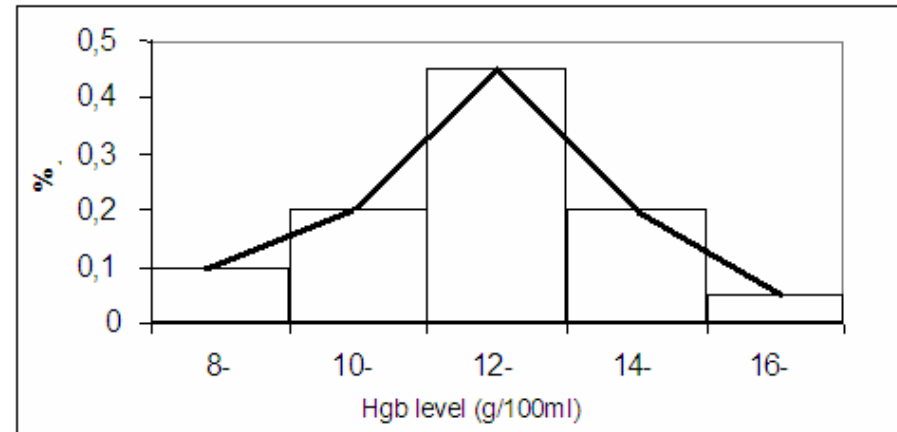
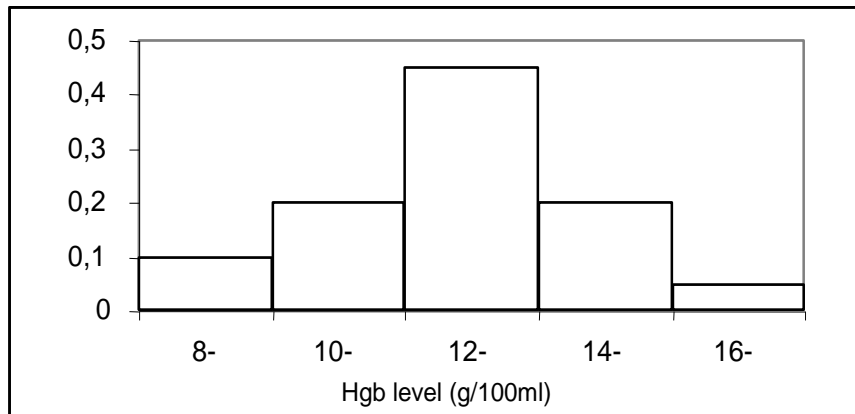


t-test για ανεξάρτητα δείγματα

Recall: Ιστόγραμμα

Hgb levels	
8,8	12,9
9,3	12,9
10,5	12,9
10,6	13,3
11,1	13,4
11,4	14,5
12	14,6
12	14,6
12,1	15,1
12,1	16,1

Hgb	Frequency	Proportion
8-	2	0,1
10-	4	0,2
12-	9	0,45
14-	4	0,2
16-	1	0,05
Σύνολο	20	

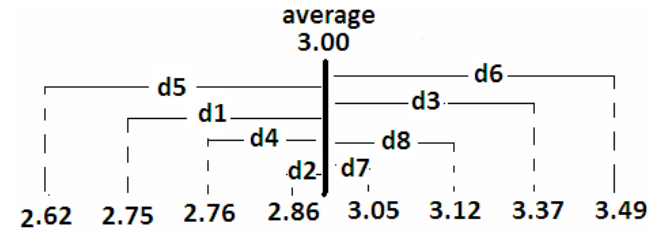


Παράδειγμα: Οι όγκοι πλάσματος (x) από 8 υγιείς ενήλικες άνδρες είναι:

2.75, 2.86, 3.37, 2.76, 2.62, 3.49, 3.05, 3.12 lt

Μέση τιμή: $\bar{x} = \Sigma \mathbf{x} / \mathbf{n} = (x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8) / \mathbf{n} =$

$$(2.75 + 2.86 + 3.37 + 2.76 + 2.62 + 3.49 + 3.05 + 3.12) / 8 = 24.02 / 8 = 3.00$$



Διακύμανση $= s^2 = \Sigma (\mathbf{x} - \bar{x})^2 / (\mathbf{n} - 1) =$

$$\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2 + (x_7 - \bar{x})^2 + (x_8 - \bar{x})^2\} / (\mathbf{n} - 1) =$$

$$(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 + d_7^2 + d_8^2) / (\mathbf{n} - 1) =$$

$$\{(2.75 - 3)^2 + (2.86 - 3)^2 + (3.37 - 3)^2 + (2.76 - 3)^2 + (2.62 - 3)^2 + (3.49 - 3)^2 + (3.05 - 3)^2 + (3.12 - 3)^2\} / (8 - 1) = 0.0961$$

Τυπική απόκλιση $= \sqrt{\text{διακύμανση}} = s = 0.311$

Τυπικό σφάλμα μέσης τιμής: $SE = s / \sqrt{\mathbf{n}} = 0.311 / \sqrt{8} = 0.111$

Αν το μέγεθος του δείγματος προσεγγίζει το μέγεθος του πληθυσμού τότε το SE τείνει στο 0

Για να αξιολογήσουμε την αποτελεσματικότητα του Tysabri, 24 ασθενείς με RRMS τυχαιοποιήθηκαν στο Tysabri ή στο placebo. Η μεταβλητή που μετρήθηκε ήταν ο ετήσιος ρυθμός υποτροπών (ARR). Τα αποτελέσματα ήταν τα ακόλουθα:

Tysabri	Placebo
(n=12)	(n=12)
0.38	0.80
0.37	0.70
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
0.20	0.81

Είναι το Tysabri αποτελεσματικό? Είναι η μέση τιμή του ARR στο Tysabri διαφορετική από το Placebo?

