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.

# Έλεγχος υπόθεσης για το μέσο ενός πληθυσμού

# Ερευνητικό Ερώτημα

Μια παιδίατρος θέλει να διερευνήσει κατά πόσο παιδιά 2 χρόνων που έχουν μεγαλώσει δεχόμενα πολλές αγκαλιές από τους γονείς τους παρουσιάζουν μεγαλύτερη ανάπτυξη από τα άλλα παιδιά.

# Διαδικασία Παρέμβασης

Δίνει οδηγίες στους γονείς 36 βρεφών που δέχτηκαν να συμμετέχουν στο πρόγραμμά της, σχετικά με το πόσο συχνά θα αγκαλιάζουν τα παιδιά τους.

Μετά από 2 χρόνια εφαρμογής του προγράμματος από τους γονείς, μετρά το βάρος των 36 παιδιών ηλικίας 2 ετών. Το μέσο βάρος του δείγματος είναι $\overline{X} = 12$ κιλά.

Ο έλεγχος των στατιστικών υποθέσεων δίνεται με τη μορφή μιας διαδικασίας τεσσάρων βημάτων.

# Βήμα 1ο (1 από 3)

Διατυπώνονται οι στατιστικές υποθέσεις (μηδενική, εναλλακτική) και επιλέγεται το **επίπεδο σημαντικότητας α**. (συνήθως α=0,05)

Η **μηδενική υπόθεση** δηλώνει ότι η μέση τιμή $\mu$ του πληθυσμού για τον οποίο γίνεται ο έλεγχος είναι ίση με μια τιμή $\mu_0$.

$$H_0 : \mu = \mu_0$$

Για το παράδειγμα μας, $H_0 : \mu = 10,5$ , όπου $\mu$ είναι η μέση τιμή βάρους του πληθυσμού των παιδιών με «πολλές αγκαλιές» μέχρι 2 ετών.

# Βήμα 1ο (2 από 3)

**Η εναλλακτική υπόθεση** είναι το συμπλήρωμα της μηδενικής, δηλώνει διαφορά και συνήθως εκφράζει την επιθυμία του ερευνητή.

$$H_A : \mu \neq \mu_0$$

Για το παράδειγμα μας, $H_A : \mu \neq 10,5$

Από προηγούμενες έρευνες είναι γνωστό ότι το μέσο βάρος γενικά των παιδιών σ' αυτή την ηλικία είναι **10,5 κιλά** και η τυπική απόκλιση του είναι **σ = 2 κιλά**.

Στον στατιστικό έλεγχο υποθέσεων τα στοιχεία του δείγματος οδηγούν σε δύο δυνατές αποφάσεις:

- Απόρριψη της μηδενικής υπόθεσης που ισοδυναμεί με αποδοχή της εναλλακτικής
- Μη απόρριψη της μηδενικής υπόθεσης (αποδοχή)

# Βήμα 1ο (3 από 3)

Στη περίπτωσή μας, απόρριψη της μηδενικής υπόθεσης και υιοθέτηση της εναλλακτικής οδηγεί στο συμπέρασμα ότι τα παιδιά με «πολλές αγκαλιές» έχουν διαφορετικό μέσο βάρος από 10,5 κιλά, δηλαδή διαφορετικό βάρος από τα τυπικά παιδιά. Αν, δε, το δείγμα μας έχει μέση τιμή μεγαλύτερη από 10,5 κιλά τότε έχουμε απάντηση στο ερευνητικό ερώτημα:

**Τα παιδιά 2 χρόνων που έχουν μεγαλώσει δεχόμενα πολλές αγκαλιές από τους γονείς τους, παρουσιάζουν μεγαλύτερη ανάπτυξη από τα άλλα παιδιά.**

# Βήμα 2ο (1 από 5)

Σ’ αυτό το στάδιο **επιλέγουμε το Στατιστικό ή κριτήριο του ελέγχου**. Υποθέτουμε ότι ισχύει η μηδενική υπόθεση, δηλ. η μέση τιμή του πληθυσμού είναι $\mu_0$ και η τυπική του απόκλιση $\sigma$ γνωστή.

Τότε η δειγματοληπτική κατανομή της μέσης τιμής για δείγματα με N=36 **είναι κανονική**, με μέση τιμή $\mu_0$ και τυπική απόκλιση $\sigma/\sqrt{\mathrm{N}}$ σύμφωνα με το Κ.Ο.Θ.

