Hellenic Republic
INTERNATIONAL HELLENIC UNIVERSITY
University Center for International Programmes of Studies

Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας
Πρόγραμμα Μεταπτυχιακών Σπουδών:
Διοίκηση Επιχειρήσεων και Οργανισμών για Στελέχη

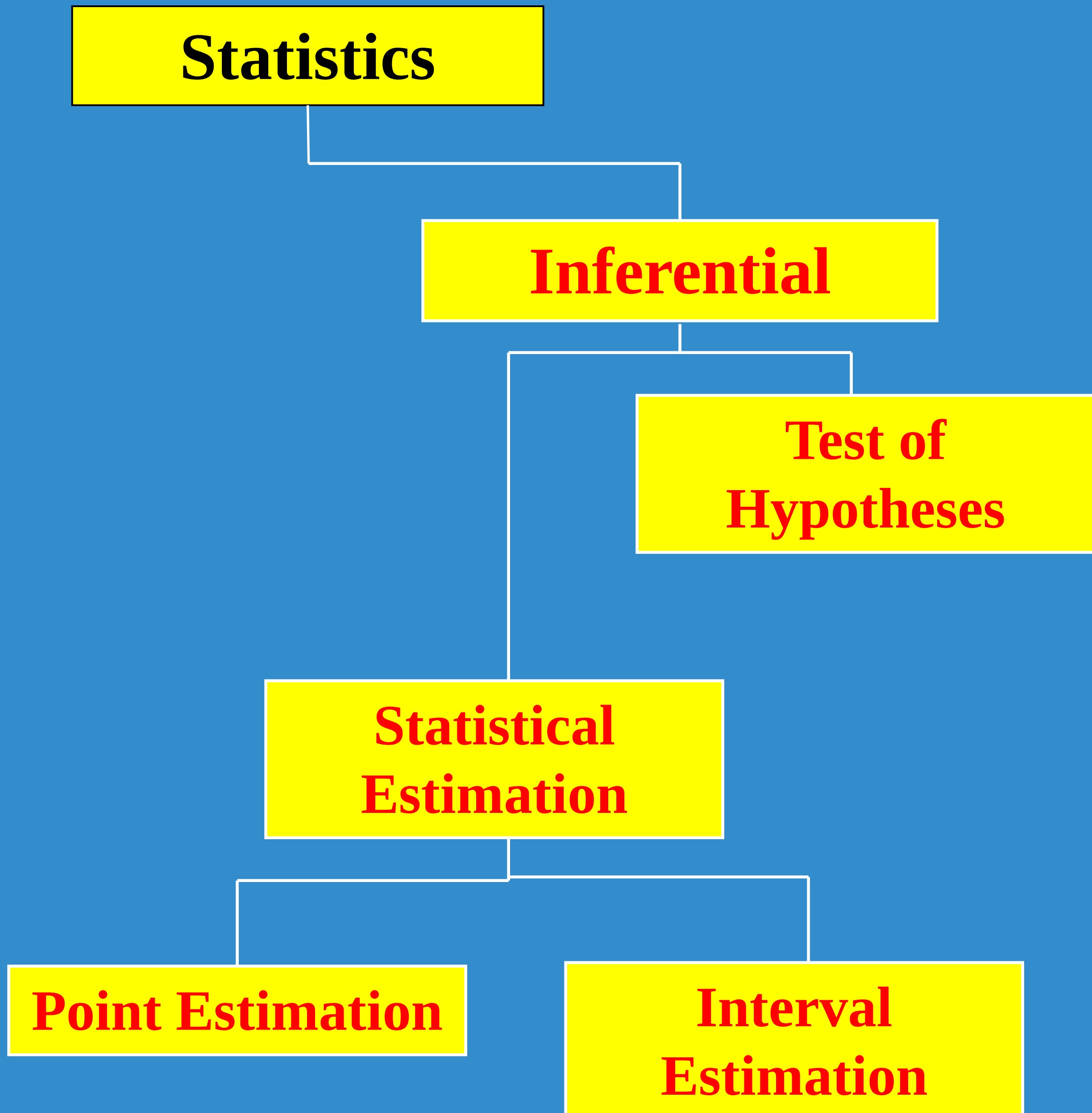
mba executive

Ποσοτικές Μέθοδοι για Στελέχη Επιχειρήσεων
Quantitative Methods for Managers
Διαστήματα Εμπιστοσύνης-Confidence Intervals

by
Dr. Efstathios Dimitriadis
Mathematic
Ph.D in Applied Statistics
M.Sc in Statistics and Demography
M.Sc in Quality Assurance

Διαστήματα Εμπιστοσύνης

Confidence Intervals


Statistics
Inferential
Test of Hypotheses
Statistical Estimation
Point Estimation
Interval Estimation

Παράδειγμα

Φανταστείτε ότι είστε αναλυτής δεδομένων σε ένα πολύ μεγάλο e-shop το οποίο έχει 100 εκατομμύρια πελάτες. Φανταστείτε ότι τα 50 εκατομμύρια είναι άνδρες πελάτες και τα 50 εκατομμύρια είναι γυναίκες πελάτες.

Θέλετε να δείτε:

1. Αν το ποσό των δαπανών διαφέρει μεταξύ των δύο φύλων. Η συγκεκριμένη ερώτηση για την οποία χρειάζεστε απάντηση είναι: Κατά μέσο όρο οι γυναίκες ξοδεύουν περισσότερα ανά συναλλαγή από τους άνδρες; (t- test of equality of means)
2. Το ποσό των δαπανών των γυναικών.

Για τους ανωτέρω ελέγχους πάρθηκε ένα δείγμα 30.000 γυναικών και 30.000 ανδρών και καταγράφηκε το ποσό των δαπανών.

