



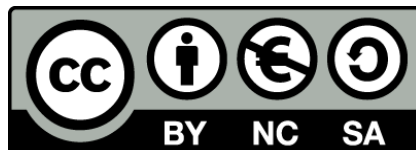
Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ Αθήνας



Βιοστατιστική (Θ)

Ενότητα 6: Έλεγχοι υποθέσεων - Διαστήματα εμπιστοσύνης (2)

Δρ.Ευσταθία Παπαγεωργίου, Καθηγήτρια
Τμήμα Ιατρικών Εργαστηρίων



Το περιεχόμενο του μαθήματος διατίθεται με άδεια Creative Commons εκτός και αν αναφέρεται διαφορετικά



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

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.

Έλεγχος υποθέσεων (συν.)

- Στην προηγούμενη ενότητα αναφερθήκαμε στον έλεγχο μέσω των τιμών 2 ανεξάρτητων πληθυσμών
 - Στην περίπτωση που οι τυπικές τους αποκλίσεις είναι ίσες
- Επίσης αναφερθήκαμε στην κατασκευή διαστήματος εμπιστοσύνης (Δ.Ε.) για τη διαφορά των 2 μέσων τιμών
- Η επόμενη διαφάνεια ανακεφαλαιώνει τους τύπους που συζητήσαμε

Ελεγχοι υποθέσεων και δ.ε. για διαφορά μέσων τιμών σε ανεξάρτητους πληθυσμούς σε μικρά δείγματα και με ισότητα διασπορών ($\sigma_1 = \sigma_2 = \sigma$):

| | | |
|--|--|---|
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$ | $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$ | $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ |
| $R = \{t > t_{n_1+n_2-2; a}\}$ | $R = \{t < -t_{n_1+n_2-2; a}\}$ | $R = \{ t > t_{n_1+n_2-2; \frac{a}{2}}\}$ |

Για τον υπολογισμό του Δ.Ε. έχουμε:

$$(\bar{x}_1 - \bar{x}_2 - s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n_1+n_2-2; \frac{a}{2}}, \bar{x}_1 - \bar{x}_2 + s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n_1+n_2-2; \frac{a}{2}}), \text{ όπου } s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Το στατιστικό κριτήριο t δίνεται από τον τύπο:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Παράδειγμα στο SPSS

- Στο επόμενο παράδειγμα, θέλουμε να ελέγξουμε αν ο μέσος ημερήσιος αριθμός τσιγάρων διαφέρει μεταξύ ανδρών και γυναικών.
- Ο πληθυσμός αναφοράς από τον οποίο προέρχονται τα δεδομένα είναι οι κάτοικοι της Περιφέρειας Αττικής το 2003.



17: address [ATHSIA

| | code | date | entry | lname | address | area | telephon | age | agegroup | agedecad |
|----|------|----------|-------|----------|----------|------|------------|-----|----------|----------|
| 1 | 2998 | 18.06.03 | 2003 | DELHMPAS | LIOSIA | 3 | 2470834 | 35 | 2 | 2 |
| 2 | 2253 | 02.07.0 | | | | | | 36 | 2 | 2 |
| 3 | 2800 | 04.10.0 | | | | | | 18 | 1 | 1 |
| 4 | 2796 | 04.10.0 | | | | | | 19 | 1 | 1 |
| 5 | 2726 | 04.10.0 | | | | | | 18 | 1 | 1 |
| 6 | 2260 | 02.07.0 | | | | | | 35 | 2 | 2 |
| 7 | 2736 | 29.11.0 | | | | | | 21 | 1 | 1 |
| 8 | 1861 | 04.04.0 | | | | | | 33 | 1 | 2 |
| 9 | 1909 | 16.04.0 | | | | | | 37 | 2 | 2 |
| 10 | 1629 | 05.02.0 | | | | | | 28 | 1 | 1 |
| 11 | 1887 | 04.04.0 | | | | | | 42 | 2 | 3 |
| 12 | 2346 | 04.07.0 | | | | | | 23 | 1 | 1 |
| 13 | 2252 | 02.07.0 | | | | | | 45 | 3 | 3 |
| 14 | 1850 | 03.04.0 | | | | | | 55 | 4 | 3 |
| 15 | 2758 | 29.11.02 | 2002 | LIADAKHS | ANTHOUP | 4 | 5732808 | 23 | 1 | 1 |
| 16 | 2708 | 26.11.02 | 2002 | LIADAKHS | ANTHOUP | 4 | 5732808 | 33 | 1 | 2 |
| 17 | 2755 | 29.11.02 | 2002 | KATAKALO | [ATHSIA | 1 | 2231707 | 19 | 1 | 1 |
| 18 | 2733 | 29.11.02 | 2002 | NTAOYTAK | AGELEYUR | 1 | 0942872524 | 19 | 1 | 1 |
| 19 | 1544 | 15.01.02 | 2002 | LALAOYNH | CYXIKO | 1 | 6205725 | 37 | 2 | 2 |
| 20 | 2002 | 21.05.02 | 2002 | ALEJANDR | PEIRAIAS | 2 | 4628258 | 57 | 4 | 3 |
| 21 | 2106 | 13.06.02 | 2002 | APOSTOLI | GYXIKO | 6 | 6712921 | 57 | 4 | 3 |
| 22 | 1918 | 18.04.02 | 2002 | SKARLAUI | PEIRAIAS | 2 | 4635377 | 35 | 2 | 2 |

