



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Στατιστική στην Εκπαίδευση II

Συμπληρωματικές σημειώσεις για το p-value

Διδάσκων: Μιχάλης Λιναρδάκης

ΠΑΙΔΑΓΩΓΙΚΟ ΤΜΗΜΑ ΔΗΜΟΤΙΚΗΣ ΕΚΠΑΙΔΕΥΣΗΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΑΓΩΓΗΣ



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται στην άδεια χρήσης **Creative Commons** και ειδικότερα ***Αναφορά – Μη εμπορική Χρήση – Όχι Παράγωγο Έργο 3.0 Ελλάδα*** (***Attribution – Non Commercial – Non-derivatives 3.0 Greece***)



[ή επιλογή ενός άλλου από τους έξι συνδυασμούς]

[και αντικατάσταση λογότυπου άδειας όπου αυτό έχει μπει (σελ. 1, σελ. 2 και τελευταία)]

- Εξαιρείται από την ως άνω άδεια υλικό που περιλαμβάνεται στις διαφάνειες του μαθήματος, και υπόκειται σε άλλου τύπου άδεια χρήσης. Η άδεια χρήσης στην οποία υπόκειται το υλικό αυτό αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη 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



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Συμπληρωματικές σημειώσεις στο μάθημα Στατιστική II

Η έννοια του p-value: Ορισμός και παραδείγματα

Η έννοια του p-value (παρατηρούμενο επίπεδο στατιστικής σημαντικότητας) είναι ιδιαίτερα σημαντική στη στατιστική συμπερασματολογία, ειδικά δε όταν αυτή γίνεται με χρήση στατιστικών πακέτων. Το p-value είναι εργαλείο που συναντάται σε όλες σχεδόν τις στατιστικές αναλύσεις και τα αποτελέσματα των ελέγχων. Για το λόγο αυτό η βαθύτερη κατανόηση του καθίσταται αναγκαία, πριν ο ερευνητής ξεκινήσει διεξαγωγή κάποιας στατιστικής ανάλυσης σε στατιστικό πακέτο.

Για την καλύτερη κατανόηση της έννοιας του p-value, η παρουσίαση γίνεται στα παρακάτω με χρήση γνωστών παραδειγμάτων από την Παραμετρική Στατιστική. Για την επίλυση αυτών έχει χρησιμοποιηθεί το στατιστικό πακέτο SPSS (Statistical Package for Social Sciences).

Παράδειγμα 1

Ένας παραγωγός σταφυλιών ισχυρίζεται ότι τα κιβώτια σταφυλιών που συσκευάζει έχουν βάρος 19 κιλά κατά μέσο όρο. Για να ελεγχθεί ο ισχυρισμός του, λαμβάνουμε ένα τυχαίο δείγμα από 10 κιβώτια. Οι μετρήσεις δίδονται στον πίνακα που ακολουθεί.

Βάρος κιβωτίων του δείγματος (σε κιλά)									
12,0	20,0	18,5	17,6	15,0	14,8	16,7	21,6	15,2	16,0

Βάσει των πληροφοριών του δείγματος, μπορεί ο ισχυρισμός του να θεωρηθεί εύλογος σε $\alpha=5\%$; (σε όλες τις ασκήσεις υποθέτουμε κανονικότητα, όπου αυτή απαιτείται)

Επίλυση

Απαιτείται να ελεγχθεί το ζεύγος υποθέσεων

$$H_0 : \mu = 19$$

versus

$$H_A : \mu \neq 19$$

Εφόσον έχουμε άγνωστη την πληθυσμιακή διακύμανση και μικρό μέγεθος δείγματος, θα χρησιμοποιήσουμε το “one sample t test”.

Τα αποτελέσματα του SPSS δίδονται στους παρακάτω πίνακες.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
βάρος κιβωτίων	10	16,7400	2,79412	,88358

One-Sample Test

	Test Value = 19					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
βάρος κιβωτίων	-2,558	9	,031	-2,2600	-4,2588	-,2612

Το p-value (0.031) είναι μικρότερο του προκαθορισμένου επιπέδου σημαντικότητας (0.031<0.05) και, κατά συνέπεια, η αρχική υπόθεση απορρίπτεται. Επομένως, το βάρος του πληθυσμού των κιβωτίων δεν μπορεί να είναι ίσο με 19 κιλά.

