ΑΣΚΗΣΗ 1: Θέλουμε να εξετάσουμε κατά πόσο ένα νέο πρόγραμμα ασκήσεων είναι αποτελεσματικό στην αντιμετώπιση χρόνιου πόνου στη μέση. Για το λόγο αυτό μοιράστηκε σε 10 ασθενείς που υποφέρουν από πόνο στη μέση ερωτηματολόγιο από το οποίο προέκυψε ο βαθμός στη κλίμακα πόνου, (0 έως 10). Το ίδιο ερωτηματολόγιο απαντήθηκε μετά το τέλος των συνεδριών. Τα απότελέσματα φαίνονται στον πίνακα:

ΠΡΙΝ ΤΟ ΠΡΟΓΡΑΜΜΑ	ΜΕΤΑ ΤΟ ΠΡΟΓΡΑΜΜΑ
9	7
10	5
5	3
8	5
8	4
6	3
9	4
7	5
9	6
9	4

Ν' απαντηθεί το ερώτημα σε επίπεδο σημαντικότητας α=0,05

Θα κάνουμε Έλεγχο ισότητας μέσων τιμών για ζευγαρωτές τιμές (Paired Samples t-test)

- Οι υποθέσεις του ελέγχου είναι:

- Μηδενική υπόθεση: **Ho:** $\boldsymbol{\mu_1 = \mu_2 \quad (\mu_z = 0)}$
- Εναλλακτική υπόθεση: **H1:** $\boldsymbol{\mu_1 > \mu_2 \quad (\mu_z > 0)}$

- Ορίζουμε την μεταβλητή: $\boldsymbol{z_i = x_i - y_i}$

Στον παρακάτω πίνακα υπολογίζουμε τις διαφορές:

X	Y	Z	Z^2
9	7	2	4
10	5	5	25
5	3	2	4
8	5	3	9
8	4	4	16
6	3	3	9
9	4	5	25
7	5	2	4
9	6	3	9
9	4	5	25
		ΣZ_i=34	ΣZ_i^2=130

- $\bar{z} = \frac{\sum_{i=4} z_i}{n} = \frac{34}{10} = 3,4$
- $s_z^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} z_i^2 - \frac{\left(\sum_{i=1} z_i\right)^2}{n}\right) = \frac{1}{10-1}\left(130 - \frac{34^2}{10}\right) = 1,6$
- Άρα η τυπική απόκλιση είναι: $s_z = \sqrt{1,6} = 1,265$
- Το στατιστικό κριτήριο είναι: $\frac{\bar{z}}{s_z}\sqrt{n} = \frac{3,4}{1,265}\sqrt{10}$ =8,499
- Η κρίσιμη τιμή είναι η: $t_{n-1;\alpha} = t_{9;0,05}$ =1,833

Η κρίσιμη περιοχή είναι: $R = \{\frac{\bar{z}}{s_z}\sqrt{n} > t_{n-1;a}\}$

Είναι 8,499= $\frac{\bar{z}}{s_z}\sqrt{n} > t_{n-1;a}$=1,833

Άρα είμαστε στην κρίσιμη περιοχή, οπότε **απορρίπτουμε την Ho.**

Άρα οι συνεδρίες ήταν αποτελεσματικές

Παράδειγμα 2

Ο παρακάτω πίνακας παρουσιάζει την κατανομή 80 καρκινοπαθών (με καρκίνο του μαστού) και 160 “υγειών” κατά οικογενειακή κατάσταση

	Οικογενειακή κατάσταση		
Καρκίνος του μαστού	**Ανύπαντρες**	**Παντρεμένες**	**Άλλα**
Ναι	30	44	6
Όχι	30	106	24

Σχετίζεται ο καρκίνος του μαστού με την οικογενειακή κατάσταση; (α=0,05)

- Στο παράδειγμα αυτό θα ελέγξουμε αν ο καρκίνος του μαστού(Α) σχετίζεται με την οικογενειακή κατάσταση(Β).
- Επειδή και οι 2 μεταβλητές είναι ποιοτικές, θα χρησιμοποιήσουμε τον έλεγχο X_2.
- Για τον έλεγχο αυτό έχουμε ότι:

H_0: Οι μεταβλητές Α και Β είναι ανεξάρτητες

H_1: Οι μεταβλητές Α και Β δεν είναι ανεξάρτητες

Στην επόμενη διαφάνεια υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης.

Υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης:

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	**Ανύπαντρες**	**Παντρεμένες**	**Άλλα**	**ΣΥΝΟΛΟ**
Ναι	30	44	6	80
Όχι	30	106	24	160
ΣΥΝΟΛΟ	60	150	30	240

- Οι εκτιμώμενες συχνότητες είναι:

- $\mathrm{E}_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{80 \cdot 60}{240} = 20$

- $\mathrm{E}_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{80 \cdot 150}{240} = 50$

- $\mathrm{E}_{13} = \frac{n_{1.} \times n_{.3}}{n} = \frac{80 \cdot 30}{240} = 10$

- $\mathrm{E}_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{160 \cdot 60}{240} = 40$

- $\mathrm{E}_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{160 \cdot 150}{240} = 100$

- $\mathrm{E}_{23} = \frac{n_{2.} \times n_{.3}}{n} = \frac{160 \cdot 30}{240} = 20$

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα	ΣΥΝΟΛΟ
Ναι	30	44	6	80
Όχι	30	106	24	160
ΣΥΝΟΛΟ	60	150	30	240

Παρατηρούμενες συχνότητες

	Οικογενειακή κατάσταση			
Καρκίνος του μαστού	Ανύπαντρες	Παντρεμένες	Άλλα	ΣΥΝΟΛΟ
Ναι	20	50	10	80
Όχι	40	100	20	160
ΣΥΝΟΛΟ	60	150	30	240

Αναμενόμενες συχνότητες

Το στατιστικό κριτήριο είναι: $X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$

$$\frac{(30-20)^2}{20} + \frac{(44-50)^2}{50} + \frac{(6-10)^2}{10} + \frac{(30-40)^2}{40} + \frac{(106-100)^2}{100} + \frac{(24-20)^2}{20} = 10{,}98$$

Η κρίσιμη τιμή, απο πίνακες είναι:

$X^2{}_{(r-1)*(s-1);a} = X^2{}_{(2-1)*(3-1);0.05} = X^2{}_{2;0.05} = 5{,}99$

Η κρίσιμη περιοχή: $R = \{X^2 > X^2{}_{(r-1)(s-1);a}\}$

Άρα, είμαστε στην κρίσιμη περιοχή, αφού 10,98>5,99

Επομένως απορρίπτουμε την μηδενική υπόθεση και δεχόμαστε ότι ο καρκίνος του μαστού σχετίζεται με την οικογενειακή κατάσταση.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

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 1η και ένας βαθμός στη 2η 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

(συν.)

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

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

- Η διασπορά είναι:

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

- 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: $t_{n-1;\frac{a}{2}} = t_{19;0{,}025}$
- Την τιμή αυτή θα την βρούμε από τον Πίνακα της κατανομής t, που δίνεται στην επόμενη διαφάνεια.
- Βλέπουμε ότι $t_{19;0{,}025}$=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= $\left\{\left|\frac{\bar{z}}{s_z}\sqrt{n}\right| > t_{n-1;\frac{a}{2}}\right\}$.
- Προφανώς 3,23 = $\left|\frac{\bar{z}}{s_z}\sqrt{n}\right| > t_{n-1\,;\,\frac{a}{2}}$ =2,093
- Άρα, βρισκόμαστε στην κρίσιμη περιοχή R.
- <u>Οπότε, απορρίπτουμε την H_0 και αποδεχόμαστε ότι ισχύει η H_A.</u>
- Άρα συμπεραίνουμε ότι ο μέσος όρος της βαθμολογίας της 1ης προόδου διαφέρει στατιστικά σημαντικά από το μέσο όρο της βαθμολογίας της 2ης προόδου, στον πληθυσμό αναφοράς.