t – test for independent data
t - test για ανεξάρτητα δείγματα

Ζιντζαράς Ηλίας, M.Sc., Ph.D.

Καθηγητής Βιομαθηματικών-Βιομετρίας
Εργαστήριο Βιομαθηματικών
Τμήμα Ιατρικής
Πανεπιστήμιο Θεσσαλίας

Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA

Θεόδωρος Μπρότσης, MSc, PhD Candidate
Ακαδημαϊκός Υπότροφος
(http://biomath.med.uth.gr)
Πανεπιστήμιο Θεσσαλίας
Email: tmprotsis@uth.gr

Μεταβολή αιμοσφαιρίνης δύο EPO φαρμάκων

Test	Reference
0.38	0.8
0.37	0.7
0.24	0.39
0.08	0.06
0.04	0.49
0.03	0.07
0.07	0.63
0.19	0.83
0.35	0.62
0.38	0.95
0.27	0.92
0.2	0.81

Έστω ότι θέλουμε να συγκρίνουμε την μεταβολή της αιμοσφαιρίνης δύο EPO (ερυθροποιητίνη) φαρμάκων, Test (T) και Reference (R)

Θέλουμε να δούμε αν υπάρχει διαφορά στη μέση τιμή της αιμοσφαιρίνης μεταξύ του Test και Reference φαρμάκου

Η υπόθεση είναι ότι η μέση διαφορά των τιμών της αιμοσφαιρίνης είναι μηδέν. Αυτό σημαίνει πως δεν υπάρχει διαφορά στις μέσες τιμές της αιμοσφαιρίνης μεταξύ του Test και Reference φαρμάκου

$n = 12$ ασθενείς λαμβάνουν το Test φάρμακο και $n = 12$ ασθενείς το Reference φάρμακο.

Περιγραφική Στατιστική

Test	0.38	0.37	0.24	0.08	0.04	0.03	0.07	0.19	0.35	0.38	0.27	0.2
Reference	0.8	0.7	0.39	0.06	0.49	0.07	0.63	0.83	0.62	0.95	0.92	0.81
n			12	12								
$\bar{x}$ (mean or average)			0.22	0.61								
s or SD (variability)			0.14	0.30								
CV% (=SD/mean)			64%	49%								
SE (=variability/√size)			0.04	0.09								

Μηδενική υπόθεση

Test φάρμακο

Υπόθεση;

Reference φάρμακο

Η μέση διαφορά είναι μηδέν

μ_t

μ_r

$\mu_t - \mu_r = 0$ **ή** $\mu_t = \mu_r$

$s_t = 0.1366$

$s_r = 0.3017$

Ανεξαρτησία;

Αυτές οι δύο μετρήσεις ΕΙΝΑΙ ανεξάρτητες μεταξύ τους

Μηδενική υπόθεση

Η μηδενική υπόθεση είναι ότι δεν υπάρχει διαφορά στη μέση μεταβολή της αιμοσφαιρίνης. Η διαφορά στην μέση μεταβολή είναι $= 0$

$$H_0: \mu_t - \mu_r = 0 \text{ ή } \mu_t = \mu_r$$

$$H_a: \mu_t - \mu_r \neq 0 \text{ ή } \mu_t \neq \mu_r$$

Επίπεδο σημαντικότητας: $a = 0.05$

Καθώς το $\boldsymbol{\sigma}$ δεν είναι γνωστό θα κάνουμε χρήση της $\boldsymbol{t - κατανομης}$ με $\boldsymbol{(n_T - 1) + (n_R - 1) = (12 - 1) + (12 - 1) = 22}$ βαθμούς ελευθερίας

Κανόνας απόφασης:

Επειδή το $a = 0.05$ και επειδή κάνουμε χρήση της $t -$ κατανομής, η H_0 απορρίπτεται αν το στατιστικό τεστ είναι > 2.07 (5% σημείο της $t -$ κατανομής για $\mathbf{22}$ βαθμούς ελευθερίας) ή < -2.07

Σημαντικότητα

Προκειμένου να ελεγχθεί η **σημαντικότητα** της διαφοράς μεταξύ των μέσω τιμών ($\boldsymbol{\bar{d} = 0.22 - 0.61 = -0.39}$),

δηλ., αν η διαφορά είναι **διαφορετική από το μηδέν** ή **όχι μία τυχαία τιμή**,

πρέπει να εξετάσουμε τη διαφορά ($\bar{d}$) σε **συνδυασμό** με το **σφάλμα** της διαφοράς (SE),

δηλαδή, τη συνολική **μεταβλητότητα** (SD των δύο θεραπειών) και το **μέγεθος** της δοκιμής ($n = 12 + 12$)

