

Μεθοδολογία των επιστημών του Ανθρώπου : Στατιστική

Εργαστήριο 6 :

1. Να χρησιμοποιηθεί το αρχείο gssft.sav για να γίνει έλεγχος της υπόθεσης ότι στους εργαζόμενους με πλήρη απασχόληση η τιμή του μέσου χρόνου εργασίας εβδομαδιαία είναι 40 ώρες(μεταβλητή hrs1).

ο Ποιες προϋποθέσεις απαιτούνται για να χρησιμοποιηθεί το t-test ενός δείγματος; Ικανοποιούνται στην περίπτωση μας;

Οι προϋποθέσεις είναι:

- i. Η μεταβλητή της οποίας η μέση τιμή ελέγχεται πρέπει να είναι ποσοτική. Η μεταβλητή hrs1 είναι ποσοτική, αφού η μέτρησή έχει μονάδα την ώρα.
- ii. Το δείγμα πρέπει να ικανοποιεί τον ορισμό του τυχαίου δείγματος (αυτό αναφέρεται στον τρόπο με τον οποίο έγινε η δειγματοληψία). Παρότι δεν γνωρίζουμε τον τρόπο με τον οποίο λήφθηκε το δείγμα, θεωρούμε ότι αυτοί που έκαναν την έρευνα ακολούθησαν διαδικασία που ικανοποιεί τις αρχές της τυχαίας δειγματοληψίας.
- iii. Αν το δείγμα είναι μικρό ($N < 30$), πρέπει να ελεγχθεί η κανονικότητα της μεταβλητής, π.χ. με τη βοήθεια ενός ιστογράμματος. Στην περίπτωση μας το δείγμα είναι μεγάλο ($N > 30$), άρα δεν είναι απαραίτητος ο έλεγχος της κανονικότητας της μεταβλητής hrs1.
- iv. Η τυπική απόκλιση (σ) του πληθυσμού πρέπει να είναι άγνωστη (διαφορετικά χρησιμοποιούμε το z test). Στην περίπτωση μας η τυπική απόκλιση των ωρών εργασίας του πληθυσμού είναι άγνωστη.

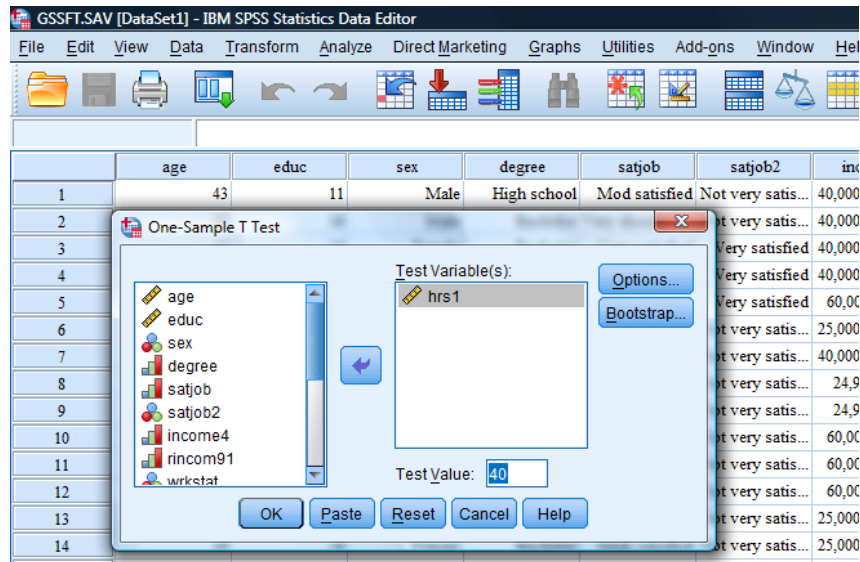
ο Ποια είναι η μηδενική υπόθεση που θέλετε να ελέγξετε; Ποια είναι η εναλλακτική υπόθεση;

Η μηδενική υπόθεση είναι: $H_0: \mu = 40$

Η εναλλακτική υπόθεση είναι: $H_A: \mu \neq 40$

- ο Να πραγματοποιηθεί ο έλεγχος της υπόθεσης και να γραφούν τα συμπεράσματα.

