στατιστική θεωρεία της δειγματοληψίας

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Εισαγωγή

δειγματοληψία

- Τα στοιχεία που απαιτούνται τόσο για την ανάλυση των μεταφορικών συστημάτων και όσο και για την ανάπτυξη των συγκοινωνιακών μοντέλων προέρχονται από **παρατηρήσεις, ανάλυση κι διερεύνηση των χαρακτηριστικών** ενός **δείγματος** του πληθυσμού που μελετάται. Ανάλυση όλου του πληθυσμού δεν εφικτή τόσο για οικονομικούς όσο και για τεχνικούς λόγους.
- Λόγω της **διακύμανσης** των τιμών / μεταβλητότητας των χαρακτηριστικών του πληθυσμού είναι απαραίτητο, το δείγμα να αναπαριστά αυτή την μεταβλητότητα να είναι δηλαδή **αντιπροσωπευτικό** του πληθυσμού.
- Ο σκοπός του σχεδιασμού της δειγματοληψίας είναι να εξασφαλίσει ότι τα στοιχεία που αναλύονται παρέχουν την **βέλτιστη πληροφορία** που απαιτείται για τον πληθυσμό που μελετάται, στο χαμηλότερο δυνατό κόστος.

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Βασικές έννοιες

διαστήματα εμπιστοσύνης

- Όταν συλλέγουμε στοιχεία από ένα δείγμα δεν αναμένουμε τα αποτελέσματα της ανάλυσης να είναι ακριβώς ίδια με εκείνα που θα υπολογίζαμε αν είχαμε στοιχεία από όλο τον πληθυσμό
- Χρησιμοποιώντας την **μεταβλητότητα** των στοιχείων του δείγματος, μπορούμε να υπολογίσουμε το **φάσμα τιμών** μέσα στο οποίο είναι πιθανό να είναι η μέση τιμή του πληθυσμού.
- Μπορούμε να μεταβάλουμε το **εύρος** αυτού του φάσματος, ανάλογα με το **πόσο σίγουροι** θέλουμε να είμαστε ότι το εύρος αυτό θα περιλαμβάνει την πραγματική μέση τιμή του πληθυσμού (συνήθως θεωρούμε επίπεδο εμπιστοσύνης το 95%).

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Διαστήματα Εμπιστοσύνης

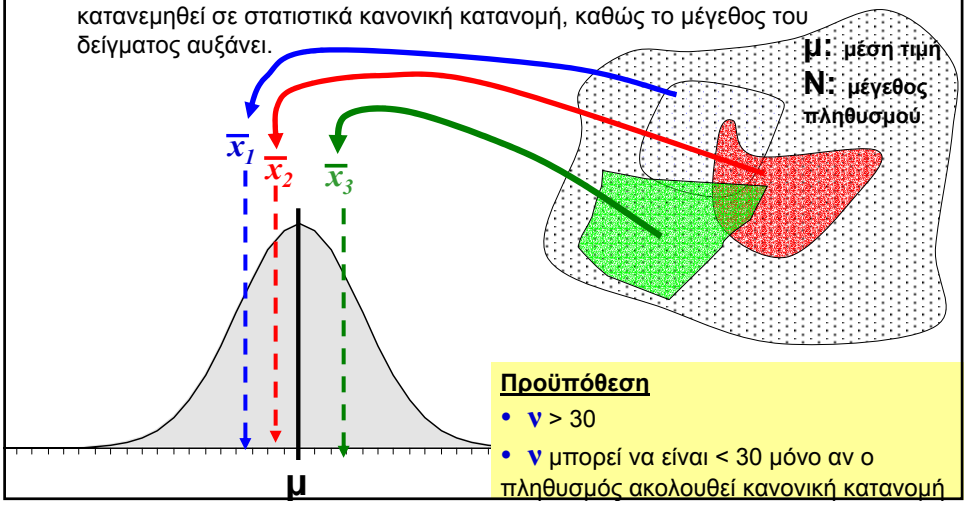
- Θεωρώντας ότι το δείγμα είναι αντιπροσωπευτικό, τα **διαστήματα εμπιστοσύνης** μπορούν να υπολογισθούν από τα δείγματα χρησιμοποιώντας την ακόλουθη σχέση:

Μέση τιμή δείγματος $\pm$ **συντελεστής επίπεδου εμπιστοσύνης** x **τυπικό σφάλμα**

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Το Θεώρημα της Κεντρικής Θέσης

Το θεώρημα της κεντρικής θέσης

Ο αριθμητικός μέσος όρος των στοιχείων τυχαίων δειγμάτων μέσου μεγέθους **(ν)**, που λαμβάνονται από ένα πληθυσμό τείνει να κατανεμηθεί σε στατιστικά κανονική κατανομή, καθώς το μέγεθος του δείγματος αυξάνει.



