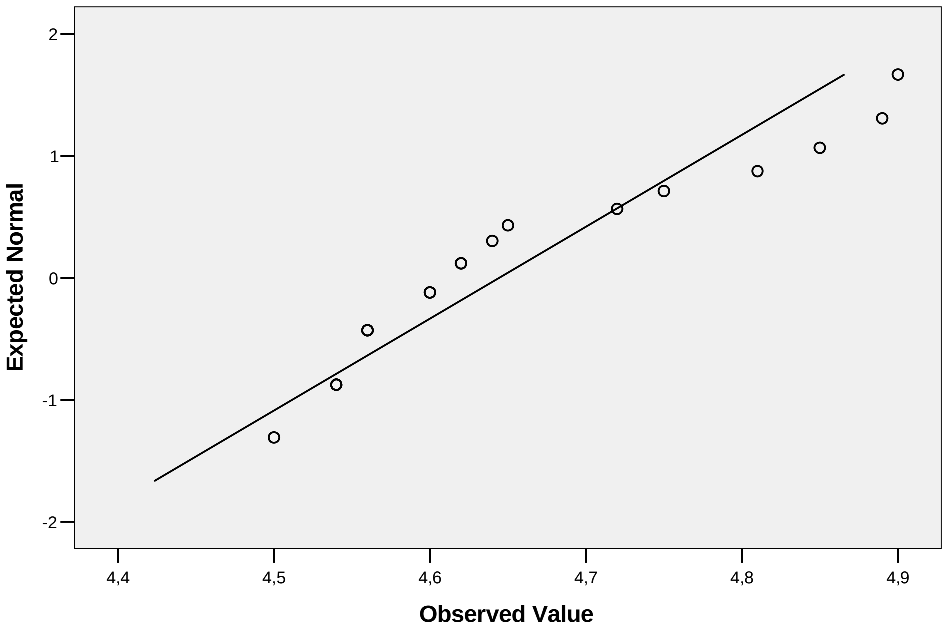
$Sig._1 = 0,076 > 0.05$

Shapiro-Wilk

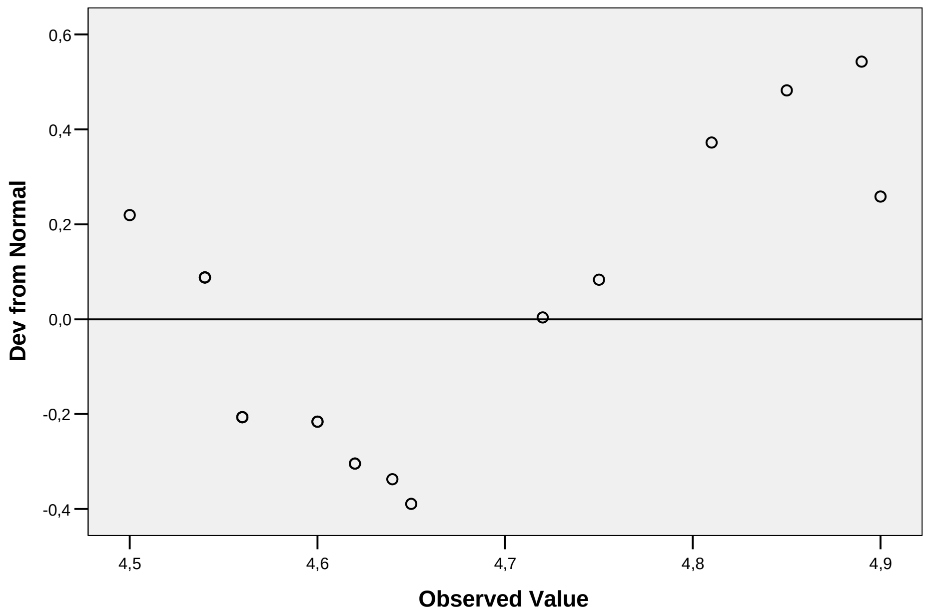
$Sig._1 = 0,065 > 0.05$

Αντιστοίχως, έχουμε αποδοχή της H_0 οπότε και για την Εμπορική Τράπεζα η μεταβλητή της χαμηλής τιμής ημέρας είναι κανονικά κατανεμημένη.

Normal Q-Q Plot of Low



Detrended Normal Q-Q Plot of Low



5. Έλεγχος κανονικότητας για τον όγκο των μετοχών.

Για τον όγκο των μετοχών της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,288	20	,000	,820	20	,002

a. Lilliefors Significance Correction

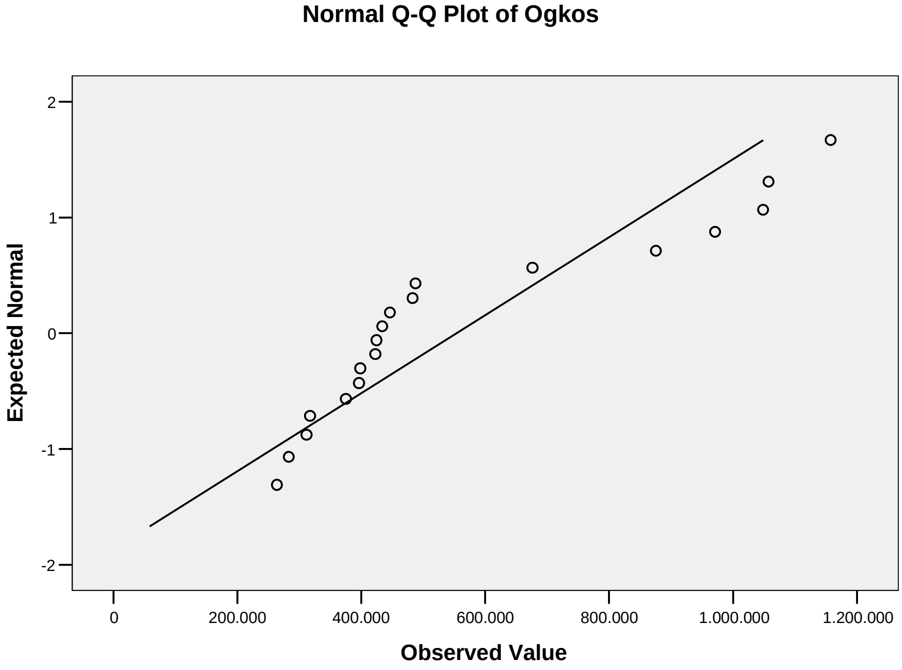
Kolmogorov – Smirnov:

$Sig._1$=0,000 < 0.05

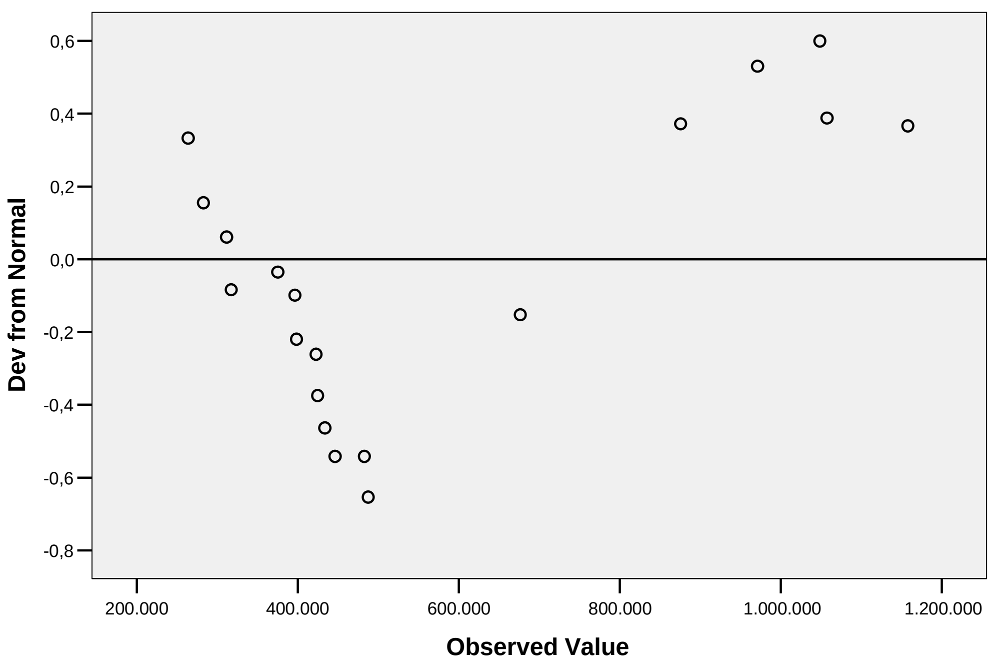
Shapiro-Wilk

$Sig._1$=0,002 < 0.05

Άρα απορρίπτουμε την H_0. Συνεπώς η μεταβλητή του όγκου των μετοχών για την Αγροτική Τράπεζα δεν είναι κανονικά κατανεμημένη.



Detrended Normal Q-Q Plot of Ogkos



Για τον όγκο των μετοχών της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,309	20	,000	,532	20	,000

a. Lilliefors Significance Correction

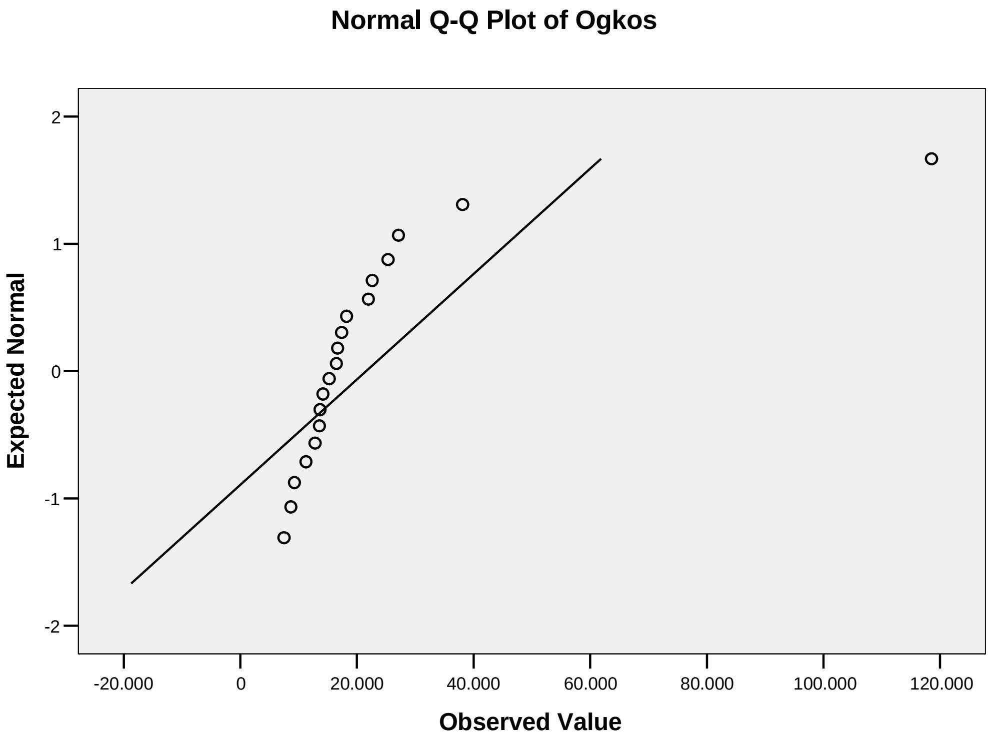
Kolmogorov – Smirnov:

$Sig._1$=0,000 < 0.05

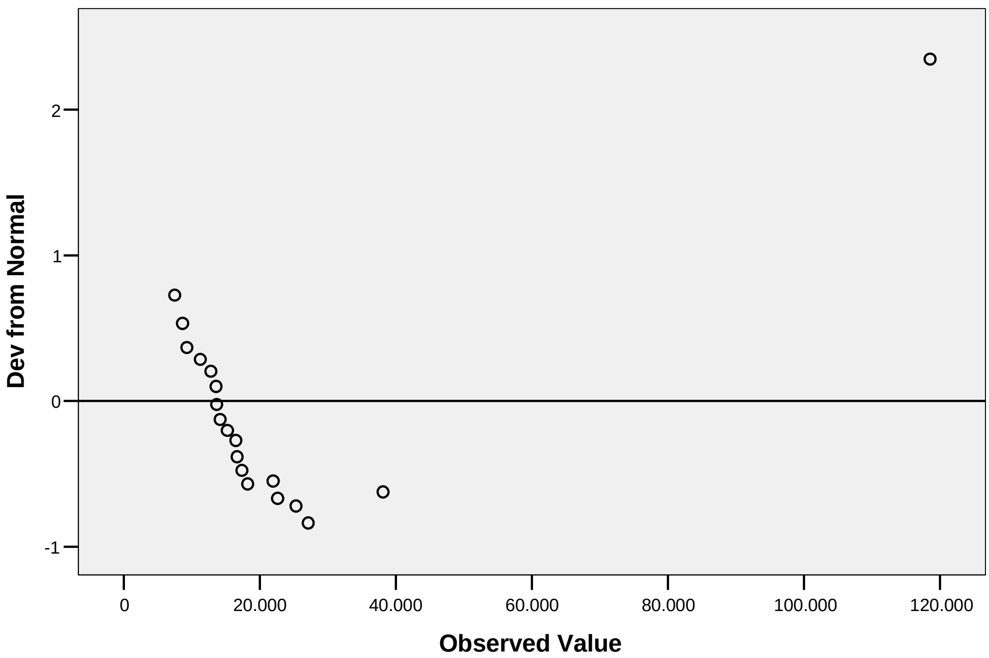
Shapiro-Wilk

$Sig._1$=0,000 < 0.05

Αντιστοίχως, έχουμε απόρριψη της H_0 οπότε και για την Εμπορική Τράπεζα η μεταβλητή του όγκου των μετοχών δεν είναι κανονικά κατανεμημένη.



Detrended Normal Q-Q Plot of Ogkos



6. Έλεγχος κανονικότητας για τον τζίρο των μετοχών (σε χιλ. €).

Για τον τζίρο των μετοχών της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,316	20	,000	,793	20	,001

a. Lilliefors Significance Correction

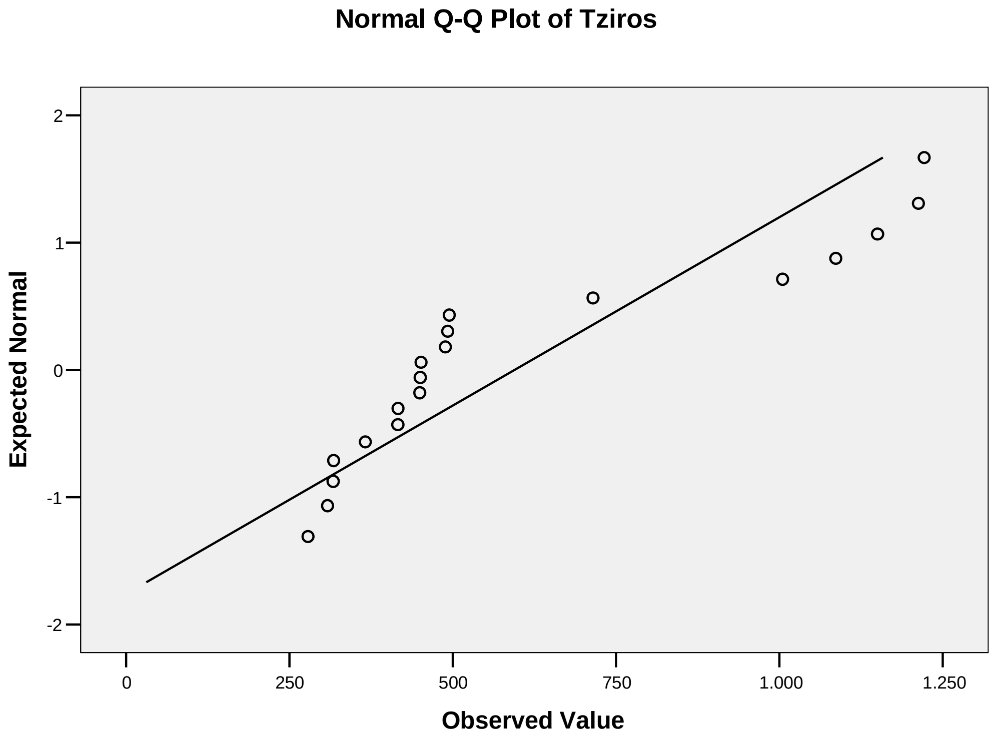
Kolmogorov – Smirnov:

$Sig._1$=0,000 < 0.05

Shapiro-Wilk

$Sig._1$=0,001 < 0.05

Άρα απορρίπτουμε την H_0. Συνεπώς η μεταβλητή του τζίρου των μετοχών για την Αγροτική Τράπεζα δεν είναι κανονικά κατανεμημένη.