Υπολογισμός του p-value (παρατηρούμενο επίπεδο στατιστικής σημαντικότητας)

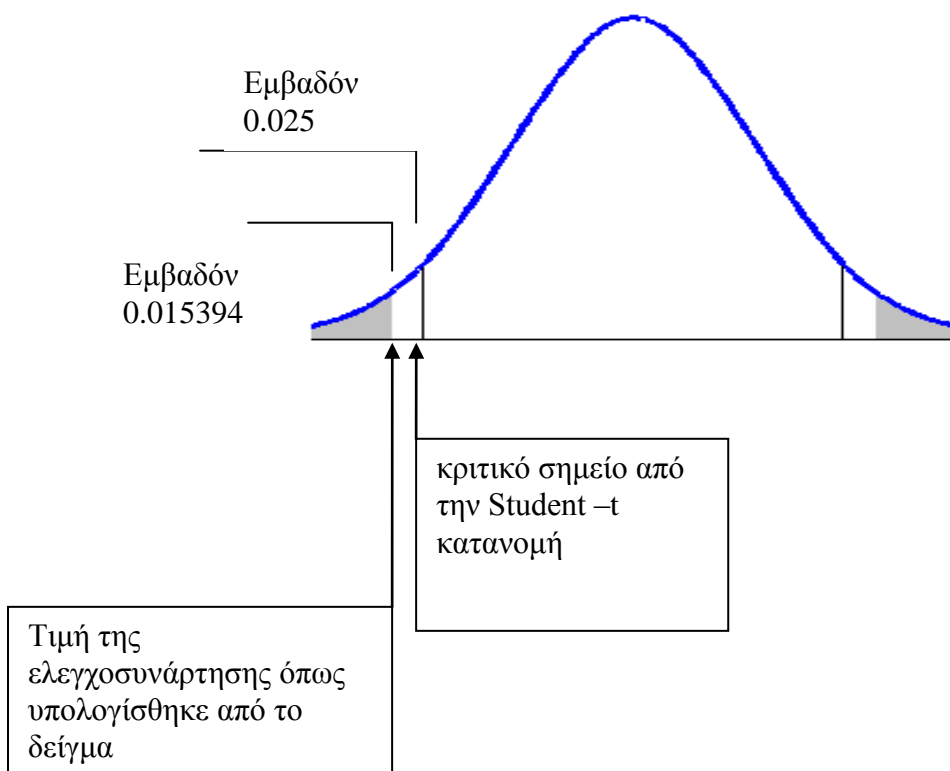
Το p-value ορίζεται ως:

η πιθανότητα να ισχύει στον πληθυσμό η τιμή της ελεγχουσυνάρτησης που υπολογίστηκε από το δείγμα μας ή να ισχύει στον πληθυσμό μια πιο ακραία από αυτήν.

Στο παράδειγμα παρατηρήσαμε τιμή της ελεγχουσυνάρτησης $t=-2.558$. Η αρνητική τιμή της υποδηλώνει ότι ο πληθυσμιακός μέσος που ελέγξαμε (το «19 κιλά») ήταν μεγαλύτερος από τη μέση τιμή που παρατηρήσαμε στο δείγμα (γιατί;). Αυτό δεν σημαίνει βέβαια ότι ήταν **στατιστικά σημαντικά μεγαλύτερος**, τούτο θα ελεγχθεί με τον υπολογισμό του p-value που ακολουθεί.

Το p-value είναι ίσο με $P(t<-2.558)$ ή $P(t>2.558)$ (αυτές οι δύο πιθανότητες υποδηλώνουν πιο ακραίες τιμές από εκείνη της ελεγχουσυνάρτησης που παρατηρήσαμε στο δείγμα. Η δίπλευρη εναλλακτική υπόθεση επιβάλλει να θεωρήσουμε ως ακραίες τις τιμές και στις δυο ουρές της κατανομής. Άρα, $p\text{-value}=P(t<-2.558)+P(t>2.558)=2 P(t<-2.558) = 2 \cdot 0.015394 = 0.0308 \approx 0.031$. (η πιθανότητα μπορεί να υπολογισθεί με το SPSS [compute, function IDF(-2.558, 9)].