Independent-Samples T Test

Test Variable(s):
of cigarettes / day [cig]

Grouping Variable:
sex(1 2)

Define Groups...

Options...

OK
Paste
Reset
Cancel
Help

badisma
barh
barh_2
omadika
Smoking habits [sm
cursmok
Type of cigarettes [c
Smoking years [yea
year_sto
reasstop
passwork

Data View Variable View

SPSS Processor is ready

Παράδειγμα, με 2 ανεξάρτητα δείγματα

Group Statistics

| Sex of Subjects | N | Mean | Std. Deviation | Std. Error Mean |
|----------------------------|-----|-------|----------------|-----------------|
| # of cigarettes / day Male | 941 | 26,05 | 16,690 | ,544 |
| Female | 655 | 18,54 | 12,510 | ,489 |

\bar{x}_1

S_1

\bar{x}_2

S_2

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|-----------------------|-----------------------------|---|------|------------------------------|----------|-----------------|-----------------|-----------------------|---|-------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| # of cigarettes / day | Equal variances assumed | 49,312 | ,000 | 9,757 | 1594 | ,000 | 7,50 | ,769 | 5,996 | 9,013 |
| | Equal variances not assumed | | | 10,261 | 1585,345 | ,000 | 7,50 | ,731 | 6,070 | 8,939 |

P-value < 0,001

(συν.)

- $H_0: \mu_1 = \mu_2$, όπου μ_1 ο μέσος ημερήσιος αριθμός τσιγάρων για τους άνδρες και μ_2 για τις γυναίκες, στον πληθυσμό αναφοράς
- Μας δίνεται ότι: $p\text{-value} < 0,001$
 - Οπότε **$p\text{-value} < 0,05$**
- Άρα, απορρίπτουμε την H_0 και συμπεραίνουμε ότι ο μέσος ημερήσιος αριθμός τσιγάρων ανδρών και γυναικών διαφέρει στατιστικά σημαντικά, στον πληθυσμό αναφοράς

(συν.)

- 95% ΔΕ: (5,996 , 9,013)
- Άρα, είμαστε 95% σίγουροι ότι η μέση ημερήσια διαφορά τσιγάρων που καπνίζουν τα 2 φύλα είναι μεταξύ (5,996 , 9,013), στον πληθυσμό αναφοράς

t-test για παρατηρήσεις κατά ζεύγη (paired t-test)

- Ο επόμενος έλεγχος που θα συζητήσουμε ονομάζεται t-test κατά ζεύγη
- Μερικές φορές οι παρατηρήσεις των 2 συγκρινόμενων ομάδων εμφανίζουν **ατομική αντιστοιχία** (δηλαδή, **δεν είναι ανεξάρτητες**).
 - Για κάθε παρατήρηση στην 1^η ομάδα υπάρχει μια αντίστοιχη παρατήρηση στη 2^η ομάδα
- Π.χ.
 - Μέτρηση της συστολικής αρτηριακής πίεσης στα ίδια άτομα, πριν και μετά από σωματική άσκηση.
 - Σύγκριση της αποτελεσματικότητας 2 φαρμάκων στους ίδιους ασθενείς.
- Τότε οι συγκρίσεις πρέπει να γίνονται **κατά ζεύγη**.

(συνέχεια)

- Αυτή η αντιστοιχία χρησιμοποιείται για να ελαττώσει την επιρροή από εξωτερικούς παράγοντες που αυξάνουν τη μεταβλητότητα των μετρήσεων
- Αν οι μετρήσεις γίνονται στα ίδια άτομα, τότε ένα σημαντικό μέρος της βιολογικής μεταβλητότητας που υπάρχει μεταξύ ανθρώπων εξαφανίζεται
- Αυτό έχει σαν αποτέλεσμα **πιο ακριβείς** συγκρίσεις

Έλεγχος υποθέσεων–Διάστημα Εμπιστοσύνης

Έλεγχοι υποθέσεων και δ.ε. για παρατηρήσεις κατά ζεύγη:

| | | |
|--|---|---|
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$ | $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$ | $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ |
| $R = \left\{ \frac{\bar{z}}{s_z} \sqrt{n} > t_{n-1; a} \right\}$ | $R = \left\{ \frac{\bar{z}}{s_z} \sqrt{n} < -t_{n-1; a} \right\}$ | $R = \left\{ \left \frac{\bar{z}}{s_z} \sqrt{n} \right > t_{n-1; \frac{a}{2}} \right\}$ |