Για τον τζίρο των μετοχών της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,316	20	,000	,541	20	,000

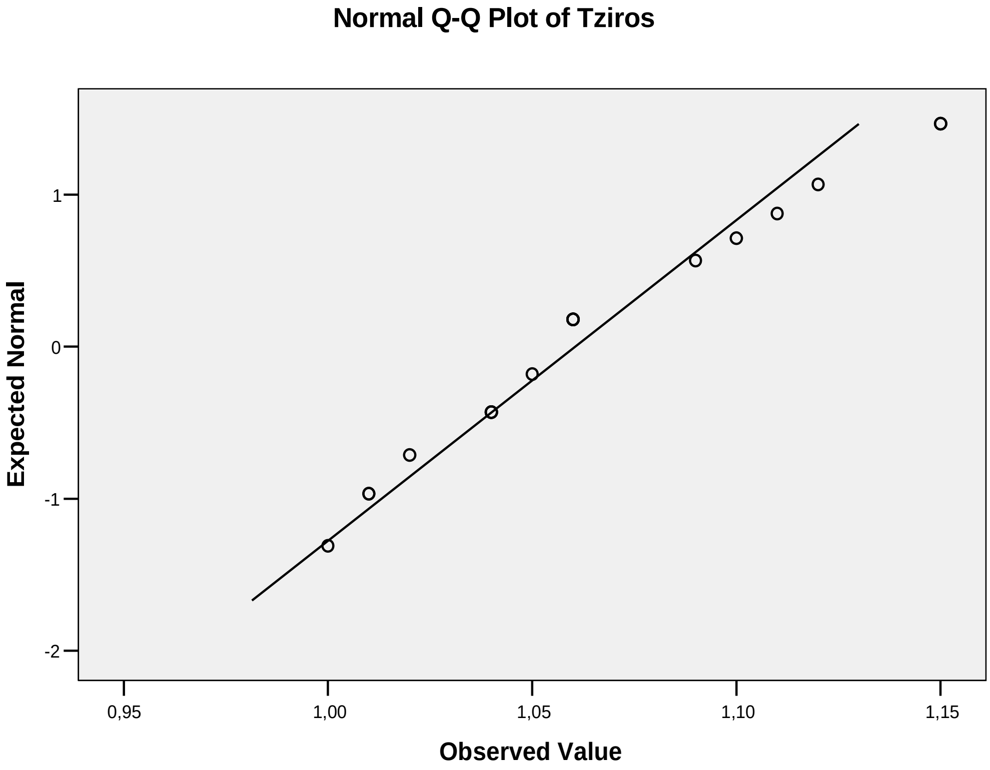
a. Lilliefors Significance Correction

<u>Kolmogorov – Smirnov:</u>

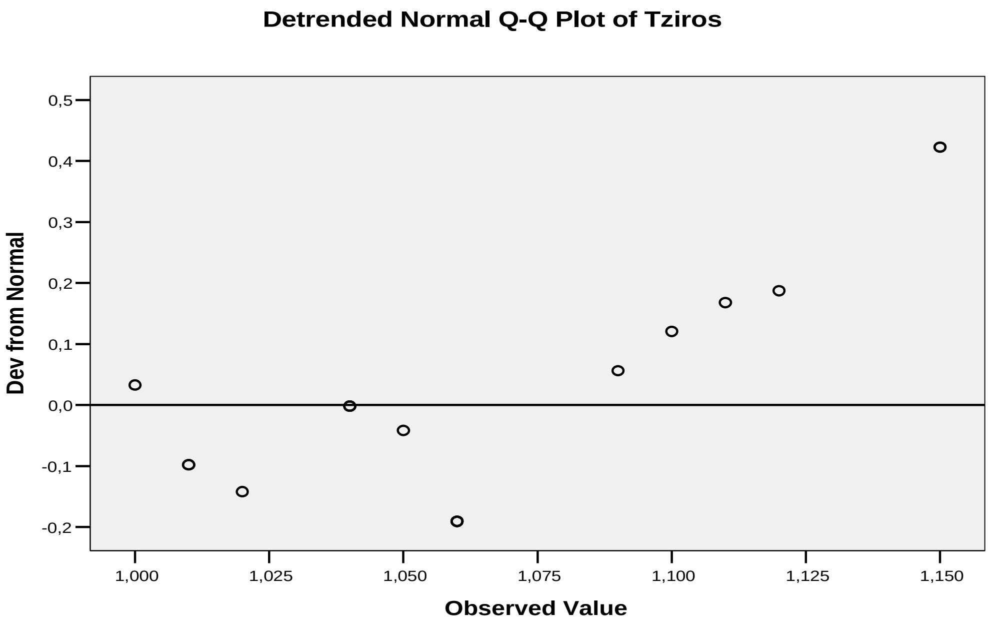
$Sig._1$=0,000 < 0.05

<u>Shapiro-Wilk</u>

$Sig._1$=0,000 < 0.05



Αντιστοίχως, έχουμε απόρριψη της H_0 οπότε και για την Εμπορική Τράπεζα η μεταβλητή του τζίρου των μετοχών δεν είναι κανονικά κατανεμημένη.



7. Έλεγχος κανονικότητας για τη μέση τιμή ημέρας των μετοχών.