Η πιθανότητα που βρέθηκε, δηλαδή να ισχύει στον πληθυσμό αυτή η τιμή της ελεγχουσυνάρτησης ή και πιο ακραία από αυτήν, είναι πολύ μικρή. Η σύγκριση αυτής της πιθανότητας με το **προκαθορισμένο α (0.05)** δίδει το αποτέλεσμα του ελέγχου. Συνεπώς, η πιθανότητα να ισχύει στον πληθυσμό η υπόθεση που ελέγχθηκε (η H_0) είναι μικρότερη του α : η αρχική υπόθεση



Σχήμα Error! No text of specified style in document.-1

απορρίπτεται υπέρ της εναλλακτικής.

Γραφικά, το p-value που υπολογίστηκε φαίνεται στο Σχήμα 1-1.

Το εμβαδόν που ορίζει το p-value (με γκρι χρώμα) ($=2 \cdot 0.015394=0.031$) είναι μικρότερο από το εμβαδόν που ορίζει το επίπεδο σημαντικότητας ($=2 \cdot 0.025=0.05$). Αυτό σημαίνει ότι, ακόμη κι αν δεν γνωρίζουμε την τιμή που έδωσε η ελεγχουσυνάρτηση ή την κριτική τιμή από την Student t κατανομή, μπορούμε, μόνο με χρήση του p-value να συμπεράνουμε ότι η τιμή της ελεγχουσυνάρτησης ήταν μικρότερη από το κριτικό σημείο στην αριστερή ουρά ή μεγαλύτερη από το κριτικό σημείο στη

δεξιά ουρά της κατανομής (εφόσον πρόκειται για δίπλευρο έλεγχο). Άρα, σίγουρα η τιμή της ελεγχουσυνάρτησης βρίσκεται στην περιοχή απόρριψης. **Για το λόγο αυτό, έχοντας μόνο το p-value, απορρίπτουμε την αρχική υπόθεση όταν το p-value είναι μικρότερο του α που επιλέγει ο ερευνητής (συνήθως 0.01, 0.05 ή 0.1), χωρίς να χρειάζεται να ανατρέχουμε σε πίνακες των κατανομών για την εύρεση της κριτικής τιμής.**

Παρατηρείστε ότι το επίπεδο σημαντικότητας α δεν εμπλέκεται στον υπολογισμό του p-value. Αυτό το επιλέγει ο ερευνητής και απλώς το συγκρίνει με το p-value.

95% διάστημα εμπιστοσύνης που υπολογίζεται από το πακέτο

Το 95% διάστημα εμπιστοσύνης (-4.2588, -0.2612) αφορά στη μεταβλητή που προκύπτει αν αφαιρέσουμε από κάθε τιμή του δείγματος την τιμή 19 (που ελέγχουμε ως πιθανό πληθυσμιακό μέσο). Το διάστημα εμπιστοσύνης δεν περιλαμβάνει την τιμή 0 και, κατά συνέπεια, η διαφορά του μέσου βάρους των κιβωτίων από το 19 δεν μπορεί να θεωρηθεί ότι μπορεί στον πληθυσμό να πάρει την τιμή 0: το μέσο βάρος του πληθυσμού συνεπώς δεν μπορεί να είναι 19 κιλά. Επίσης, οι αρνητικές τιμές που περιέχει το διάστημα υποδηλώνουν ότι ο μέσος του πληθυσμού θα είναι μικρότερος του 19.