Για τον υπολογισμό του Δ.Ε. της διαφοράς των μέσων τιμών, έχουμε:

$$\left(\bar{z} - \frac{s_z}{\sqrt{n}} t_{n-1; \frac{a}{2}}, \bar{z} + \frac{s_z}{\sqrt{n}} t_{n-1; \frac{a}{2}} \right), \text{ όπου } z = x_i - y_i.$$

Έλεγχος υποθέσεων–Διάστημα Εμπιστοσύνης

Για τον υπολογισμό του στατιστικού κριτηρίου:

- Υπολογίζουμε τη διαφορά z_i των παρατηρήσεων σε κάθε ζεύγος: $z_i = x_i - y_i$
- Υπολογίζουμε τη μέση τιμή των διαφορών: $\bar{z} = \frac{\sum_{i=1}^n z_i}{n}$
- Υπολογίζουμε την διακύμανση s_z^2 των διαφορών z_i

$$s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n z_i^2 - \frac{(\sum_{i=1}^n z_i)^2}{n} \right)$$

- **Τονίζουμε**, ότι στους παραπάνω τύπους το n αντιπροσωπεύει τον αριθμό των ζευγαριών, και όχι των συνολικών παρατηρήσεων

Παράδειγμα

- Πέντε άτομα που έπασχαν από μια νόσο θερμομετρήθηκαν το πρωί (ομάδα Α) και το βράδυ (ομάδα Β) της ίδιας μέρας. Οι τιμές της θερμοκρασίας (σε °C) ήταν:

| Άτομο | 1 ^ο | 2 ^ο | 3 ^ο | 4 ^ο | 5 ^ο |
|---------------|----------------|----------------|----------------|----------------|----------------|
| Πρωινή θερμ. | 37.1 | 37.4 | 37.2 | 37.3 | 37.0 |
| Βραδυνή θερμ. | 37.8 | 38.2 | 38.1 | 38.1 | 37.6 |

- Υπάρχει διαφορά μεταξύ **πρωινής** και **βραδινής** θερμοκρασίας;

(συν.)

- Στο παράδειγμά μας οι παρατηρήσεις εμφανίζουν ατομική αντιστοιχία (κάθε άτομο έχει 1 παρατήρηση και στις 2 ομάδες).
- Άρα, για τον έλεγχο της διαφοράς των μέσων τιμών της θερμοκρασίας μεταξύ των ομάδων A και B ενδείκνυται η εφαρμογή του **t-test κατά ζεύγη**.

(συν.)

- Η δοκιμασία **t-test κατά ζεύγη** λαμβάνει υπόψη τις διαφορές των παρατηρήσεων σε κάθε ζευγάρι.
 - Το t-test κατά ζεύγη χρησιμοποιεί τις διαφορές των παρατηρήσεων!
- Για τον έλεγχο αυτό, θα υπολογίσουμε τις **διαφορές (z)** των αντίστοιχων θερμομετρήσεων.

| Άτομο | 1 ^ο | 2 ^ο | 3 ^ο | 4 ^ο | 5 ^ο |
|-------------------|----------------|----------------|----------------|----------------|----------------|
| Διαφορά θερμ. (z) | 0.7 | 0.8 | 0.9 | 0.8 | 0.6 |

Μηδενική και εναλλακτική υπόθεση

- Η μηδενική και η εναλλακτική υπόθεση στο **t-test κατά ζεύγη** που θα δουλέψουμε είναι:
- $H_0: \mu_1 = \mu_2$ (μηδενική υπόθεση)
- $H_A: \mu_1 \neq \mu_2$ (εναλλακτική υπόθεση)

(συν.)

- Έχουμε: $\bar{z} = \frac{\sum_{i=1}^5 z_i}{n} = \frac{0,7+0,8+0,9+0,8+0,6}{5} = \frac{3,8}{5} = 0,76$

- $\sum_{i=1}^5 z_i^2 = 0,7^2 + 0,8^2 + 0,9^2 + 0,8^2 + 0,6^2 = 2,94$

- Η διασπορά είναι:

$$s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n z_i^2 - \frac{(\sum_{i=1}^k z_i)^2}{n} \right) = \frac{1}{5-1} \left(2,94 - \frac{3,8^2}{5} \right) = 0,013$$

- Άρα, η τυπική απόκλιση είναι: $s_z = \sqrt{0,013} = 0,114$

(συν.)

