

Στατιστική ανάλυση κατηγορικών δεδομένων Στατιστικό τεστ χ^2

Για να αναλύσουμε στατιστικά ποιοτικά ή κατηγορικά δεδομένα χρησιμοποιούμε τη στατιστική δοκιμασία χ^2 . Το στατιστικό τεστ χ^2 μπορεί να είναι καλής προσαρμογής ή ανεξαρτησίας ή ομοιογένειας.

Α) Δοκιμασία χ^2 καλής προσαρμογής

Τα πειραματικά δεδομένα είναι χωρισμένα σε κάποιες κατηγορίες και έχουμε στη διάθεσή μας τις συχνότητες κάθε κατηγορίας (π.χ. αριθμό ατόμων, ζώων, φυτών ανάλογα με το πείραμά μας, που ανήκουν σε κάποια κατηγορία). Με το στατιστικό τεστ χ^2 ελέγχουμε εάν τα δεδομένα βρίσκονται σε κάποια αναλογία, σε κάποιο ποσοστό ή γενικότερα εάν προσαρμόζονται σε κάποιο θεωρητικό μοντέλο ή κάποια θεωρητική κατανομή.

Παράδειγμα 1: Από τη διασταύρωση φυτών δύο τύπων παράγονται απόγονοι τριών τύπων Α, Β και Γ. Ένα θεωρητικό μοντέλο κληρονομικότητας υποστηρίζει ότι οι τρεις τύποι απογόνων βρίσκονται σε αναλογία 1:2:1. Σε ένα πείραμα με διασταυρώσεις φυτών δύο τύπων προέκυψαν 90 απόγονοι και πιο συγκεκριμένα 18 απόγονοι τύπου Α, 44 απόγονοι τύπου Β και 28 απόγονοι τύπου Γ. Σε επίπεδο σημαντικότητας 5%, μπορούμε να ισχυριστούμε ότι τα αποτελέσματα αυτά επαληθεύουν το γενετικό μοντέλο;

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Οι απόγονοι βρίσκονται σε αναλογία 1:2:1 (Αρχική ή μηδενική υπόθεση)

H_1 : Οι απόγονοι δεν βρίσκονται σε αναλογία 1:2:1 (Εναλλακτική υπόθεση)

Ως αρχική υπόθεση H_0 θέτουμε πάντοτε ότι τα πειραματικά ή δειγματοληπτικά δεδομένα βρίσκονται σε κάποια αναλογία, σε κάποιο ποσοστό ή γενικότερα εάν προσαρμόζονται σε κάποιο θεωρητικό μοντέλο ή κάποια θεωρητική κατανομή.

Στη συνέχεια σκεφτόμαστε ως εξής: Αν ίσχυε η αναλογία 1:2:1 του γενετικού μοντέλου θα αναμέναμε το $1/4$ των απογόνων να είναι τύπου Α, τα $2/4$ να είναι τύπου Β και το $1/4$ τύπου Γ. Επομένως στους 90 απογόνους του πειράματός μας θα αναμέναμε: $1/4 \cdot 90 = 22.5$ απογόνους τύπου Α, $2/4 \cdot 90 = 45$ απογόνους τύπου Β και $1/4 \cdot 90 = 22.5$ απογόνους τύπου Γ. Στον

παρακάτω πίνακα έχουμε τις παρατηρηθείσες και τις αναμενόμενες ή θεωρητικές συχνότητες $\theta_i = n \cdot p_i$, $i = 1, 2, \dots, k$:

	A	B	Γ	Άθροισμα
Παρατηρηθείσες συχνότητες π_i	18	44	28	90
Αναμενόμενες ή θεωρητικές συχνότητες $\theta_i = n \cdot p_i$	$1/4 \cdot 90 =$ 22.5	$2/4 \cdot 90 =$ 45	$1/4 \cdot 90$ $= 22.5$	90
$\pi_i - \theta_i$	-4.5	-1	5.5	
$(\pi_i - \theta_i)^2$	20.25	1	30.25	
$\frac{(\pi_i - \theta_i)^2}{\theta_i}$	0.9	0.02	1.34	2.26