Το πηλίκο $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{\mathrm{N}}}$ ακολουθεί την τυπική κανονική κατανομή και είναι το κριτήριο του ελέγχου που χρησιμοποιείται για τον κανόνα απόφασης.

Όπου $\mu_0$ είναι η τιμή της μηδενικής υπόθεσης για τη μέση του πληθυσμού και $\sigma$ η τυπική απόκλιση του πληθυσμού που είναι γνωστή.

# Βήμα 2ο (2 από 5)

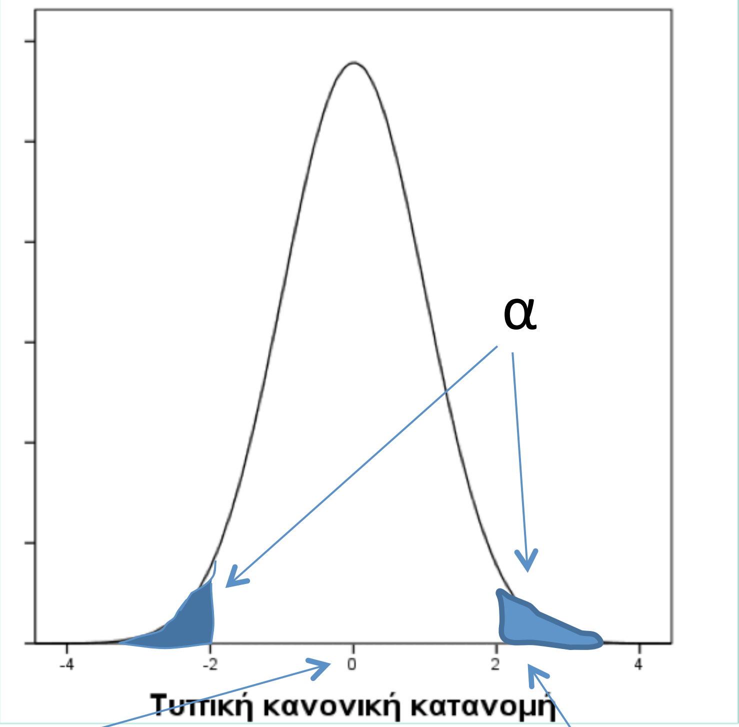
Επειδή θεωρητικά ένα τυχαίο δείγμα από τον πληθυσμό με μέση τιμή την $\mu_0$ μπορεί να δώσει οποιαδήποτε τιμή z, πρέπει να αποφασιστεί ποιες τιμές Z απορρίπτουν την μηδενική υπόθεση.

Προφανώς αυτές είναι στα άκρα της τυπικής κανονικής κατανομής αφού όσο πιο πολύ αποκλίνει η μέση τιμή του δείγματος από την $\mu_0$ τόσο μεγαλύτερη ή απόκλιση του z του δείγματος από το 0.

Η περιοχή αποδοχής ή απόρριψης της μηδενικής καθορίζεται από το **επίπεδο σημαντικότητας α** που συνήθως είναι α = 0,05 ή 5%.

Δηλ. αν ένα δείγμα δίνει μια τιμή z που ανήκει στο 5% των πιο ακραίων τιμών της τυπικής κανονικής κατανομής τότε απορρίπτεται η μηδενική υπόθεση.