Από το μενού Analyze => Compare Means => One Sample T Test εισάγουμε στο πλαίσιο Test Variable (s) τη μεταβλητή hrs1 και στο πλαίσιο Test_Value την τιμή της μηδενικής υπόθεσης 40.



Μετά το «κλικ» στο OK του παραθύρου παίρνουμε στο output τους εξής πίνακες:

Πίνακας 1:

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
hrs1 Number of Hours Worked Last Week	741	46,29	11,269	,414

Πίνακας 2:

One-Sample Test						
	Test Value = 40					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
hrs1 Number of Hours Worked Last Week	15,191	740	,000	6,289	5,48	7,10

Στον πίνακα 1 εμφανίζονται τα στοιχεία του δείγματος: το μέγεθος του δείγματος είναι 741 συμμετέχοντες ($N = 741$), η μέση τιμή του δείγματος (Mean) είναι 46,29, η τυπική απόκλιση του δείγματος (Std. Deviation) είναι $S = 11,27$ και το τυπικό σφάλμα του μέσου (Std. Error Mean) (S/\sqrt{N}) είναι 0,41.

Στον πίνακα 2 εμφανίζονται τα αποτελέσματα του ελέγχου T ενός δείγματος: $t = 15,19$, βαθμοί ελευθερίας ($df = 740$ ($N-1$)), p value (Sig. (2-tailed)) = 0,000, και η διαφορά του δείγματος από την μηδενική υπόθεση (Mean Difference) = 6,29.

Προκειμένου να αποφασίσουμε αν θα απορριφθεί η H_0 ή όχι σε αμφίπλευρο έλεγχο, συγκρίνουμε το p value με το επίπεδο σημαντικότητας α (συνήθως 0,05). Εάν $p \leq \alpha$ απορρίπτεται η H_0 , διαφορετικά ($p > \alpha$) η H_0 δεν απορρίπτεται. Στην περίπτωσή μας $p = 0,000$, δηλαδή $p < 0,0005$ και κατά μείζονα λόγο $p < 0,05$. Εάν ο έλεγχος είναι μονόπλευρος τότε για να απορρίψουμε τη μηδενική υπόθεση πρέπει $p/2 \leq \alpha$. Εάν $p/2 > \alpha$ τότε η μηδενική υπόθεση δεν απορρίπτεται.

Όταν απορριφθεί η μηδενική υπόθεση σε αμφίπλευρο έλεγχο, για να αποφασίσουμε την κατεύθυνση της απόρριψης, δηλαδή εάν $\mu > 40$ ή $\mu < 40$, συγκρίνουμε τη μέση τιμή του δείγματός μας με την τιμή της μηδενικής υπόθεσης, δηλ. εδώ το 40. Πράγματι επειδή η μέση τιμή του δείγματός μας είναι 46,29, δηλ. μεγαλύτερη του 40 αποφασίζουμε ότι η μέση τιμή των ωρών εργασίας του πληθυσμού είναι μεγαλύτερη του 40.

Το παραπάνω αποτέλεσμα μπορεί να γραφτεί στα πλαίσια μιας εργασίας ως εξής: Ελέγξαμε την υπόθεση ότι ο μέσος Αμερικανός εργαζόμενος εργάζεται 40 ώρες την εβδομάδα, με τη βοήθεια του ελέγχου T ενός δείγματος. Από το αποτέλεσμα του ελέγχου προκύπτει ότι ο μέσος Αμερικανός εργαζόμενος σημαντικά περισσότερο από 40 ώρες εβδομαδιαίως [$t(740) = 15,19$, $p < 0,0005$].

○ Να δοθούν η τυπική απόκλιση του δείγματος και το τυπικό σφάλμα του μέσου. Να εξηγήσετε τη διαφορά τους.

Η τυπική απόκλιση του δείγματος και το τυπικό σφάλμα του μέσου είναι 11,27 και 0,414 αντίστοιχα (πίνακας 1). Η διαφορά τους είναι ότι η τυπική απόκλιση της μέσης τιμής είναι η τυπική απόκλιση της δειγματοληπτικής κατανομής των μέσων τιμών δειγμάτων μεγέθους $N = 741$.

- ο Ποιο είναι το 95% διάστημα εμπιστοσύνης για το μέσο χρόνο εβδομαδιαίας απασχόλησης σε εργαζόμενους με πλήρη απασχόληση.

