ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Παναγιώτα Λάλου, Αλέξανδρος Γρυπάρης

Παράδειγμα:

Η αρτηριακή πίεση 10 ασθενών πρίν (X) και μετά (Y) τη χορήγηση φαρμάκου κατά της πίεσης είναι:

X	13	15	18	14	12	13	15	16	18	19
Y	12	13	15	15	14	13	13	14	14	13

Να ελεγχθεί σε επίπεδο σημαντικότητας 5%, εαν το συγκεκριμένο φάρμακο είναι αποτελεσματικό κατά της πίεσης.

- Ελαττώνει το φάρμακο αυτό την πίεση στους ασθενείς;

ΛΥΣΗ

- Οι υποθέσεις του ελέγχου είναι:

- Μηδενική υπόθεση: $\mathbf{H_0}$: $\boldsymbol{\mu_1 = \mu_2}$ ($\boldsymbol{\mu_z = 0}$)
- Εναλλακτική υπόθεση: $\mathbf{H_1}$: $\boldsymbol{\mu_1 > \mu_2}$ ($\boldsymbol{\mu_z > 0}$)

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

Στον παρακάτω πίνακα υπολογίζουμε τις διαφορές:

x_i	y_i	$z_i = x_i - y_i$	z_i^2
13	12	1	1
15	13	2	4
18	15	3	9
14	15	-1	1
12	14	-2	4
13	13	0	0
15	13	2	4
16	14	2	4
18	14	4	16
19	13	6	36
		Σz_i=17	Σx_i^2=79

Έλεγχος υποθέσεων–Διάστημα Εμπιστοσύνης

Έλεγχοι υποθέσεων και δ.ε. για παρατηρήσεις κατά ζεύγη:

$H_0: \mu_1=\mu_2$ $H_1: \mu1>\mu2$	$H_0: \mu_1=\mu_2$ $H_1: \mu1<\mu2$	$H_0: \mu_1=\mu_2$ $H_1: \mu1\neq\mu2$
$R=\{\frac{\bar{z}}{s_z}\sqrt{n}>t_{n-1;a}\}$	$R=\{\frac{\bar{z}}{s_z}\sqrt{n}<-t_{n-1;a}\}$	$R=\{\left\lvert\frac{\bar{z}}{s_z}\sqrt{n}\right\rvert>t_{n-1;\frac{a}{2}}\}$

Για τον υπολογισμό του Δ.Ε. της διαφοράς των μέσων τιμών, έχουμε:

$$(\bar{z}-\frac{s_z}{\sqrt{n}}t_{n-1;\frac{a}{2}}, \bar{z}+\frac{s_z}{\sqrt{n}}t_{n-1;\frac{a}{2}}), \text{όπου } z=x_i-y_i .$$

Θα χρειαστεί να υπολογίσουμε την μέση τιμή $\bar{z}$ και την διασπορά s_z^2:

$$\bar{z} = \frac{\sum_{i=10} z_i}{n} = \frac{17}{10} = 1{,}7$$

$$s_z^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} z_i^2 - \frac{\left(\sum_{i=1} z_i\right)^2}{n}\right) = \frac{1}{10-1}\left(79 - \frac{17^2}{10}\right) = 5{,}57$$

Άρα η τυπική απόκλιση είναι: $s_z = \sqrt{5{,}57} = 2{,}36$

Το στατιστικό κριτήριο είναι: $\frac{\bar{z}}{s_z}\sqrt{n} = \frac{1,7}{2,36}\sqrt{10} = 2,278$

Η κρίσιμη τιμή (όπως φαίνεται από τον πίνακα της επόμενης σελίδας) 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 2,278>1,833 δηλαδή $\frac{\bar{z}}{s_z}\sqrt{n} > t_{n-1;a}$ οπότε είμαστε στην κρίσιμη περιοχή κι επομένως **απορρίπτουμε την Ho.**

Άρα, συμπεραίνουμε ότι η μέση της πίεσης μετά τη χρήση του φαρμάκου είναι μικρότερη από τη μέση τιμή της πίεσης πριν τη χρήση.

Δηλαδή η δράση του φαρμάκου είναι αποτελεσματική κατά της πίεσης.

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

Παράδειγμα 2:

- Σε τέσσερα άτομα με αυξημένες τιμές των τριγλυκεριδίων του ορού, χορηγήθηκε για ένα μήνα φάρμακο που πιστεύεται ότι ελαττώνει τα επίπεδα των τριγλυκεριδίων. Οι τιμές των τριγλυκεριδίων στα τέσσερα αυτά άτομα πριν και μετά την χορήγηση του φαρμάκου

Πριν τη χορήγηση	Μετά την χορήγηση
180	120
200	220
240	130
230	160

- Ελαττώνει τα επίπεδα των τρυγλυκεριδίων το φάρμακο αυτό; (α=0,05)

ΛΥΣΗ

- 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_i	y_i	$z_i = x_i - y_i$	z_i^2
180	120	60	3600
200	220	-20	400
240	130	110	12100
230	160	70	4900
		Σx_i=220	Σx_i^2=21000

$$\bar{z} = \frac{\sum_{i=4} z_i}{n} = \frac{220}{4} = 55$$

$$s_z^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} z_i^2 - \frac{\left(\sum_{i=1} z_i\right)^2}{n}\right) = \frac{1}{4-1}\left(21000 - \frac{220^2}{4}\right) = 2966{,}67$$

Άρα η τυπική απόκλιση είναι: $s_z = \sqrt{2966{,}67} = 54{,}4671$

Έλεγχος υποθέσεων–Διάστημα Εμπιστοσύνης

Έλεγχοι υποθέσεων και δ.ε. για παρατηρήσεις κατά ζεύγη:

H_0: $\mu_1=\mu_2$ H_1: μ1>μ2	H_0: $\mu_1=\mu_2$ H_1: μ1<μ2	H_0: $\mu_1=\mu_2$ H_1: μ1≠μ2
$R=\{\frac{\bar{z}}{s_z}\sqrt{n}>t_{n-1;a}\}$	$R=\{\frac{\bar{z}}{s_z}\sqrt{n}<-t_{n-1;a}\}$	$R=\{\left\lvert\frac{\bar{z}}{s_z}\sqrt{n}\right\rvert>t_{n-1;\frac{a}{2}}\}$

Για τον υπολογισμό του Δ.Ε. της διαφοράς των μέσων τιμών, έχουμε:

$$(\bar{z}-\frac{s_z}{\sqrt{n}}t_{n-1;\frac{a}{2}}, \bar{z}+\frac{s_z}{\sqrt{n}}t_{n-1;\frac{a}{2}}), \text{ όπου } z=x_i-y_i\,.$$

Το στατιστικό κριτήριο είναι: $\frac{\bar{z}}{s_z}\sqrt{n} = \frac{55}{54,4671}\sqrt{4} = 2,019$

Η κρίσιμη τιμή είναι η: $t_{n-1;\alpha} = t_{3;0,05} = 2,353$

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

Είναι 2,019<2,353 δηλαδή $\frac{\bar{z}}{s_z}\sqrt{n} < t_{n-1;a}$ οπότε δεν είμαστε στην κρίσιμη περιοχή κι επομένως **δεν απορρίπτουμε την Ho.**

Δηλαδή το φάρμακο δεν ελαττώνει τα επίπεδα των τριγλυκεριδίων.

Έλεγχος ανεξαρτησίας 2 ποιοτικών χαρακτηριστικών

Παράδειγμα

- Έστω ότι θέλουμε να ελέγξουμε αν:
  - η παρουσία καρκίνου εξαρτάται από το φύλο, ή αν
  - η παρουσία καπνίσματος εξαρτάται από το μορφωτικό επίπεδο: καθόλου – δημοτικό – γυμνάσιο - λύκειο – ΑΕΙ/ΤΕΙ

- Μπορούμε να χρησιμοποιήσουμε το t-test;

- Πότε χρησιμοποιούμε το t-test;

Έλεγχος ανεξαρτησίας 2 ποιοτικών χαρακτηριστικών

- Στα προηγούμενα παραδείγματα δεν έχει νόημα η μέση τιμή
- Στις περιπτώσεις αυτές ο στατιστικός έλεγχος γίνεται με σύγκριση των συχνοτήτων των παρατηρήσεων στις διάφορες κατηγορίες
- Διαξονική ταξινόμηση (two-way classification) ή ταξινόμηση δύο διευθύνσεων: η ταυτόχρονη ταξινόμηση των δεδομένων ως προς δύο χαρακτηριστικά (μεταβλητές).

Διαξονική ταξινόμηση

- Την χρησιμοποιούμε όταν θέλουμε να μελετήσουμε την ύπαρξη συσχέτισης μεταξύ 2 ποιοτικών χαρακτηριστικών
- Παράδειγμα: Κατανομή 300 ατόμων σύμφωνα με το κάπνισμα και την ύπαρξη ή όχι καρκίνου του εντέρου

	Κάπνισμα		
Καρκίνος εντέρου	Ναί	Όχι	Σύνολο
Ναί	50	50	100
Όχι	130	70	200
Σύνολο	180	120	300

Έλεγχος X^2

- Ο έλεγχος αυτός πραγματοποιείται όταν έχουμε 2 ποιοτικές μεταβλητές

- Η υποθέσεις υπό έλεγχο είναι οι:

  H_0: Δεν υπάρχει σχέση μεταξύ των 2 μεταβλητών μας, δηλαδή οι μεταβλητές μας είναι ανεξάρτητες

  H_1: Υπάρχει σχέση μεταξύ των 2 μεταβλητών μας, δηλαδή οι μεταβλητές μας δεν είναι ανεξάρτητες

- Έτσι, στο προηγούμενο παράδειγμα, η μηδενική υπόθεση λέει ότι δεν υπάρχει σχέση μεταξύ καπνίσματος και καρκίνου του εντέρου

(συν.)

- Υποθέστε ότι μπορούμε να κατασκευάσουμε ένα πίνακα που να αντιστοιχεί στο προηγούμενο παράδειγμά μας, και να εκφράζει τη μηδενική υπόθεση

- Δηλαδή, τον πίνακα που εκφράζει την μηδενική υπόθεση, με βάση τα δεδομένα μας
  - Δηλαδή, τον πίνακα με τις συχνότητες που θα είχαμε σε κάθε κελί όταν δεν υπάρχει σχέση μεταξύ καπνίσματος και καρκίνου του εντέρου

(συν.)

- Έτσι, θα έχουμε 2 διαφορετικούς πίνακες!
  - Αφενός τον πίνακα που πήραμε από το δείγμα μας (και είδαμε προηγουμένως). Τον πίνακα δηλαδή με τις **παρατηρούμενες** τιμές:

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	50	50	100
Όχι	130	70	200
Σύνολο	180	120	300

(συν.)

- … και έναν άλλο πίνακα που ισχύει όταν ισχύει η μηδενική υπόθεση. Αυτός είναι ο πίνακας με τις αναμενόμενες τιμές:

(θα δούμε στη συνέχεια πως τον κατασκευάζουμε)

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	60	40	100
Όχι	120	80	200
Σύνολο	180	120	300

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	50	50	100
Όχι	130	70	200
Σύνολο	180	120	300

Παρατηρούμενες συχνότητες

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	60	40	100
Όχι	120	80	200
Σύνολο	180	120	300

Αναμενόμενες συχνότητες

(συν.)

- Παρατηρήστε ότι ο <u>συνολικός</u> αριθμός παρατηρήσεων ανά γραμμή και ανά στήλη είναι ίδιος, και στους 2 πίνακες

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	50	50	100
Όχι	130	70	200
Σύνολο	180	120	300

	Κάπνισμα		Σύνολο
Καρκίνος εντ.	Ναί	Όχι	
Ναί	60	40	100
Όχι	120	80	200
Σύνολο	180	120	300

(συν.)

- Όσο πιο «κοντά» είναι ο πίνακας με τα στοιχεία από το δείγμα μας στον πίνακα της μηδενικής υπόθεσης, τόσο πιο βέβαιοι είμαστε ότι ισχύει η μηδενική υπόθεση

- Όσο πιο «μακριά» είναι ο πίνακας με τα στοιχεία από το δείγμα μας στον πίνακα της μηδενικής υπόθεσης, τόσο πιο βέβαιοι είμαστε ότι ΔΕΝ ισχύει η μηδενική υπόθεση

- Ο έλεγχος X^2 βασίζεται στην απόσταση των παρατηρούμενων τιμών από τις αναμενόμενες τιμές.
  - Οι παρακάτω διαφάνειες παρουσιάζουν τη γενική μορφή του ελέγχου

Έλεγχος Ανεξαρτησίας δυο κατηγορικών μεταβλητών: Έλεγχος X^2

- Με τον έλεγχο αυτό, ελέγχουμε αν δύο κατηγορικές μεταβλητές είναι ανεξάρτητες ή όχι, δηλαδή αν επηρεάζει η μία την άλλη.
  - Αν βρεθεί εξάρτηση δεν είμαστε σε θέση να πούμε ποια είναι η αιτία και ποιο το αποτέλεσμα.
- Εστω ότι ένας πληθυσμός εξετάζεται ως προς 2 χαρακτηριστικά (μεταβλητές) Α και Β.

A \ B	**B1**	**B2**	**. . .**	**Bs**	**Σύνολο**
A1	n_{11}	n_{12}	…	n_{1s}	$n_{1.}$
A2	n_{21}	n_{22}	…	n_{2s}	$n_{2.}$
⋮					
Ar	n_{r1}	n_{r2}	…	n_{rs}	$n_{r.}$
Σύνολο	$n_{.1}$	$n_{.2}$	…	$n_{.s}$	n

H_0: Οι μεταβλητές A και B είναι ανεξάρτητες
H_1: Οι μεταβλητές A και B δεν είναι ανεξάρτητες

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

Στατιστικό κριτήριο: $\boldsymbol{X^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}}$

όπου: $\boldsymbol{O_{ij} = n_{ij}}$ οι παρατηρούμενες τιμές (*Observed*)

$\boldsymbol{E_{ij} = \frac{n_{i.}\ n_{.j}}{n}}$ οι αναμενόμενες τιμές (*Expected*)

Ή αλλιώς:

$$\boldsymbol{E_{ij} = \frac{Aθροισμα\ i\ γραμμης \times Άθροισμα\ \ j\ στηλης}{n}}$$

Στην **κρίσιμη τιμή** $X^2{}_{(r-1)*(s-1);a}$ **r: αριθμός γραμμών**

s: αριθμός στηλών

Προϋποθέσεις εφαρμογής του ελέγχου:

1. Τα δείγματα είναι τυχαία.

2. Οι παρατηρήσεις είναι ανεξάρτητες.

3. Όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 1.

4. Το πολύ 20% από τις αναμενόμενες συχνότητες είναι μικρότερες από 5.

- Δηλαδή, πάνω από το 80% των αναμενόμενων συχνοτήτων είναι μεγαλύτερες του 5.

Παράδειγμα 1

Μια έρευνα σε δείγμα 500 ατόμων έδειξε οτι οι 200 ήταν καπνιστές. Μετά από ειδικό φόρο που επιβλήθηκε, σε καινούρια έρευνα, σε δείγμα 600 ατόμων, οι 210 ήταν καπνιστές. Τα αποτελέσματα αυτά φαίνονται στο παρακάτω πίνακα:

	Κάπνισμα	
Φορολογία	Καπνιστές	Μη καπνιστές
Πριν το φόρο	200	300
Μετά τον φόρο	210	390

Επέδρασε η φορολογία στη χρήση του τσιγάρου; (α=0,05)

- Αν Α: Κάπνισμα (καπνιστές – μη καπνιστές)

  Β: Φορολογία (πριν το φόρο – μετά τον φόρο)

H_0: Οι μεταβλητές Α και Β είναι ανεξάρτητες

H_1: Οι μεταβλητές Α και Β δεν είναι ανεξάρτητες

Υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης:

	Κάπνισμα		
Φορολογία	Καπνιστές	Μη καπνιστές	ΣΥΝΟΛΟ
Πριν το φόρο	200	300	500
Μετά τον φόρο	210	390	600
ΣΥΝΟΛΟ	410	690	1100

- Οι εκτιμώμενες συχνότητες:

- $E_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{500 \cdot 410}{1100} = 186{,}36$

- $E_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{500 \cdot 690}{1100} = 313{,}64$

- $E_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{600 \cdot 410}{1100} = 223{,}64$

- $E_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{600 \cdot 690}{1100} = 376{,}36$

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

$$\frac{(200-186{,}36)^2}{186{,}36} + \frac{(300-313{,}64)^2}{313{,}64} + \frac{(210-223{,}64)^2}{223{,}64} + \frac{(390-376{,}36)^2}{376{,}36} = \cdots = 2{,}92$$

Η κρίσιμη τιμή, απο πίνακες είναι:

$X^2{}_{(r-1)*(s-1);a} = X^2{}_{(2-1)*(2-1);0.05} = X^2{}_{1;0.05} = 3{,}84$

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

Όμως, είναι: $2{,}92 = X^2 < X^2{}_{1;0.05} = 3{,}84$

Άρα, δεν είμαστε στην κρίσιμη περιοχή, οπότε δεν απορρίπτουμε την H_0.

Επομένως οι μεταβλητές είναι ανεξάρτητες, δηλαδή ο φόρος δεν επιδρά στην κατανάλωση τσιγάρου

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 2

Σε 500 μαθητές δημοτικού σχολείου μελετήθηκε η σχέση της υγείας του στόματος τους με τη χλωρίωση του νερού στην περιοχή διαμονής τους. Η κατανομή των 500 μαθητών ανάλογα με την υγεία του στόματος και τη χλωρίωση του νερού ήταν:

	Υγιεινή στόματος		
Χλωρίωση νερού	**Κακή**	**Μέτρια**	**Καλή**
Ανεπαρκής	80	120	75
Επαρκής	40	80	105

Σχετίζεται η υγιεινή του στόματος των μαθητών με τη χλωρίωση του νερού;

- Στο παράδειγμα αυτό θα ελέγξουμε αν η υγιεινή του στόματος (A) σχετίζεται με τη χλωρίωση (B).
- Επειδή και οι 2 μεταβλητές είναι ποιοτικές, θα χρησιμοποιήσουμε τον έλεγχο X_2.
- Για τον έλεγχο αυτό έχουμε ότι:

H_0: Οι μεταβλητές A και B είναι ανεξάρτητες

H_1: Οι μεταβλητές A και B δεν είναι ανεξάρτητες

Στην επόμενη διαφάνεια υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης.

Υπολογίζουμε το άθροισμα κάθε γραμμής και στήλης:

	Υγιεινή στόματος			
Χλωρίωση νερού	Κακή	Μέτρια	Καλή	ΣΥΝΟΛΟ
Ανεπαρκής	80	120	75	275
Επαρκής	40	80	105	225
ΣΥΝΟΛΟ	120	200	180	500

- Οι εκτιμώμενες συχνότητες είναι:

- $E_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{275 \cdot 120}{500} = 66$

- $E_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{275 \cdot 200}{500} = 110$

- $E_{13} = \frac{n_{1.} \times n_{.3}}{n} = \frac{275 \cdot 180}{500} = 99$

- $E_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{225 \cdot 120}{500} = 54$

- $E_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{225 \cdot 200}{500} = 90$

- $E_{23} = \frac{n_{2.} \times n_{.3}}{n} = \frac{225 \cdot 180}{500} = 81$

$$X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$$

$$\frac{(80-66)^2}{66} + \frac{(120-110)^2}{110} + \frac{(75-95)^2}{95} + \frac{(40-54)^2}{54} + \frac{(80-90)^2}{90} + \frac{(105-81)^2}{81} = 21{,}55$$

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 21,55>5,99

Επομένως απορρίπτουμε την μηδενική υπόθεση και δεχόμαστε ότι η χλωρίωση του νερού επηρεάζει την κατάσταση του σώματος των μαθητών.

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 3

Στον παρακάτω πίνακα φαίνονται τα αποτελέσματα από μία έρευνα που μελετά τη σχέση της νόσου Αλτσχάιμερ με τη σωματική άσκηση.

	Επίπεδο σωματικής άσκησης		
Νόσος	**Αυξημένο**	**Μέτριο**	**Χαμηλό**
Ασθενείς	85	105	110
Υγιείς	125	120	100

Σχετίζεται η σωματική άσκηση (Α) με την νόσο Αλτσχάιμερ (Β);

- 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

	Επίπεδο σωματικής άσκησης			
Νόσος	**Αυξημένο**	**Μέτριο**	**Χαμηλό**	**ΣΥΝΟΛΟ**
Ασθενείς	85	105	110	300
Υγιείς	125	120	100	345
ΣΥΝΟΛΟ	210	225	220	645

- $\mathrm{E}_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{300 \cdot 210}{645} = 97{,}674$
- $\mathrm{E}_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{300 \cdot 225}{645} = 104{,}651$
- $\mathrm{E}_{13} = \frac{n_{1.} \times n_{.3}}{n} = \frac{300 \cdot 220}{645} = 102{,}326$
- $\mathrm{E}_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{345 \cdot 210}{645} = 112{,}326$
- $\mathrm{E}_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{345 \cdot 225}{645} = 120{,}349$
- $\mathrm{E}_{23} = \frac{n_{2.} \times n_{.3}}{n} = \frac{345 \cdot 220}{645} = 117{,}674$

$$X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$$

$$\frac{(85-97{,}674)^2}{97{,}674} + \frac{(105-104{,}651)^2}{104{,}651} + \frac{(110-102{,}326)^2}{102{,}326} + \frac{(125-112{,}326)^2}{112{,}326} +$$

$$\frac{(120-120{,}349)^2}{120{,}349} + \frac{(100-117{,}674)^2}{117{,}674} = 4{,}991$$

Η κρίσιμη τιμή, απο πίνακες είναι:

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

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

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

Άρα, ΔΕΝ είμαστε στην κρίσιμη περιοχή, αφού
X^2=4,991 και $X^2{}_{2;0.05} = 5{,}99$

Επομένως δεν απορρίπτουμε την μηδενική υπόθεση, δηλαδή η σωματική άσκηση δεν σχετίζεται με τη νόσο Αλτσχάιμερ