# Βήμα 2ο (3 από 5)




Z δείγματος που υποστηρίζει την $H_0$

Z δείγματος που απορρίπτει την $H_0$

# Βήμα 2ο (4 από 5)

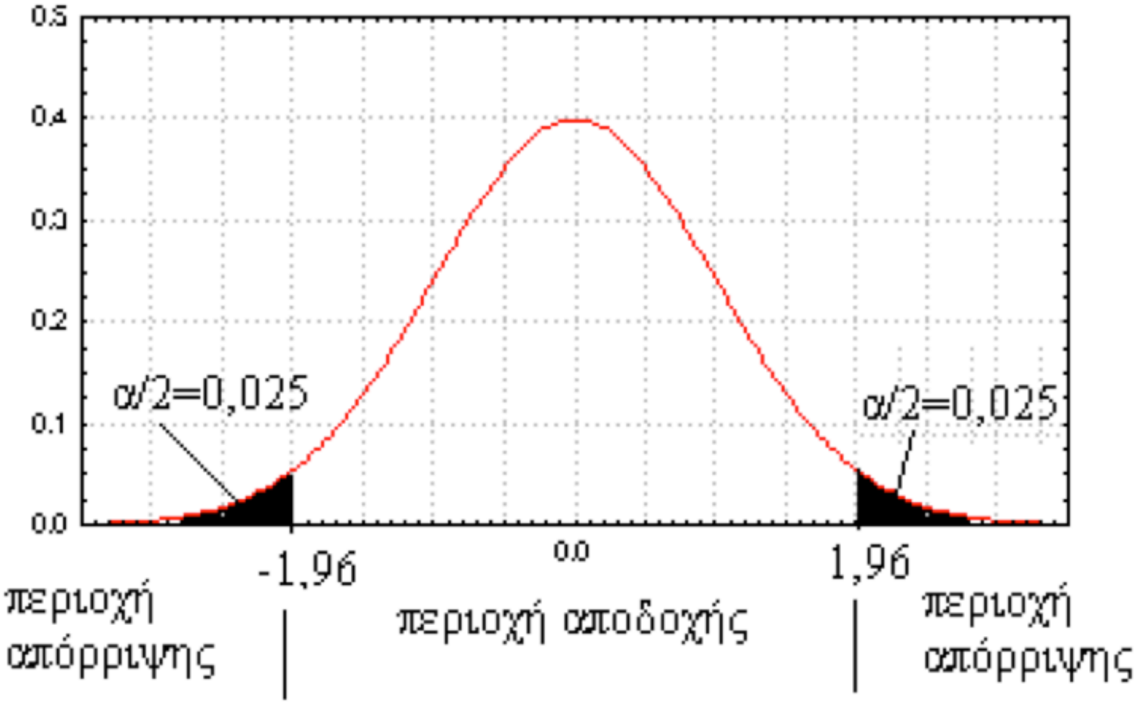
**Βρίσκουμε τις κρίσιμες με τη βοήθεια του α και δημιουργούμε τον κανόνα απόφασης σχετικά με την απόρριψη της $H_0$.**

Συνεπώς σύμφωνα από τον πίνακα των τιμών της τυπικής κανονικής και το επίπεδο σημαντικότητας α=0,05 **οι κρίσιμες τιμές** (μεταξύ των οποίων βρίσκεται το 1-α=0,95 των τιμών z) **είναι οι -1,96 και 1,96**.

**Κανόνας απόφασης**

Αν η τιμή z βρίσκεται έξω από το διάστημα (-1,96 και 1,96) απορρίπτεται η μηδενική υπόθεση. Διαφορετικά αποφασίζουμε ότι η μηδενική υπόθεση μπορεί να ισχύει.

# Βήμα 2ο (5 από 5)




Κρίσιμες τιμές, περιοχές «αποδοχής» και απόρριψης της μηδενικής υπόθεσης για αμφίπλευρο έλεγχο με κανονική κατανομή σε επίπεδο σημαντικότητας α=0,05.

# Βήμα 3ο

Στο στάδιο αυτό **υπολογίζεται η τιμή z για το δείγμα που διαθέτουμε και τη συγκρίνουμε με τις κρίσιμες τιμές του κανόνα απόφασης**.

Στο παράδειγμα μας θεωρούμε ότι ο πληθυσμός των παιδιών με «πολλές αγκαλιές» έχει την ίδια τυπική απόκλιση σ με τον υπαρκτό πληθυσμό.

$$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{\mathrm{N}}} = \frac{12 - 10,5}{2/\sqrt{36}} = \frac{1,5}{0,33} = 4,55$$

Το z ανήκει στην περιοχή απόρριψης αφού z >1,96 και απορρίπτουμε την υπόθεση ότι $\mu$=10,5. Επειδή δε η τιμή του z είναι θετική αποφασίζουμε ότι $\mu$>10,5.

# Βήμα 4ο

**Εκφράζουμε με μη στατιστικούς όρους το αποτέλεσμα του ελέγχου.**

Τα παιδιά που δέχονται «πολλές αγκαλιές» από τους γονείς τους στην βρεφική ηλικία έχουν σημαντικά ταχύτερη ανάπτυξη σε ηλικία 2 ετών.