Για τον δεύτερο έλεγχο υπάρχουν δύο προσεγγίσεις:

1. Καταλήγετε στο συμπέρασμα ότι η μέση δαπάνη 30.000 γυναικών πελατών, (μέσος όρος δείγματος=2.350€) είναι ίση με τη μέση δαπάνη και των 50 εκατομμυρίων γυναικών πελατών (μέσος όρος πληθυσμού). Αυτό ονομάζεται **εκτίμηση σημείου** όπου χρησιμοποιήσατε τα δείγματα δεδομένων για να καταλήξετε στην καλύτερη εικασία μιας παραμέτρου άγνωστου πληθυσμού.
2. Ένας καλύτερος και πιο πειστικός τρόπος είναι να χρησιμοποιήσετε αυτόν τον μέσο όρο του δείγματος για να βρείτε ένα **διάστημα** εντός του οποίου θα βρίσκεται ο μέσος όρος του πληθυσμού (2.340-2.360€). Χρησιμοποιώντας αυτό το δείγμα 30.000 γυναικών πελατών θα υπολογίσετε το διάστημα εντός του οποίου μπορεί να βρίσκεται η μέση δαπάνη 50 εκατομμυρίων γυναικών πελατών.

Εισαγωγή

Όταν κάνετε μια εκτίμηση για μια στατιστική παράμετρο (π.χ μέση τιμή ή τυπική απόκλιση), υπάρχει πάντα αβεβαιότητα γύρω από αυτήν την εκτίμηση επειδή βασίζεται (η εκτίμηση) σε ένα δείγμα του πληθυσμού που μελετάτε.

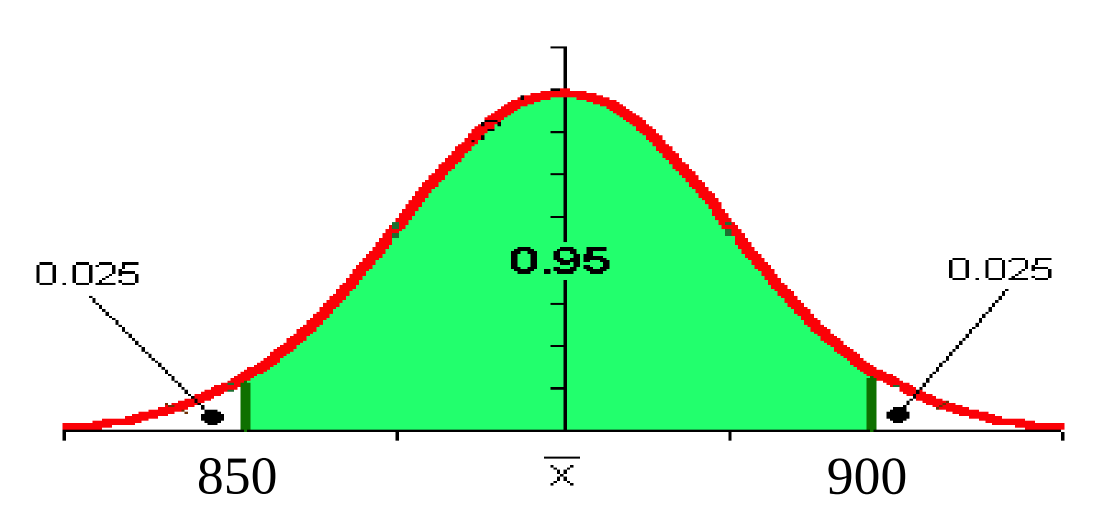
Το **διάστημα εμπιστοσύνης** είναι το εύρος των τιμών εντός του οποίου αναμένεται ότι θα βρίσκεται η μέση τιμή του πληθυσμού της παραμέτρου την οποία εκτιμήσατε. Αν δηλαδή είχατε τη δυνατότητα να πάρετε πληροφορίες από ολόκληρο τον πληθυσμό και να υπολογίσετε την μέση τιμή, αυτή θα βρίσκεται εντός των ορίου του διαστήματος εμπιστοσύνης, με συγκεκριμένο επίπεδο εμπιστοσύνης.

Επίπεδο εμπιστοσύνης (1-α) είναι η % πιθανότητα, η υπό εξέταση παράμετρος να βρίσκεται εντός του διαστήματος εμπιστοσύνης.

Επίπεδο σημαντικότητας (α) είναι η % πιθανότητα η εκτίμησή σας να μην είναι σωστή και τελικά η παράμετρος του πληθυσμού να βρίσκεται εκτός των ορίων του διαστήματος εμπιστοσύνης.

Παράδειγμα:

Έστω ότι θέλουμε να εκτιμήσουμε το μέσο μηνιαίο εισόδημα των εργαζομένων σε μια επιχείρηση. Από ένα δείγμα εργαζομένων προέκυψε ότι το διάστημα αυτό είναι (850€ - 900€), σε επίπεδο σημαντικότητας α=5%.



Ερμηνεία: Εκτιμάται ότι το μέσο μηνιαίο εισόδημα του συνόλου των εργαζομένων (πληθυσμού) θα κυμαίνεται μεταξύ 850€ και 900€, με πιθανότητα επιτυχούς εκτίμησης 95%.
Η πιθανότητα η μέση τιμή του εισοδήματος να είναι μεγαλύτερη των 900€ είναι 2,5%
Η πιθανότητα η μέση τιμή του εισοδήματος να είναι μικρότερη των 850€ είναι 2,5%
Τελικά, η πιθανότητα λάθους της εκτίμησης είναι συνολικά 5% (α).
Η πιθανότητα σωστής εκτίμησης είναι 95% (1-α).

Εφαρμογές των Διαστημάτων Εμπιστοσύνης/1

Βιολογία

Τα διαστήματα εμπιστοσύνης χρησιμοποιούνται συχνά στη βιολογία για την εκτίμηση του μέσου ύψους, βάρους, πλάτους, διαμέτρου κ.λπ. διαφορετικών φυτικών και ζωικών ειδών.