- Το στατιστικό κριτήριο είναι: $\left| \frac{\bar{z}}{s_z} \sqrt{n} \right| = \left| \frac{0,76}{0,114} \sqrt{5} \right| = 14,91$
- Η κρίσιμη τιμή είναι η: $t_{n-1; \frac{\alpha}{2}} = t_{4; 0,025}$
- Την τιμή αυτή θα την βρούμε από τον Πίνακα της κατανομής t, που δίνεται στην επόμενη διαφάνεια.
- Βλέπουμε ότι $t_{4; 0,025} = 2,776$

| DF | A = 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|----------|------------|-------|--------|--------|--------|---------|---------|
| ∞ | ta = 1.282 | 1.645 | 1.96 | 2.326 | 2.576 | 3.091 | 3.291 |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.656 | 318.289 | 636.578 |
| 2 | 1.886 | 2.92 | 4.303 | 6.965 | 9.925 | 22.328 | 31.6 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.214 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.61 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.894 | 6.869 |
| 6 | 1.44 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.86 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.25 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.93 | 4.318 |
| 13 | 1.35 | 1.771 | 2.16 | 2.65 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.14 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.12 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.74 | 2.11 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.33 | 1.734 | 2.101 | 2.552 | 2.878 | 3.61 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.85 |
| 21 | 1.323 | 1.721 | 2.08 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.319 | 1.714 | 2.069 | 2.5 | 2.807 | 3.485 | 3.768 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.316 | 1.708 | 2.06 | 2.485 | 2.787 | 3.45 | 3.725 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.689 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.66 |
| 30 | 1.31 | 1.697 | 2.042 | 2.457 | 2.75 | 3.385 | 3.646 |
| 60 | 1.296 | 1.671 | 2 | 2.39 | 2.66 | 3.232 | 3.46 |
| 120 | 1.289 | 1.658 | 1.98 | 2.358 | 2.617 | 3.16 | 3.373 |
| 1000 | 1.282 | 1.646 | 1.962 | 2.33 | 2.581 | 3.098 | 3.3 |

Υπολογισμοί (συν.)

- Η κρίσιμη περιοχή είναι η:
- $R = \left\{ \left| \frac{\bar{z}}{s_z} \sqrt{n} \right| > t_{n-1; \frac{\alpha}{2}} \right\}$.
- Προφανώς $14,91 = \left| \frac{\bar{z}}{s_z} \sqrt{n} \right| > t_{n-1; \frac{\alpha}{2}} = 2,776$
- Άρα, βρισκόμαστε στην κρίσιμη περιοχή R.
- Οπότε, απορρίπτουμε την H_0 και αποδεχόμαστε ότι ισχύει η H_A .
- Άρα συμπεραίνουμε ότι η μέση πρωινή θερμοκρασία διαφέρει στατιστικά σημαντικά από τη μέση βραδινή θερμοκρασία, στον πληθυσμό αναφοράς.

Υπολογισμοί για το 95% Δ.Ε.

- Για το 95% Δ.Ε. γνωρίζουμε ότι δίνεται από την σχέση:

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

- Αντικαθιστώντας στην παραπάνω σχέση παίρνουμε:

$$\left(0,76 - \frac{0,114}{\sqrt{5}} * 2,776, 0,76 + \frac{0,114}{\sqrt{5}} * 2,776 \right)$$

- Κάνοντας τις πράξεις έχουμε ότι το 95% Δ.Ε. για τη διαφορά των 2 μέσων τιμών στον πληθυσμό αναφοράς είναι το: (0,618 , 0,902)
- Άρα, είμαστε 95% σίγουροι ότι στον πληθυσμό αναφοράς η διαφορά της μέσης πρωινής από τη μέση βραδινή θερμοκρασία βρίσκεται στο (0,618 , 0,902)
 - Παρατηρήστε ότι το 95% Δ.Ε. δεν περιέχει το 0.

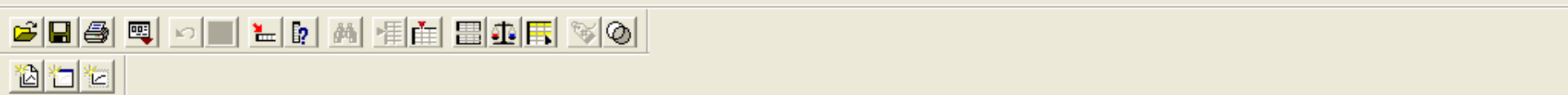
Παράδειγμα

- Στη συνέχεια θα ελέγξουμε αν ο δείκτης μάζας σώματος (ΔΜΣ)) στον πληθυσμό αναφοράς μεταβλήθηκε από το 2001 στο 2006.
- Για το λόγο αυτό επιλέξαμε ένα δείγμα 1615 ατόμων από τον πληθυσμό αυτό, και μετρήσαμε το ΔΜΣ το 2001 και το 2006.
- Το δείγμα θα μας βοηθήσει να βγάλουμε συμπεράσματα για τον πληθυσμό αναφοράς.