# Μονόπλευρος έλεγχος (1 από 5)

- Ο έλεγχος που παρουσιάστηκε στις προηγούμενες διαφάνειες ονομάζεται **αμφίπλευρος** αφού απορρίπτοντας τη μηδενική υπόθεση μπορεί να αποφασίσουμε ότι η μέση τιμή είναι είτε μεγαλύτερη είτε μικρότερη από την $\mu_0$.
- Σε πολλές περιπτώσεις ο ερευνητής θέλει να υποστηρίξει ότι η μέση τιμή είναι μεγαλύτερη (ή σε άλλες περιπτώσεις μικρότερη) από την τιμή $\mu_0$. Τότε ο έλεγχος ονομάζεται **μονόπλευρος** και οι υποθέσεις γράφονται:

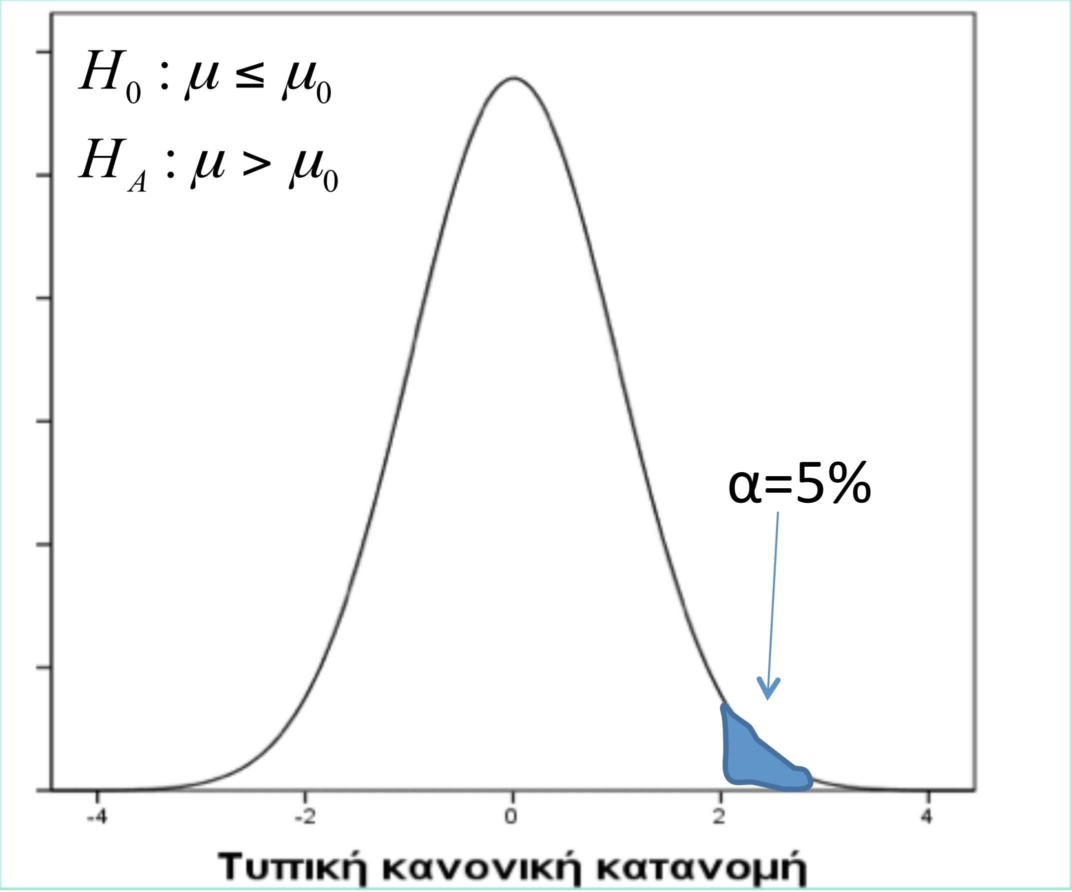
$$H_0 : \mu \le \mu_0 \qquad H_0 : \mu \ge \mu_0$$