Για παράδειγμα, ένας βιολόγος μπορεί να ενδιαφέρεται να μετρήσει το μέσο βάρος ενός συγκεκριμένου είδους βατράχου στην Αυστραλία. Δεδομένου ότι θα χρειαζόταν πολύς χρόνος για να περιηγηθούν και να ζυγιστούν χιλιάδες μεμονωμένα βατράχια, ο βιολόγος μπορεί να συλλέξει ένα απλό τυχαίο δείγμα 50 βατράχων και να μετρήσει τη μέση και τυπική απόκλιση των βατράχων στο δείγμα.

Στη συνέχεια θα μπορούσε να χρησιμοποιήσει τον μέσο όρο του δείγματος και την τυπική απόκλιση του δείγματος για να δημιουργήσει ένα διάστημα για τον πραγματικό μέσο όρο των βατράχων σε ολόκληρο τον πληθυσμό.

Εφαρμογές των Διαστημάτων Εμπιστοσύνης/2

Κλινικές δοκιμές

Τα διαστήματα εμπιστοσύνης χρησιμοποιούνται συχνά σε κλινικές δοκιμές για τον προσδιορισμό της μέσης αλλαγής στην αρτηριακή πίεση, τον καρδιακό ρυθμό, τη χοληστερόλη κ.λπ. που παράγεται από κάποιο νέο φάρμακο ή θεραπεία.

Για παράδειγμα, ένας γιατρός μπορεί να πιστεύει ότι ένα νέο φάρμακο είναι ικανό να μειώσει την αρτηριακή πίεση στους ασθενείς. Για να το δοκιμάσει, μπορεί να στρατολογήσει 20 ασθενείς για να συμμετάσχουν σε μια δοκιμή στην οποία χρησιμοποίησαν το νέο φάρμακο για ένα μήνα. Στο τέλος του μήνα, ο γιατρός μπορεί να καταγράψει τη μέση μείωση της αρτηριακής πίεσης και την τυπική απόκλιση της μείωσης σε κάθε ασθενή στο δείγμα.

Στη συνέχεια, θα μπορούσε να χρησιμοποιήσει τον μέσο όρο του δείγματος και την τυπική απόκλιση του δείγματος για να δημιουργήσει ένα διάστημα για την πραγματική μέση μεταβολή της αρτηριακής πίεσης που είναι πιθανό να βιώσουν οι ασθενείς στον πληθυσμό.

Εφαρμογές των Διαστημάτων Εμπιστοσύνης/3

Διαφήμιση

Τα διαστήματα εμπιστοσύνης χρησιμοποιούνται συχνά από τα τμήματα μάρκετινγκ εντός των εταιρειών για να καθορίσουν εάν κάποια νέα διαφημιστική τεχνική, μέθοδος, τακτική κ.λπ. παράγει σημαντικά υψηλότερα έσοδα.

Για παράδειγμα, μια ομάδα μάρκετινγκ σε έναν λιανοπωλητή παντοπωλείου μπορεί να εκτελέσει δύο διαφορετικές διαφημιστικές καμπάνιες σε 20 διαφορετικά καταστήματα το καθένα κατά τη διάρκεια ενός τριμήνου και να μετρήσει τις μέσες πωλήσεις που παράγονται από κάθε καμπάνια σε κάθε κατάστημα στο τέλος του τριμήνου.

Στη συνέχεια, θα μπορούσαν να χρησιμοποιήσουν τη μέση τιμή δείγματος και την τυπική απόκλιση των πωλήσεων από κάθε καμπάνια για να δημιουργήσουν ένα διάστημα εμπιστοσύνης για τη διαφορά μεταξύ των μέσων πωλήσεων. Αυτό θα πει στην ομάδα μάρκετινγκ εάν υπάρχει κάποια σημαντική διαφορά στις πωλήσεις που προκύπτει ως αποτέλεσμα των δύο καμπανιών.

Εφαρμογές των Διαστημάτων Εμπιστοσύνης/4

Βιομηχανία

Τα διαστήματα εμπιστοσύνης χρησιμοποιούνται συχνά από τους μηχανικούς στα εργοστάσια παραγωγής για να προσδιορίσουν εάν κάποια νέα διαδικασία, τεχνική, μέθοδος κ.λπ. προκαλεί σημαντική αλλαγή στον αριθμό των ελαττωματικών προϊόντων που παράγονται από το εργοστάσιο.

Για παράδειγμα, ένας μηχανικός μπορεί να πιστεύει ότι μια νέα διαδικασία θα αλλάξει τον αριθμό των ελαττωματικών widget που παράγονται ανά ημέρα, που είναι επί του παρόντος 50. Για να το δοκιμάσει, μπορεί να εφαρμόσει τη νέα διαδικασία και να καταγράψει τον αριθμό των ελαττωματικών προϊόντων που παράγονται κάθε μέρα για ένα μήνα στο εργοστάσιο.

Στη συνέχεια, θα μπορούσε να χρησιμοποιήσει τον μέσο όρο του δείγματος και την τυπική απόκλιση δείγματος του αριθμού των ημερήσιων ελαττωμάτων για να δημιουργήσει ένα διάστημα εμπιστοσύνης για τον πραγματικό μέσο αριθμό ελαττωματικών προϊόντων που παράγονται από τη νέα διαδικασία.

Εάν το διάστημα εμπιστοσύνης δεν περιέχει την τιμή "50", τότε ο μηχανικός μπορεί να είναι βέβαιος ότι η νέα διαδικασία παράγει διαφορετικό αριθμό καθημερινών ελαττωματικών προϊόντων σε σύγκριση με την τρέχουσα διαδικασία.

Εκτίμηση διαστήματος (Ένας πληθυσμός)

Α. Διάστημα Εμπιστοσύνης για την μέση τιμή μ του πληθυσμού

Β. Διάστημα Εμπιστοσύνης για την αναλογία p του πληθυσμού

Για να δημιουργήσετε ένα διάστημα εμπιστοσύνης θα πρέπει να λάβετε υπόψη:

1. Τη μορφή της κατανομής του πληθυσμού (Κανονική ή Κατά προσέγγιση κανονική).
2. Το μέγεθος του δείγματος (Μικρό ή Μεγάλο).
3. Την τυπική απόκλιση του πληθυσμού (Γνωστή ή Άγνωστη)

Μαθησιακά αποτελέσματα

Σε αυτό το κεφάλαιο θα μάθετε:

1. Πώς να υπολογίζετε το περιθώριο σφάλματος που σχετίζεται με ένα μέσο δείγματος και μια αναλογία δείγματος.

2. Πώς να χρησιμοποιήσετε αυτές τις πληροφορίες για να κατασκευάσετε και να ερμηνεύσετε τις εκτιμήσεις του διαστήματος εμπιστοσύνης ενός μέσου πληθυσμού και μιας αναλογίας πληθυσμού.

3. Πώς να καθορίζετε το μέγεθος του δείγματος που απαιτείται για να διασφαλίσετε ότι το περιθώριο σφάλματος θα είναι εντός αποδεκτών ορίων.

A1. Διάστημα Εμπιστοσύνης για την μέση τιμή μ του πληθυσμού

Μεγάλο Δείγμα (n>30) με γνωστό σ

a = Επίπεδο Σημαντικότητας
1-a = Επίπεδο Εμπιστοσύνης

$$\text{Δειγματοληπτικό Σφάλμα} = \left|\bar{\chi} - \mu\right|$$

Περιθώριο Σφάλματος: $E = Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Κατώτερο όριο: $\bar{x} - Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Ανώτερο όριο: $\bar{x} + Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Διάστημα Εμπιστοσύνης: $\bar{x} \pm Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Εύρος: $R = 2 \cdot Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Παράδειγμα 1ο

Από πληθυσμό 4.000 νοικοκυριών μίας περιοχής επιλέγεται τυχαίο δείγμα 250 με σκοπό να εκτιμηθεί το μέσο μηνιαίο εισόδημα των νοικοκυριών της περιοχής. Το δείγμα έδωσε μέσο εισόδημα 940 Euro. Αν είναι γνωστό ότι η τυπική απόκλιση του πληθυσμού είναι 21 Euro, να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για το μέσο εισόδημα των νοικοκυριών.

Λύση:

$$\mathrm{N} = 4000 \quad \mathrm{n} = 250 \quad \bar{x} = 940€ \quad \sigma = 21 \quad \alpha = 5\%$$

$$\bar{x} \pm Z_{a/2} \cdot \frac{\sigma}{\sqrt{\mathrm{n}}} = 940 \pm 1{,}96 \cdot \frac{21}{\sqrt{250}} = 940 \pm 2{,}6 = (937{,}4 - 942{,}6)$$

A2. Διάστημα Εμπιστοσύνης για την μέση τιμή μ του πληθυσμού

Μεγάλο Δείγμα (n>30) με άγνωστο σ (εκτιμώμενο από το s)

a = Επίπεδο Σημαντικότητας
1-a = Επίπεδο Εμπιστοσύνης

$$\text{Δειγματοληπτικό Σφάλμα} = \left|\bar{\chi} - \mu\right|$$

Περιθώριο Σφάλματος: $E = Z_{a/2} \cdot \frac{s}{\sqrt{n}}$

Κατώτερο όριο: $\bar{x} - Z_{a/2} \cdot \frac{s}{\sqrt{n}}$

Ανώτερο όριο: $\bar{x} + Z_{a/2} \cdot \frac{s}{\sqrt{n}}$

Διάστημα Εμπιστοσύνης: $\bar{x} \pm Z_{a/2} \cdot \frac{s}{\sqrt{n}}$

Εύρος: $R = 2 \cdot Z_{a/2} \cdot \frac{s}{\sqrt{n}}$

Παράδειγμα 2ο

Από πληθυσμό 4.000 νοικοκυριών μίας περιοχής επιλέγεται τυχαίο δείγμα 250 με σκοπό να εκτιμηθεί το μέσο μηνιαίο εισόδημα των νοικοκυριών της περιοχής. Το δείγμα έδωσε μέσο εισόδημα 940 Euro και τυπική απόκλιση 21 Euro. Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για το μέσο εισόδημα των νοικοκυριών.

Λύση:

$$\mathrm{N} = 4000 \quad \mathrm{n} = 250 \quad \bar{x} = 940€ \quad \mathrm{s} = 21 \quad \alpha = 5\%$$

$$\bar{x} \pm Z_{a/2} \cdot \frac{s}{\sqrt{\mathrm{n}}} = 940 \pm 1{,}96 \cdot \frac{21}{\sqrt{250}} = 940 \pm 2{,}6 = (937{,}4 - 942{,}6)$$

A3. Διάστημα Εμπιστοσύνης για την μέση τιμή μ του πληθυσμού

Μικρό Δείγμα (n<=30) με γνωστό σ

a = Επίπεδο Σημαντικότητας
1-a = Επίπεδο Εμπιστοσύνης

$$\text{Δειγματοληπτικό Σφάλμα} = |\bar{\chi} - \mu|$$

Περιθώριο Σφάλματος: $E = Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Κατώτερο όριο: $\bar{x} - Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Ανώτερο όριο: $\bar{x} + Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Διάστημα Εμπιστοσύνης: $\bar{x} \pm Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

Εύρος: $R = 2 \cdot Z_{a/2} \cdot \frac{\sigma}{\sqrt{n}}$

A4. Διάστημα Εμπιστοσύνης για την μέση τιμή μ του πληθυσμού:

Μικρό Δείγμα (n=<30) με άγνωστο σ (εκτιμώμενο από το s)

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

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

Περιθώριο Σφάλματος: $E = t_{a/2} \cdot \frac{s}{\sqrt{n}}$

Κατώτερο όριο: $\bar{x} - t_{a/2} \cdot \frac{s}{\sqrt{n}}$

Δειγματοληπτικό Σφάλμα $= |\bar{\chi} - \mu|$

Ανώτερο όριο: $\bar{x} + t_{a/2} \cdot \frac{s}{\sqrt{n}}$

Διάστημα Εμπιστοσύνης: $\bar{x} \pm t_{a/2} \cdot \frac{s}{\sqrt{n}}$

$d.f$: n-1

Εύρος: $R = 2 \cdot t_{a/2} \cdot \frac{s}{\sqrt{n}}$

Παράδειγμα 3ο

Στη δοκιμή μιας νέας μεθόδου παραγωγής, επιλέχθηκαν τυχαία 18 εργαζόμενοι και τους ζητήθηκε να δοκιμάσουν τη νέα μέθοδο. Ο μέσος ρυθμός παραγωγής του δείγματος για τους 18 υπαλλήλους ήταν 80 μέρη ανά ώρα και η τυπική απόκλιση του δείγματος ήταν 10 μέρη ανά ώρα. Δώστε ένα διαστήματα εμπιστοσύνης 90% και 95% για το μέσο ρυθμό παραγωγής του πληθυσμού για τη νέα μέθοδο, υποθέτοντας ότι ο πληθυσμός έχει κανονική κατανομή πιθανοτήτων.

Λύση

$$n = 18 \quad \bar{x} = 80 \quad s = 10 \quad \alpha = 10\% \quad \alpha = 5\%$$

t- student κατανομή με n-1=18-1=17 d.f (βαθμοί ελευθερίας= β.ε)

Για 90% confidence interval και 17 d.f $t_{0,1/2}=t_{0,05}$= 1,740 : $80 \pm 1{,}740 \cdot \frac{10}{\sqrt{18}} = 80 \pm 4{,}1$

(75,89-84,1) 90% δ.ε

For 95% confidence interval και 17 d.f $t_{0,05/2}=t_{0,025}$= 2,110: $80 \pm 2{,}110 \cdot \frac{10}{\sqrt{18}} = 80 \pm 4{,}97$

(75,03-84,97) 95% δ.ε

Μέγεθος Δείγματος για τη Μέση Τιμή του Πληθυσμού

$$\mathbf{n} = \frac{(\mathbf{Z}_{a/2})^2 \cdot \boldsymbol{\sigma}^2}{\mathbf{E}^2}$$

όπου E = Το επιθυμητό εύρος του σφάλματος

$$n = \frac{(2 \cdot Z_{a/2})^2 \cdot \sigma^2}{R^2}$$

όπου R = Το επιθυμητό εύρος του διαστήματος

!!! Όσο μικραίνουν το E και το R, μεγαλώνει το μέγεθος του δείγματος

!!! Όσο μικραίνει η τιμή του επιπέδου σημαντικότητας (α) μεγαλώνει το μέγεθος του δείγματος

!!! Όσο μικραίνει η τιμή του επιπέδου σημαντικότητας (α) μεγαλώνει το εύρος του διαστήματος εμπιστοσύνης

Παράδειγμα 4ο

Από πληθυσμό 4.000 νοικοκυριών μίας περιοχής επιλέγεται τυχαίο δείγμα με σκοπό να εκτιμηθεί το μέσο μηνιαίο εισόδημα των νοικοκυριών της περιοχής. Το δείγμα έδωσε μέσο εισόδημα 940 Euro και τυπική απόκλιση 21 Euro. α=5%

1. Να υπολογισθεί το μέγεθος του δείγματος αν το επιθυμητό εύρος του σφάλματος είναι 1,5 €.
2. Να υπολογισθεί το μέγεθος του δείγματος αν το επιθυμητό εύρος του διαστήματος είναι 3 €.

Λύση:

$$n = \frac{(Z_{a/2})^2 \cdot \sigma^2}{\mathrm{E}^2} = \frac{1{,}96^2 \cdot 21^2}{1{,}5^2} = 753$$

$$n = \frac{(2Z_{a/2})^2 \cdot \sigma^2}{R^2} = \frac{(2 \cdot 1{,}96)^2 \cdot 21^2}{3^2} = 753$$

Β1. Διάστημα Εμπιστοσύνης για την Αναλογία p του Πληθυσμού

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

$\mathbf{p}$ = Αναλογία Πληθυσμού

$\overline{\mathbf{p}}$ = Αναλογία Δείγματος

$$\overline{\mathbf{p}} = \frac{\mathbf{x}}{\mathbf{n}}$$ = Αναλογία Δείγματος

όπου: **χ** το πλήθος των επιτυχιών,

n Μέγεθος δείγματος

Με στόχο την επίτευξη των καλλίτερων αποτελεσμάτων, είναι απαραίτητο να λάβουμε υπόψη τις παρακάτω συνθήκες:

1. Το μέγεθος του δείγματος πρέπει να είναι μεγαλύτερο των 20.
2. $np>5$ και $n(1-p)>5$
3. Το δείγμα πρέπει να είναι μικρότερο του 10% του πληθυσμού.

Αν το p είναι γνωστό:

Περιθώριο Σφάλματος: $E = Z_{a/2} \cdot \sigma_{\overline{p}}$ Όπου: $\sigma_{\overline{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Κατώτερο όριο: $\overline{p} - Z_{a/2}\sqrt{\dfrac{p(1-p)}{n}}$

Ανώτερο όριο: $\overline{p} + Z_{a/2}\sqrt{\dfrac{p(1-p)}{n}}$

Διάστημα Εμπιστοσύνης: $\overline{p} \pm Z_{a/2}\sqrt{\dfrac{p(1-p)}{n}}$

Εύρος: $2 \cdot Z_{a/2}\sqrt{\dfrac{p(1-p)}{n}}$

$\textit{Sampling}\ \text{Error} = \left|\overline{p} - p\right|$

Αν το p είναι άγνωστο στη θέση του χρησιμοποιούμε το $\overline{p}$

Διάστημα Εμπιστοσύνης: $\overline{p} \pm Z_{a/2}\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

Παράδειγμα 5ο

Έρευνα έδειξε ότι 104 από τους 400 χρήστες του ιστότοπου ESPN ήταν γυναίκες. Ποιο είναι το 95% διάστημα εμπιστοσύνης για το ποσοστό των γυναικών χρηστών στον πληθυσμό;

Λύση

Ποσοστό γυναικών στο δείγμα: $\overline{p} = \dfrac{104}{400} = 0,26$

Ποσοστό γυναικών στο δείγμα: $1 - \overline{p} = 1 - 0,26 = 0,74$

$Z_{0,05/2} = Z_{0,025} = 1,96$

$$\overline{p} \pm Z_{a/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0,26 \pm 1,96\sqrt{\frac{0,26(1-0,26)}{400}} = 0,26 \pm 0,043 = (0,217 - 0,303)$$

Το Ποσοστό γυναικών στον πληθυσμό κυμαίνεται από 21,7% έως 30,3%.