Με τη δοκιμασία X^2 ελέγχουμε εάν υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στις παρατηρηθείσες και τις θεωρητικές συχνότητες. Εάν υπάρχει στατιστικά σημαντική διαφορά απορρίπτουμε την H_0 , διαφορετικά την δεχόμαστε. Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$X^2 = \sum_{i=1}^k \frac{(\pi_i - \theta_i)^2}{\theta_i} > \chi_{k-1, \alpha}^2 \quad \text{απορρίπτουμε την } H_0$$

όπου π_i οι παρατηρηθείσες συχνότητες, θ_i οι θεωρητικές και k ο αριθμός των κατηγοριών. Για να εφαρμόσουμε το τεστ X^2 , θα πρέπει για όλες τις θεωρητικές συχνότητες να ισχύει: $\theta_i \geq 5$, $i = 1, 2, \dots, k$.

Αντικαθιστούμε για να ελέγξουμε εάν ισχύει η ανισότητα του στατιστικού κριτηρίου:

$$X^2 = \frac{(18 - 22.5)^2}{22.5} + \frac{(44 - 45)^2}{45} + \frac{(28 - 22.5)^2}{22.5} = 2.26 < \chi_{k-1, \alpha}^2 = \chi_{3-1, 0.05}^2 = \chi_{2, 0.05}^2 = 5.99$$

Άρα σε επίπεδο σημαντικότητας 5%, δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς δεν μπορούμε να απορρίψουμε την H_0 . Επομένως τα αποτελέσματα αυτά επαληθεύουν το γενετικό μοντέλο.

B) Δοκιμασία X^2 ανεξαρτησίας ή ομοιογένειας

Πολύ συχνά σε πειραματικές ή δειγματοληπτικές έρευνες τα δεδομένα ταξινομούνται όχι μόνο ως προς ένα χαρακτηριστικό, αλλά και ως προς ένα δεύτερο και έτσι προκύπτουν δισδιάστατα δεδομένα που παρουσιάζονται σε δισδιάστατους πίνακες συχνότητας, τους πίνακες συνάφειας.

Παράδειγμα 2: Δύο εμβόλια (A και B) που δοκιμάζονται για κάποια ασθένεια που προσβάλλει τα πρόβατα, συγκρίθηκαν με ένα εμβόλιο που περιείχε αποσταγμένο νερό (Γ). Ο αριθμός των προβάτων που αρρώστησαν ή δεν αρρώστησαν από τη συγκεκριμένη ασθένεια ήταν:

	Εμβόλια		
	A	B	Γ
Αρρώστησαν	23	27	50
Δεν αρρώστησαν	147	153	100

Σε στάθμη σημαντικότητας 5% να εξετάσετε: **α)** Αν τα εμβόλια A, B, Γ έχουν την ίδια αποτελεσματικότητα. **β)** Αν τα εμβόλια A, B έχουν την ίδια αποτελεσματικότητα.

α) Πρόκειται για X^2 στατιστικό έλεγχο ομοιογένειας.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Τα εμβόλια έχουν την ίδια αποτελεσματικότητα

$$\text{ή } p_A = p_B = p_\Gamma \quad (\text{Υπάρχει ομοιογένεια ανάμεσα στις κατηγορίες})$$

H_1 : Τα εμβόλια δεν έχουν την ίδια αποτελεσματικότητα ή

$$\text{τουλάχιστον κάποιο } p_i \neq p_j, \quad i, j = A, B, \Gamma \quad (\text{Δεν υπάρχει ομοιογένεια στις κατηγορίες})$$

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_{ij} οι παρατηρηθείσες συχνότητες, θ_{ij} οι θεωρητικές, κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας. Για να εφαρμόσουμε το τεστ X^2 , θα πρέπει για όλες τις θεωρητικές συχνότητες να ισχύει: $\theta_i \geq 5, \forall (i, j)$.

Οι θεωρητικές συχνότητες υπολογίζονται από τη σχέση:

$$\theta_{ij} = \frac{(\text{άθροισμα } i \text{ γραμμής}) \times (\text{άθροισμα } j \text{ στήλης})}{\text{συνολικό άθροισμα}} \quad \forall (i, j).$$