$$H_A : \mu > \mu_0 \qquad H_A : \mu < \mu_0$$

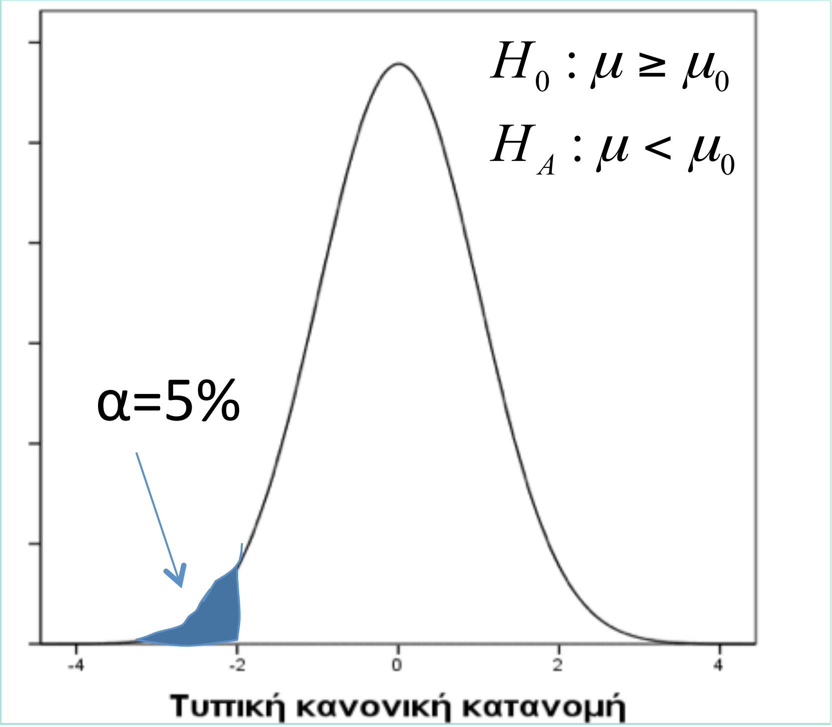
# Μονόπλευρος έλεγχος (2 από 5)

- Στον μονόπλευρο έλεγχο το επίπεδο σημαντικότητας τοποθετείται στο ένα άκρο της τυπικής κανονικής κατανομής.

**Κανόνας απόφασης:** η μηδενική υπόθεση απορρίπτεται αν $z \geq 1{,}64$ .



# Μονόπλευρος έλεγχος (3 από 5)




**Κανόνας απόφασης:** η μηδενική υπόθεση απορρίπτεται αν $z \leq -1{,}64$.

# Μονόπλευρος έλεγχος (4 από 5)

**Αν η παιδίατρος επιθυμούσε να ελέγξει την ερευνητική υπόθεση ότι τα παιδιά με «πολλές αγκαλιές» σε ηλικία 2 ετών έχουν μεγαλύτερο βάρος από τα τυπικά παιδιά θα έπρεπε να ακολουθήσει τα παρακάτω βήματα:**

1. Διατύπωση στατιστικών υποθέσεων και επιλογή επιπέδου σημαντικότητας

$$H_0 : \mu \leq 10,5$$
$$H_A : \mu > 10,5$$

$$\alpha = 0,05$$

# Μονόπλευρος έλεγχος (5 από 5)

2. Επειδή το δείγμα είναι μεγάλο Ν=36>30 και ισχύει το Κ.Ο.Θ. για τη δειγματοληπτική κατανομή της μέσης τιμής, το πηλίκο Ζ ακολουθεί την τυπική κανονική κατανομή και κανόνας απόφασης είναι: Η μηδενική υπόθεση απορρίπτεται αν $Z \geq 1{,}64$ .

3. Επειδή $z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{N}} = \dfrac{12 - 10{,}5}{2 / \sqrt{36}} = \dfrac{1{,}5}{0{,}33} = 4{,}55 > 1{,}64$

απορρίπτεται η μηδενική υπόθεση και γίνεται αποδεκτή η εναλλακτική.

4. Επαληθεύεται η ερευνητική υπόθεση.

# Κατανομή t (student) (1 από 4)

Στον προηγούμενο έλεγχο θεωρήσαμε ότι η τυπική απόκλιση του πληθυσμού είναι γνωστή.

Στην πράξη όμως σπάνια γνωρίζουμε την τυπική απόκλιση ενός αγνώστου κατά τα άλλα πληθυσμού και πρέπει να εκτιμήσουμε την τιμή της $\sigma$ από τα δειγματικά δεδομένα.

Ένας πολύ καλός εκτιμητής της $\sigma$ είναι η δειγματική τυπική απόκλιση S.