Προϋπόθεση

- **ν** > 30
- **ν** μπορεί να είναι < 30 μόνο αν ο πληθυσμός ακολουθεί κανονική κατανομή

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Το Τυπικό Σφάλμα

	Πληθυσμός	Δείγμα
μέγεθος	N	ν
μέση τιμή (mean)	μ	$\bar{x}$
διακύμανση (variance)	σ^2	S^2

παράδειγμα

Εάν χρησιμοποιούμε ένα **μόνο δείγμα** η καλύτερη εκτίμηση του μ είναι το $\bar{x}$ και η καλύτερη εκτίμηση του σ^2 είναι το S^2

Σε αυτή την περίπτωση η τυπική απόκλιση δηλ. **το τυπικό σφάλμα** του μ είναι

$$\text{se}(\bar{x}) = \sqrt{\frac{(N-\nu).S^2}{\nu.N}}$$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Το Τυπικό Σφάλμα

Πότε το τυπικό σφάλμα τείνει να μηδενισθεί ?

$$se(\bar{x}) = \sqrt{\frac{(N-\nu).S^2}{\nu.N}}$$

$$\nu \longrightarrow N$$

Στη πράξη όμως έχουμε συνήθως μεγάλους πληθυσμούς και μικρό δείγμα

$$\Rightarrow \frac{(N-\nu)}{N} \approx 1 \Rightarrow se(\bar{x}) = \frac{S}{\sqrt{\nu}}$$

Επιλύνοντας μπορούμε να προσδιορίσουμε το **μέγεθος του δείγματος** , δηλ.

$$\nu = \frac{\nu'}{1+\frac{\nu'}{N}}$$

Διόρθωση για δείγματα πεπερασμένου μεγέθους

$$\nu' = \frac{S^2}{se(\bar{x})^2}$$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Το Τυπικό Σφάλμα

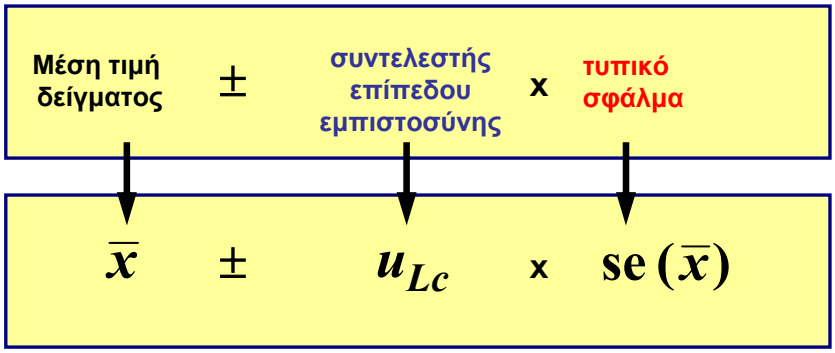
Προβλήματα εφαρμογής:

- Η ***εκτίμηση της διακύμανσης του δείγματος*** (S^2) που μπορεί να υπολογισθεί αφού πρώτα έχουν συλλεχθεί τα στοιχεία => πρέπει να εκτιμηθεί από άλλες πηγές (π.χ. πιλοτική έρευνα)
- Ο ***επιθυμητός βαθμός εμπιστοσύνης*** που συνδέεται με την χρήση της μέσης τιμής του δείγματος σαν εκτίμηση της μέσης τιμής του πληθυσμού. Ο βαθμός εμπιστοσύνης, στην πράξη συνήθως καθορίζεται σαν ένα **διάστημα γύρω από την μέση τιμή** του πληθυσμού για ένα δεδομένο **επίπεδο εμπιστοσύνης**. Επομένως:
  - Το **επίπεδο εμπιστοσύνης** για το διάστημα θα πρέπει να καθορισθεί, δηλ. η αποδεκτή συχνότητα εμφάνισης σφάλματος που οφείλεται στην παραδοχή ότι η μέση τιμή του δείγματος είναι η πραγματική μέση τιμή του πληθυσμού (δηλ. το τυπικό επίπεδο εμπιστοσύνης 95% σημαίνει ότι δεχόμαστε ότι στο 5% των περιπτώσεων θα υπάρχει σφάλμα)
  - Θα πρέπει καθορισθούν τα **όρια του διαστήματος γύρω από την μέση τιμή**

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Διαστήματα Εμπιστοσύνης

Υπολογισμός των διαστημάτων εμπιστοσύνης

- Θεωρώντας ότι το δείγμα είναι αντιπροσωπευτικό, τα **διαστήματα εμπιστοσύνης** μπορούν να υπολογισθούν από τα δείγματα χρησιμοποιώντας την ακόλουθη σχέση:



ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Διαστήματα Εμπιστοσύνης

Τι είναι το Διάστημα Εμπιστοσύνης

- Αν θεωρήσουμε άπειρα δείγματα μεγέθους ν από ένα πληθυσμό
- Ένα διάστημα εμπιστοσύνης 95% για την μέση τιμή, μπορεί να υπολογισθεί για κάθε ένα από τα δείγματα :

Διαστήματα εμπιστοσύνης 95%

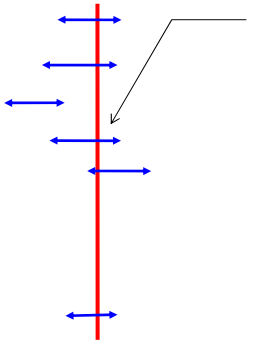
$$\bar{x}_1 \pm u_{95\%} \times \left(s_1 / \sqrt{n}\right),$$
$$\bar{x}_2 \pm u_{95\%} \times \left(s_2 / \sqrt{n}\right),$$
$$\vdots$$
$$\bar{x}_\infty \pm u_{95\%} \times \left(s_\infty / \sqrt{n}\right).$$

95% αυτών των διαστημάτων θα περιλαμβάνουν την μέση τιμή του πληθυσμού μ, ενώ το 5% από αυτά τα διαστήματα δεν θα περιλαμβάνουν την μέση τιμή του πληθυσμού.