- Reports
- Descriptive Statistics
- Tables
- Compare Means**
 - Means...
 - One-Sample T Test...
 - Independent-Samples T Test...
 - Paired-Samples T Test...**
 - One-Way ANOVA...
- General Linear Model
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Data Reduction
- Scale
- Nonparametric Tests
- Time Series
- Survival
- Multiple Response
- Missing Value Analysis...
- Complex Samples

| code | interven | sameschl | reexam98 | present9 | sex_2004 | sex | var0003 |
|------|----------|----------|-------------|-----------|----------|-----|-------------|
| 1 | 20001 | 1,0 | . | . | . | . | ΤΣΑΠΑΛΟΥΜΗΣ |
| 2 | 10024 | 1,0 | . | . | . | . | |
| 3 | 20019 | 1,0 | . | . | . | . | SPIRIDAKIS |
| 4 | 20024 | 1,0 | . | . | . | . | |
| 5 | 20035 | 1,0 | . | . | . | . | MARHS |
| 6 | 20038 | 1,0 | . | . | . | . | STIBAKTAKHS |
| 7 | 20040 | 1,0 | . | . | . | . | KOKOLAKHS |
| 8 | 20044 | 1,0 | . | . | . | . | KARYDHS |
| 9 | 30002 | 1,0 | . | . | . | . | GEORGOPOULO |
| 10 | 30004 | 1,0 | . | . | . | . | |
| 11 | 30013 | 1,0 | . | . | . | . | MOULAKAKIS |
| 12 | 30015 | 1,0 | . | . | . | . | |
| 13 | 30021 | 1,0 | . | . | . | . | |
| 14 | 30024 | 1,0 | 12-NOV-1992 | 06-OCT-86 | 3,00 | 3 | STEFANAKIS |
| 15 | 40010 | 1,0 | . | . | 4,00 | 4 | |
| 16 | 40011 | 1,0 | . | . | 4,00 | 4 | |
| 17 | 40013 | 1,0 | . | . | 4,00 | 4 | |
| 18 | 40044 | 1,0 | 18-NOV-1992 | 14-JUN-86 | 4,00 | 4 | XIMERIS |
| 19 | 50002 | 1,0 | 13-NOV-1992 | 27-OCT-86 | 5,00 | 5 | SKOULAS |
| 20 | 50003 | 1,0 | . | . | 5,00 | 5 | |
| 21 | 50005 | 1,0 | 13-NOV-1992 | 20-FEB-87 | 5,00 | 5 | GIANAKAKIS |
| 22 | 50006 | 1,0 | 13-NOV-1992 | 25-JAN-87 | . | . | ORFANOS |
| 23 | 50007 | 1,0 | 13-NOV-1992 | 16-OCT-86 | 5,00 | 5 | KANAKIS |
| 24 | 50010 | 1,0 | . | . | 5,00 | 5 | |
| 25 | 50012 | 1,0 | . | . | 5,00 | 5 | |
| 26 | 50017 | 1,0 | . | . | 5,00 | 5 | |
| 27 | 60015 | 1,0 | . | . | 6,00 | 6 | |
| 28 | 70039 | 1,0 | 19-NOV-1992 | 12-OCT-86 | 7,00 | 7 | ZERBAKIS |
| 29 | 70042 | 1,0 | . | . | 7,00 | 7 | |
| 30 | 100009 | 1,0 | . | . | 10,00 | 10 | |
| 31 | 100018 | 1,0 | 26-NOV-1992 | 15-JUL-86 | 10,00 | 10 | MPOYCIKAS |



| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|----|------------|---------|-------|----------|-----------------------------|----------------|---------|---------|-------|---------|
| 1 | code | Numeric | 8 | 0 | Code number | None | None | 8 | Right | Scale |
| 2 | date | Date | 10 | 0 | Date of entry | None | None | 16 | Right | Scale |
| 3 | entry | Numeric | 8 | 0 | | None | None | 8 | Right | Scale |
| 4 | lname | String | 8 | 0 | | None | None | 18 | Left | Nominal |
| 5 | address | String | 8 | 0 | | None | None | 12 | Left | Nominal |
| 6 | area | Numeric | 4 | 0 | Region of Subi (1 - Athens) | None | None | 16 | Right | Ordinal |
| 7 | telephon | String | 12 | | | | | | | Nominal |
| 8 | age | Numeric | 4 | | | | | | | Scale |
| 9 | agegroup | Numeric | 8 | | | | | | | Ordinal |
| 10 | agedecad | Numeric | 8 | | | | | | | Scale |
| 11 | age_group3 | Numeric | 8 | | | | | | | Scale |
| 12 | pc_age | Numeric | 3 | | | | | | | Ordinal |
| 13 | elderly | Numeric | 8 | | | | | | | Scale |
| 14 | sex | Numeric | 4 | | | | | | | Ordinal |
| 15 | ep | Numeric | 2 | | | | | | | Ordinal |
| 16 | years_ep | Numeric | 8 | | | | | | | Scale |
| 17 | weight | Numeric | 8 | | | | | | | Scale |
| 18 | height | Numeric | 8 | | | | | | | Scale |
| 19 | bmi | Numeric | 8 | | | | | | | Scale |
| 20 | bmi_grou | Numeric | 8 | 2 | Bmi group | {,00, Underwei | None | 12 | Right | Scale |
| 21 | bmi_25 | Numeric | 8 | 2 | | None | None | 10 | Right | Scale |
| 22 | obese | Numeric | 8 | 2 | Obesity | None | None | 8 | Right | Scale |
| 23 | | | | | | | | | | |
| 24 | | | | | | | | | | |
| 25 | | | | | | | | | | |
| 26 | | | | | | | | | | |
| 27 | height_est | Numeric | 8 | 2 | Self reported h | None | None | 8 | Right | Scale |
| 28 | bmi_est | Numeric | 8 | 2 | Self reported B | None | None | 8 | Right | Scale |
| 29 | bmi_grou_e | Numeric | 8 | 2 | Self reported B | {1,00, <25 kg/ | None | 11 | Right | Scale |
| 30 | d_bmi | Numeric | 8 | 2 | Dif between ob | None | None | 8 | Right | Scale |