Το στατιστικό του ελέγχου που γίνεται $\dfrac{\overline{X}-\mu_0}{S/\sqrt{N}}$ δεν ακολουθεί πλέον την τυπική κανονική κατανομή αλλά την κατανομή t.

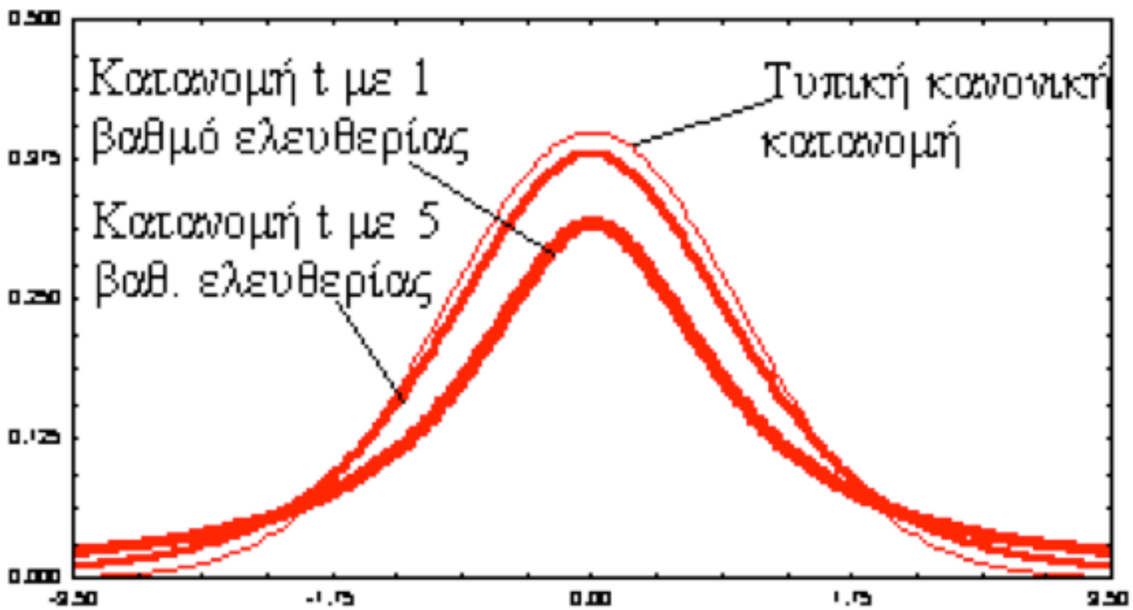
# Κατανομή t (student) (2 από 4)

Η t είναι μια κωδωνοειδής κατανομή με μικρή διαφορά από την τυπική κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση λίγο μεγαλύτερη από τη μονάδα. Οι κρίσιμες τιμές της t δίνονται από πίνακα. Για κάθε μέγεθος δείγματος N η ποσότητα N-1 εκφράζει τους βαθμούς ελευθερίας της t. Για κάθε τιμή N-1 έχουμε διαφορετική κατανομή t και διαφορετικές κρίσιμες τιμές. Όσο το μέγεθος του δείγματος ή οι βαθμοί ελευθερίας αυξάνονται τόσο η κατανομή t την τυπική κατανομή.

Αν στο παράδειγμά μας δεν γνωρίζουμε την τυπική απόκλιση του πληθυσμού τότε θα έπρεπε να εκτιμηθεί υπολογίζοντας την τ.απόκλιση του δείγματος.

Οι κρίσιμες τιμές του ελέγχου, με τη βοήθεια του πίνακα της επόμενης διαφάνειας για δείγμα μεγέθους N=18 (βαθμοί ελευθερίας 17) και επίπεδο σημαντικότητας α= 0,05, είναι -2,11 και 2,11.

# Κατανομή t (student) (3 από 4)




Τυπική κανονική κατανομή και κατανομές t-student με 1 και 5 βαθμούς ελευθερίας.