1.1 Μονόπλευροι έλεγχοι και p-value

Παράδειγμα 2

Με χρήση των δεδομένων της προηγούμενης άσκησης, να ελεγχθούν οι υποθέσεις

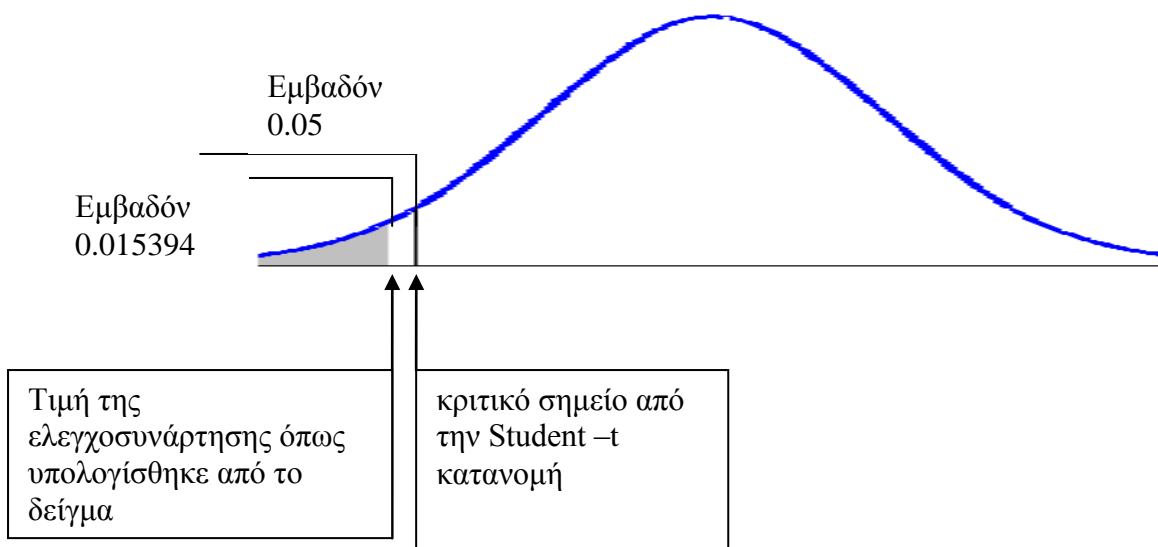
$$\begin{array}{ll} H_0 : \mu = 19 & H_0 : \mu = 19 \\ \text{A) versus} & \text{και} & \text{B) versus} \\ H_A : \mu < 19 & & H_A : \mu > 19 \end{array}$$

Επίλυση

Το p-value που υπολογίστηκε προηγουμένως ήταν 0.031, το οποίο αφορούσε στο δίπλευρο έλεγχο. Το SPSS δεν υπολογίζει p-values για μονόπλευρους ελέγχους, μπορούν όμως αυτά να υπολογισθούν γνωρίζοντας το p-value του δίπλευρου.

A) Η περιοχή απόρριψης σε αυτή την περίπτωση ορίζεται μόνο στην αριστερή ουρά της κατανομής (όπως ορίζει η εναλλακτική υπόθεση). Άρα, το p-value (πιο ακραία τιμή από αυτή που παρατηρήθηκε στο δείγμα...) περιλαμβάνει τις ακραίες τιμές, οι οποίες προφανώς βρίσκονται στην αριστερή ουρά της κατανομής (βλ. Σχήμα 1-2).