Paired-Samples T Test

Paired Variables:
bmi -- bmi06

Current Selections:
Variable 1:
Variable 2:

OK, Paste, Reset, Cancel, Help, Options...

Σύγκριση του ΔΜΣ των ίδιων ατόμων σε 2 διαφορετικές χρονικές περιόδους (2001 και 2006).

Παράδειγμα

Διαφορά μέσων τιμών δύο εξαρτημένων δειγμάτων.

Paired Samples Statistics

| | | Mean | N | Std. Deviation | Std. Error Mean |
|--------|-------------------------|---------|------|----------------|-----------------|
| Pair 1 | Body Mass Index (kg/m2) | 25,7829 | 1615 | 3,69540 | ,09195 |
| | bmi06 | 26,1831 | 1615 | 3,59815 | ,08954 |

Paired Samples Test

| | | Paired Differences | | | | | | | |
|--------|---------------------------------|--------------------|----------------|-----------------|---|---------|--------|------|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | Sig. (2-tailed) |
| | | | | | Lower | Upper | | | |
| Pair 1 | Body Mass Index (kg/m2) - bmi06 | -,40015 | 4,21030 | ,10477 | -,60565 | -,19466 | -3,819 | 1614 | ,000 |

95% ΔΕ για τη διαφορά των μέσων τιμών ΔΜΣ

p-value

Παράδειγμα

- Ο επόμενος Πίνακας παρουσιάζει τα αποτελέσματα 20 φοιτητών σε 2 προόδους που έδωσαν στο μάθημα της Βιοστατιστικής.
- Μας ενδιαφέρει να ελέγξουμε αν ο μέσος βαθμός της 1^{ης} προόδου διαφέρει από τον μέσο βαθμό της 2^{ης} προόδου.

| Φοιτητής | Αποτελέσματα 1ης προόδου (x) | Αποτελέσματα 2ης προόδου (y) |
|-----------------|---|---|
| 1 | 18 | 22 |
| 2 | 21 | 25 |
| 3 | 16 | 17 |
| 4 | 22 | 24 |
| 5 | 19 | 16 |
| 6 | 24 | 29 |
| 7 | 17 | 20 |
| 8 | 21 | 23 |
| 9 | 23 | 19 |
| 10 | 18 | 20 |
| 11 | 14 | 15 |
| 12 | 16 | 15 |
| 13 | 16 | 18 |
| 14 | 19 | 26 |
| 15 | 18 | 18 |
| 16 | 20 | 24 |
| 17 | 12 | 18 |
| 18 | 22 | 25 |
| 19 | 15 | 19 |
| 20 | 17 | 16 |

(συν.)

- Επειδή οι παρατηρήσεις εμφανίζουν κατά ζεύγη αντιστοιχία (σε κάθε φοιτητή αντιστοιχεί ένας βαθμός στην 1^η και ένας βαθμός στη 2^η πρόοδο), θα χρησιμοποιήσουμε τον έλεγχο t-test κατά ζεύγη.
- Ο έλεγχος αυτός χρειάζεται τις διαφορές των 2 μεταβλητών, οπότε προσθέτουμε μια ακόμη στήλη στον πίνακα με τις διαφορές αυτές.