Για τη μέση τιμή των μετοχών της μετοχής της Αγροτικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Agrotiki_March	,204	20	,028	,953	20	,418

a. Lilliefors Significance Correction

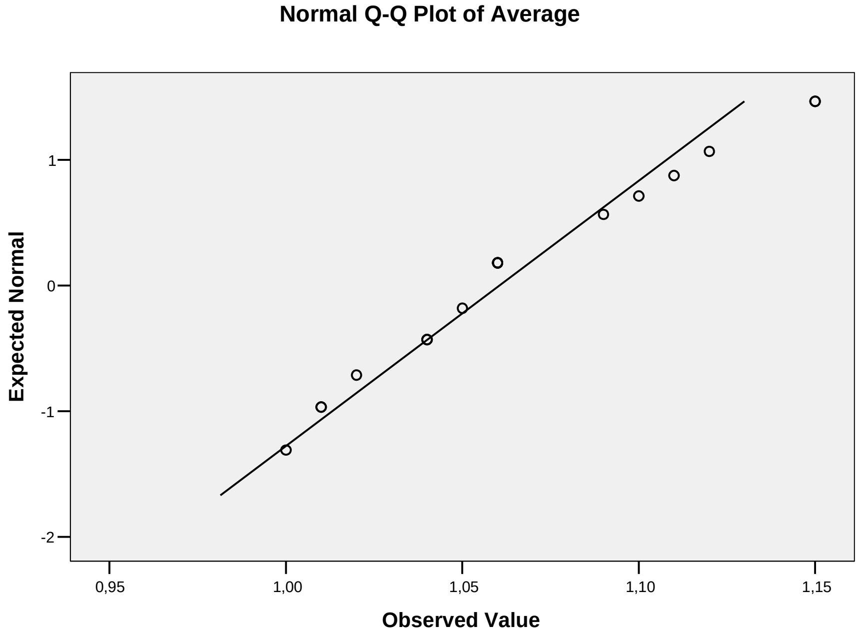
<u>Kolmogorov – Smirnov:</u>

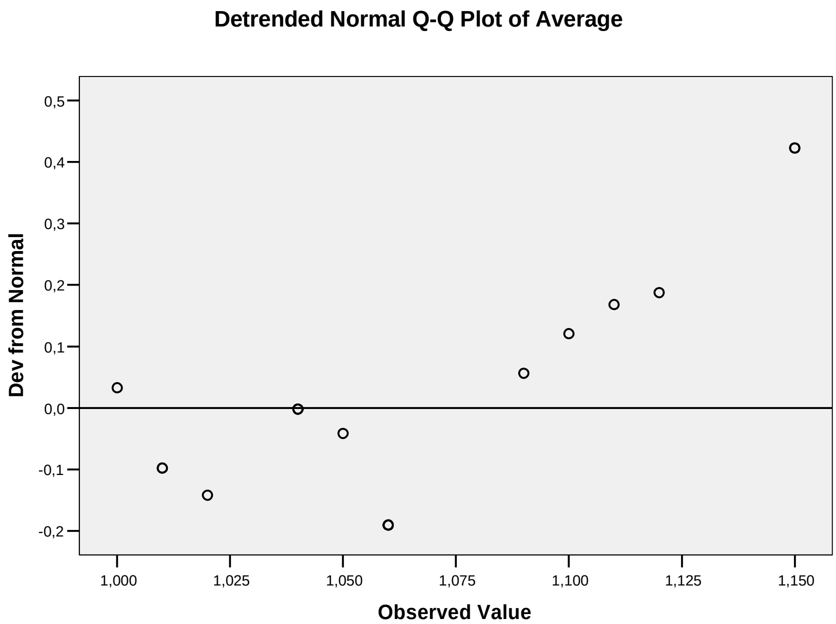
$Sig._1=0,028 < 0.05$ Επειδή $n=20 < 50$

<u>Shapiro-Wilk</u>

$Sig._1=0,418 > 0.05$

Άρα αποδεχόμαστε την H_0. Συνεπώς η μεταβλητή της μέσης τιμής ημέρας για την Αγροτική Τράπεζα είναι κανονικά κατανεμημένη.





Για τη μέση τιμή των μετοχών της μετοχής της Εμπορικής Τράπεζας έχουμε:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Emporiki_Trapeza	,256	20	,001	,862	20	,008