ARR: Annualized relapse rate is calculated as the number of relapses for each subject divided by the number of years followed in the study for that subject.

Περίληπτικά Στατιστικά

ARR

Tysabri Placebo

0.38	0.80
0.37	0.70
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
0.20	0.81

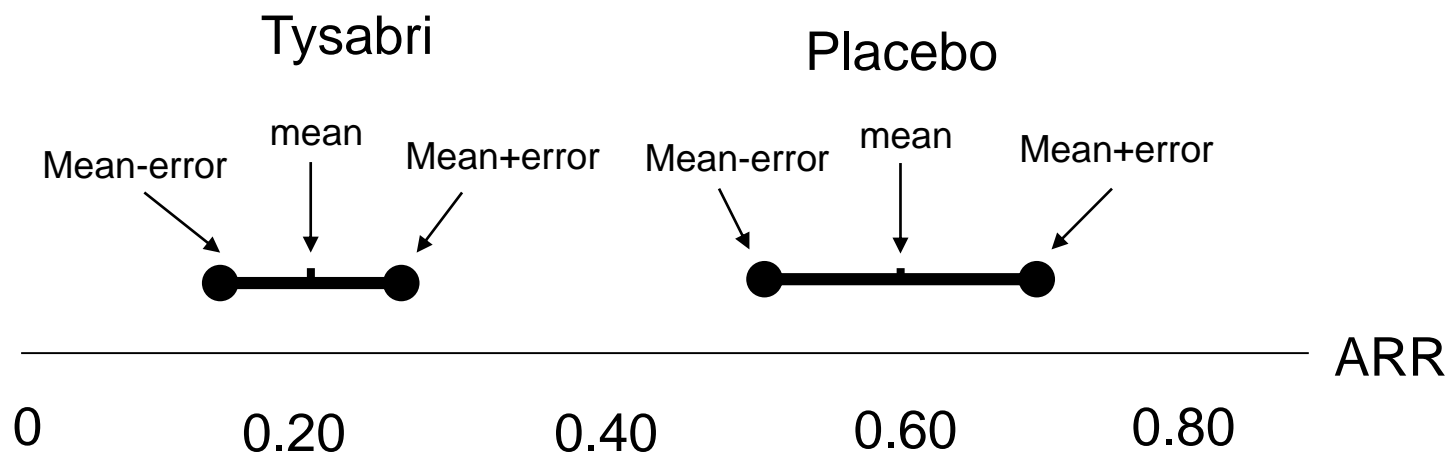
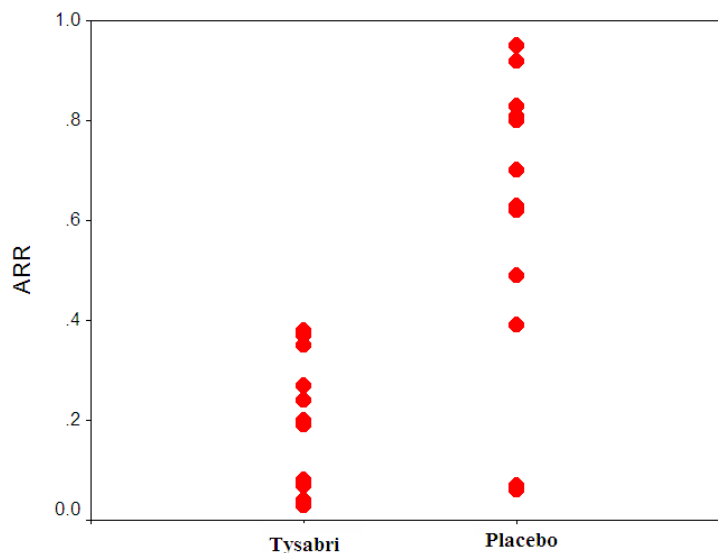
n	12	12
Mean or average (\bar{x})	0.22	0.61
SD (variability)	0.14	0.30
SE (error=variability/size)	0.04	0.09

	Tysabri	Placebo
n	12	12
\bar{X} (Mean)	0.22	0.61
S, SD (variability)	0.14	0.30
SE (error=variability/size)	0.04	0.09

Κατά μέσο όρο, το Tysabri εμφάνισε λιγότερες ARR (μέση τιμή=0.22) από ότι το Placebo (mean=0.61)

Το SE για το Tysabri (SE=0.04) είναι μικρότερο από το Placebo (SE=0.09)

**Μπορούμε να
διερευνήσουμε τα
αποτελέσματα ως εξής:**



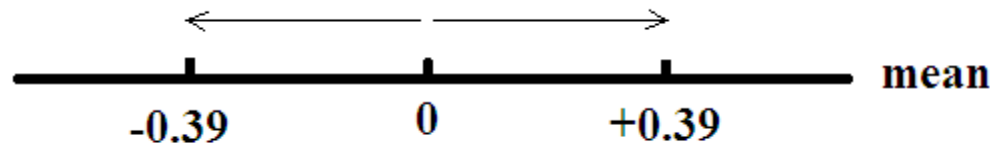
Αν πάρουμε υπόψη τη μέση τιμή και το τυπικό σφάλμα, SE, (δηλ. τη μεταβλητότητα των δεδομένων και το μέγεθος της μελέτης) του κάθε φαρμάκου, το Tysabri φαίνεται να είναι πιο αποτελεσματικό από το Placebo όσον αφορά το ARR

Το μέγεθος της Αποτελεσματικότητας

Μπορούμε να συγκρίνουμε το Tysabri vs. Placebo υπολογίζοντας την διαφορά μεταξύ των δύο μέσων τιμών

$$\text{Difference (D)} = \text{Mean T} - \text{Mean P} = 0.22 - 0.61 = -0.39$$

Αλλά, με τι σιγουριά μπορώ να ισχυριστώ ότι η διαφορά μεταξύ των δύο μέσων τιμών ($D=-0.39$) είναι σημαντική, δηλ. ότι δεν είναι τυχαία και συνεπώς, ότι είναι διαφορετική από το 0?