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Διαστήματα Εμπιστοσύνης

Μέση τιμή πληθυσμού
και διαστήματα εμπιστοσύνης από τα δείγματα

δείγμα 1
δείγμα 2
δείγμα 3
δείγμα 4
δείγμα 5
⋮
δείγμα v



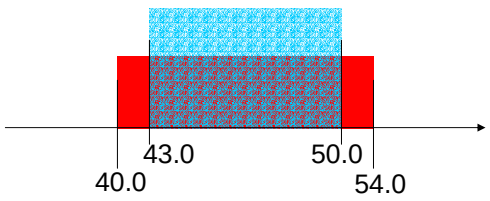
Πραγματική μέση τιμή
του πληθυσμού (μ, π,..)

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Διαστήματα Εμπιστοσύνης

- Αύξηση του επιπέδου εμπιστοσύνης από 95% σε 99% αυξάνει την βεβαιότητα ότι το διάστημα εμπιστοσύνης περιλαμβάνει την μέση τιμή του πληθυσμού, αλλά μειώνει την ακρίβεια της εκτίμησης, δεδομένου ότι το διάστημα είναι πιο ευρύ.

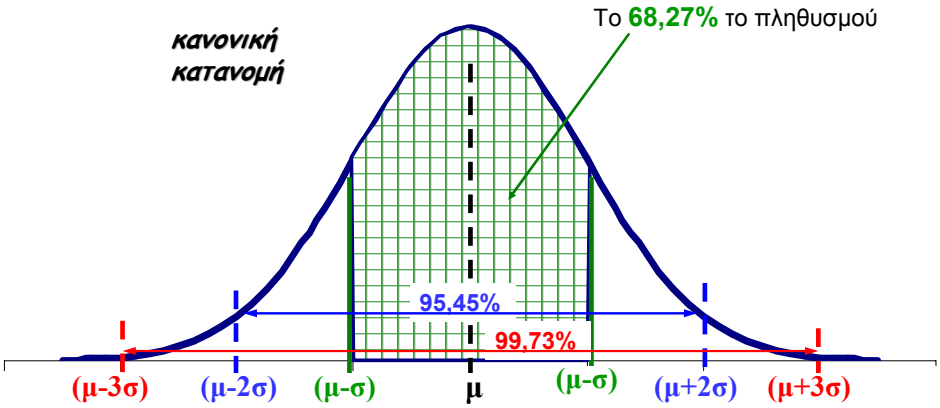
π.χ.

- Με επίπεδο εμπιστοσύνης 99% ο χρόνος διαδρομής θα είναι μεταξύ 40 και 54 λεπτών
- Με επίπεδο εμπιστοσύνης 95% ο χρόνος διαδρομής θα είναι μεταξύ 43 και 50 λεπτών



ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Όρια ακρίβειας της μέσης τιμής του δείγματος

Ακρίβεια της εκτίμησης : Η πιθανότητα που υπάρχει ο πραγματικός μέσος όρος (δηλ. ο μ.ο. του πληθυσμού) να βρίσκεται μέσα σε ορισμένα όρια



ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Όρια ακρίβειας της μέσης τιμής του δείγματος

Υπάρχει πιθανότητα

68,27% $\bar{x} - se(\bar{x}) < \mu < \bar{x} + se(\bar{x})$

95,45% $\bar{x} - 2.se(\bar{x}) < \mu < \bar{x} + 2.se(\bar{x})$

99,73% $\bar{x} - 3.se(\bar{x}) < \mu < \bar{x} + 3.se(\bar{x})$

Όπου :

$\bar{x}$: ο μέσος όρος του δείγματος

μ : ο μέσος όρος του πληθυσμού

$se(\bar{x}) = \frac{S}{\sqrt{\nu}}$ το τυπικό σφάλμα και

ν το μέγεθος του δείγματος

S η τυπική απόκλιση του δείγματος

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Υπολογισμός πιθανότητας

- Για να υπολογίσουμε την πιθανότητα

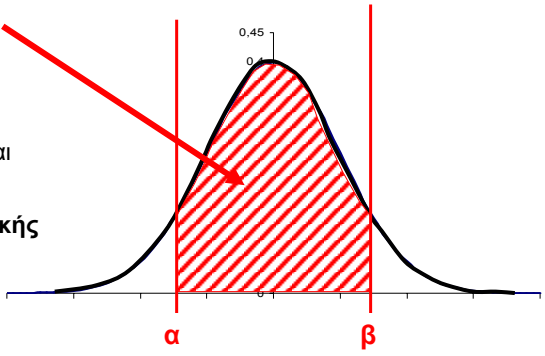
  η τιμή μιας μεταβλητής να είναι μεταξύ δύο συγκεκριμένων ορίων,

  θα πρέπει να υπολογίσουμε το εμβαδόν της περιοχής κάτω από την καμπύλη και ανάμεσα στα δυο όρια.