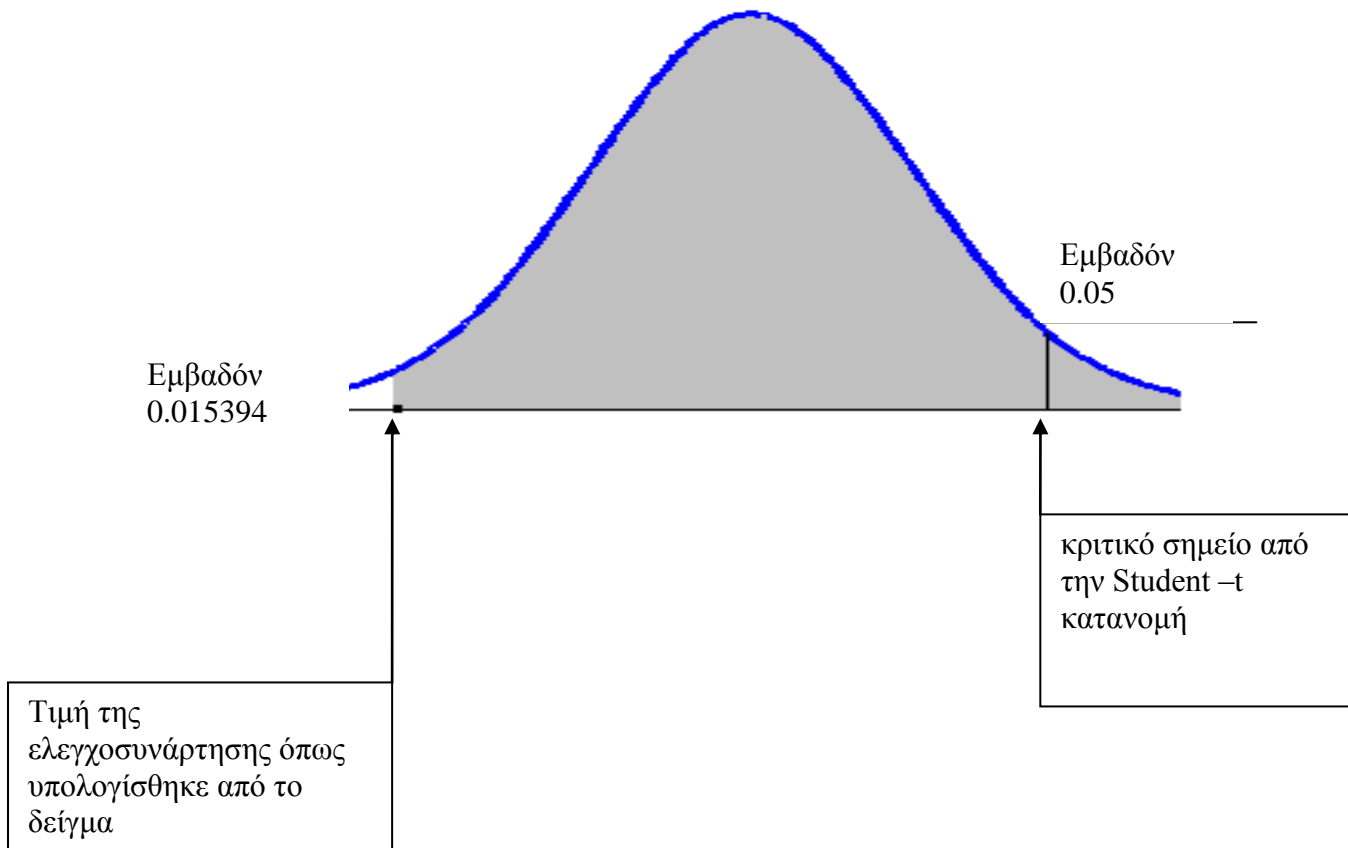
Συνεπώς, $p\text{-value} = \frac{0.031}{2} = 0.015394$. Η πιθανότητα αυτή είναι μικρότερη από το α (έστω 0.05), και επομένως, η H_0 απορρίπτεται. Ο πληθυσμιακός μέσος μπορεί να θεωρηθεί μικρότερος των 19 κιλών (ό,τι εκφράζει η εναλλακτική υπόθεση).



Σχήμα Error! No text of specified style in document.-2

B) Η περιοχή απόρριψης στην περίπτωση αυτή ορίζεται μόνο στη δεξιά ουρά της κατανομής (όπως και πάλι ορίζει η εναλλακτική υπόθεση). Άρα, το p-value (...πιο ακραία τιμή από αυτή που παρατηρήθηκε στο δείγμα....) περιλαμβάνει τις ακραίες τιμές, οι οποίες προφανώς βρίσκονται στη δεξιά ουρά της κατανομής, όλες δηλαδή οι τιμές που βρίσκονται δεξιότερα της τιμής της ελεγχουσυνάρτησης.

Ο υπολογισμός του p-value φαίνεται στο Σχήμα 1-3. Το «λευκό» εμβαδόν ισούται με 0.015394 όπως υπολογίσθηκε στο προηγούμενο ερώτημα. Συνεπώς, $p\text{-value} = 1 - 0.031/2 = 0.9846$. Η πιθανότητα αυτή είναι μεγαλύτερη από το α (έστω 0.05), και επομένως, η H_0 δεν απορρίπτεται. Ο πληθυσμιακός μέσος μπορεί να θεωρηθεί ότι δεν διαφέρει από 19 κιλά (ό,τι εκφράζει η αρχική υπόθεση). Έτσι, σε αντίθεση, με τις προηγούμενες περιπτώσεις, δεν απορρίπτουμε την αρχική υπόθεση, αφού η εναλλακτική εκφράζει κάτι που έρχεται σε ακόμη μεγαλύτερη αντίθεση με την πληροφορία που παίρνουμε από τα δεδομένα.