| Φοιτητής | Αποτελέσματα 1ης προόδου (x) | Αποτελέσματα 2ης προόδου (y) | Διαφορά (z=x-y) |
|----------|------------------------------------|------------------------------------|--------------------|
| 1 | 18 | 22 | -4 |
| 2 | 21 | 25 | -4 |
| 3 | 16 | 17 | -1 |
| 4 | 22 | 24 | -2 |
| 5 | 19 | 16 | 3 |
| 6 | 24 | 29 | -5 |
| 7 | 17 | 20 | -3 |
| 8 | 21 | 23 | -2 |
| 9 | 23 | 19 | 4 |
| 10 | 18 | 20 | -2 |
| 11 | 14 | 15 | -1 |
| 12 | 16 | 15 | 1 |
| 13 | 16 | 18 | -2 |
| 14 | 19 | 26 | -7 |
| 15 | 18 | 18 | 0 |
| 16 | 20 | 24 | -4 |
| 17 | 12 | 18 | -6 |
| 18 | 22 | 25 | -3 |
| 19 | 15 | 19 | -4 |
| 20 | 17 | 16 | 1 |

Μηδενική και εναλλακτική υπόθεση

- Η μηδενική και η εναλλακτική υπόθεση στο **t-test κατά ζεύγη** που θα δουλέψουμε είναι:
- $H_0: \mu_1 = \mu_2$ (μηδενική υπόθεση)
- $H_A: \mu_1 \neq \mu_2$ (εναλλακτική υπόθεση)
- Όπου μ_1 και μ_2 ο μέσος όρος της βαθμολογίας στην 1^η και την 2^η πρόοδο, αντίστοιχα.

(συν.)

- Έχουμε: $\bar{z} = \frac{\sum_{i=1}^{20} z_i}{n} = \frac{-4-4-1+\dots+1}{20} = \frac{-41}{20} = -2,05$

- $\sum_{i=1}^{20} z_i^2 = (-4)^2 + (-4)^2 + (-1)^2 + \dots + 1^2 = 237$

- Η διασπορά είναι:

$$s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n z_i^2 - \frac{(\sum_{i=1}^k z_i)^2}{n} \right) = \frac{1}{20-1} \left(237 - \frac{(-41)^2}{20} \right) = 8,05$$

- Άρα, η τυπική απόκλιση είναι: $s_z = \sqrt{8,05} = 2,84$

(συν.)

- Το στατιστικό κριτήριο είναι: $\left| \frac{\bar{z}}{s_z} \sqrt{n} \right| = \left| \frac{-2,05}{2,84} \sqrt{20} \right| = 3,23$
- Η κρίσιμη τιμή είναι η: $t_{n-1; \frac{\alpha}{2}} = t_{19; 0,025}$
- Την τιμή αυτή θα την βρούμε από τον Πίνακα της κατανομής t, που δίνεται στην επόμενη διαφάνεια.
- Βλέπουμε ότι $t_{4; 0,025} = 2,093$

| DF | A = 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|----------|------------|-------|--------|--------|--------|---------|---------|
| ∞ | ta = 1.282 | 1.645 | 1.96 | 2.326 | 2.576 | 3.091 | 3.291 |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.656 | 318.289 | 636.578 |
| 2 | 1.886 | 2.92 | 4.303 | 6.965 | 9.925 | 22.328 | 31.6 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.214 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.61 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.894 | 6.869 |
| 6 | 1.44 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.86 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.25 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.93 | 4.318 |
| 13 | 1.35 | 1.771 | 2.16 | 2.65 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.14 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.12 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.74 | 2.11 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.33 | 1.734 | 2.101 | 2.552 | 2.878 | 3.61 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.85 |
| 21 | 1.323 | 1.721 | 2.08 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.319 | 1.714 | 2.069 | 2.5 | 2.807 | 3.485 | 3.768 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.316 | 1.708 | 2.06 | 2.485 | 2.787 | 3.45 | 3.725 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.689 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.66 |
| 30 | 1.31 | 1.697 | 2.042 | 2.457 | 2.75 | 3.385 | 3.646 |
| 60 | 1.296 | 1.671 | 2 | 2.39 | 2.66 | 3.232 | 3.46 |
| 120 | 1.289 | 1.658 | 1.98 | 2.358 | 2.617 | 3.16 | 3.373 |
| 1000 | 1.282 | 1.646 | 1.962 | 2.33 | 2.581 | 3.098 | 3.3 |

Υπολογισμοί (συν.)

- Η κρίσιμη περιοχή είναι η:
- $R = \left\{ \left| \frac{\bar{z}}{s_z} \sqrt{n} \right| > t_{n-1; \frac{\alpha}{2}} \right\}$.
- Προφανώς $3,23 = \left| \frac{\bar{z}}{s_z} \sqrt{n} \right| > t_{n-1; \frac{\alpha}{2}} = 2,093$
- Άρα, βρισκόμαστε στην κρίσιμη περιοχή R.
- Οπότε, απορρίπτουμε την H_0 και αποδεχόμαστε ότι ισχύει η H_A .
- Άρα συμπεραίνουμε ότι ο μέσος όρος της βαθμολογίας της 1^{ης} προόδου διαφέρει στατιστικά σημαντικά από το μέσο όρο της βαθμολογίας της 2^{ης} προόδου, στον πληθυσμό αναφοράς.