a. Lilliefors Significance Correction

<u>Kolmogorov – Smirnov:</u>

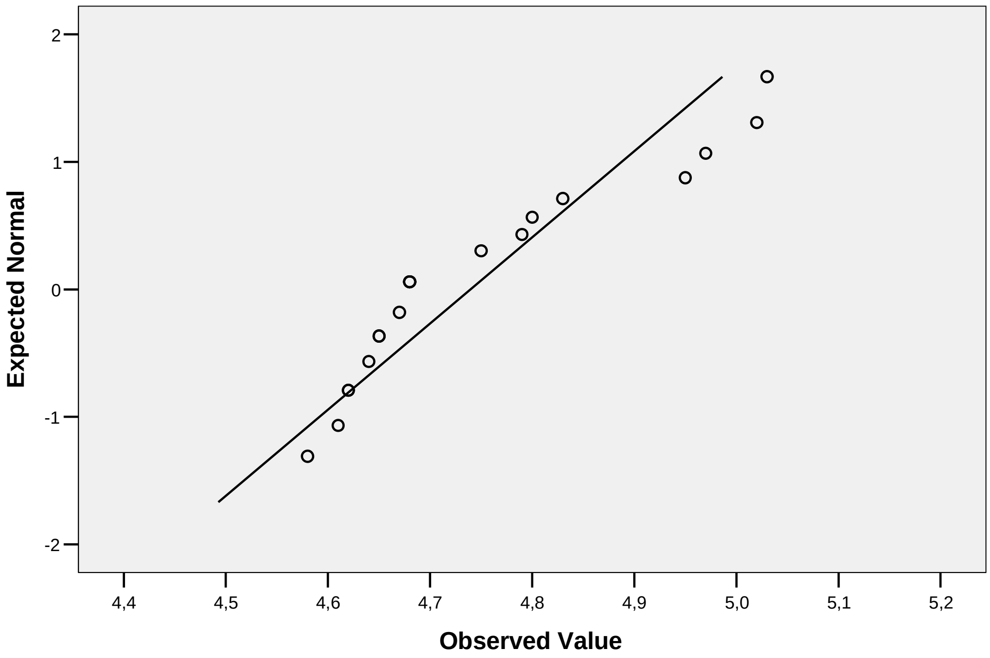
$Sig._1$=0,001 < 0.05

<u>Shapiro-Wilk</u>

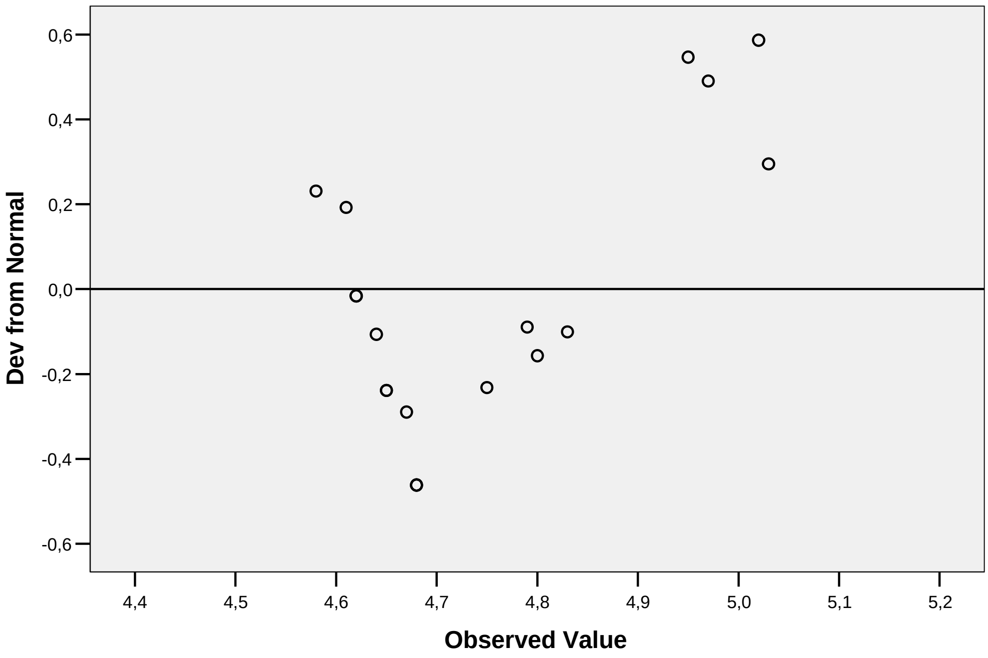
$Sig._1$=0,008 < 0.05

Αντιστοίχως, έχουμε αποδοχή της H_1 οπότε και για την Εμπορική Τράπεζα η μεταβλητή της μέσης τιμής ημέρας δεν είναι κανονικά κατανεμημένη.

Normal Q-Q Plot of Average



Detrended Normal Q-Q Plot of Average



Από τον έλεγχο κανονικότητας που κάναμε καταλήξαμε στο συμπέρασμα πως μόνο τρεις από τις μεταβλητές μας είναι κανονικά κατανεμημένες. Συνεπώς μόνο σε αυτές μπορούμε να διενεργήσουμε έλεγχο υποθέσεως για τη μέση τιμή.

1. Έλεγχος υποθέσεων για τη μέση τιμή του ανοίγματος των μετοχών.

$1^η$ περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1 \neq \mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας δεν είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας

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
Open	Equal variances assumed	14,547	,000	-98,138	38	,000	-3,67200	,03742	-3,74775	-3,59625
	Equal variances not assumed			-98,138	22,865	,000	-3,67200	,03742	-3,74943	-3,59457

Group Statistics

	Onoma_Metoxis	N	Mean	Std. Deviation	Std. Error Mean
Open	Agrotiki	20	1,0610	,05108	,01142
	Emporiki	20	4,7330	,15934	,03563

Θα κάνουμε έλεγχο διασπορών:

H_0: $s_1^2 = s_2^2$ Οι διασπορές είναι ίσες.

vs

H_1: $s_1^2 \neq s_2^2$ Οι διασπορές είναι άνισες.

Levene's Test $p=0,000 < \alpha=0,05$

Άρα απορρίπτουμε την H_0. Δηλαδή οι διακυμάνσεις είναι άνισες (συνεχίζουμε με τα δεδομένα της δεύτερης γραμμής από τον πίνακα).Οπότε:

$p=0,000<0,05$ απόρριψη H_0

$2^η$ περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1>\mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας μεγαλύτερη από την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε τους μέσους. Δηλαδή:

$Mean_1$= 1,0610 < $Mean_2$= 4,7330

δεν ικανοποιείται η H_1.

Οπότε $P=1-\frac{p-value}{2}=1-\frac{0}{2}=\sim 1$ > 0,05 άρα έχουμε αποδοχή της H_0.

3η περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1<\mu_2$ Η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας είναι μικρότερη από την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε και πάλι τους μέσους. Δηλαδή

$Mean_1$=1,0610 < $Mean_2$=4,7330

Άρα ικανοποιείται η H_1.

$$P=\frac{p-value}{2}=\frac{0}{2}=0<0{,}05$$ άρα έχουμε απόρριψη της H_0

Επομένως, η μέση τιμή του ανοίγματος της Αγροτικής Τράπεζας είναι μικρότερη από την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

2. **Έλεγχος υποθέσεων για τη μέση τιμή της υψηλής τιμής ημέρας των μετοχών.**

$1^η$ περίπτωση:

H_0: $\mu_1=\mu_2$ — Η μέση τιμή της υψηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1 \neq \mu_2$ — Η μέση τιμή της υψηλής τιμής της Αγροτικής Τράπεζας δεν είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας

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
High	Equal variances assumed	26,821	,000	-101,778	38	,000	-3,75100	,03685	-3,82561	-3,67639
	Equal variances not assumed			-101,778	22,979	,000	-3,75100	,03685	-3,82724	-3,67476

Group Statistics

	Onoma_Metoxis	N	Mean	Std. Deviation	Std. Error Mean
High	Agrotiki	20	1,0870	,05100	,01140
	Emporiki	20	4,8380	,15673	,03505

Πραγματοποιούμε έλεγχο διασπορών:

H_0: $s_1^2 = s_2^2$ Οι διασπορές είναι ίσες.

vs

H_1: $s_1^2 \neq s_2^2$ Οι διασπορές είναι άνισες.