Μέγεθος Δείγματος για την Εκτίμηση της Αναλογίας p του Πληθυσμού

$$n = \frac{(Z_{a/2})^2 \cdot p(1-p)}{E^2}$$

όπου E = Επιθυμητό Περιθώριο Σφάλματος

$$n = \frac{(2 \cdot Z_{a/2})^2 \cdot p(1-p)}{R^2}$$

όπου R = Επιθυμητό Εύρος

!!! Όσο μικραίνουν το E και το R, μεγαλώνει το μέγεθος του δείγματος

!!! Όσο μικραίνει η τιμή του επιπέδου σημαντικότητας (α) μεγαλώνει το μέγεθος του δείγματος

!!! Όσο μικραίνει η τιμή του επιπέδου σημαντικότητας (α) μεγαλώνει το εύρος του διαστήματος εμπιστοσύνης

Παράδειγμα 6ο

Στο προηγούμενο παράδειγμα, πόσο μεγάλο πρέπει να είναι το μέγεθος του δείγματος αν το επιθυμητό σφάλμα είναι 3%;

$$n = \frac{(Z_{a/2})^2 \cdot p(1-p)}{\mathrm{E}^2} = \frac{1{,}96^2 \cdot 0.26(1-0.26)}{0{,}03^2} = 821$$

Πόσο μεγάλο πρέπει να είναι το μέγεθος του δείγματος αν το επιθυμητό εύρος του διαστήματος είναι 6%;

$$n = \frac{(2 \cdot Z_{a/2})^2 \cdot p(1-p)}{R^2} = \frac{(2 \cdot 1{,}96)^2 \cdot 0{,}26 \cdot 0{,}74}{0{,}06^2} = 821$$

Διάστημα Εμπιστοσύνης (Δύο Πληθυσμοί)

A. Διάστημα Εμπιστοσύνης για τη διαφορά των Μέσων Τιμών δύο Πληθυσμών: **Ανεξάρτητα Δείγματα**

B. : Διάστημα Εμπιστοσύνης για τη διαφορά των Μέσων Τιμών δύο Πληθυσμών: **Εξαρτημένα Δείγματα**

C. Διάστημα Εμπιστοσύνης για τη διαφορά των Αναλογιών δύο Πληθυσμών

Μαθησιακά Αποτελέσματα

Σε αυτό το κεφάλαιο θα μάθετε:

1. Πώς να δημιουργείτε ένα διάστημα εμπιστοσύνης για τις μέσες τιμές και τις αναλογίες δύο πληθυσμών..
2. Πώς να αναλύετε ανεξάρτητα τυχαία δείγματα όπως επίσης και εξαρτημένα δείγματα.

Για να δημιουργήσετε ένα διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών θα πρέπει να λάβετε υπόψη:

1. Τη μορφή της κατανομής των πληθυσμών (Κανονική ή Κατά προσέγγιση κανονική).
2. Το μέγεθος των δειγμάτων (Μικρά ή Μεγάλα).
3. Την τυπική απόκλιση των πληθυσμών (Γνωστή ή Άγνωστη)

A1. Εκτίμηση της διαφοράς των μέσων τιμών δύο πληθυσμών

μ_1-μ_2: Ανεξάρτητα Δείγματα

Περίπτωση Μεγάλων Δειγμάτων με σ_1 and σ_2 Γνωστές

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

$n_1>30$ και $n_2>30$

n_1 = Δείγμα από τον πληθυσμό 1

n_2 = Δείγμα από τον πληθυσμό 2

σ_1 = Τυπική Απόκλιση από τον πληθυσμό 1

σ_2 = Τυπική Απόκλιση από τον πληθυσμό 2

Περίπτωση Μεγάλων Δειγμάτων με σ_1 and σ_2 Γνωστές

Τυπική Απόκλιση της Διαφοράς $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Περιθώριο Σφάλματος $E = Z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Διάστημα Εμπιστοσύνης: $(\bar{x}_1 - \bar{x}_2) \pm Z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Εύρος: $R = 2 \cdot Z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Παράδειγμα 7ο

Το Υπουργείο Μεταφορών των ΗΠΑ παρέχει τον αριθμό των μιλίων που διανύουν οι κάτοικοι των 75 μεγαλύτερων μητροπολιτικών περιοχών την ημέρα με ένα αυτοκίνητο. Από ένα απλό τυχαίο δείγμα 50 κατοίκων του Buffalo ο μέσος όρος είναι 22,5 μίλια την ημέρα και από το ανεξάρτητο απλό τυχαίο δείγμα 100 κατοίκων της Βοστώνης ο μέσος όρος είναι 18,6 μίλια την ημέρα. Αν οι τυπικές αποκλίσεις των πληθυσμών είναι αντίστοιχα 8,4 μίλια την ημέρα και 7,4 μίλια την ημέρα, ποιο είναι το 95% διάστημα εμπιστοσύνης για τη διαφορά μεταξύ των δύο μέσων του πληθυσμού;

Λύση

n_1=50 n_2=100 σ_1=8,4 σ_2=7,4 $\bar{x}_1 = 22.5 \quad \bar{x}_2 = 18.6$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{8,4^2}{50}+\frac{7,4^2}{100}} = 1,399$$

$$(\bar{x}_1-\bar{x}_2) \pm Z_{a/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = (22.5-18.6) \pm 1.96 \cdot 1.399 =$$

$$= (1.58 - 6.64)$$

Α2. Εκτίμηση της διαφοράς των μέσων τιμών δύο πληθυσμών μ_1-μ_2: Ανεξάρτητα Δείγματα