Υπολογισμοί για το 95% Δ.Ε.

- Για το 95% Δ.Ε. γνωρίζουμε ότι δίνεται από την σχέση:

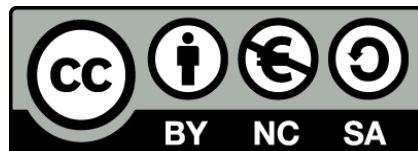
$$\left(\bar{z} - \frac{s_z}{\sqrt{n}} t_{n-1; \frac{\alpha}{2}}, \bar{z} + \frac{s_z}{\sqrt{n}} t_{n-1; \frac{\alpha}{2}} \right)$$

- Αντικαθιστώντας στην παραπάνω σχέση παίρνουμε:

$$\left(3,23 - \frac{2,84}{\sqrt{20}} * 2,093, 3,23 + \frac{2,84}{\sqrt{20}} * 2,093 \right)$$

- Κάνοντας τις πράξεις έχουμε ότι το 95% Δ.Ε. για τη διαφορά των 2 μέσων τιμών στον πληθυσμό αναφοράς είναι το: (1,90 , 4,56)
- Άρα, είμαστε 95% σίγουροι ότι στον πληθυσμό αναφοράς η διαφορά του μέσου όρου της βαθμολογίας της 1ης προόδου από το μέσο όρο της βαθμολογίας της 2ης προόδου βρίσκεται στο διάστημα (1,90 , 4,56).
 - Παρατηρείστε ότι το 95% Δ.Ε. δεν περιέχει το 0.

Τέλος Ενότητας



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα Αναφοράς

Copyright Τεχνολογικό Εκπαιδευτικό Ίδρυμα Αθήνας, Ευσταθία Παπαγεωργίου 2014. Ευσταθία Παπαγεωργίου. «Βιοστατιστική (Θ). Ενότητα 6: Έλεγχοι υποθέσεων - Διαστήματα εμπιστοσύνης». Έκδοση: 1.0. Αθήνα 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: ocp.teiath.gr.

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. Οι όροι χρήσης των έργων τρίτων επεξηγούνται στη διαφάνεια «Επεξήγηση όρων χρήσης έργων τρίτων».

Τα έργα για τα οποία έχει ζητηθεί και δοθεί άδεια αναφέρονται 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.

Επεξήγηση όρων χρήσης έργων τρίτων

© Δεν επιτρέπεται η επαναχρησιμοποίηση του έργου, παρά μόνο εάν ζητηθεί εκ νέου άδεια από το δημιουργό.

διαθέσιμο με άδεια CC-BY
Επιτρέπεται η επαναχρησιμοποίηση του έργου και η δημιουργία παραγώγων αυτού με απλή αναφορά του δημιουργού.

διαθέσιμο με άδεια CC-BY-SA
Επιτρέπεται η επαναχρησιμοποίηση του έργου με αναφορά του δημιουργού, και διάθεση του έργου ή του παράγωγου αυτού με την ίδια άδεια.

διαθέσιμο με άδεια CC-BY-ND
Επιτρέπεται η επαναχρησιμοποίηση του έργου με αναφορά του δημιουργού.
Δεν επιτρέπεται η δημιουργία παραγώγων του έργου.

διαθέσιμο με άδεια CC-BY-NC
Επιτρέπεται η επαναχρησιμοποίηση του έργου με αναφορά του δημιουργού.
Δεν επιτρέπεται η εμπορική χρήση του έργου.

διαθέσιμο με άδεια CC-BY-NC-SA
Επιτρέπεται η επαναχρησιμοποίηση του έργου με αναφορά του δημιουργού, και διάθεση του έργου ή του παράγωγου αυτού με την ίδια άδεια.
Δεν επιτρέπεται η εμπορική χρήση του έργου.

διαθέσιμο με άδεια CC-BY-NC-ND
Επιτρέπεται η επαναχρησιμοποίηση του έργου με αναφορά του δημιουργού.
Δεν επιτρέπεται η εμπορική χρήση του έργου και η δημιουργία παραγώγων του.

διαθέσιμο με άδεια CCO Public Domain
Επιτρέπεται η επαναχρησιμοποίηση του έργου, η δημιουργία παραγώγων αυτού και η εμπορική του χρήση, χωρίς αναφορά του δημιουργού.

διαθέσιμο ως κοινό κτήμα
Επιτρέπεται η επαναχρησιμοποίηση του έργου, η δημιουργία παραγώγων αυτού και η εμπορική του χρήση, χωρίς αναφορά του δημιουργού.

χωρίς σήμανση
Συνήθως δεν επιτρέπεται η επαναχρησιμοποίηση του έργου.

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.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ 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.