Σχήμα Error! No text of specified style in document.-3

1.2 Άλλα παραδείγματα

Παράδειγμα 3

Σε κάποια μεταβλητή, ο έλεγχος

$$H_0 : \mu = 110$$

versus

$$H_A : \mu \neq 110$$

έδωσε τιμή της ελεγχουσυνάρτησης $t=1.985$ και p-value 0.07. Τα υπόλοιπα αποτελέσματα του ελέγχου δεν είναι γνωστά.

Στα ίδια δεδομένα, να βρεθεί το p-value του ελέγχου:

$$H_0 : \mu = 110$$

versus

$$H_A : \mu < 110$$

Επίλυση

Το ζητούμενο p-value είναι $1-0.035=0.965$ (γιατί:). Το p-value αυτό είναι μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας (10%, 5% ή 1%) και, κατά συνέπεια, δεν απορρίπτουμε την αρχική υπόθεση. Ο πληθυσμιακός μέσος λοιπόν μπορεί να πάρει την τιμή των 110 μονάδων, έναντι της εναλλακτικής υπόθεσης.

Παράδειγμα 4

Ένας παραγωγός σταφυλιών ισχυρίζεται ότι τα κιβώτια σταφυλιών που συσκευάζει είναι βαρύτερα από τα κιβώτια που συσκεύασε την προηγούμενη χρονιά. Για να ελεγχθεί ο ισχυρισμός του, λαμβάνουμε ένα τυχαίο δείγμα από 9 κιβώτια της φετινής χρονιάς και 9 κιβώτια της προηγούμενης χρονιάς που είχαμε καταγράψει το βάρος τους. Οι μετρήσεις δίδονται στον πίνακα που ακολουθεί.

Βάρος κιβωτίων του δείγματος (σε κιλά)									
Φέτος	20,0	18,5	17,6	15,0	14,8	16,7	21,6	15,2	16,0
Πέρυσι	19,0	16,0	18,0	15,5	15,0	16,0	20,0	16,0	15,0

Βάσει των πληροφοριών του δείγματος, μπορεί ο ισχυρισμός του να θεωρηθεί εύλογος;

Επίλυση

Ο έλεγχος που θα πρέπει να χρησιμοποιηθεί είναι ο “2 independent samples t-test” για ισότητα των μέσων δυο ανεξάρτητων δειγμάτων:

$$H_0 : \mu_{\Phi} = \mu_{\Pi}$$

versus

$$H_A : \mu_{\Phi} \neq \mu_{\Pi}$$

Για να διεξαχθεί ο έλεγχος, θα πρέπει πρώτα να γνωρίζουμε αν οι διακυμάνσεις των δυο πληθυσμών μπορούν να θεωρηθούν ίσες ή όχι (έτσι ώστε να χρησιμοποιηθεί ο

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
VAR00003	Equal variances assumed	,641	,435	,545	16	,593	,5444	,9990	-1,5733	2,6622
	Equal variances not assumed			,545	14,981	,594	,5444	,9990	-1,5851	2,6740

κατάλληλος τύπος της ελεγχουσυνάρτησης). Το SPSS δίνει τον παρακάτω πίνακα αποτελεσμάτων.

Παρατηρούμε ότι ο έλεγχος των διακυμάνσεων δίνει p-value **0.435**, μεγαλύτερο από κάθε σύνηθες επίπεδο σημαντικότητας. Άρα, τα στοιχεία δεν παρέχουν ενδείξεις για να απορρίψουμε την ισότητα των διακυμάνσεων στους δυο πληθυσμούς. Συνεχίζουμε λοιπόν, εξετάζοντας μόνο την πρώτη γραμμή του πίνακα αποτελεσμάτων (όπου δίδεται ο έλεγχος κάτω από την υπόθεση της ισότητας των διακυμάνσεων). Το p-value του ελέγχου για την ισότητα των μέσων είναι **0.593** και, επομένως, τα στοιχεία δεν παρέχουν επαρκείς ενδείξεις ώστε να θεωρήσουμε τους μέσους των δυο πληθυσμών άνισους. Κατά συνέπεια, ο ισχυρισμός του παραγωγού δεν ευσταθεί: οι μέσες τιμές των βαρών των κιβωτίων μπορούν να θεωρηθούν ίσες στις δυο υπό εξέταση χρονιές.