Levene's test $p=0,000 < \alpha=0,05$

Άρα απορρίπτουμε την H_0. Οι διακυμάνσεις είναι άνισες (επιλέγουμε και πάλι τη δεύτερη γραμμή)

Οπότε:

$P=0,0000<0,05$ απόρριψη H_0

$2^{η}$ περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή της υψηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1>\mu_2$ Η μέση τιμή της υψηλής τιμής της Αγροτικής Τράπεζας είναι μεγαλύτερη από τη μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε τους μέσους $Mean_1=1{,}0870 < Mean_2=4{,}8380$. Άρα δεν ικανοποιείται η H_1.

$P=1-\frac{p-value}{2}=1-\frac{0}{2}=\sim 1 > 0{,}05$ άρα έχουμε αποδοχή της H_0.

$3^{η}$ περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή της υψηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1 < \mu_2$ Η μέση τιμή της υψηλής τιμής της Αγροτικής Τράπεζας είναι μικρότερη από τη την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε τους μέσους $Mean_1=1{,}0870 < Mean_2=4{,}8380$. Συνεπώς ικανοποιείται η H_1.

$$P=\frac{p-value}{2}=\frac{0}{2}=0<0{,}05$$ άρα έχουμε απόρριψη της H_0.

Επομένως, η μέση τιμή της υψηλής τιμής της Αγροτικής Τράπεζας είναι μικρότερη από τη την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

3. Έλεγχος υποθέσεων για τη μέση τιμή της χαμηλής τιμής ημέρας των μετοχών.

1η περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή της χαμηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1 \neq \mu_2$ Η μέση τιμή της χαμηλής τιμής της Αγροτικής Τράπεζας δεν είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας

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
Low	Equal variances assumed	17,022	,000	-116,319	38	,000	-3,61150	,03105	-3,67435	-3,54865
	Equal variances not assumed			-116,319	22,631	,000	-3,61150	,03105	-3,67579	-3,54721

Group Statistics

	Onoma_Metoxis	N	Mean	Std. Deviation	Std. Error Mean
Low	Agrotiki	20	1,0330	,04118	,00921
	Emporiki	20	4,6445	,13260	,02965

Έλεγχος διασπορών:

H_0: $s_1^2 = s_2^2$ Οι διασπορές είναι ίσες.

vs

H_1: $s_1^2 \neq s_2^2$ Οι διασπορές είναι άνισες.

Levene's test p=0,000 < α=0,05

Άρα απορρίπτουμε την H_0. Οι διακυμάνσεις είναι άνισες οπότε (επιλέγουμε και πάλι τη δεύτερη γραμμή του πίνακα):

p=0,000<0,05 Απόρριψη H_0

$2^η$ περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή της χαμηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: μ1>μ2 Η μέση τιμή της χαμηλής τιμής της Αγροτικής Τράπεζας είναι μεγαλύτερη από τη μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε τους μέσους

$Mean_1=1{,}0330 < Mean_2=4{,}6445$

Δεν ικανοποιείται η H_1.

$P=1-\frac{p-value}{2}=1-\frac{0}{2}=\sim 1 > 0{,}05$ άρα έχουμε αποδοχή της H_0.

3η περίπτωση:

H_0: $\mu_1=\mu_2$ Η μέση τιμή της χαμηλής τιμής ημέρας της Αγροτικής Τράπεζας είναι ίση με την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Vs

H_1: $\mu_1 < \mu_2$ — Η μέση τιμή της χαμηλής τιμής της Αγροτικής Τράπεζας είναι μικρότερη από τη την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

Συγκρίνουμε τους μέσους

$Mean_1=1{,}0330 < Mean_2=4{,}6445$

Συνεπώς ικανοποιείται η H_1.

$P=\frac{p-value}{2}=\frac{0}{2}=0<0{,}05$ άρα έχουμε απόρριψη της H_0.

Επομένως, η μέση τιμή της χαμηλής τιμής της Αγροτικής Τράπεζας είναι μικρότερη από τη την μέση τιμή του ανοίγματος της Εμπορικής Τράπεζας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΗ

1. ΠΕΤΡΟΣ Α. ΚΙΟΧΟΣ, «ΣΤΑΤΙΣΤΙΚΗ», INTERBOOKS, ΑΘΗΝΑ 1993.
2. ΓΕΩΡΓΙΟΥ Κ.ΓΕΩΡΓΙΟΥ, «ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΟΙΚΟΝΟΜΕΤΡΙΑ», GUTENBERG, ΑΘΗΝΑ 2002.
3. ΓΙΑΝΝΗΣ ΜΠΑΦΟΥΝΗΣ, «Η ΕΛΛΗΝΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΤΟΝ $19^{Ο}$ ΑΙΩΝΑ», ΕΤΑΙΡΕΙΑ ΜΕΛΕΤΗΣ ΝΕΟΥ ΕΛΛΗΝΙΣΜΟΥ, ΑΘΗΝΑ 2006.
4. ΖΑΪΡΗΣ, ΠΟΣΕΙΔΩΝΑΣ Ε., «ΣΤΑΤΙΣΤΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ», ΚΡΙΤΙΚΗ, ΑΘΗΝΑ 2007.
5. ΧΑΡΑΛΑΜΠΟΣ ΓΝΑΡΔΕΛΛΗΣ, «ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ», ΠΑΠΑΖΗΣΗΣ, ΑΘΗΝΑ 2003.
6. ΠΑΠΑΔΗΜΗΤΡΙΟΥ ΓΙΑΝΝΗΣ, «Η ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ», ΤΥΠΩΘΗΤΩ, ΑΘΗΝΑ 2007.
7. ΠΕΤΡΟΣ Α. ΚΙΟΧΟΣ, «ΕΦΑΡΜΟΓΕΣ ΕΠΑΓΩΓΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ», INTERBOOKS, ΑΘΗΝΑ 2007.
8. ΗΛΙΟΠΟΥΛΟΣ ΓΙΩΡΓΟΣ, «ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΠΑΡΑΜΕΤΡΩΝ», ΣΤΑΜΟΥΛΗ Α.Ε., ΑΘΗΝΑ 2006.
9. ΑΠΟΣΤΟΛΟΠΟΥΛΟΣ ΘΕΟΔΩΡΟΣ Η.,ΑΠΟΣΤΟΛΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ Θ., «ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ: ΠΕΡΙΓΡΑΦΙΚΗ ΚΑΙ ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ», ΣΥΓΧΡΟΝΗ ΕΚΔΟΤΙΚΗ, 2003.
10. ΧΑΛΙΚΙΑΣ ΙΩΑΝΝΗΣ Γ., «ΣΤΑΤΙΣΤΙΚΗ : ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΓΙΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΑΠΟΦΑΣΕΙΣ», ROSILI, ΑΘΗΝΑ 2003.
11. ΜΠΕΝΟΣ ΒΑΣΙΛΕΙΟΣ Κ.. «ΕΦΑΡΜΟΓΕΣ ΕΠΑΓΩΓΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ», ΣΤΑΜΟΥΛΗ Α.Ε.

12.ΡΟΥΣΣΑΣ ΓΕΩΡΓΙΟΣ Γ.. «ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ», ΖΗΤΗ, ΑΘΗΝΑ 1994.
13.ΠΑΡΑΣΚΕΥΟΠΟΥΛΟΣ ΙΩΑΝΝΗΣ Ν., «ΣΤΟΙΧΕΙΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΚΑΙ ΕΠΑΓΩΓΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ», ΙΔΙΩΤΙΚΗ ΈΚΔΟΣΗ, ΑΘΗΝΑ 1984.
14.ΣΤΑΜΑΤΕΛΟΣ ΓΕΩΡΓΙΟΣ, «ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ II», ΖΗΤΗ, ΑΘΗΝΑ 1998.
15.ΔΡΟΣΟΣ ΓΕΩΡΓΙΟΣ, «ΣΤΑΤΙΣΤΙΚΗ & ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ», ΑΝΙΚΟΥΛΑ, ΘΕΣΣΑΛΟΝΙΚΗ 2006.

ΞΕΝΗ.

1. RICHARDSON CLIVE, ΒΑΣΙΛΑΙΝΑΣ ΑΘΑΝΑΣΙΟΣ, «ΣΤΟΙΧΕΙΑ ΕΠΑΓΩΓΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ», ΚΑΚΤΟΣ.
2. GRAIS BERNARD, «ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ», ΤΥΠΩΘΗΤΩ, ΑΘΗΝΑ 2005.

ΠΑΡΑΡΤΗΜΑ ΠΙΝΑΚΕΣ

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.878	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large	2.575	2.327	1.960	1.645	1.282	.675

Chi Square Distribution Table

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	32.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	32.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Table from Ronald J. Wonnacott and Thomas H. Wonnacott, *Statistics: Discovering Its Power*, New York: John Wiley and Sons, 1982, p.352.

F - Distribution (α = 0.01 in the Right Tail)

Numerator Degrees of Freedom

Denominator Degrees of Freedom

df_2 \ df_1	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

F - Distribution ($\alpha = 0.01$ in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	10	12	15	20	24	30	40	60	120	∞
1	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.9
2	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	10.051	9.8883	9.7222	9.5526	9.4665	9.3793	9.2912	9.2020	9.1118	9.0204
6	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0567	6.9690	6.8800
7	6.6201	6.4691	6.3143	6.1554	6.0743	5.9920	5.9084	5.8236	5.7373	5.6495
8	5.8143	5.6667	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9461	4.8588
9	5.2565	5.1114	4.9621	4.8080	4.7290	4.6486	4.5666	4.4831	4.3978	4.3105
10	4.8491	4.7059	4.5581	4.4054	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.6904	3.6024
12	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	4.1003	4.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	3.8049	3.6662	3.5222	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	3.6909	3.5527	3.4089	3.2587	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8348	2.7459	2.6530
18	3.5082	3.3706	3.2273	3.0771	2.9990	2.9185	2.8354	2.7493	2.6597	2.5660
19	3.4338	3.2965	3.1533	3.0031	2.9249	2.8442	2.7608	2.6742	2.5839	2.4893
20	3.3682	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	3.3098	3.1730	3.0300	2.8796	2.8010	2.7200	2.6359	2.5484	2.4568	2.3603
22	3.2576	3.1209	2.9779	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	3.2106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2558
24	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4923	2.4035	2.3100	2.2107
25	3.1294	2.9931	2.8502	2.6993	2.6203	2.5383	2.4530	2.3637	2.2696	2.1694
26	3.0941	2.9578	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	3.0618	2.9256	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1985	2.0965
28	3.0320	2.8959	2.7530	2.6017	2.5223	2.4397	2.3535	2.2629	2.1670	2.0642
29	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1379	2.0342
30	2.9791	2.8431	2.7002	2.5487	2.4689	2.3860	2.2992	2.2079	2.1108	2.0062
40	2.8005	2.6648	2.5216	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	2.3209	2.1847	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000