Τα δεδομένα εμφανίζουν μεταβλητότητα και μέγεθος του δείγματος είναι μικρό, $n=12+12=24$ ασθενείς, (δηλ. υπάρχει σφάλμα)

	Tysabri (T)	Placebo (P)
n	12	12
Mean \bar{x}	0.22	0.61
S, SD (variability)	0.14	0.30

Συνεπώς, για να **ελέγξουμε** ότι η ότι η διαφορά μεταξύ των δύο μέσων τιμών (**$D=-0.39$**) είναι **σημαντική**, πρέπει να λάβουμε υπόψη την **μεταβλητότητα των δεδομένων** (δηλ. την τυπική απόκλιση, s ή SD, των δύο φαρμάκων) και το **μέγεθος της μελέτης** ($n=12+12$), δηλ. το τυπικό σφάλμα (SE) της διαφοράς.

	Tysabri (T)	Placebo (P)
n	12	12
Mean \bar{x}	0.22	0.61
S, SD (variability)	0.14	0.30

Συνεπώς, ελέγχουμε αν η διαφορά μεταξύ των δύο μέσων τιμών (**D=-0.39**) είναι **σημαντική** (δηλ. αν διαφέρει από το 0 ή ότι δεν είναι τυχαία) χρησιμοποιώντας το **t-test**:

$$t = \frac{(\text{mean T}) - (\text{mean P})}{SE} = \frac{\bar{x}_T - \bar{x}_P}{SE}$$

$$\text{where } SE = \sqrt{s_{\text{pooled}} \left(\frac{1}{n_T} + \frac{1}{n_P} \right)} \text{ and } s_{\text{pooled}}^2 = \frac{(n_T - 1)s_T^2 + (n_P - 1)s_P^2}{(n_T - 1) + (n_P - 1)}$$

	Tysabri (T)	Placebo (P)
n	12	12
Mean \bar{X}	0.22	0.61
S, SD (variability)	0.14	0.30

$$t = \frac{(\text{mean T}) - (\text{mean P})}{\text{SE}} = \frac{\bar{X}_T - \bar{X}_P}{\text{SE}} = \frac{0.61 - 0.22}{0.096} = -4.07$$

$$\text{where SE} = \sqrt{s_{\text{pooled}} \left(\frac{1}{12} + \frac{1}{12} \right)}$$

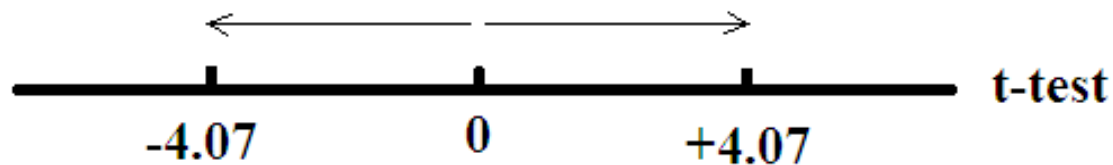
$$\text{and } s_{\text{pooled}}^2 = \frac{(12-1)0.14^2 + (12-1)0.30^2}{(12-1) + (12-1)}$$

Significance of the difference, P-value

Τώρα έχουμε να απαντήσουμε στο ερώτημα:

είναι η τιμή $t=4.07$ (το πρόσημο αγνοείται) **σημαντική** (δηλ. **όχι τυχαία**), και συνεπώς, **μακριά από το 0**?

ή με τι **πιθανότητα λάθους (P-value)** μπορώ να ισχυριστώ ότι το $t=4.07$ είναι σημαντικό?



Αν το $t=4.07$ είναι σημαντικό σημαίνει ότι η διαφορά μεταξύ των δύο μέσων τιμών ($D=-0.39$) λαμβάνοντας υπόψη την μεταβλητότητα και το μέγεθος των δεδομένων (δηλ. το SE), είναι **σημαντική**, δηλαδή **διαφορετική από το μηδέν**.

Μπορούμε να απαντήσουμε στο προηγούμενο ερώτημα συγκρίνοντας την τιμή $t=4.07$ με το **5% σημείο της t-κατανομής για $(n_T-1)+(n_P-1)=(12-1)+(12-1)=22$ df** (που από τον πίνακα της t-κατανομής είναι 2.07) .

Table Percentage points of the t distribution.

<u>d.f. (=n-1)</u>	<u>P=0.05</u>	<u>d.f. (=n-1)</u>	<u>P=0.05</u>	<u>d.f. (=n-1)</u>	<u>P=0.05</u>
1	12.71	13	2.16	25	2.06
2	4.30	14	2.14	26	2.06
3	3.18	15	2.13	27	2.05
4	2.78	16	2.12	28	2.05
5	2.57	17	2.11	29	2.04
6	2.45	18	2.10	30	2.04
7	2.36	19	2.09	40	2.02
8	2.31	20	2.09	60	2.00
9	2.26	21	2.08	120	1.98
10	2.23	22	2.07	∞	1.96
11	2.20	23	2.07		
12	2.18	24	2.06		

Επειδή το $t=4.07$ είναι **μεγαλύτερο** από το **2.07** συμπεραίνουμε ότι η **διαφορά** των δύο μέσων τιμών είναι **διαφορετική από το μηδέν** με πιθανότητα λάθους $P\text{-value} < 0.05$ ($P < 0.05$).