Παράδειγμα 5

Ένας παραγωγός σταφυλιών ισχυρίζεται ότι μπορεί να υπολογίσει ικανοποιητικά το βάρος των κιβωτίων σταφυλιών που συσκευάζει χωρίς να χρησιμοποιεί ζυγαριά. Για να ελεγχθεί ο ισχυρισμός του, λαμβάνουμε ένα τυχαίο δείγμα από 9 κιβώτια τα οποία ζυγίζουμε και στη συνέχεια του τα δίνουμε για να υπολογίσει το βάρος τους, χωρίς να γνωρίζει το πραγματικό βάρος που έδειξε η ζυγαριά. Οι μετρήσεις δίδονται στον πίνακα που ακολουθεί.

Βάρος κιβωτίων του δείγματος (σε κιλά)									
Ζυγαριά	20,0	18,5	17,6	15,0	14,8	16,7	21,6	15,2	16,0
Παραγωγός	19,0	16,0	18,0	15,5	15,0	16,0	20,0	16,0	15,0

Βάσει των πληροφοριών του δείγματος, μπορεί ο ισχυρισμός του να θεωρηθεί εύλογος;

Επίλυση

Τα δείγματα προφανώς είναι εξαρτημένα και ο έλεγχος που θα πρέπει να χρησιμοποιηθεί είναι ο “paired samples t-test” για ισότητα των μέσων των δυο δειγμάτων:

$$H_0 : \mu_z = \mu_{\Pi}$$

versus

$$H_A : \mu_z \neq \mu_{\Pi}$$

ο έλεγχος ισοδύναμι με τον :

$$H_0 : \mu_z - \mu_{\Pi} = 0$$

versus

$$H_A : \mu_z - \mu_{\Pi} \neq 0$$

(άρα, ισοδύναμα, θα μπορούσαμε να δημιουργήσουμε τη διαφορά των δυο μεταβλητών και κατόπιν να ελέγξουμε αν η μεταβλητή αυτή μπορεί να έχει μέσο το 0 στον πληθυσμό ή όχι).

Τα αποτελέσματα δίδονται στον παρακάτω πίνακα:

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 ζυγαριά - παραγωγός	,5444	1,1024	,3675	-,3029	1,3918	1,482	8	,177

Το p-value 0.177 είναι μεγαλύτερο από τα συνήθη επίπεδα σημαντικότητας και, κατά συνέπεια, **η διαφορά του μέσου των βαρών** (υπολογισμός ζυγαριάς-παραγωγού) **δεν είναι στατιστικά σημαντική. Ο παραγωγός πράγματι εκτιμά μέσο βάρος όσο και η ζυγαριά.**

Παρατηρείστε ότι τα δείγματα είναι ακριβώς τα ίδια με την προηγούμενη άσκηση, η πληροφορία όμως περί εξαρτημένων δειγμάτων, δίνει εντελώς διαφορετικά αποτελέσματα.

Παράδειγμα 6

Στα δεδομένα της προηγούμενης άσκησης προσθέτουμε ένα ακόμη κιβώτιο πραγματικού βάρους 17.5 κιλών, του οποίου το βάρος ο παραγωγός εκτίμησε 16 κιλά. Πως αλλάζει το αποτέλεσμα του ελέγχου;

Επίλυση

Το p-value του ελέγχου είναι ίσο με 0.094 (να επαληθευτεί). Συνεπώς, η απόφαση που θα ληφθεί είναι διαφορετική σε $\alpha=5\%$ και σε $\alpha=10\%$.

Σε $\alpha=10\%$, τα στοιχεία παρέχουν ενδείξεις ότι ο παραγωγός υποεκτιμά το μέσο βάρος των κιβωτίων (σε δίπλευρο έλεγχο). Δεν συμβαίνει το ίδιο σε $\alpha=5\%$. Αν εφαρμόσουμε τον πιο ακριβή στην περίπτωση αυτή μονόπλευρο έλεγχο με εναλλακτική υπόθεση ότι η μέση διαφορά είναι μεγαλύτερη του 0 (που σημαίνει ότι ο παραγωγός πράγματι υποεκτιμά το μέσο βάρος), τότε το p-value είναι ίσο με 0.047. Άρα, ακόμη και σε $\alpha=5\%$, η μηδενική ισότητα απορρίπτεται.

Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