# Κατανομή t (student) (4 από 4)

| | ποσοστά της καμπύλης | | | |
|---|---|---|---|---|
| | 0,950 | 0,975 | 0,990 | 0,995 |
| Βαθμοί ελευθερίας | | | | |
| 1 | 6,314 | 12,706 | 31,821 | 63,657 |
| 2 | 2,290 | 4,303 | 6,965 | 9,925 |
| 3 | 2,353 | 3,182 | 4,541 | 5,841 |
| 4 | 2,132 | 2,776 | 3,747 | 4,604 |
| 5 | 2,015 | 2,571 | 3,365 | 4,032 |
| 6 | 1,943 | 2,447 | 3,143 | 3,707 |
| 7 | 1,895 | 2,365 | 2,998 | 3,499 |
| 8 | 1,860 | 2,306 | 2,896 | 3,355 |
| 9 | 1,833 | 2,262 | 2,821 | 3,350 |
| 10 | 1,812 | 2,228 | 2,764 | 3,169 |
| 11 | 1,796 | 2,201 | 2,718 | 3,106 |
| 12 | 1,782 | 2,179 | 2,681 | 3,055 |
| 13 | 1,771 | 2,160 | 2,650 | 3,012 |
| 14 | 1,761 | 2,145 | 2,624 | 2,977 |
| 15 | 1,753 | 2,131 | 2,602 | 2,947 |
| 16 | 1,746 | 2,120 | 2,583 | 2,921 |
| 17 | 1,740 | 2,110 | 2,567 | 2,898 |
| 18 | 1,734 | 2,101 | 2,552 | 2,878 |
| 19 | 1,729 | 2,093 | 2,539 | 2,861 |
| 20 | 1,725 | 2,086 | 2,528 | 2,845 |

Πίνακας με κρίσιμες τιμές της t-κατανομής.

# Κατανομή t (student): πρόβλημα (1 από 3)

**Να διερευνηθεί ο ισχυρισμός των μαθητών ότι χρειάζονται τουλάχιστον 30 λεπτά κάθε πρωί για να φθάσουν στο σχολείο τους, όταν σε ένα τυχαίο δείγμα 16 μαθητών ο μέσος χρόνος προσέλευσης βρέθηκε 25,6 λεπτά και η τυπική απόκλιση 5 λεπτά, αν είναι γνωστό ότι οι τιμές του χρόνου προσέλευσης κατανέμονται περίπου κανονικά (α=0,05).**

# Κατανομή t (student): πρόβλημα (2 από 3)

1. Υποθέσεις $H_0 : \mu \geq 30$

   $H_A : \mu < 30$

2. Θα χρησιμοποιηθεί το στατιστικό t επειδή η τυπική απόκλιση του πληθυσμού των χρόνων προσέλευσης είναι άγνωστη και εκτιμάται από την τυπική απόκλιση του δείγματος S=5. Από τον πίνακα της t για βαθμούς ελευθερίας N-1=15 στην πρώτη στήλη (ποσοστό 0,95): $t_{κρ}$ = 1,75 που στην περίπτωση αυτή γίνεται $t_{κρ}$ = -1,75 λόγω της κατεύθυνσης του ελέγχου. Κανόνας απόφασης: αν t < -1,75 απορρίπτεται η μηδενική υπόθεση.

# Κατανομή t (student): πρόβλημα (3 από 3)

3. Από τα δεδομένα προκύπτει η απόρριψη της μηδενικής υπόθεσης:

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{N}} = \frac{25{,}6 - 30}{5/\sqrt{16}} = \frac{-4{,}4}{1{,}25} = -3{,}52 < -1{,}75$$

4. Ο ισχυρισμός των μαθητών δεν ευσταθεί.

# 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.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

# Σημειώματα

# Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

# Σημείωμα Αναφοράς

Copyright Εθνικόν και Καποδιστριακόν Πανεπιστήμιον 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

# 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 (1/3)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

**Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες**

**Εικόνα 1, Σελίδα 11, 17-18:** Εικόνα γραφικής παράστασης τυπικής κανονικής κατανομής / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

**Εικόνα 2, Σελίδα 13:** Εικόνα γραφικής παράστασης με κανονική κατανομή / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

# Σημείωμα Χρήσης Έργων Τρίτων (2/3)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

**Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες**

**Εικόνα 3, Σελίδα 23:** Εικόνα με γραφικές παραστάσεις τυπικής κανονικής κατανομής και κατανομών t-student με 1 και 5 βαθμούς ελευθερίας / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

# Σημείωμα Χρήσης Έργων Τρίτων (3/3)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

**Πίνακες**

**Πίνακας 1, Σελίδα 24:** Πίνακας με κρίσιμες τιμές της t-κατανομής / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη