



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική

Ενότητα 2: Επαγωγική Στατιστική

Βασίλης Γιαλαμάς
Σχολή Επιστημών της Αγωγής
Τμήμα Εκπαίδευσης και Αγωγής στην Προσχολική
Ηλικία

Περιεχόμενα ενότητας

Παρουσιάζονται οι βασικές έννοιες του ελέγχου υποθέσεων που αφορούν στη σύγκριση μέσων τιμών και τη συνάφεια μεταξύ δυο μεταβλητών. Γίνεται επίσης εισαγωγή στην εκτιμητική της μέσης τιμής με τη βοήθεια των διαστημάτων εμπιστοσύνης.



Εκτιμητική

Σημειακές εκτιμήσεις παραμέτρων

Σε πολλές περιπτώσεις το ενδιαφέρον των ερευνητών είναι η εκτίμηση της τιμής μιας παραμέτρου ενός πληθυσμού.

Η παιδιάτρος του παραδείγματος που είδαμε στον έλεγχο υποθέσεων που προτείνει το μεγάλωμα των παιδιών με «πολλές αγκαλιές», ενδιαφέρεται για το μέσο βάρος των παιδιών αυτών. Δηλαδή, ενδιαφέρεται για μια εκτίμηση της μέσης τιμής μ του πληθυσμού των βαρών των παιδιών αυτών με τη βοήθεια του δείγματος που διαθέτει.

Ο βέλτιστος σημειακός εκτιμητής της μέσης τιμής ενός πληθυσμού είναι η μέση τιμή ενός τυχαίου δείγματος από τον πληθυσμό: $\hat{\mu} = \bar{X}$

Ονομάζεται **σημειακός** επειδή δίνει **μια μόνο τιμή** για τη μέση τιμή του πληθυσμού.

Στην περίπτωση των παιδιών «με πολλές αγκαλιές» $\hat{\mu} = 12 \text{ Kgr}$ αφού η μέση τιμή του δείγματος είναι: $\bar{X} = 12 \text{ Kgr}$



Διαστήματα εμπιστοσύνης (1 από 8)

Η βεβαιότητα ότι η εκτίμηση σημείου έδωσε την πραγματική μέση τιμή του πληθυσμού είναι πολύ μικρή. Οι μέσες τιμές των δειγμάτων έχουν μια φυσική μεταβλητότητα που εκφράζεται από το τυπικό σφάλμα της μέσης τιμής.

Αυτό οδήγησε σε μεθόδους εκτίμησης της μέσης τιμής ενός πληθυσμού με τη βοήθεια ενός διαστήματος τιμών για το οποίο η βεβαιότητα να περιέχει την μέση τιμή του πληθυσμού είναι γνωστή και υψηλή.



Διαστήματα εμπιστοσύνης (2 από 8)

Η λογική της κατασκευής ενός διαστήματος εμπιστοσύνης είναι απλή:

Για ένα πληθυσμό με άγνωστη μέση τιμή μ και γνωστή τυπική απόκλιση σ , όταν ισχύει το Κ.Ο.Θ. η δειγματοληπτική κατανομή της μέσης τιμής είναι κανονική με μέση τιμή μ και τυπική απόκλιση σ/\sqrt{N} .

Συνεπώς το 95% των δειγματικών μέσων \bar{X} ανήκει στο παρακάτω διάστημα:

$$\left[\mu - 1,96 \cdot \frac{\sigma}{\sqrt{N}}, \mu + 1,96 \cdot \frac{\sigma}{\sqrt{N}} \right]$$

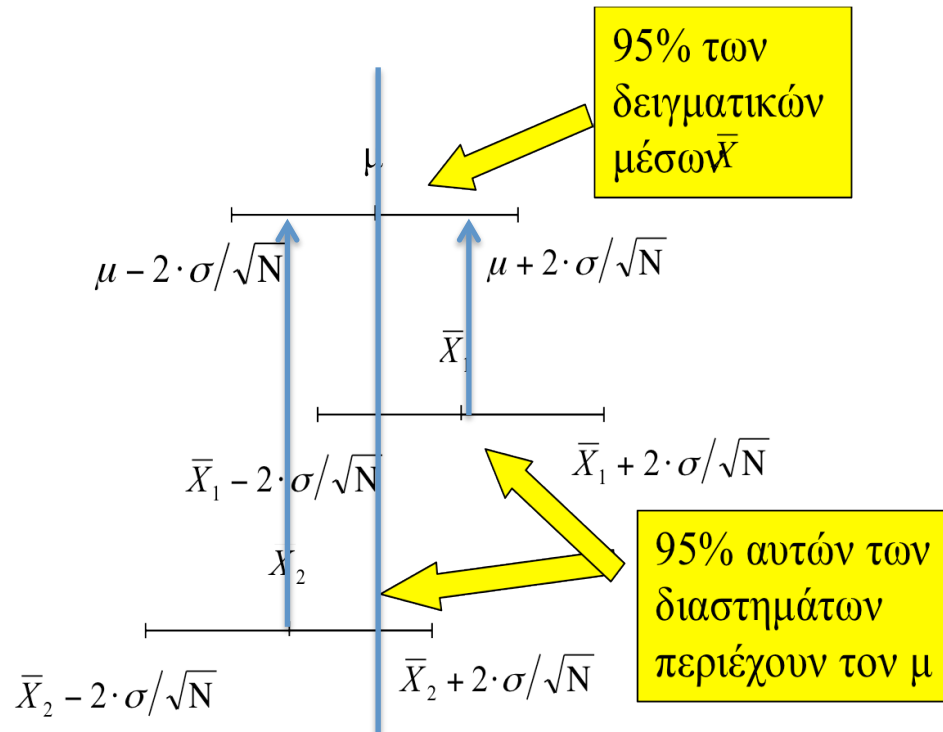
Ισοδύναμα η μέση τιμή μ του πληθυσμού περιλαμβάνεται στο 95% των διαστημάτων της μορφής:

$$\left[\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{N}} \right]$$

Το παραπάνω διάστημα ονομάζεται **95%-διάστημα εμπιστοσύνης** και το ποσοστό **95%** είναι ο **συντελεστής εμπιστοσύνης**.



Διαστήματα εμπιστοσύνης (3 από 8)



Διαστήματα εμπιστοσύνης (4 από 8)

Το 95%-διαστήματος εμπιστοσύνης για τη μέση τιμή του πληθυσμού των βρεφών με «πολλές αγκαλιές» με τη βοήθεια των στοιχείων του δείγματος ($N=36$) με $\bar{X} = 12 \text{ Kgr}$ και δεδομένο ότι η τυπική απόκλιση του συγκεκριμένου πληθυσμού είναι γνωστή $\sigma=2 \text{ Kgr}$ είναι:

$$\begin{aligned} \left[\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{N}} \right] &= \left[12 \pm 1,96 \cdot \frac{2}{\sqrt{36}} \right] \\ &= [12 \pm 0,65] = [11,35 \quad 12,65] \end{aligned}$$



Διαστήματα εμπιστοσύνης (5 από 8)

Η τιμή 1,96 είναι η τιμή της τυπική κανονικής κατανομής και θα μπορούσε να είναι μια οποιαδήποτε τιμή $Z_{(1-\alpha/2)}$ αυτής της κατανομής που ονομάζεται **συντελεστής αξιοπιστίας** καθορίζοντας το ποσοστό των $100(1-\alpha)\%$ διαστημάτων της μορφής:

$$\left[\bar{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{N}} \right]$$

που περιέχουν την άγνωστη μέση τιμή μ του πληθυσμού.

Αν ο στόχος μας είναι η κατασκευή ενός 90%-διαστήματος τότε $\alpha = 0,1$ αφού $100(1-0,1)\% = 90\%$. Άρα η στον πίνακα της τυπικής κανονικής κατανομής θα αναζητηθεί η $Z_{1-0,05}$ ή $Z_{0,95}$. Από τον πίνακα προκύπτει $Z_{0,95} = 1,64$ (στη στήλη B για $B=0,95$).

Το 90%-διάστημα εμπιστοσύνης είναι :

$$\left[\bar{X} - 1,64 \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + 1,64 \cdot \frac{\sigma}{\sqrt{N}} \right]$$



Διαστήματα εμπιστοσύνης (6 από 8)

Το 90%-διαστήματος εμπιστοσύνης για τη μέση τιμή του πληθυσμού των βρεφών με «πολλές αγκαλιές» με τη βοήθεια των στοιχείων του δείγματος ($N=36$) με $\bar{X} = 12 \text{ Kgr}$ και δεδομένο ότι η τυπική απόκλιση του συγκεκριμένου πληθυσμού είναι γνωστή $\sigma=2 \text{ Kgr}$ είναι:

$$\begin{aligned} & \left[\bar{X} - 1,64 \cdot \frac{\sigma}{\sqrt{N}}, \bar{X} + 1,64 \cdot \frac{\sigma}{\sqrt{N}} \right] = \left[12 \pm 1,64 \cdot \frac{2}{\sqrt{36}} \right] \\ & = [12 \pm 0,55] = [11,45 \ 12,55] \end{aligned}$$

Το 95% διάστημα $[11,35 \ 12,65]$ είχε το ίδιο κέντρο με το 90%-διάστημα, την μέση τιμή $\bar{X} = 12 \text{ Kgr}$ του δείγματος, αλλά είναι ευρύτερο. Το εύρος το πρώτου είναι $12,65 - 11,35 = 1,3 \text{ Kgr}$ ενώ του τελευταίου $1,1 \text{ Kgr}$.

Δηλαδή όσο αυξάνεται η εμπιστοσύνη μας ότι το διάστημα περιέχει τη μέση τιμή του πληθυσμού τόσο μειώνεται η ακρίβεια της εκτίμησης.



Διαστήματα εμπιστοσύνης (7 από 8)

Αν η κατανομή του πληθυσμού δεν είναι κανονική η μορφή του διαστήματος εμπιστοσύνης που είδαμε:

$$[\bar{X} \pm Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{N}}]$$

μπορεί να χρησιμοποιηθεί μόνο αν $N \geq 30$.

Στην πράξη σπάνια είναι γνωστή η τυπική απόκλιση του πληθυσμού. Τότε, όπως είδαμε στον έλεγχο υποθέσεων αν ο πληθυσμός ακολουθεί κανονική κατανομή το ηλίκο $\frac{\bar{X} - \mu}{S/\sqrt{N}}$ ακολουθεί κατανομή t με $N-1$ βαθμούς ελευθερίας.

Άρα το διάστημα $100(1-\alpha)\%$ διάστημα εμπιστοσύνης είναι:

$$[\bar{X} \pm t_{(1-\alpha/2)} \cdot \frac{S}{\sqrt{N}}]$$

Όπου η τιμή $t_{(1-\alpha/2)}$ είναι η τιμή του πίνακα της t για $N-1$ βαθμούς ελευθερίας κάτω από την οποία βρίσκεται το $\alpha/2$ ποσοστό των τιμών της κατανομής.



Διαστήματα εμπιστοσύνης (8 από 8)

Για την κατασκευή ενός 95% διαστήματος για τη μέση επίδοση νηπίων που διδάσκονται αρμόνιο με την μέθοδο «Θεοχαράτου» χρησιμοποιώντας τυποποιημένο test για το οποίο η μέση επίδοση ενός δείγματος μαθητών ($N=17$) ήταν $\bar{X} = 6,8$ και $S=1,8$.

Από την επισκόπηση του ιστογράμματος των βαθμών αρμονίου δεν προέκυψε σοβαρή απόκλιση από τη κανονική κατανομή συνεπώς μπορεί να χρησιμοποιηθεί η κατανομή t και συγκεκριμένα η τιμή της $t_{(1-\alpha/2)}$ και το διάστημα:

$$\left[\bar{X} \pm t_{(1-\alpha/2)} \cdot \frac{S}{\sqrt{N}} \right]$$

Η τιμή $t_{(1-0,05/2)}$ ή $t_{0,975}$ για 16 βαθμούς ελευθερίας από τον πίνακα της t αντιστοιχεί στην τιμή 2,12 και το 95% διάστημα εμπιστοσύνης:

$$\left[\bar{X} - 2,12 \cdot \frac{S}{\sqrt{N}}, \bar{X} + 2,12 \cdot \frac{S}{\sqrt{N}} \right] = \left[6,8 \pm 2,12 \cdot \frac{1,8}{\sqrt{17}} \right] = \left[6,8 \pm 0,93 \right] = \left[5,87 \quad 7,73 \right]$$



Διαστήματα εμπιστοσύνης με το SPSS (1 από 4)

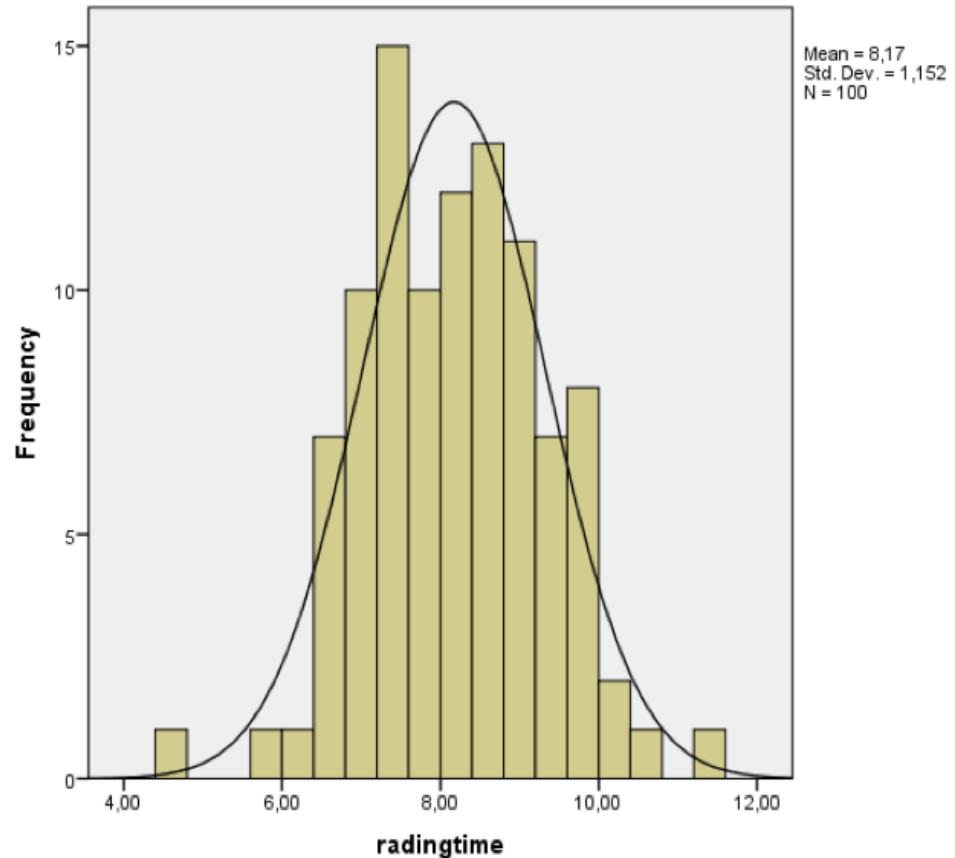
Στόχοι έρευνας:

- Να κατασκευαστεί ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνο μελέτης των σημερινών μαθητών στο σπίτι.
- Να διερευνηθεί κατά πόσο ο χρόνος μελέτης των σημερινών μαθητών διαφέρει από το χρόνο που μελετούσαν οι μαθητές της ίδιας ηλικίας πριν 20 έτη (σε έρευνα της εποχής βρέθηκε $\mu=7,5$ ώρες).

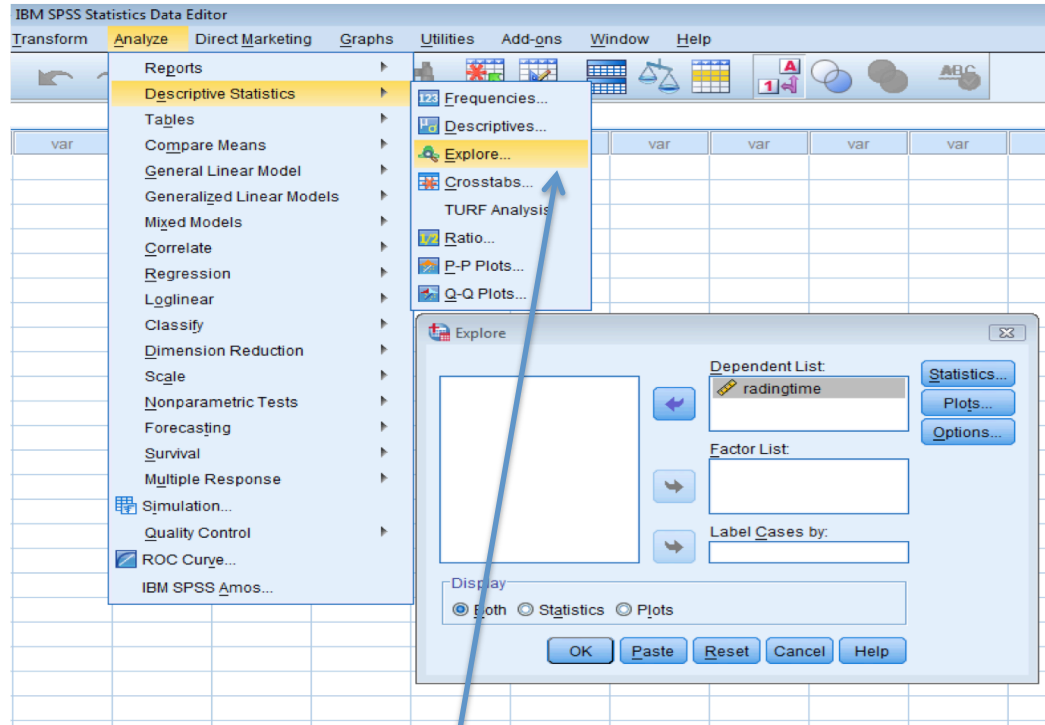


Διαστήματα εμπιστοσύνης με το SPSS (2 από 4)

Από το ιστόγραμμα δεν παρατηρούμε σοβαρή απόκλιση της κατανομής των χρόνων μελέτης δείγματος από την κανονική κατανομή. Συνεπώς για την κατασκευή του διαστήματος θα χρησιμοποιηθεί η t αφού είναι άγνωστη η διακύμανση του χρόνου μελέτης στον πληθυσμό των σημερινών μαθητών.



Διαστήματα εμπιστοσύνης με το SPSS (3 από 4)



Επιλέγουμε
Descriptive Statistics=> Explore



Διαστήματα εμπιστοσύνης με το SPSS (4 από 4)

Το ζητούμενο 95% διάστημα εμπιστοσύνης.

Απορρίπτεται η μηδενική υπόθεση $\mu=7,5$ σε επίπεδο σημαντικότητας $\alpha=0,05$ επειδή η τιμή 7,5 δεν ανήκει στο διάστημα 7,9424-8,3996. Επειδή $7,5 < 7,94$ συμπεραίνουμε ότι οι σημερινοί μαθητές μελετούν σημαντικά περισσότερο σε σχέση με το παρελθόν.

Descriptives					
		Statistic	Std. Error		
readingtime	Mean	8,1710	,11520		
	95% Confidence Interval for Mean	Lower Bound	7,9424		
		Upper Bound	8,3996		
	5% Trimmed Mean	8,1714			
	Median	8,0821			
	Variance	1,327			
	Std. Deviation	1,15200			
	Minimum	4,58			
	Maximum	11,28			
	Range	6,69			
	Interquartile Range	1,63			



Τέλος Ενότητας

Επαγωγική Στατιστική

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.



Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Βασίλης Γιαλαμάς 2015. Βασίλης Γιαλαμάς. «Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική. Επαγωγική Στατιστική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/ECD102/>.



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.



Σημείωμα Χρήσης Έργων Τρίτων (1/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Σχήμα 1, Σελίδα 7: Σχήμα με διαστήματα εμπιστοσύνης / Copyrighted

Εικόνα 1, Σελίδα 14: Ιστόγραμμα / Copyrighted

Εικόνα 2, Σελίδα 15: Εικόνα οθόνης ηλεκτρονικού υπολογιστή όπου γίνεται χρήση του SPSS / Copyrighted



Σημείωμα Χρήσης Έργων Τρίτων (2/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Πίνακες

Πίνακας 1, Σελίδα 16: Πίνακας δεδομένων / Copyrighted