Το 95% διάστημα εμπιστοσύνης μπορεί να υπολογιστεί από το 95% διάστημα εμπιστοσύνης για την διαφορά (5,48 – 7,10), εάν προσθέσουμε σε κάθε ένα από τα όριά του την τιμή 40 (45,48 – 74,10). Αυτό σημαίνει ότι υπάρχει 95% βεβαιότητα ότι οι ώρες εργασίας του μέσου Αμερικανού βρίσκονται μέσα σε αυτό το διάστημα.

- ο Βασιζόμενοι στο 95% διάστημα εμπιστοσύνης μπορείτε να απορρίψετε την μηδενική υπόθεση ότι ο πληθυσμιακός μέσος είναι 43 ώρες την εβδομάδα

Ένας εναλλακτικός τρόπος ελέγχου υποθέσεων ο οποίος στηρίζεται στη χρήση του διαστήματος εμπιστοσύνης για τη μέση τιμή του πληθυσμού που κατασκευάσαμε στο προηγούμενο ερώτημα, είναι ο εξής:

Εάν η τιμή της μηδενικής υπόθεσης μ_0 δεν ανήκει στο 95% διάστημα εμπιστοσύνης για τη μέση τιμή του πληθυσμού, απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο $\alpha = 0,05$, και διακρίνουμε δύο περιπτώσεις: 1). Εάν η μ_0 είναι μικρότερη από το κατώτερο όριο του διαστήματος συμπεραίνουμε ότι η μέση τιμή του πληθυσμού είναι σημαντικά μεγαλύτερη από τη μ_0 . 2). Εάν η μ_0 είναι μεγαλύτερη από το ανώτερο όριο του διαστήματος συμπεραίνουμε ότι η μέση τιμή του πληθυσμού είναι σημαντικά μικρότερη από τη μ_0 .

Στην περίπτωσή μας η $\mu_0 = 40$, δεν ανήκει στο 95% διάστημα εμπιστοσύνης (45,48 – 47,10) και είναι μικρότερη από το κατώτερο όριο (45,48) και επομένως ο μέσος Αμερικανός εργάζεται σημαντικά περισσότερο από 40 ώρες την εβδομάδα.

Να χρησιμοποιηθεί το αρχείο **schools.sav** για να απαντηθούν τα παρακάτω:

2. Ο υπεύθυνος των σχολείων μιας περιοχής ισχυρίζεται ότι έγιναν δραματικές αλλαγές από το 1993 έως το 1994. Εξετάσετε τις μεταβλητές που δείχνουν μεταβολή στο ποσοστό των μαθητών του σχολείου που ξεπερνούν τα standard της πολιτείας (**mathch94**, **readch94** και **scich94**). Να ελεγχθεί η υπόθεση ότι η πραγματική αλλαγή στο ποσοστό των

μαθητών που ξεπερνούν τα standards είναι 0. Θετική τιμή για κάθε μια από τις παραπάνω μεταβλητές δηλώνει βελτίωση των επιδόσεων του σχολείου στα μαθηματικά, γλώσσα και φυσικές επιστήμες (mathch94, readch94 και scich94) από το 1993 στο 1994 ενώ αντιθέτως αρνητική τιμή για κάθε μια από τις παραπάνω μεταβλητές δηλώνει επιδείνωση των επιδόσεων του σχολείου. Να γραφεί μια μικρή αναφορά με τα ευρήματά σας.

3. Να επιλεγεί ένα δείγμα $N = 25$ μαθητών από τα άτομα του αρχείου «χρόνος μετάβασης». Στη συνέχεια, να ελεγχθεί ο ισχυρισμός των μαθητών χρειάζονται τουλάχιστον 30 λεπτά κάθε πρωί για να φθάσουν στο σχολείο τους. Η κατανομή του χρόνου θα θεωρηθεί κατά προσέγγιση κανονική. Το επίπεδο σημαντικότητας είναι $\alpha=0,05$.

(Οι ερωτήσεις 2 και 3 δίνονται ως δραστηριότητα στις φοιτήτριες/τες).

Η ύλη του ελέγχου T ενός δείγματος βρίσκεται στο αρχείο διαφανειών «Επαγωγική Στατιστική_6.pptx» στο φάκελο «Παρουσιάσεις».

Στο video με τίτλο «one sample t-test.swf» θα βρείτε την εκτέλεση του ελέγχου T ενός δείγματος στο SPSS.