Οπότε το φάρμακο είναι αποτελεσματικό.

Table Percentage points of the t distribution.

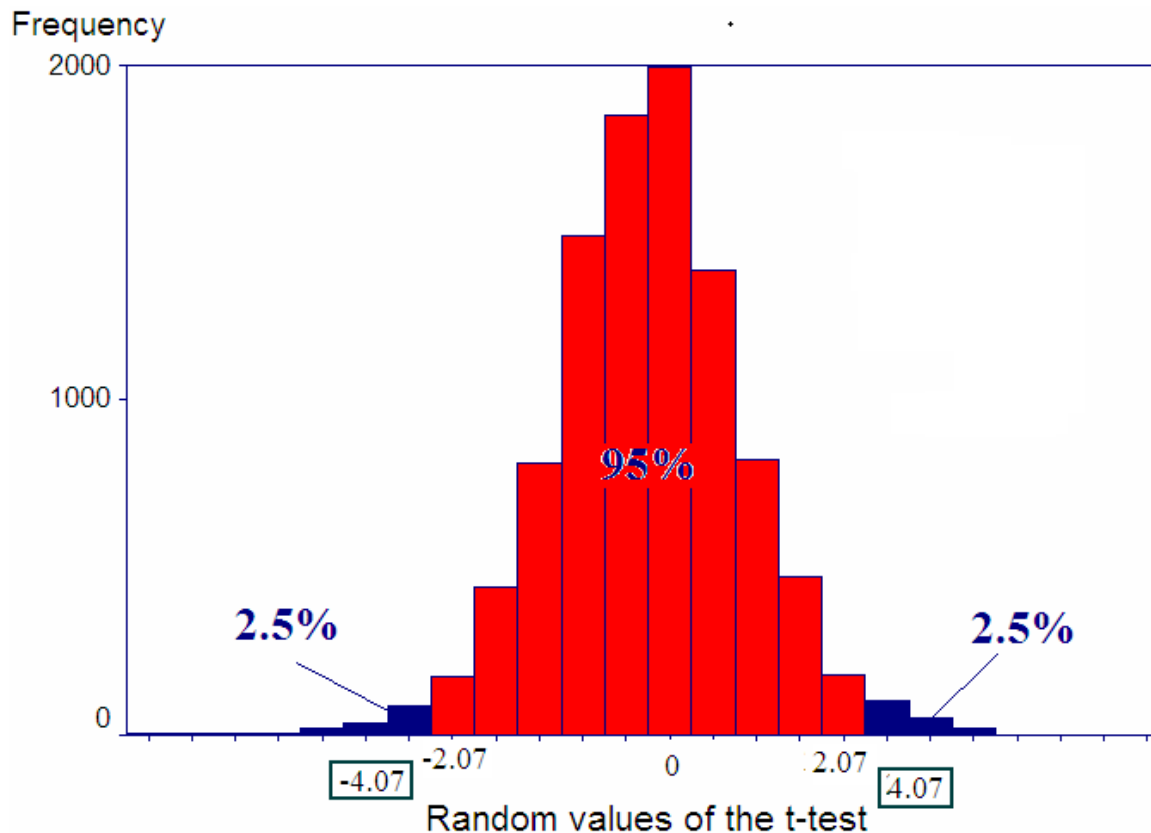
<u>d.f. (=n-1)</u>	<u>P=0.05</u>	<u>d.f. (=n-1)</u>	<u>P=0.05</u>	<u>d.f. (=n-1)</u>	<u>P=0.05</u>
1	12.71	13	2.16	25	2.06
2	4.30	14	2.14	26	2.06
3	3.18	15	2.13	27	2.05
4	2.78	16	2.12	28	2.05
5	2.57	17	2.11	29	2.04
6	2.45	18	2.10	30	2.04
7	2.36	19	2.09	40	2.02
8	2.31	20	2.09	60	2.00
9	2.26	21	2.08	120	1.98
10	2.23	22	2.07	∞	1.96
11	2.20	23	2.07		
12	2.18	24	2.06		

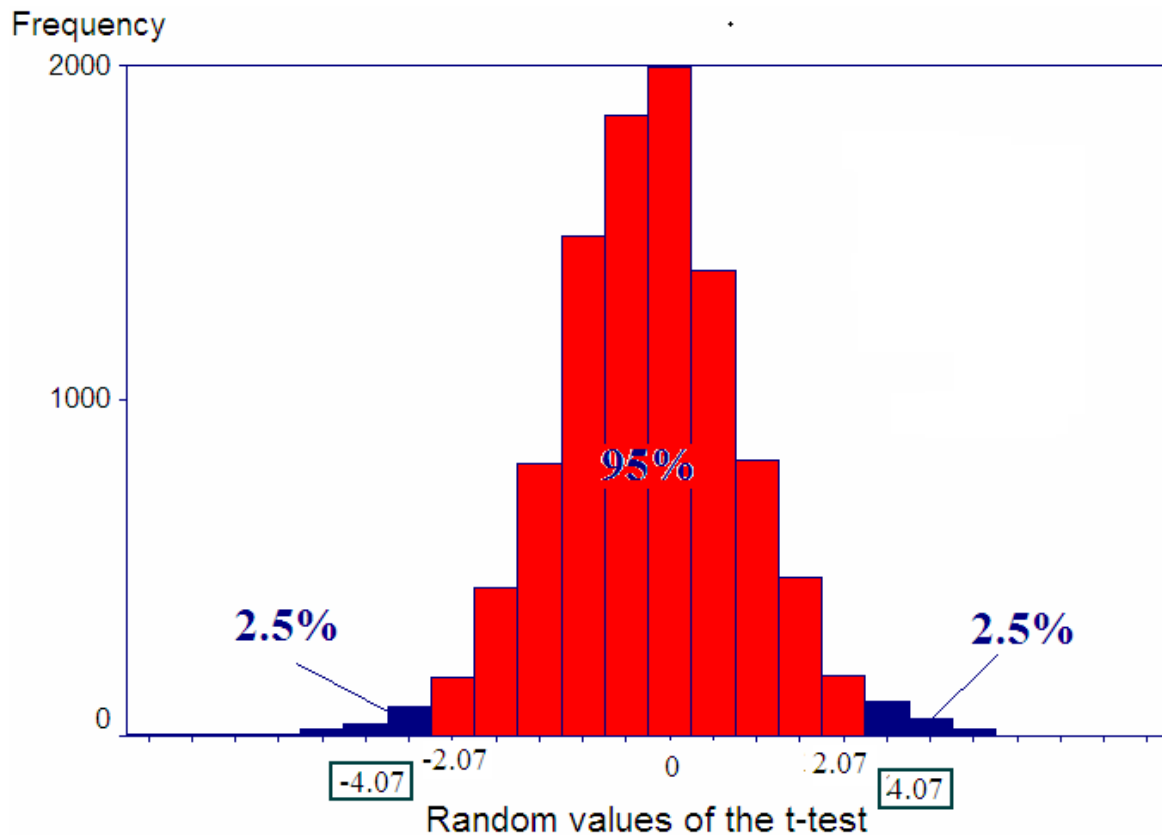
Το P-value λέγεται και στάθμη σημαντικότητας

Τι είναι το 2.07?

Αν επαναλάβουμε (**προσομοιώσουμε**) με τυχαίο τρόπο την μελέτη **10000** φορές (πχ ανακατεύοντας τυχαία τα νούμερα) και υπολογίσουμε κάθε φορά το t-test, τότε το ποσοστό της κατανομής (δηλ. των τυχαίων t-tests) που είναι μεγαλύτερο από **2.07** (ή μικρότερο από **-2.07**) είναι **5%** (**P-value, $P=0.05$**).

Το **2.07** λέγεται **5% σημείο της t-κατανομής** για 22β.ε.





Το $t=4.07$ είναι μεγαλύτερο από 2.07.

Οπότε, το $t=4.07$ δεν είναι τυχαία τιμή και είναι διαφορετική (μακριά) από το μηδέν με μία πιθανότητα λάθους μικρότερη του 5% ($P < 0.05$), δηλαδή **μικρή πιθανότητα λάθους, και συγκεκριμένα, με πιθανότητα λάθους 0.1% ($P = 0.001$) (βρίσκεται μόνο με Στατιστικό Πακέτο)**

Συνεπώς, η **διαφορά** των δύο μέσων τιμών είναι **σημαντική**, δηλ. διαφορετική από το μηδέν, με **πιθανότητα λάθους $P < 0.05$** (ή $P = 0.001$).

Συμπέρασμα

Υπάρχει **σημαντική διαφορά** μεταξύ των μέσων τιμών του **ARR** μεταξύ **Tysabri** και **Placebo** (με το Tysabri να εμφανίζει **καλύτερη ανταπόκριση**) με μικρή πιθανότητα λάθους (**$P < 0.05$**) ή (**$P = 0.001$**)

Διάστημα εμπιστοσύνης (δε) μέσης τιμής των διαφορών

Το 95% δ.ε. για την διαφορά δύο μέσων τιμών είναι:

$$(D - t*SE, D + t*SE)$$

t είναι το 5% σημείο της t-κατανομής για 22df, και είναι $t=2.07$ (see previous Table).

$$(D - t*SE, D+ t*SE)$$

$$(-0.39-1.96*0.096, -0.39+1.96*0.096) = (-0.59, -0.19)$$

Συνεπώς, υπάρχει **95% πιθανότητα** η διαφορά των δύο μέσων τιμών να βρίσκεται στο **διάστημα** αυτό.

Επειδή το **0** δεν συμπεριλαμβάνεται στο **95% δ.ε.** συμπεραίνουμε ότι το υπάρχει σημαντική διαφορά μεταξύ των δύο φαρμάκων με το Tysabri να εμφανίζει καλύτερη ανταπόκριση στο ARR

Τα αποτελέσματα

	Tysabri (T)	Placebo (P)
N	12	12
Mean	0.22	0.61
SD (variability)	0.14	0.30
SE (error=variability/size)	0.04	0.09
Difference (D) = MeanT -MeanP	-0.39	
t-test (D / error)	-4.07	
P-value	0.001	
95% CI	(-0.59, -0.19)	

The FEV1 was measured in patients with asthma and in controls:

FEV	
Asthma	Controls
2.12	2.22
1.63	2.20
2	3.71
2.3	2.49
2.02	2.64
1.74	1.38
2.49	2.95
2.3	2.49
2.19	2.68
	2.42
	1.96

Test the difference between the two groups in terms of FEV1.



- Output
 - T-Test
 - Title
 - Notes
 - Group Statistics
 - Independent Samples Test

T-Test

Group Statistics

GROUP	N	Mean	Std. Deviation	Std. Error Mean
FEV Asthma	9	2.0878	.27526	.09175
FEV Controls	11	2.4673	.58548	.17653

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
FEV	Equal variances assumed	1.469	.241	-1.784	18	.091	-.3795	.21278	-.82653	.06754
	Equal variances not assumed			-1.907	14.784	.076	-.3795	.19895	-.80408	.04509

Έλεγχος για ένα ποσοστό

Έλεγχος για ένα ποσοστό

Για τον έλεγχο ενός ποσοστού χρησιμοποιείται το z-test.

Παράδειγμα: Σε ένα νοσοκομείο συλλέχθηκε ένα τυχαίο δείγμα από $n=215$ γυναίκες από τους καταλόγους νοσηλευομένων και βρέθηκε ότι $r=39$ από αυτές έχουν ιστορικό **άσθματος** (δηλ. το παρατηρούμενο ποσοστό άσθματος είναι $p=r/n=39/215=18\%$). Είναι γνωστό ότι ο επιπολασμός (ποσοστό) της νόσου είναι $\pi=15\%$. Συμφωνεί το ποσοστό των γυναικών που βρέθηκε με άσθμα με αυτό του γενικού πληθυσμού;

Για να ελέγξουμε αν το ποσοστό των περιστατικών δεν διαφέρει από τον επιπολασμό της νόσου χρησιμοποιούμε το z-test:

$$z = \frac{(\text{παρατηρούμενο ποσοστό}) - (\text{θεωρητικό ποσοστό})}{\text{τυπικό σφάλμα}} = \frac{p - \pi}{SE}$$

where $SE = \sqrt{\frac{\pi(1 - \pi)}{n}}$

Το παρατηρούμενο ποσοστό των περιστατικών με άσθμα είναι:

$$p = r / n = 39 / 215 = 0.18$$

Ο επιπολασμός της νόσου (το θεωρητικό ποσοστό) είναι:

$$\pi = 0.15$$

$$z = \frac{\text{(παρατηρούμενο ποσοστό)} - \text{(θεωρητικό ποσοστό)}}{\text{τυπικό σφάλμα}} = \frac{p - \pi}{SE} = \frac{0.18 - 0.15}{0.024} = 1.23$$

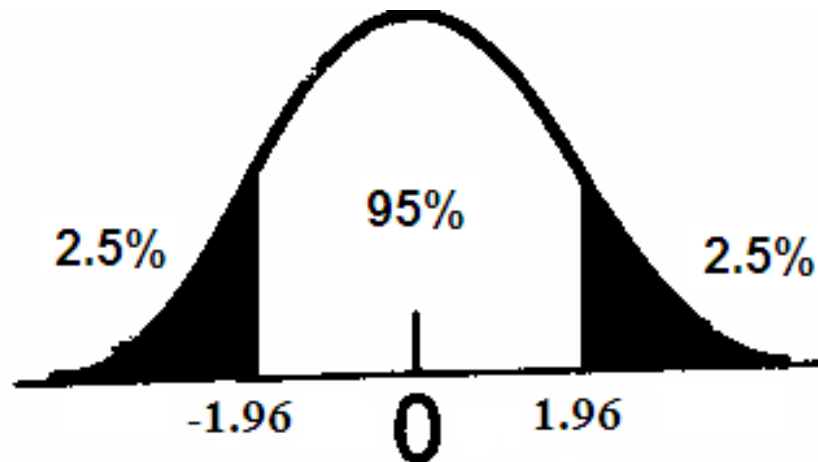
(Το πρόσημο αγνοείται)

$$\text{where } SE = \sqrt{\frac{\pi(1 - \pi)}{n}} = \sqrt{\frac{0.15(1 - 0.15)}{215}} = 0.024$$

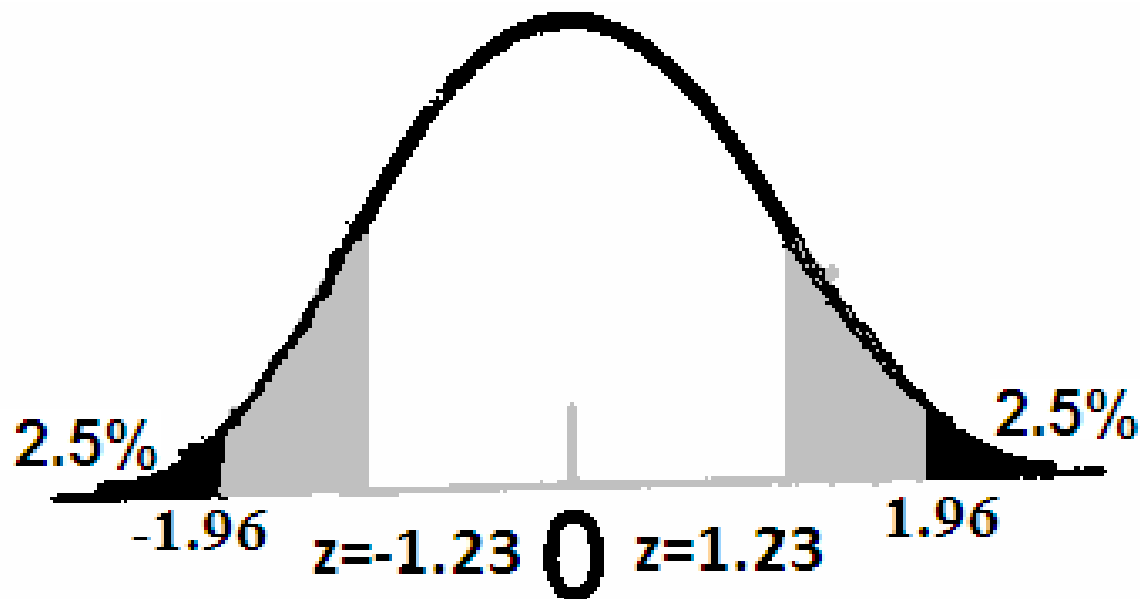
Αν προσομοιώσουμε (με τυχαίο τρόπο) την μελέτη 10000 φορές και υπολογίσουμε κάθε φορά το z-test, τότε η κατανομή των z-tests είναι η τυπική κανονική κατανομή

Το 5% σημείο της τυπικής κανονικής κατανομής είναι 1.96, δηλ. το 5% των 10000 z-tests είναι μεγαλύτερα (ή μικρότερα) του 1.96 (ή του -1.96)

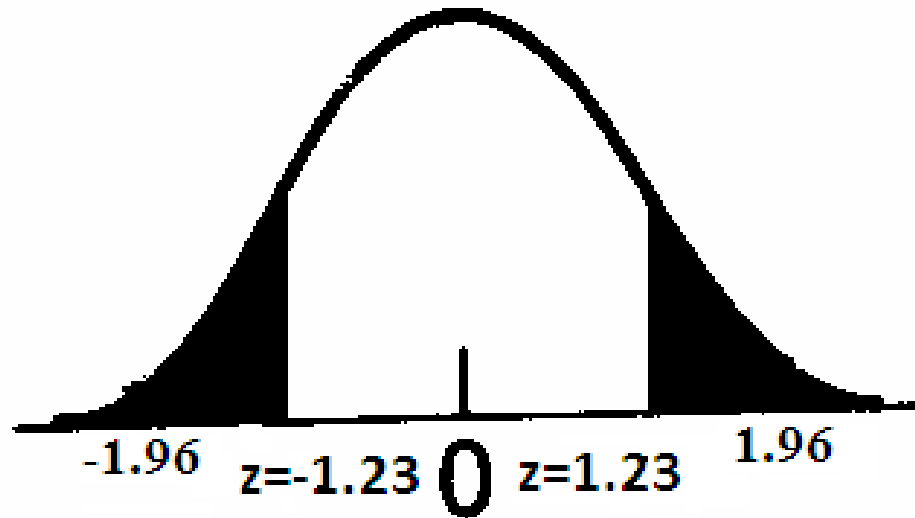
Τυπική Κανονική κατανομή



- Για να ελέγξουμε αν το $z=1.23$ είναι σημαντικό, συγκρίνουμε το $z=1.23$ με το 5% σημείο της τυπικής κανονικής κατανομής που είναι 1.96
- Επειδή, το $z=1.23$ είναι μικρότερο από το 1.96, το ποσοστό των τυχαίων z -tests που είναι μεγαλύτερα του $z=1.23$ (ή μικρότερα του $z=-1.23$) είναι μεγαλύτερο από 5% (P-value, $P \geq 0.05$)



Συνεπώς, συμπεραίνουμε ότι το $z=1.23$ είναι διαφορετικό (μακριά) από το μηδέν με πιθανότητα λάθους μεγάλη, μεγαλύτερη από το 5%, δηλ. το z δεν διαφέρει από το μηδέν

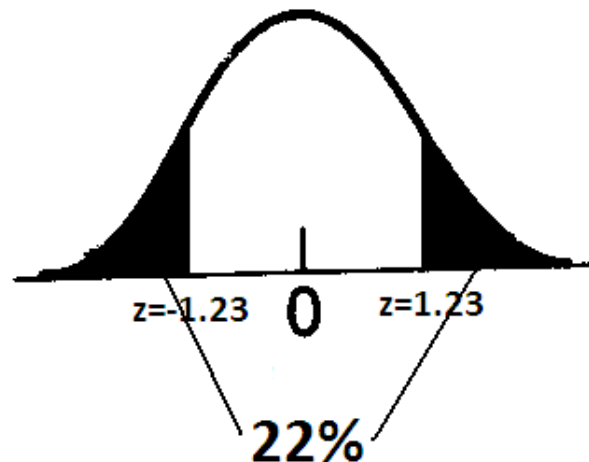


Το ποσοστό των περιστατικών με άσθμα στο νοσοκομείο δεν διαφέρει από τον επιπολασμό της νόσου ($P \geq 0.05$)

Εναλλακτικά, για να ελέγξουμε αν το $z=1.23$ είναι σημαντικό, υπολογίζουμε το ποσοστό των προσομοιωμένων z -tests που είναι μεγαλύτερα του $z=1.23$ (ή μικρότερα του $z=-1.23$), το οποίο είναι $P=0.22$

Δηλαδή, το $z=1.23$ είναι μία τυχαία τιμή και συνεπώς δεν είναι σημαντική, δηλ. δεν είναι διαφορετική από το μηδέν με πιθανότητα λάθους μεγάλη, μεγαλύτερη από το 5% ($P \geq 0.05$), $P=0.22$

Τυπική
Κανονική
κατανομή



Διάστημα εμπιστοσύνης για ένα ποσοστό

Το 95% δ.ε. ενός παρατηρούμενου ποσοστού δίνεται
από:

(παρατηρούμενο ποσοστό) \pm 1.96*(τυπικό σφάλμα)

95% δ.ε.

(παρατηρούμενο ποσοστό) \pm 1.96*(τυπικό σφάλμα)

(0.18–1.96*0.024, 0.18+1.96*0.024) ή

(0.13, 0.23)

Αν υποθέσουμε ότι το δείγμα των γυναικών από το νοσοκομείο είναι αντιπροσωπευτικό όλων των γυναικών στη χώρα τότε μπορούμε να ισχυριστούμε με βεβαιότητα 95% ότι ο επιπολασμός του άσθματος στη χώρα είναι μεταξύ 13 και 23%

Έλεγχος ποσοστού με διόρθωση

Για σχετικά μικρά δείγματα το παρακάτω z-test δίνει πιο ακριβή αποτελέσματα:

$$z_c = \frac{|\mathbf{p} - \pi| - \frac{1}{2n}}{\mathbf{SE}}$$

Στο παράδειγμά μας, το $z_c = 1.14$ είναι λίγο μικρότερο από το $z = 1.23$ (που βρήκαμε χωρίς τη διόρθωση) επειδή το μέγεθος του δείγματος είναι μεγάλο

Έλεγχος για τη σύγκριση δύο ποσοστών

Έλεγχος για τη σύγκριση δύο ποσοστών

Για τον έλεγχο της σύγκρισης δύο ποσοστών χρησιμοποιείται το z-test

Παράδειγμα: Είκοσι ($r_1=20$) από τα 240 ($n_1=240$) άτομα που εμβολιάσθηκαν με πραγματικό εμβόλιο προσβλήθηκαν από γρίπη σε σύγκριση με τα 80 ($r_2=80$) από τα 220 ($n_2=220$) που εμβολιάσθηκαν με placebo. Υπάρχει ένδειξη ότι το εμβόλιο ήταν αποτελεσματικό;

Τότε, για να ελέγξουμε αν το ποσοστό προσβολής από γρίπη όταν γίνεται εμβολιασμός δεν είναι διαφορετικό από το ποσοστό με το placebo, χρησιμοποιείται το z-test:

$$z = \frac{(\text{ποσοστό γρίπης με εμβολιασμό}) - (\text{ποσοστό γρίπης με placebo})}{\text{τυπικό σφάλμα}} = \frac{p_1 - p_2}{SE}$$

$$\text{όπου } p_1 = \frac{r_1}{n_1}, \quad p_2 = \frac{r_2}{n_2},$$

$$SE = \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad \text{και} \quad p = \frac{r_1 + r_2}{n_1 + n_2}$$

Τότε $p_1=20/240=0.083$, $p_2=80/220=0.364$

Το **συνολικό** ποσοστό με γρίπη είναι: $p=(20+80)/(240+220) = 0.217$

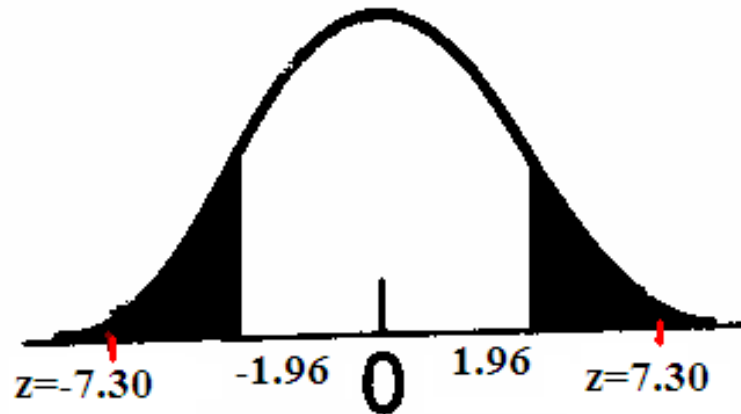
$$SE = \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0.217(1-0.217) \left(\frac{1}{240} + \frac{1}{220} \right)} = 0.038$$

οπότε

$$z = \frac{p_1 - p_2}{SE} = \frac{0.083 - 0.364}{0.038} = -7.30 \text{ (το πρόσημο αγνοείται)}$$

$$z = \frac{p_1 - p_2}{SE} = \frac{0.083 - 0.364}{0.038} = -7.30 \text{ (το πρόσημο αγνοείται)}$$

Αν δεν υπάρχει διαφορά μεταξύ των δύο ποσοστών τότε η ποσότητα της z είναι μία τυπική τιμή της τυπικής κανονικής κατανομής.



- Η τιμή $z=7.30$ είναι μεγαλύτερη από το 5% σημείο της τυπικής κανονικής κατανομής που είναι 1.96.
- **Συνεπώς, υπάρχει ένδειξη ($P < 0.05$) ότι ο πραγματικός εμβολιασμός ελαττώνει τον κίνδυνο προσβολής από γρίπη.**

Διάστημα εμπιστοσύνης για τη διαφορά δύο ποσοστών

Το 95% δ.ε. για τη διαφορά μεταξύ των δύο ποσοστών δίνεται προσεγγιστικά από:

(διαφορά δύο ποσοστών) \pm 1.96(τυπικό σφάλμα)

$((p_1-p_2)-1.96 \times SE, (p_1-p_2)+1.96 \times SE)$

$$\text{όπου } SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Στο παράδειγμά μας, το 95% δ.ε. αυτής της ελάττωσης του ποσοστού των ατόμων με γρίπη λόγω του εμβολίου είναι:

$$(-0.281 - 1.96 * 0.037, -0.281 + 1.96 * 0.037) \text{ ή}$$

$$(-0.354, -0.208)$$

Συνεπώς, η πιθανότητα προσβολής από γρίπη όταν γίνεται πραγματικός εμβολιασμός είναι χαμηλότερος μεταξύ 20.8% και 35.4% σε σχέση με τη χορήγηση placebo