

Κεφάλαιο 7

Διαστήματα Εμπιστοσύνης

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

- Το άθροισμα και -επομένως- ο μέσος όρος, μεγάλου αριθμού ανεξάρτητων μεταβλητών, που ακολουθούν την ίδια κατανομή, ακολουθεί κατά προσέγγιση κανονική κατανομή, ανεξαρτήτως από το ποια κατανομή ακολουθούν οι μεταβλητές.
- Κατά συνέπεια οι δειγματικοί μέσοι όροι (ως αθροίσματα που είναι) που σχηματίζονται από τα δείγματα σταθερού μεγέθους n , ακολουθούν την Κανονική Κατανομή.

Κεντρικό Οριακό Θεώρημα

- Αν θα μπορούσαμε να πάρουμε όλα τα δυνατά τυχαία δείγματα μεγέθους n και να υπολογίσουμε τους δειγματικούς μέσους όρους, τότε αυτοί έχουν μέσο όρο την άγνωστη σε εμάς μέση τιμή του πληθυσμού (μ) και κατανέμονται συμμετρικά εκατέρωθέν της σε σχήμα καμπάνας (από την κανονική κατανομή). Έχουν τυπική απόκλιση ίση με την τυπική απόκλιση του πληθυσμού (σ) διαιρεμένη με την τετραγωνική ρίζα του n

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

Διαστήματα Εμπιστοσύνης

- Επειδή δεν γνωρίζουμε την τυπική απόκλιση του πληθυσμού, προσεγγίζουμε την τυπική απόκλιση του πληθυσμού με την εκτίμησή της s .

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

- Η ποσότητα αυτή ονομάζεται *Τυπικό Σφάλμα του Μέσου Όρου* ή απλά *Τυπικό Σφάλμα*.

Διαστήματα Εμπιστοσύνης

- Το διάστημα

$$(\mu_{\bar{x}} - 2 s_{\bar{x}}, \mu_{\bar{x}} + 2 s_{\bar{x}}) = \left(\mu - 2 \frac{s}{\sqrt{n}}, \mu + 2 \frac{s}{\sqrt{n}} \right)$$

περιλαμβάνει το 95% των δειγματικών μέσων
όρων

Διαστήματα Εμπιστοσύνης

- Το διάστημα

$$\left(\bar{x} - 2 \frac{s}{\sqrt{n}}, \bar{x} + 2 \frac{s}{\sqrt{n}} \right)$$

ονομάζεται 95% διάστημα εμπιστοσύνης (95% δ.ε.). Μπορούμε να συμπεράνουμε με πιθανότητα 95% ότι η μέση τιμή του πληθυσμού δεν θα απέχει πάνω από $2 \frac{s}{\sqrt{n}}$ από τον μέσο όρο του δείγματος.

Διαστήματα Εμπιστοσύνης

- Αν εφαρμοστούν οι τύποι και η μεθοδολογία που είδαμε προηγουμένως, το 95% διάστημα εμπιστοσύνης (95% δε) για την αναλογία p του πληθυσμού είναι

$$\left(p - 2 \sqrt{\frac{p(1-p)}{n}}, p + 2 \sqrt{\frac{p(1-p)}{n}} \right)$$

Παράδειγμα 7.1

- Απλή τυχαία δειγματοληψία με $n=1000$ και $p=0,30$. Ποιο θα ήταν το 95% δ.ε. για την περίπτωση αυτή;

- $$\left(p - 2\sqrt{\frac{p(1-p)}{n}}, p + 2\sqrt{\frac{p(1-p)}{n}} \right) =$$

$$\left(0,3 - 2\sqrt{\frac{0,3(1-0,3)}{1000}}, 0,3 + 2\sqrt{\frac{0,3(1-0,3)}{1000}} \right) =$$

$$(0,3 - 0,029, 0,3 + 0,029) = (0,271, 0,329)$$

Παράδειγμα 7.1

- Λάθος εκτίμησης e

$$e = 2 \sqrt{\frac{p(1 - p)}{n}}$$

- Μέγιστο λάθος εκτίμησης. Εμφανίζεται όταν $p=0,50$

$$e = 2 \sqrt{\frac{0,25}{n}} = \frac{1}{\sqrt{n}}$$

Παράδειγμα 7.1

- Για τη μέση τιμή το 95% δ.ε. Είναι

$$\left(\bar{x} - 1,96 \frac{s}{\sqrt{n}}, \bar{x} + 1,96 \frac{s}{\sqrt{n}} \right)$$

- Για το ποσοστό το 95% δ.ε. είναι

$$\left(p - 1,96 \sqrt{\frac{p(1-p)}{n}}, p + 1,96 \sqrt{\frac{p(1-p)}{n}} \right)$$

Παράδειγμα 7.1

- Για οποιοδήποτε $100(1-\alpha)\%$ δ.ε. οι τύποι γίνονται:
- Για τη μέση τιμή το δ.ε. είναι

$$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

- Για το ποσοστό το δ.ε. είναι

$$\left(p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}, p + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \right)$$

Παράδειγμα 7.2

- Ποιο θα ήταν το 98% δ.ε.;
- $100(1-\alpha)\%=98\%$ τότε $\alpha=2\%=0,02$. Άρα $\alpha/2=0,01$ και $1-\alpha/2=0,99$
- $z_{0,99} = 2,33$
 - Για τη μέση τιμή το 98% δ.ε. είναι

$$\left(\bar{x} - 2,33 \frac{s}{\sqrt{n}}, \bar{x} + 2,33 \frac{s}{\sqrt{n}} \right)$$

- Για το ποσοστό το 98% δ.ε. είναι

$$\left(p - 2,33 \sqrt{\frac{p(1-p)}{n}}, p + 2,33 \sqrt{\frac{p(1-p)}{n}} \right)$$

Διάστημα εμπιστοσύνης για τη μέση τιμή για μικρά δείγματα

- Το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για τη μέση τιμή

$$\left(\bar{X} - t_{n-1; \frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1; \frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

Παράδειγμα 7.3

- Ας υποθέσουμε ότι ο μέσος όρος της βαθμολογίας τυχαίου δείγματος 20 φοιτητών στο μάθημα της στατιστικής, είναι 7 και η τυπική απόκλιση 3. Υπολογίστε ένα 98% δ.ε. για τη μέση τιμή της βαθμολογίας των φοιτητών στο μάθημα της στατιστικής.

Παράδειγμα 7.3

$$\begin{aligned} & \left(\bar{X} - t_{19;0,01} \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_{19;0,01} \cdot \frac{s}{\sqrt{n}} \right) \\ &= \left(7 - t_{19;0,01} \cdot \frac{3}{\sqrt{20}}, 7 + t_{19;0,01} \cdot \frac{3}{\sqrt{20}} \right) \\ &= \left(7 - 2,539 \cdot \frac{3}{\sqrt{20}}, 7 + 2,539 \cdot \frac{3}{\sqrt{20}} \right) \\ &= (5,27, \quad 8,7) \end{aligned}$$

Διάστημα εμπιστοσύνης για τη διαφορά μέσων τιμών-Μεγάλα δείγματα

- **Παράδειγμα 7.4**

Ένας καθηγητής στατιστικής διδάσκει το ίδιο μάθημα με τον ίδιο τρόπο σε δύο τμήματα. Στο τέλος του εξαμήνου παίρνει δύο τυχαία δείγματα από 100 φοιτητές από κάθε τμήμα και μελετά τους βαθμούς που έγραψαν οι φοιτητές των δύο τμημάτων στις εξετάσεις του μαθήματος. Βρίσκει λοιπόν ο καθηγητής ότι το 1^ο τμήμα έχει μέσο όρο 7 και τυπική απόκλιση 2 και το 2^ο τμήμα μέσο όρο 8 και τυπική απόκλιση 3. Τι μπορεί να συμπεράνει συγκριτικά για την επίδοση των φοιτητών των δύο τμημάτων;

Παράδειγμα 7.4

- Όταν τα δείγματα είναι μεγάλα, το $100(1-\alpha)\%$ δ.ε. δίνεται από

- $$\left(\bar{X}_1 - \bar{X}_2 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

όπου οι συμβολισμοί που χρησιμοποιούνται είναι

	1ο δείγμα	2ο δείγμα
Μέγεθος δείγματος	$n_1=100$	$n_2=100$
Μέσος όρος	$\bar{X}_1=7$	$\bar{X}_2=8$
Διασπορά	$s_1^2=4$	$s_2^2=9$

Παράδειγμα 7.4

- $z_{1-\frac{\alpha}{2}} = z_{1-0,025} = z_{0,975} = 1,96$
- $\left(7 - 8 - 1,96 \cdot \sqrt{\frac{4}{100} + \frac{9}{100}}, 7 - 8 + 1,96 \cdot \sqrt{\frac{4}{100} + \frac{9}{100}} \right) = (-1,7, -0,29)$
- Το 95% δ.ε. που υπολογίσαμε έχει και τα δύο του άκρα αρνητικά, άρα όλες οι τιμές του διαστήματος είναι αρνητικές. Άρα με σιγουριά 95% μπορούμε να συμπεράνουμε ότι το 2^ο τμήμα έχει καλύτερη επίδοση, αφού οι δική του μέση βαθμολογία αφαιρείται από αυτήν του 1^{ου} και παράγει αποτέλεσμα που είναι αρνητικός αριθμός.

Μικρά δείγματα

- Προϋποθέσεις: τα δείγματα που χρησιμοποιούμε είναι τυχαία, ανεξάρτητα μεταξύ τους και προέρχονται από κανονικούς πληθυσμούς. Για την συγκεκριμένη περίπτωση που εξετάζουμε όμως θα πρέπει να υποθέσουμε και ότι οι διασπορές των δύο αυτών πληθυσμών είναι μεταξύ τους ίσες.
- Η εκτίμηση της κοινής διασποράς των δύο πληθυσμών δίνεται από τον τύπο

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

Μικρά δείγματα

- 100(1-α)% δ.ε. για τη διαφορά των δύο μέσων τιμών των δύο πληθυσμών ($\mu_1 - \mu_2$)

$$\left(\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2; \frac{\alpha}{2}} \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2; \frac{\alpha}{2}} \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Παράδειγμα 7.5

- Ο καθηγητής στατιστικής διδάσκει το ίδιο μάθημα με τον ίδιο τρόπο σε δύο τμήματα. Στο τέλος του εξαμήνου παίρνει δύο τυχαία δείγματα από 15 φοιτητές από κάθε τμήμα και μελετά τους βαθμούς που έγραψαν οι φοιτητές των δύο τμημάτων στις εξετάσεις του μαθήματος. Βρίσκει λοιπόν ο καθηγητής ότι το 1^ο τμήμα έχει μέσο όρο 7 και τυπική απόκλιση 2 και το 2^ο τμήμα μέσο όρο 8 και τυπική απόκλιση 3. Τι μπορεί να συμπεράνει συγκριτικά για την επίδοση των φοιτητών των δύο τμημάτων;

Παράδειγμα 7.5

- $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(15-1)4 + (15-1)9}{15+15-2} = 6,5$
- $t_{n_1+n_2-2; \frac{\alpha}{2}} = t_{28; 0,025} = 2,048$

$$\begin{aligned} & \left(\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2; \frac{\alpha}{2}} \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2; \frac{\alpha}{2}} \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \\ &= \left(7 - 8 - 2,048 \cdot 2,6 \cdot \sqrt{\frac{1}{15} + \frac{1}{15}}, \right. \\ & \quad \left. 7 - 8 + 2,048 \cdot 2,6 \cdot \sqrt{\frac{1}{15} + \frac{1}{15}} \right) = (-1 - 1,94, \quad -1 + 1,94) \end{aligned}$$

Παράδειγμα 7.5

- Το 95% δ.ε. που υπολογίσαμε έχει το ένα άκρο του αρνητικό και το άλλο θετικό, άρα περιλαμβάνει την τιμή μηδέν, η οποία θα δήλωνε την ισότητα των δύο μέσων τιμών. Άρα δεν μπορούμε να πούμε, με βάση τα δεδομένα που έχουμε, ότι οι δύο μέσες τιμές διαφέρουν.

Διάστημα εμπιστοσύνης για παρατηρήσεις κατά ζεύγη

- Οι παρατηρήσεις κατά ζεύγη, ή αλλιώς ζευγαρωτές παρατηρήσεις, αναφέρονται σε μετρήσεις στο ίδιο δείγμα σε δύο διαφορετικές χρονικές στιγμές ή σε μετρήσεις σε δύο δείγματα που είναι μεταξύ τους εξαρτημένα.
- Συνήθως το δείγμα είναι μικρό ($n < 30$) και έτσι χρησιμοποιούμε την κατανομή t του Student.

Παρατηρήσεις κατά ζεύγη

Αριθμός παρατήρησης	Πριν (X)	Μετά (Y)	Διαφορά $Z=X-Y$
1	x_1	y_1	$z_1 = x_i - y_i$
2	x_2	y_2	$z_2 = x_i - y_i$
3	x_3	y_3	$z_3 = x_i - y_i$
.	.	.	.
.	.	.	.
n-1	x_{n-1}	y_{n-1}	$z_{n-1} = x_i - y_i$
n	x_n	y_n	$z_n = x_i - y_i$

Παρατηρήσεις κατά ζεύγη

- $\bar{z} = \frac{\sum_{i=1}^n z_i}{n}$

- $s_z = \sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}} = \sqrt{\frac{\sum z_i^2 - \frac{(\sum z_i)^2}{n}}{n-1}}$

- $100(1-\alpha)\% \delta\varepsilon:$

$$\left(\bar{z} - t_{n-1; \frac{\alpha}{2}} \cdot \frac{s_z}{\sqrt{n}}, \bar{z} + t_{n-1; \frac{\alpha}{2}} \cdot \frac{s_z}{\sqrt{n}} \right)$$

Παράδειγμα 7.6

- Ένας γιατρός έχει 10 υπερτασικούς ασθενείς και θέλει να δοκιμάσει ένα φάρμακο για την πίεση. Μετρά την πίεση των ασθενών και στη συνέχεια οι ασθενείς αρχίζουν θεραπεία με καθημερινή χορήγηση του φαρμάκου για ένα μήνα. Στο τέλος του μήνα ο γιατρός ξαναμετρά την πίεσή τους. Οι μετρήσεις περιγράφονται στον παρακάτω πίνακα. Τι συμπεράσματα μπορεί να βγάλει ο γιατρός;

Παράδειγμα 7.6

ασθενής	Μέτρηση πριν	Μέτρηση μετά
1	15	12
2	16	15
3	15	13
4	15	12
5	16	13
6	16	16
7	15	16
8	18	15
9	16	13
10	17	12

Παράδειγμα 7.6

ασθενής	Μέτρηση πριν	Μέτρηση μετά	Διαφορά
1	15	12	3
2	16	15	1
3	15	13	2
4	15	12	3
5	16	13	3
6	16	16	0
7	15	16	-1
8	18	15	3
9	16	13	3
10	17	12	5

Παράδειγμα 7.6

- $\bar{Z} = \frac{\sum_{i=1}^n z_i}{n} = \frac{22}{10} = 2,2$

- $s_Z = \sqrt{\frac{\sum z_i^2 - \frac{(\sum z_i)^2}{n}}{n-1}} =$
 $\sqrt{\frac{76 - \frac{22^2}{10}}{9}} = 1,75$

ασθενής	Διαφορά z	z ²
1	3	9
2	1	1
3	2	4
4	3	9
5	3	9
6	0	0
7	-1	1
8	3	9
9	3	9
10	5	25
Άθροισμα	$\sum_{i=1}^n z_i = 22$	$\sum z_i^2 = 76$

Παράδειγμα 7.6

- $\left(\bar{z} - t_{n-1; \frac{\alpha}{2}} \cdot \frac{s_z}{\sqrt{n}}, \bar{z} + t_{n-1; \frac{\alpha}{2}} \cdot \frac{s_z}{\sqrt{n}} \right) =$
 $\left(\bar{z} - t_{9; 0,025} \cdot \frac{s_z}{\sqrt{n}}, \bar{z} + t_{9; 0,025} \cdot \frac{s_z}{\sqrt{n}} \right) =$
 $\left(2,2 - 2,262 \cdot \frac{1,75}{\sqrt{10}}, 2,2 + 2,262 \cdot \frac{1,75}{\sqrt{10}} \right) = (0,95, 3,45)$
- Το διάστημα περιέχει θετικές τιμές, άρα η πίεση έπεσε με τη χορήγηση του φαρμάκου.

Διάστημα εμπιστοσύνης για τη διαφορά αναλογιών

- Γενικά το $100(1-\alpha)\%$ δ.ε. για τη διαφορά ποσοστών είναι
- $\left(p_1 - p_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, p_1 - p_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$

Παράδειγμα 7.7

- Μια εταιρία δημοσκοπήσεων πραγματοποιεί μια πανελλαδική δημοσκόπηση με συνολικό δείγμα $n=1000$. Το δείγμα είναι τυχαίο. Το δείγμα χωρίζεται σε δύο τμήματα, τα 600 άτομα είναι το δείγμα που αντιστοιχεί στις αστικές περιοχές και τα υπόλοιπα 400 στις αγροτικές. Στην έρευνα διερευνάται ποιο είναι το ποσοστό που θα έπαιρνε το κόμμα Α αν αύριο γινόταν εκλογές. Το ποσοστό αυτό είναι για το δείγμα των αστικών περιοχών 30% και για το δείγμα των αγροτικών περιοχών 27%. Υπάρχει διαφορά ανάμεσα στο ποσοστό των αστικών και των αγροτικών περιοχών;

Παράδειγμα 7.7

- Το 95% δ.ε. για τη διαφορά των ποσοστών είναι
- $$\left(p_1 - p_2 - 1,96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, p_1 - p_2 + 1,96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) =$$
$$\left(0,30 - 0,27 - 1,96 \sqrt{\frac{0,30(1-0,30)}{600} + \frac{0,27(1-0,27)}{400}}, 0,30 - 0,27 + 1,96 \sqrt{\frac{0,30(1-0,30)}{600} + \frac{0,27(1-0,27)}{400}} \right) =$$
$$(-0,0269, 0,0869)$$
- Παρατηρούμε ότι τα άκρα του διαστήματος είναι ετερόσημα. Άρα τα ποσοστά διαφέρουν.