Περίπτωση Μεγάλων Δειγμάτων με σ_1 and σ_2 *εκτιμώμενες από* s_1 και s_2

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

$n_1>30$ και $n_2>30$

n_1 = Δείγμα από τον πληθυσμό 1

n_2 = Δείγμα από τον πληθυσμό 2

s_1 = Τυπική Απόκλιση από το δείγμα 1

s_2 = Τυπική Απόκλιση από το δείγμα 2

Περίπτωση Μεγάλων Δειγμάτων με *σ_1* and *σ_2 εκτιμώμενες από* s_1 και s_2

Τυπική Απόκλιση της Διαφοράς $$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Περιθώριο Σφάλματος $$E = Z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Διάστημα Εμπιστοσύνης: $$(\bar{x}_1 - \bar{x}_2) \pm Z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Εύρος: $$R = 2 \cdot Z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Παράδειγμα 8ο

Το Υπουργείο Μεταφορών των ΗΠΑ παρέχει τον αριθμό των μιλίων που διανύουν οι κάτοικοι των 75 μεγαλύτερων μητροπολιτικών περιοχών την ημέρα με ένα αυτοκίνητο. Αν υποθέσουμε ότι για ένα απλό τυχαίο δείγμα 50 κατοίκων του Buffalo ο μέσος όρος είναι 22,5 μίλια την ημέρα και η τυπική απόκλιση 8,4 μίλια την ημέρα και από το ανεξάρτητο απλό τυχαίο δείγμα 100 κατοίκων της Βοστώνης ο μέσος όρος είναι 18,6 μίλια την ημέρα και η τυπική απόκλιση 7,4 μίλια την ημέρα, ποιο είναι το 95% διάστημα εμπιστοσύνης για τη διαφορά μεταξύ των δύο μέσων του πληθυσμού;

Λύση

n_1=50 n_2=100 s_1=8,4 s_2=7,4 $\bar{x}_1 = 22.5 \quad \bar{x}_2 = 18.6$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{8{,}4^2}{50} + \frac{7{,}4^2}{100}} = 1{,}399$$

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (22.5 - 18.6) \pm 1.96 \cdot 1.399 =$$

$$= (1.58 - 6.64)$$

Α3. Εκτίμηση της διαφοράς των μέσων τιμών δύο πληθυσμών μ_1-μ_2: Ανεξάρτητα Δείγματα Περίπτωση μικρών δειγμάτων ($n_1 =< 30$ και/ή $n_2 =< 30$) με σ_1 και σ_2 εκτιμώμενες από s_1 και s_2

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

n_1 = Δείγμα από τον πληθυσμό 1

n_2 = Δείγμα από τον πληθυσμό 2

s_1 = Τυπική Απόκλιση από το δείγμα 1

s_2 = Τυπική Απόκλιση από το δείγμα 2

Ξεκινώντας με τις εξής υποθέσεις:

1. Και οι δύο πληθυσμοί ακολουθούν την κανονική κατανομή
2. Οι Διακυμάνσεις των πληθυσμών είναι ίσες $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

Περίπτωση μικρών δειγμάτων (n_1=<30 και/ή n_2=<30) με σ_1 **και** σ_2 **εκτιμώμενες από** s_1 **και** s_2

Τυπική Απόκλιση της Διαφοράς: $$\sigma_{\bar{x}_1-\bar{x}_2}=\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Εκτίμηση της σ^2: $$s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

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

Διάστημα Εμπιστοσύνης: $$(\bar{x}_1-\bar{x}_2)\pm t_{a/2}\cdot\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

d.f : n_1+n_2-2

Παράδειγμα 9ο

Μια ομάδα πολεοδομικού σχεδιασμού ενδιαφέρεται να εκτιμήσει τη διαφορά μεταξύ του μέσου εισοδήματος των νοικοκυριών για δύο γειτονιές σε μια μεγάλη μητροπολιτική περιοχή. Ανεξάρτητα τυχαία δείγματα νοικοκυριών στις γειτονιές έδωσαν τα ακόλουθα αποτελέσματα.

Γειτονιά 1	**Γειτονιά 2**
$n_1 = 8$	$n_2 = 12$
Μέση τιμή 1 =15.700$	Μέση τιμή 2 =14.500$
Τυπική απόκλιση 1=700$	Τυπική απόκλιση 2=850$

Λύση

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(8-1)700^2 + (12-1)850^2}{8+12-2} = 632083{,}33$$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{632083{,}33\left(\frac{1}{8}+\frac{1}{12}\right)} = 362{,}88$$

$t_{a/2} = t_{0,025}$ με 8+12-2=18 d.f είναι 2,101.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{a/2} \cdot \sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = (15700-14500) \pm 2{,}101 \cdot 362{,}88 = 1200 \pm 762{,}41$$

Β. Εκτίμηση της διαφοράς των μέσων τιμών δύο πληθυσμών μ_1-μ_2: : Εξαρτημένα Δείγματα

a = Επίπεδο Σημαντικότητας

1-a = Επίπεδο Εμπιστοσύνης

Ζευγαρωτή Διαφορά : $\boldsymbol{d_i = x_{1i} - x_{2i}}$

όπου: χ_{1i}, χ_{2i} είναι οι τιμές από το 1ο και το 2ο δείγμα

Μέσος της Διαφοράς $$\bar{d} = \frac{\sum d_i}{n}$$

Τυπική Απόκλιση $$s_d = \sqrt{\frac{\sum d_i^2 - \left[\frac{(\sum d_i)^2}{n}\right]}{n-1}}$$

Περιθώριο Σφάλματος: $$E = t_{a/2} \cdot \frac{s_d}{\sqrt{n}}$$

Διάστημα Εμπιστοσύνης: $$\bar{d} \pm t_{a/2} \cdot \frac{s_d}{\sqrt{n}}$$ *d.f : n-1*

Παράδειγμα 9ο

Επιλέχτηκε τυχαίο δείγμα 6 αυτοκινήτων και σε καθένα από αυτά τοποθετήθηκε βενζίνη τύπου Α και μετρήθηκε η απόσταση που διανύθηκε. Στη συνέχεια, τα ίδια αυτοκίνητα φορτώθηκαν με την ίδια ποσότητα βενζίνης τύπου Β και μετρήθηκε ξανά η απόσταση που διένυσαν. Υποθέτοντας ότι οι δοκιμές πραγματοποιήθηκαν υπό τις ίδιες τυπικές συνθήκες, τα αποτελέσματα δίνονται στον πίνακα (σε Km). Να δώσετε ένα 95% δ.ε για την διαφορά στην κατανάλωση βενζίνης.

Αυτοκίνητο	1	2	3	4	5	6
Τύπος Α	125	64	94	38	90	106
Τύπος Β	133	65	103	37	102	115

Λύση

Αυτοκίνητο	1	2	3	4	5	6	
Τύπος Α	125	64	94	38	90	106	
Τύπος Β	133	65	103	37	102	115	
d_i (B-A)	8	1	9	-1	12	9	38
d_i^2 (B-A)^2	64	1	81	1	144	81	372

$$\bar{d} = \frac{\sum d_i}{n} = \frac{38}{6} = 6{,}33$$

$$s_d = \sqrt{\frac{\sum d_i^2 - \left[\frac{(\sum d_i)^2}{n}\right]}{n-1}} = \sqrt{\frac{372 - \frac{38^2}{6}}{6-1}} = 5{,}1$$

$$\bar{d} \pm t_{a/2} \cdot \frac{s_d}{\sqrt{n}} = 6{,}33 \pm 2{,}57 * \frac{5{,}1}{\sqrt{6}} = 6{,}33 \pm 5{,}35$$

Η διαφορά στα χιλιόμετρα, με τους δύο διαφορετικούς τύπους βενζίνης, κυμαίνεται από 0,98 έως 11,68 χιλιόμετρα. Δηλαδή με τον τύπο Β διήνυσαν περισσότερα χιλιόμετρα.

C. Εκτίμηση της Διαφοράς μεταξύ των Αναλογιών δύο πληθυσμών p_1-p_2

a = Επίπεδο Σημαντικότητας και 1-a = Επίπεδο Εμπιστοσύνης

p_1 = Αναλογία Πληθυσμού 1

p_2 = Αναλογία Πληθυσμού 2

$\bar{p}_1$ = Αναλογία Δείγματος από τον πληθυσμό 1 $\quad \bar{p}_1 = \frac{x}{n_1}$

$\bar{p}_2$ = Αναλογία Δείγματος από τον πληθυσμό 2 $\quad \bar{p}_2 = \frac{x}{n_2}$

Για την επίτευξη καλύτερων αποτελεσμάτων πρέπει να ικανοποιούνται οι παρακάτω υποθέσεις:

1. Κάθε δείγμα πρέπει να είναι μεγαλύτερο του 20.
2. $n_1p_1>5$, $n_1(1-p_1)>5$, $n_2p_2>5$, $n_2(1-p_2)>5$
3. Κάθε δείγμα δεν πρέπει να είναι μεγαλύτερο του 10% του αντιστοίχου πληθυσμού.

Τυπική Απόκλιση: $\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$

Περιθώριο Σφάλματος: $E = Z_{a/2}\cdot\sqrt{\frac{p_1(1-p)_1}{n_1}+\frac{p_2(1-p_2)}{n_2}}$

Διάστημα Εμπιστοσύνης: $(\bar{p}_1-\bar{p}_2)\pm Z_{a/2}\cdot\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$

όταν p_1 και p_2 δεν είναι γνωστές, χρησιμοποιούνται $\bar{p}_1$ και $\bar{p}_2$

Τυπική Απόκλιση: $s_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

Περιθώριο Σφάλματος : $E = Z_{a/2}\cdot\sqrt{\frac{\bar{p}_1(1-\bar{p})_1}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

Διάστημα Εμπιστοσύνης : $(\bar{p}_1-\bar{p}_2)\pm Z_{a/2}\cdot\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

Παράδειγμα 10ο

Ανεξάρτητα απλά τυχαία δείγματα φορολογικών δηλώσεων από δύο γραφεία παρέχουν τις ακόλουθες πληροφορίες:

Γραφείο 1	Γραφείο 2
n_1=250	n_2= 300
Επιστροφές λόγω λαθών=35	Επιστροφές λόγω λαθών=27

Δώστε ένα 95% διάστημα εμπιστοσύνης για τη διαφορά μεταξύ των δύο ποσοστών σφάλματος στον πληθυσμό.

Λύση

$$\overline{p}_1 = \frac{x}{n_1} = \frac{35}{250} = 0{,}14 \quad \overline{p}_2 = \frac{x}{n_2} = \frac{27}{300} = 0{,}09$$

$$s_{\overline{p}_1-\overline{p}_2} = \sqrt{\frac{\overline{p}_1(1-\overline{p}_1)}{n_1} + \frac{\overline{p}_2(1-\overline{p}_2)}{n_2}} = \sqrt{\frac{0{,}14*0{,}86}{250} + \frac{0{,}09*0{,}91}{300}} = 0{,}0275$$

$$(\overline{p}_1 - \overline{p}_2) \pm Z_{a/2} \cdot \sqrt{\frac{\overline{p}_1(1-\overline{p}_1)}{n_1} + \frac{\overline{p}_2(1-\overline{p}_2)}{n_2}} = (0{,}14-0{,}09) \pm 1{,}96*0{,}0275 = 0{,}05 \pm 0{,}0539$$

Η διαφορά στον πληθυσμό, μεταξύ των δύο γραφείων κυμαίνεται από 5% έως 5,39%.