$$P(\alpha < x < \beta)$$

Το εμβαδόν αυτό υπολογίζεται εύκολα με χρήση της

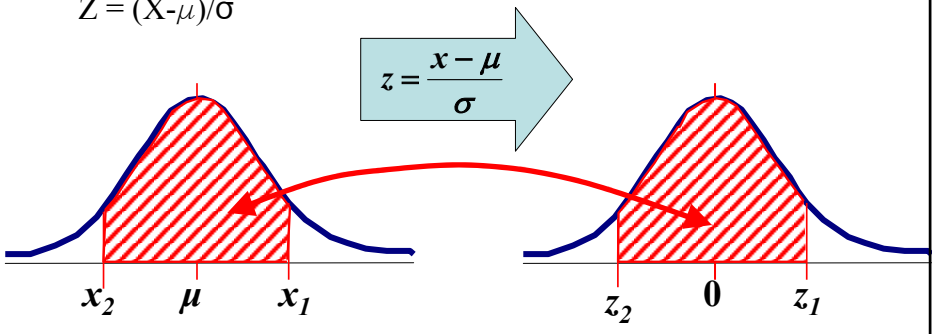
Τυπικής/μοναδιαίας κανονικής κατανομής



ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Τυπική/Μοναδιαία Κανονική Κατανομή -

- Η **Τυπική/Μοναδιαία Κανονική Κατανομή** είναι μια κανονική κατανομή πιθανότητας πού έχει μέση τιμή (μ) = 0, και τυπική απόκλιση (σ)= 1.
- Τα περισσότερα μεγέθη που ακολουθούν Κανονική Κατανομή δεν έχουν μέση τιμή = 0 και τυπική απόκλιση =1.
- Είναι δυνατό όμως να τυποποιήσουμε τις μη τυπικές περιπτώσεις χρησιμοποιώντας την σχέση :

$$Z = (X-\mu)/\sigma$$



ΔΕΙΓΜΑΤΟΛΗΨΙΑ : ο συντελεστής z μετατροπής σε μοναδιαία κατανομή

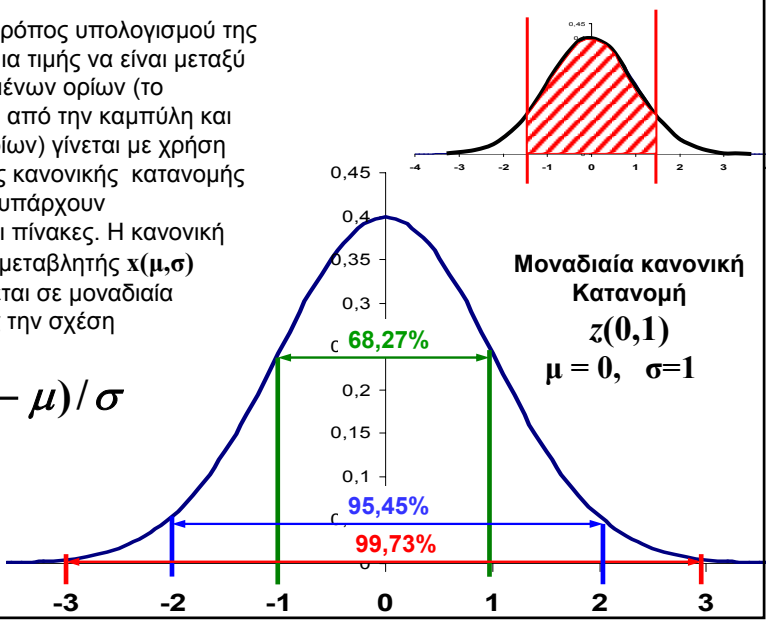
$$z = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + z.\sigma$$

Οι τιμές του συντελεστή z μετρούν τον αριθμό των τυπικών αποκλίσεων από απέχει μια τιμή από την μέση τιμή

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Μοναδιαία Κανονική Κατανομή

Ο κλασσικός τρόπος υπολογισμού της πιθανότητας μια τιμής να είναι μεταξύ δύο συγκεκριμένων ορίων (το εμβαδόν κάτω από την καμπύλη και μεταξύ των ορίων) γίνεται με χρήση της μοναδιαίας κανονικής κατανομής για την οποία υπάρχουν τυποποιημένοι πίνακες. Η κανονική κατανομή της μεταβλητής **x(μ,σ)** μετασχηματίζεται σε μοναδιαία εφαρμόζοντας την σχέση

$$z = (x-\mu)/\sigma$$



Μοναδιαία κανονική Κατανομή

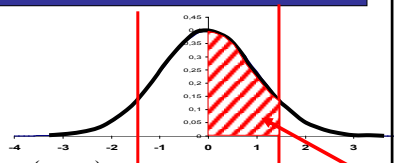
$z(0,1)$

μ = 0, σ=1

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Μοναδιαία Κανονική Κατανομή

Appendix F Table of Areas of the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3233
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3718	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.398		4015
1.3	.4032	.4049	.4066	.4083	.4099	.4115	.4131	.414		4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.429		4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441



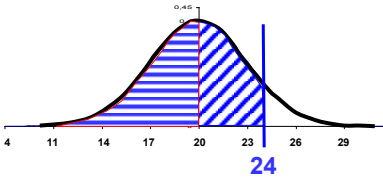
$$z = \frac{(x_i - \mu)}{\sigma}$$

Ο πίνακας δίνει το εμβαδόν κάτω από την μοναδιαία κανονική κατανομή και μεταξύ μιας τεταγμένης στο 0 και μιας στο z.

Παράδειγμα 1

X : $\mathbf{N}(\mu,\sigma) = \mathbf{N}(20, 3)$

Ποια η πιθανότητα $x < 24$?