T-Test

Στη συνέχεια, θα μπορούσαμε να ελέγξουμε στατιστικά εάν η διαφορά μεταξύ των δύο μέσων τιμών αποκλίνει από το μηδέν (δηλ. τη σημαντικότητά του) χρησιμοποιώντας το **t-test**:

$$t = \frac{(\bar{x}_t - \bar{x}_r)}{SE}$$

$$SE = \sqrt{s_{pooled}^2 \left(\frac{1}{n_t} + \frac{1}{n_r}\right)}$$

$$s_{pooled}^2 = \frac{(n_t - 1)s_t^2 + (n_r - 1)s_r^2}{n_t + n_r - 2}$$

s_t = τυπική απόκλιση των παρατηρήσεων φαρμάκου T

s_r = τυπική απόκλιση των παρατηρήσεων φαρμάκου R

n_t = αριθμός ασθενών που έλαβαν το T φάρμακο

n_r = αριθμός ασθενών που έλαβαν το R φάρμακο

T-Test

$$t = \frac{(\bar{x}_t - \bar{x}_r)}{SE}$$

	T	R
n	12	12
$\bar{x}$ (mean or average)	0.22	0.61
s or SD (variability)	0.14	0.30

$$SE = \sqrt{\left(\frac{(11)0.1366^2 + (11)0.3017^2}{12+12-2}\right)\left(\frac{1}{12}+\frac{1}{12}\right)} = \sqrt{0.055(0.167)} = \sqrt{0.0092} = 0.096$$

$$t = \frac{(0.22 - 0.61)}{SE} = \frac{-0.39}{0.096} = -4.06$$

Το πρόσημο αγνοείται

Σημαντικότητα

Η τιμή του t-test ($\boldsymbol{t = 4.06}$) είναι μεγαλύτερη από το 5% σημείο της **t-κατανομής** με

$$(n_T - 1) + (n_R - 1) = (12 - 1) + (12 - 1) = 22$$

βαθμούς ελευθερίας που είναι 2.07 (δες πίνακα)

Έτσι, η $t = 4.06$ τιμή είναι **σημαντική** με πιθανότητα σφάλματος $\boldsymbol{P < 0.05}$

(δηλαδή μια μικρή πιθανότητα σφάλματος)

	Percentage points of the t distribution		
	p-value		
df (=n-1)	0.05	0.01	0.001
1	12.71	63.66	636.62
2	4.3	9.92	31.6
3	3.18	5.84	12.92
4	2.78	4.6	8.61
5	2.57	4.03	6.87
6	2.45	3.71	5.96
7	2.36	3.5	5.41
8	2.31	3.36	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59
20	2.09	2.85	3.85
22	2.07	2.82	3.79
30	2.04	2.75	3.65
120	1.98	2.62	3.37
∞	1.96	2.58	3.29

Συμπέρασμα

Έτσι, μπορούμε να υποστηρίξουμε ότι υπάρχει σημαντική διαφορά στην **μέση μεταβολή** της αιμοσφαιρίνης **μεταξύ του T και R φαρμάκου** με μία μικρή πιθανότητα σφάλματος ($\boldsymbol{P < 0.05}$)

95% διάστημα εμπιστοσύνης

Η σημαντικότητα της διαφοράς μεταξύ των δύο μέσων, τιμή $\bar{d}$, μπορεί να αξιολογηθεί και από το 95% διάστημα εμπιστοσύνης

Το 95% CI ορίζεται ως:

$$(\bar{d} - t * SE, \bar{d} + t * SE)$$

Όπου t είναι το 5% σημείο της t-κατανομής για $(n_t - 1) + (n_r - 1) = (12 - 1) + (12 - 1) = 22$ βαθμούς ελευθερίας
και είναι 2.07
(δες πίνακα της t-κατανομής)

	Percentage points of the t distribution		
	p-value		
df (=n-1)	**0.05**	**0.01**	**0.001**
1	12.71	63.66	636.62
2	4.3	9.92	31.6
3	3.18	5.84	12.92
4	2.78	4.6	8.61
5	2.57	4.03	6.87
6	2.45	3.71	5.96
7	2.36	3.5	5.41
8	2.31	3.36	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59
20	2.09	2.85	3.85
22	2.07	2.82	3.79
30	2.04	2.75	3.65
120	1.98	2.62	3.37
∞	1.96	2.58	3.29

95% διάστημα εμπιστοσύνης

$$(\bar{d} - \ t * SE, \bar{d} + \ t * SE)$$

$$(-0.39 - 2.07 * 0.096, -0.39 + 2.07 * 0.096)$$

$$(-0.59, -0.19)$$

Έτσι, με **95% εμπιστοσύνη (πιθανότητα)** μπορούμε να ισχυριστούμε ότι η πραγματική διαφορά μεταξύ των δύο μέσων τιμών βρίσκεται εντός του διαστήματος **(-0.59, -0.19)**

Το **0 δεν περιλαμβάνεται** στο **95% CI**, οπότε υπάρχει διαφορά στατιστικά σημαντική μεταξύ του Test και Reference φαρμάκου σε ότι αφορά την αιμοσφαιρίνη