$\Pr(20<x<24) = \Pr(20 < \mu + z.\sigma < 24) =$

$= \Pr(20 < 20 + z.3 < 24) = \Pr(0 < 3z < 4) =$

$= \Pr(\mathbf{0 < z < 1,33}) = \mathbf{0,4083} \Rightarrow$

$\Rightarrow \Pr(20<x<24) = 0,4083$

$\Pr(x < 24) = \Pr(x<20) + \Pr(20 < x < 24)$

$\Pr(x<20) = 0,5$

$\Rightarrow$

$\Rightarrow \Pr(x < 24) = 0,5 + 0,4083 = 0,9083$

$\Pr(16 < x < 24) = 2 \times 0,4083 = 0,8166$

$\Pr(x < 16) = 0,5 - 0,4083 = 0,0917$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Μοναδιαία Κανονική Κατανομή

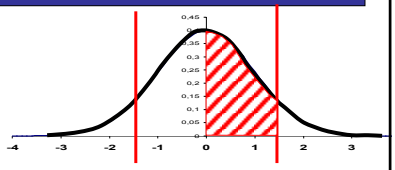
Appendix F Table of Areas of the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3233
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3718	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4083	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4698	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4758	.4762	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4984	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.4987	.4987	.4988	.4988	.4988	.4989	.4989	.4989	.4990
3.1	.49903	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Μοναδιαία Κανονική Κατανομή

Appendix F Table of Areas of the Normal Curve

u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3233
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3718	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4083	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4698	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4758	.4762	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4984	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.4987	.4987	.4988	.4988	.4988	.4989	.4989	.4989	.4990
3.1	.49903	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.49931									
3.3	.49952									
3.4	.49966									
3.5	.49977									
3.6	.49984									
3.7	.49980									
3.8	.49993									
3.9	.49995									



Παράδειγμα 2

X : $\mathbf{N}(\mu,\sigma) = \mathbf{N}(20, 3)$

Μεταξύ ποιών ορίων μπορούμε να πούμε ότι κυμαίνεται η μεταβλητή X, με ακρίβεια (επίπεδο εμπιστοσύνης) 95%?

Pr (μ-*u*.σ < x < μ+ *u*.σ) = 0,95 =>

Pr (μ-*u*.σ < μ+*z*.σ < μ+ *u*.σ) = 0,95 =>

Pr (-*u*.σ < *z*.σ < *u*.σ) = 0,95 =>

Pr (0,5 < *z* < *u*) = 0,475 =>

u = **1,96**

Xmin = 20 - 1,96x3 = 14,12

Xmax = 20 + 1,96x3 = 25,88

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Όρια ακρίβειας της μέσης τιμής του δείγματος

$$\Pr\left(\bar{x} - 1\,.se(\bar{x}) < \mu < \bar{x} + 1\,.se(\bar{x})\right) = 68{,}27\%$$
$$\Pr\left(\bar{x} - 2\,.se(\bar{x}) < \mu < \bar{x} + 2\,.se(\bar{x})\right) = 95{,}45\%$$
$$\Pr\left(\bar{x} - 3\,.se(\bar{x}) < \mu < \bar{x} + 3\,.se(\bar{x})\right) = 99{,}73\%$$

$$\Pr\left(\bar{x} - z\,.se(\bar{x}) < \mu < \bar{x} + z\,.se(\bar{x})\right) = L$$

Τα όρια διακύμανσης των τιμών για διαφορετικά επίπεδα εμπιστοσύνης προσδιορίζονται από τον σχετικό πίνακα του παραδείγματος 2. Ενδεικτικά αναφέρονται ότι οι συντελεστές z για διαφορετικά επίπεδα εμπιστοσύνης, L.

Οι τιμές του συντελεστή z για διαφορετικά επίπεδα εμπιστοσύνης είναι:

Επίπεδο εμπιστοσύνης	z
90%	1,65
95%	1,96
98%	2,33
99%	2,58

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Ανάλυση Μεγεθών εκφρασμένων σε Ποσοστά

- Σε περίπτωση που τα μεγέθη που αναλύουμε, εκφράζονται σε ποσοστά, π.χ. % νοικοκυριών με ιδιοκτησία Ι.Χ. αυτοκινήτου 2 ή υψηλότερο
  % μετακινούμενων που χρησιμοποιούν Μ.Μ.Μ.

- Η μέση τυπική απόκλιση υπολογίζεται από την σχέση:

$$\mathbf{se}\,(p) = \sqrt{\frac{p.q}{\nu}}$$

Όπου :

$\mathbf{se}\,(p)$	η προσέγγιση της τυπικής απόκλισης
p	το ποσοστιαίο αποτέλεσμα της μετρήσεως
q	(100 – p)
ν	το μέγεθος του δείγματος

Προϋποθέσεις για ικανοποιητικά αποτελέσματα $p \geq 10\%$ $\nu \geq 30$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Σύγκριση αποτελεσμάτων δύο δειγματοληψιών

	δείγμα 1	δείγμα 2
μέγεθος	ν_1	ν_2
μέση τιμή (mean)	$\bar{x}_1$	$\bar{x}_2$
διακύμανση (variance)	S_1^2	S_2^2

Ερώτημα

- τα δύο δείγματα προέρχονται από δύο διαφορετικούς πληθυσμούς με διαφορετικό μέσο όρο (πραγματική διαφορά)

ή

- από τον ίδιο πληθυσμό αλλά με διαφορετικές διακυμάνσεις (τυχαία διαφορά)

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Σύγκριση αποτελεσμάτων δύο δειγματοληψιών

Σύμφωνα με το θεώρημα κεντρικής θέσης

- η καλύτερη εκτίμηση του μ_1 είναι το $\bar{x}_1$
- και η καλύτερη εκτίμηση του σ_1^2 είναι το S_1^2

(και αντίστοιχα για το δείγμα 2)

Υπόθεση προς έλεγχο: Οι δύο πληθυσμοί είναι στην ουσία ίδιοι δηλ. $\mu_1 = \mu_2$

Αποδεικνύεται στατιστικά ότι:

- Η διαφορά $x_1 - x_2$ ακολουθεί μια κατά προσέγγιση κανονική κατανομή με μέση τιμή 0
- Το τυπικό σφάλμα της κατανομής της διαφοράς των δύο μέσων όρων υπολογίζεται από την σχέση

$$S_D(\bar{x}) = \sqrt{\frac{S_1^2}{\nu_1} + \frac{S_2^2}{\nu_2}}$$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Σύγκριση αποτελεσμάτων δύο δειγματοληψιών

Εάν η υπόθεση είναι σωστή:

- Με επίπεδο εμπιστοσύνης 95,45% η διαφορά $\bar{x}_1 - \bar{x}_2$ θα βρίσκεται μεταξύ $\pm 3.S_D(\bar{x})$

- Εάν η διαφορά $\bar{x}_1 - \bar{x}_2$ είναι μεγαλύτερη από $\pm 3.S_D(\bar{x})$
  Η διαφορά είναι σημαντική, και άρα με επίπεδο 99,73% εμπιστοσύνης, τα δείγματα προέρχονται από διαφορετικούς πληθυσμούς με διαφορετικούς μέσους όρους

- Γενικά, για δείγματα με $\nu > 30$ συγκρίνεται η διαφορά $\bar{x}_1 - \bar{x}_2$ με το $z.S_D(\bar{x})$ για το επίπεδο εμπιστοσύνης που αντιστοιχεί το z

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Σύγκριση ποσοστιαίων αποτελεσμάτων

Σύγκριση ποσοστιαίων αποτελεσμάτων από δύο δείγματα

- Ακολουθείται η ίδια διαδικασία με την περίπτωση των μέσων όρων
- Το τυπικό σφάλμα υπολογίζεται από την σχέση:

$$S_D(p) = \sqrt{p_o \cdot q_o \cdot \left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)}$$

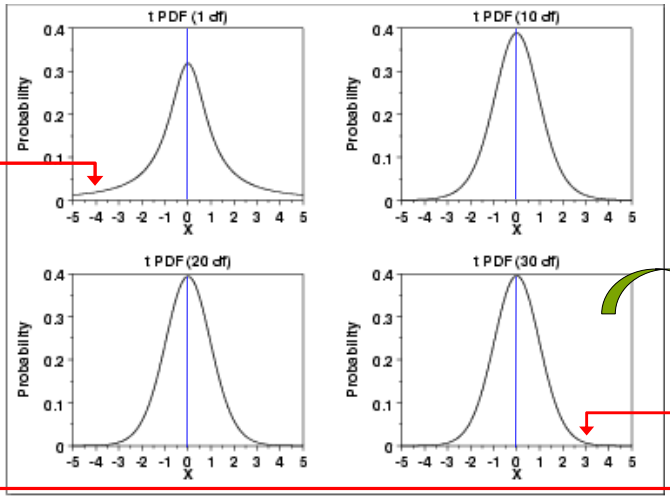
- Η αναλογική μέση τιμή των δύο ποσοστών είναι ίση με τον λόγο

$$p_o = \frac{p_1 \cdot \nu_1 + p_2 \cdot \nu_2}{\nu_1 + \nu_2}$$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ: Αξιοπιστία μικρών Δειγμάτων – ο συντελεστής t STUDENT

- Ο έλεγχος αξιοπιστίας του δείγματος, με βάση την υπόθεση της κανονικής κατανομής ισχύει για τις περιπτώσεις που το μέγεθος του δείγματος είναι μεγάλο, δηλ., τουλάχιστον 25 – 30.
- Για μικρά δείγματα αντί για τον συντελεστή z της μοναδιαίας κανονικής κατανομής χρησιμοποιείται ο συντελεστής t του Student
- Για μεγάλα δείγματα, οι τιμές του συντελεστή t ταυτίζονται με τις τιμές του συντελεστή z. Καθώς το μέγεθος του δείγματος ελαττώνεται, η διαφορά των τιμών των δύο συντελεστών αυξάνεται.
- Οι τιμές του συντελεστή t δίνονται σε πίνακες για διαφορετικά επίπεδα εμπιστοσύνης και διαφορετικά βαθμούς ελευθερίας (ο βαθμός ελευθερίας είναι $\nu - 1$: το μέγεθος του δείγματος μείον ένα)

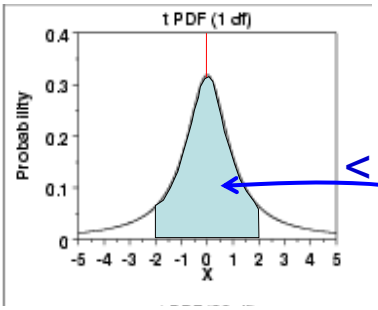
ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Κατανομή t-Student και βαθμοί ελευθερίας



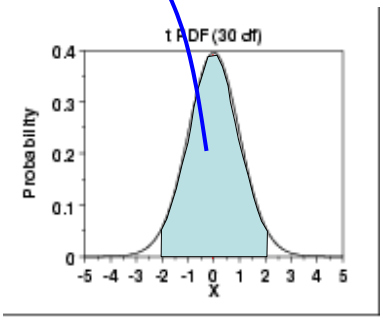
Προσέγγιση Κανονικής Κατανομής

Οι κατανομές έχουν παρόμοια μορφή. Η διαφορές εντοπίζονται στο **πάχος** των «ουρών» κάθε κατανομής, που είναι μεγαλύτερο για χαμηλότερους βαθμούς ελευθερίας δηλ. μικρότερο δείγμα. Καθώς ο βαθμός ελευθερίας αυξάνεται η κατανομή t-Student, προσεγγίζει την κανονική κατανομή.

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Κατανομή t-Student και βαθμοί ελευθερίας



Για το ίδιο διάστημα
$\pm 2.\sigma$



Το επίπεδο εμπιστοσύνης (πιθανότητα = εμβαδόν) είναι πολύ μεγαλύτερο

όταν

ο βαθμός ελευθερίας δηλ. το δείγμα είναι μεγαλύτερο

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Συμπέρασμα - Μεθοδολογία

επομένως

- Είναι δυνατό να υπολογίσουμε το μέγεθος του δείγματος, εάν θέλουμε να πετύχουμε ένα συγκεκριμένο επίπεδο ακρίβειας

- Η ακρίβεια των εκτιμήσεων μπορεί να αυξηθεί όταν ελαττώσουμε το τυπικό σφάλμα

- Το μέγεθος του τυπικού σφάλματος εξαρτάται από το μέγεθος του δείγματος

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Γενική Μεθοδολογία υπολογισμού μεγέθους δείγματος

Υπολογισμός μεγέθους δείγματος με βάση την επιθυμητή ακρίβεια για συγκεκριμένο επίπεδο εμπιστοσύνης,
π.χ. ακρίβεια χρόνου διαδρομής ± 0,5 λεπτά με πιθανότητα 95%

***e* : επιθυμητή ακρίβεια = μέγιστο επιτρεπτό σφάλμα**

***L* : επίπεδο εμπιστοσύνης, δηλ. η πιθανότητα σφάλματος = (100% - *L*)**

1. Προ-εκτίμηση του μέσης τυπικής απόκλισης του δείγματος, S, ή του ποσοστού p, από πιλοτική έρευνα/μετρήσεις, με δείγμα μεγέθους $\nu > 30$
   (Παραδοχή : το πιλοτικό δείγμα είναι αντιπροσωπευτικό του πληθυσμού)

2. Υπολογισμός του τυπικού σφάλματος με βάση το ν

Μεγάλος πληθυσμός	Πληθυσμός πεπερασμένου μεγέθους	Μεγέθη που εκφράζονται σε ποσοστά
$se(\bar{x}) = \frac{S}{\sqrt{\nu}}$	$se(\bar{x}) = \sqrt{\frac{(N-\nu).S^2}{\nu.N}}$	$se(p) = \sqrt{\frac{p.q}{\nu}}$

ΔΕΙΓΜΑΤΟΛΗΨΙΑ : Γενική Μεθοδολογία υπολογισμού μεγέθους δείγματος

3. Υπολογισμός των ορίων διακύμανσης των τιμών του σφάλματος για διαφορετικά επίπεδα εμπιστοσύνης / ακρίβειας με βάση το δείγμα της πιλοτικής εφαρμογής

4. Υπολογισμός του συντελεστή z, (μοναδιαίας κανονικής κατανομής) για την επίτευξη του απαιτούμενου επίπεδου εμπιστοσύνης, $z = z(L)$

5. Υπολογισμός του μεγέθους του δείγματος, ν, έτσι ώστε το σφάλμα του τελικού δείγματος να είναι μικρότερο από το μέγιστο επιτρεπτό

$$z \cdot \mathrm{se}(\bar{x}) \le e \Rightarrow$$

$$\Rightarrow z \cdot \frac{S}{\sqrt{\nu}} \le e \Rightarrow \nu = \left(\frac{z}{e}\right)^2 \cdot S$$

$$z \cdot \mathrm{se}(p) \le e \Rightarrow$$

$$\Rightarrow z \cdot \sqrt{\frac{p.q}{\nu}} \le e \Rightarrow \nu = \left(\frac{z}{e}\right)^2 \cdot p.q$$

Άσκηση 4: Υπολογισμός μεγέθους δείγματος

Υπολογισμός μεγέθους δείγματος όταν δίνεται η επιθυμητή ακρίβεια (ανεκτό σφάλμα) για ορισμένο επίπεδο εμπιστοσύνης

Για την εκτίμηση του χρόνου διαδρομής μεταξύ δύο σημείων μιας αστικής περιοχής έχουν γίνει μετρήσεις με παρατηρητές που κάνουν την ίδια πάντα διαδρομή με αυτοκίνητο.

Έχουν γίνει 32 μετρήσεις και οι χρόνοι διαδρομής παρουσιάζονται στο πίνακα.

Εάν επιθυμούμε ο χρόνος διαδρομής να εκτιμηθεί με ακρίβεια $\pm 0{,}5$ λεπτών στο επίπεδο εμπιστοσύνης 95%, να υπολογισθεί ο απαιτούμενος αριθμός των μετρήσεων

Συχνότητα	Χρόνος Διαδρομής
2	24,0
3	24,3
4	25,1
6	26,3
5	27,2
4	27,9
3	28,5
3	29,2
2	32,3

Δίδονται: z = 1,96 για επίπεδο εμπιστοσύνης 95%

t = 2,04 για επίπεδο εμπιστοσύνης 95% και 31 βαθμούς ελευθερίας

Άσκηση 4: Υπολογισμός μεγέθους δείγματος

$$\bar{x} = \frac{\sum_i f_i \cdot x_i}{\nu} = \frac{864,4}{32} = 27,01 \quad \text{λεπτά}$$

$$S = \sqrt{\frac{\sum_i f_i \cdot (x_i - \bar{x})^2}{\nu - 1}} = \sqrt{\frac{138,13}{31}} = 2,11 \quad \text{λεπτά}$$

$$se(x) = \frac{S}{\sqrt{\nu}} = \frac{2,11}{\sqrt{32}} = 0,37 \quad \text{λεπτά}$$

- Για επίπεδο εμπιστοσύνης 95% ο συντελεστής z=1,96 και το σφάλμα που προκύπτει από τις 32 μετρήσεις είναι 1,96x0,373 =0,73 > 0,5 δηλ. από το επιτρεπτό σφάλμα.
- Με τις 32 μετρήσεις προκύπτει ότι το 95% των περιπτώσεων ο πραγματικός μέσος χρόνος διαδρομής θα είναι σε ένα εύρος $\pm$ 0,73 λεπτά από τον μέσο όρο του δείγματος

Άσκηση 4: Υπολογισμός μεγέθους δείγματος

Επομένως θα πρέπει να αυξηθεί το μέγεθος του δείγματος έτσι ώστε το σφάλμα για επίπεδο εμπιστοσύνης 95% να είναι μικρότερο από το επιτρεπτό.

$$z \cdot se(x) < \text{επιτρεπτό σφάλμα}$$

$$1,96 \times \frac{2,11}{\sqrt{N}} < 0,5 \quad \Rightarrow N > \left(1,96 \times \frac{2,11}{0,5}\right)^2 \Rightarrow N > 68$$

Το πρόβλημα μπορεί να επιλυθεί και με χρήση της κατανομής t-Student.

Με αυτή την μέθοδο το απαιτούμενο δείγμα θα είναι μεγαλύτερο.

Άσκηση 5: Σύγκριση Δειγμάτων

Για να αξιολογηθούν τα αποτελέσματα κυκλοφοριακών ρυθμίσεων που εφαρμόσθηκαν σε κυκλοφοριακό διάδρομο αστικής περιοχής, έγιναν μετρήσεις χρόνου διαδρομής μεταξύ δύο σημείων, προ και μετά την εφαρμογή των μέτρων.

Τα αποτελέσματα από την ανάλυση των μετρήσεων παρουσιάζονται στον πίνακα.

Μετρήσεις πριν και μετά την εφαρμογή του νέου συστήματος Φωτεινής Σηματοδότησης

	Δείγμα - Πριν	**Δείγμα - Μετά**
μέση τιμή	22,6	21,2
τυπική απόκλιση	2,1	1,8
Μέγεθος δείγματος	50	60

Ζητείται να εξετασθεί αν η παρατηρούμενη μείωση του χρόνου διαδρομής οφείλεται σε τυχαία διακύμανση των συνθηκών της κυκλοφορίας ή αν είναι αποτέλεσμα των εφαρμοσθέντων ρυθμίσεων.

Άσκηση 5: Σύγκριση Δειγμάτων

- Η διαφορά των μέσων όρων των δειγμάτων είναι:

$$\bar{x}_1 - \bar{x}_2 = 22{,}6 - 21.2 = 1{,}4 \text{ λεπτά}$$

- Το τυπικό σφάλμα των διαφορών των μέσων όρων των δειγμάτων είναι:

$$sd = \sqrt{\frac{2{,}1^2}{50} + \frac{1{,}8^2}{60}} = 0{,}377 \text{ λεπτά}$$

Για επίπεδο εμπιστοσύνης 99,75%, (οπότε ο σχετικός συντελεστής z = 3),

$$\bar{x}_1 - \bar{x}_2 > z \times sd \quad \Leftrightarrow \quad 1{,}4 > 3 \times 0{,}377$$

Επομένως συμπεραίνουμε ότι πρόκειται για πραγματική διαφορά που οφείλεται στις νέες ρυθμίσεις.

Άσκηση 6 : Μέγεθος Δείγματος ποσοστιαίων μεγεθών

Έρευνα επιλογής μεταφορικού μέσου σε 100 εργαζόμενους για τις μετακινήσεις μεταξύ κατοικίας και χώρου εργασίας, έδωσε τα εξής αποτελέσματα:

Μεταφ. Μέσο	**Ποσοστό**
Αυτοκίνητο	40%
Λεωφορείο	35%
Μετρό	25%

1. **Ποια είναι η ακρίβεια των παραπάνω μεριδίων αγοράς με πιθανότητας 99% (δηλ. μεταξύ ποιών ορίων κυμαίνονται τα μερίδια, για επίπεδο εμπιστοσύνης 99%)**
2. **Ποιο είναι το απαιτούμενο μέγεθος του δείγματος έτσι ώστε τα μερίδια αγοράς να μην απέχουν από τα πραγματικά μερίδια περισσότερο από ± 2%, με πιθανότητα 95%?**

spreadsheet