Στη συνέχεια αφού υπολογίσουμε τα αθροίσματα των γραμμών και των στηλών του πίνακα συνάφειας υπολογίζουμε τις θεωρητικές συχνότητες:

$$\begin{aligned} \theta_{11} &= \frac{100 \cdot 170}{500} = 34 & \theta_{12} &= \frac{100 \cdot 180}{500} = 36 & \theta_{13} &= \frac{100 \cdot 150}{500} = 30 \\ \theta_{21} &= \frac{400 \cdot 170}{500} = 136 & \theta_{22} &= \frac{400 \cdot 180}{500} = 144 & \theta_{23} &= \frac{400 \cdot 150}{500} = 120 \end{aligned}$$

	Εμβόλια			Άθροισμα
	A	B	Γ	
Αρρώστησαν	23 (34)	27 (36)	50 (30)	100
Δεν αρρώστησαν	147 (136)	153 (144)	100 (120)	400
Άθροισμα	170	180	150	500

Με τη δοκιμασία χ^2 ελέγχουμε εάν υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στις παρατηρηθείσες και τις θεωρητικές συχνότητες και πιο συγκεκριμένα, όπως αναφέραμε παραπάνω, ελέγχουμε εάν ισχύει το στατιστικό κριτήριο: Εάν

$$\chi^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας.

Αντικαθιστώντας έχουμε:

$$\begin{aligned} \chi^2 &= \frac{(23 - 34)^2}{34} + \frac{(27 - 36)^2}{36} + \dots + \frac{(100 - 120)^2}{120} = 23.93 > \chi_{(2-1)(3-1), \alpha}^2 \\ &= \chi_{(2-1)(3-1), 0.05}^2 = \chi_{2, 0.05}^2 = 5.99 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς απορρίπτουμε την H_0 . Επομένως τα εμβόλια A, B, Γ δεν έχουν την ίδια αποτελεσματικότητα.

β) Πρόκειται για χ^2 στατιστικό έλεγχο ομοιογένειας, αλλά σε αυτήν την περίπτωση θα

συγκρίνουμε μόνο τα εμβόλια A και B.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Τα εμβόλια A και B έχουν την ίδια αποτελεσματικότητα ή $p_A = p_B$.

H_1 : Τα εμβόλια A και B δεν έχουν την ίδια αποτελεσματικότητα ή $p_A \neq p_B$

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_{ij} οι παρατηρηθείσες συχνότητες και θ_{ij} οι θεωρητικές. Θα πρέπει: $\theta_{ij} \geq 5, \forall (i, j)$.

Οι θεωρητικές συχνότητες υπολογίζονται από τη σχέση:

$$\theta_{ij} = \frac{(\text{άθροισμα } i \text{ γραμμής}) \times (\text{άθροισμα } j \text{ στήλης})}{\text{συνολικό άθροισμα}} \quad \forall (i, j).$$

Στη συνέχεια αφού υπολογίσουμε τα αθροίσματα των γραμμών και των στηλών του πίνακα συνάφειας υπολογίζουμε τις θεωρητικές συχνότητες:

$$\theta_{11} = \frac{50 \cdot 170}{350} = 24.3 \quad \theta_{12} = \frac{50 \cdot 180}{350} = 25.7$$

$$\theta_{21} = \frac{300 \cdot 170}{350} = 145.7 \quad \theta_{22} = \frac{300 \cdot 180}{350} = 154.3$$

	A	B	Άθροισμα
Αρρώστησαν	23 (24.3)	27 (25.7)	50
Δεν αρρώστησαν	147 (145.7)	153 (154.3)	300
Άθροισμα	170	180	350

Καθώς για όλες τις θεωρητικές συχνότητες ισχύουν οι προϋποθέσεις $\theta_{ij} \geq 5$, ελέγχουμε εάν ισχύει το στατιστικό κριτήριο. Αντικαθιστώντας έχουμε:

$$X^2 = \frac{(23 - 24.3)^2}{24.3} + \frac{(27 - 25.7)^2}{25.7} + \frac{(147 - 145.7)^2}{145.7} + \frac{(153 - 154.3)^2}{154.3} = 0.016 <$$

$$< \chi_{(\kappa-1)(\lambda-1),\alpha}^2 = \chi_{(2-1)(2-1),0.05}^2 = \chi_{1,0.05}^2 = 3.84$$

Άρα σε επίπεδο σημαντικότητας 5%, δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς δεν μπορούμε να απορρίψουμε την H_0 . Επομένως τα εμβόλια Α και Β δεν έχουν την ίδια αποτελεσματικότητα.

Λύσεις ασκήσεων από το φυλλάδιο 7 – Ασκήσεις στο στατιστικό τεστ X^2

1. Σύμφωνα με τη θεωρία του Mendel, αν διασταυρωθούν φυτά μπιζελιών με στρογγυλούς-κίτρινους σπόρους, με φυτά μπιζελιών με ρυτιδωμένους πράσινους σπόρους, θα δώσουν σπόρους στρογγυλούς-κίτρινους, ρυτιδωμένους-κίτρινους, στρογγυλούς-πράσινους και ρυτιδωμένους-πράσινους σε αναλογία 9:3:3:1. Σε ένα πείραμα παρατηρήθηκαν τα παρακάτω αποτελέσματα:

i) Στρογγυλοί-κίτρινοι 87 ii) Ρυτιδωμένοι-κίτρινοι 35

iii) Στρογγυλοί-πράσινοι 26 iv) Ρυτιδωμένοι-πράσινοι 12

Σε επίπεδο σημαντικότητας $\alpha=0.05$, συμφωνούν οι παρατηρήσεις αυτές με τη θεωρία του Mendel;

Πρόκειται για X^2 στατιστικό έλεγχο καλής προσαρμογής. Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Οι απόγονοι βρίσκονται σε αναλογία 9:3:3:1 ή
τα αποτελέσματα συμφωνούν με τη θεωρία του Mendel

H_1 : Οι απόγονοι δεν βρίσκονται σε αναλογία 9:3:3:1 ή
τα αποτελέσματα δεν συμφωνούν με τη θεωρία του Mendel

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$X^2 = \sum_{i=1}^{\kappa} \frac{(\pi_i - \theta_i)^2}{\theta_i} > \chi_{\kappa-1,\alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_i οι παρατηρηθείσες συχνότητες, θ_i οι θεωρητικές και κ ο αριθμός των κατηγοριών. Για να εφαρμόσουμε το τεστ X^2 , θα πρέπει για όλες τις θεωρητικές συχνότητες να ισχύει: $\theta_i \geq 5$, $i = 1, 2, 3, 4$.

Αν ίσχυε η αναλογία 9:3:3:1 της θεωρίας του Mendel θα αναμέναμε τα 9/16 των απογόνων να

είναι στρογγυλοί-κίτρινοι, τα 3/16 να είναι ρυτιδωμένοι-κίτρινοι, τα 3/16 να είναι στρογγυλοί-πράσινοι και το 1/16 να είναι ρυτιδωμένοι-πράσινοι. Επομένως στους 160 απογόνους του πειράματός μας θα αναμέναμε: $9/16 \cdot 160 = 90$ απογόνους στρογγυλούς-κίτρινους, $3/16 \cdot 160 = 30$ απογόνους ρυτιδωμένους-κίτρινους, $3/16 \cdot 160 = 30$ απογόνους στρογγυλούς-πράσινους και $1/16 \cdot 160 = 10$ απογόνους ρυτιδωμένους-πράσινους. Στον παρακάτω πίνακα έχουμε τις παρατηρηθείσες και τις αναμενόμενες συχνότητες:

	Σ - Κ	P - Κ	Σ - Π	P - Π	Άθροισμα
Παρατηρηθείσες συχνότητες π_i	87	35	26	12	160
Αναμενόμενες ή θεωρητικές συχνότητες $\theta_i = n \cdot p_i$	$9/16 \cdot 160$ = 90	$3/16 \cdot 160$ = 30	$3/16 \cdot 160 = 30$	$1/16 \cdot 160$ = 10	160
$\pi_i - \theta_i$	-3	5	-4	2	
$(\pi_i - \theta_i)^2$	9	25	16	4	
$\frac{(\pi_i - \theta_i)^2}{\theta_i}$	0.1	0.83	0.53	0.4	1.86

Αντικαθιστούμε τις παρατηρηθείσες και τις αναμενόμενες συχνότητες και ελέγχουμε εάν ισχύει η ανισότητα του στατιστικού κριτηρίου:

$$\begin{aligned} \chi^2 &= \frac{(87 - 90)^2}{90} + \frac{(35 - 30)^2}{30} + \frac{(26 - 30)^2}{30} + \frac{(12 - 10)^2}{10} = 1.86 < \chi_{k-1, \alpha}^2 = \chi_{4-1, 0.05}^2 \\ &= \chi_{3, 0.05}^2 = 7.81 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς δεν μπορούμε να απορρίψουμε την H_0 . Επομένως τα αποτελέσματα αυτά επαληθεύουν τη θεωρία του Mendel.

2. Σε 600 πρόβατα μιας περιοχής βρέθηκε ότι τα 210 ήταν της φυλής Α, τα 220 της φυλής Β και τα υπόλοιπα της φυλής Γ. **α)** Σε επίπεδο σημαντικότητας $\alpha=0.05$, μπορούμε να ισχυριστούμε ότι οι τρεις φυλές Α, Β, Γ της περιοχής βρίσκονται στην ίδια αναλογία; **β)** Σε επίπεδο σημαντικότητας $\alpha=0.05$, μπορούμε να ισχυριστούμε ότι τα πραγματικά ποσοστά των τριών φυλών Α, Β και Γ είναι 35%, 35% και 30% αντίστοιχα;

α) Πρόκειται για χ^2 στατιστικό έλεγχο καλής προσαρμογής.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Οι τρεις φυλές A, B, Γ της περιοχής βρίσκονται στην ίδια αναλογία (1:1:1)

H_1 : Οι τρεις φυλές A, B, Γ της περιοχής δεν βρίσκονται στην ίδια αναλογία (1:1:1)

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$\chi^2 = \sum_{i=1}^k \frac{(\pi_i - \theta_i)^2}{\theta_i} > \chi_{k-1, \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_i οι παρατηρηθείσες συχνότητες και θ_i οι θεωρητικές. Θα πρέπει: $\theta_i \geq 5, i = 1, 2, 3$

Στον παρακάτω πίνακα παραθέτουμε τις παρατηρηθείσες και τις αναμενόμενες συχνότητες:

	Κατηγορίες			Άθροισμα
	A	B	Γ	
Παρατηρηθείσες συχνότητες π_i	210	220	170	600
Αναμενόμενες ή θεωρητικές συχνότητες $\theta_i = n \cdot p_i$	$1/3 \cdot 600 =$ 200	$1/3 \cdot 600 =$ 200	$1/3 \cdot 600 =$ 200	600
$\pi_i - \theta_i$	10	20	-30	
$(\pi_i - \theta_i)^2$	100	400	900	
$\frac{(\pi_i - \theta_i)^2}{\theta_i}$	0.5	2	4.5	7

Ελέγχουμε εάν ισχύει η ανισότητα του στατιστικού κριτηρίου:

$$\begin{aligned} \chi^2 &= \frac{(210 - 200)^2}{200} + \frac{(220 - 200)^2}{200} + \frac{(170 - 200)^2}{200} = 7 > \chi_{k-1, \alpha}^2 = \chi_{3-1, 0.05}^2 = \chi_{2, 0.05}^2 \\ &= 5.99 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς απορρίπτουμε την H_0 . Επομένως δεν μπορούμε να ισχυριστούμε ότι οι τρεις φυλές A, B, Γ της περιοχής βρίσκονται στην ίδια αναλογία.

β) Θα εφαρμόσουμε και σε αυτήν την περίπτωση το τεστ χ^2 καλής προσαρμογής.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Οι τρεις φυλές A, B, Γ βρίσκονται σε ποσοστό $p_A = 0.35$, $p_B = 0.35$, $p_\Gamma = 0.30$

H_1 : Οι τρεις φυλές A, B, Γ δεν βρίσκονται σε ποσοστό $p_A = 0.35$, $p_B = 0.35$, $p_\Gamma = 0.30$

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$\chi^2 = \sum_{i=1}^k \frac{(\pi_i - \theta_i)^2}{\theta_i} > \chi_{k-1, \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_i οι παρατηρηθείσες συχνότητες και θ_i οι θεωρητικές. Θα πρέπει: $\theta_i \geq 5$, $i = 1, 2, 3$.

Στον παρακάτω πίνακα παραθέτουμε τις παρατηρηθείσες και τις αναμενόμενες συχνότητες:

	Κατηγορίες			Άθροισμα
	A	B	Γ	
Παρατηρηθείσες συχνότητες π_i	210	220	170	600
Αναμενόμενες ή θεωρητικές συχνότητες θ_i	$0.35 \cdot 600 =$ 210	$0.35 \cdot 600 =$ 210	$0.30 \cdot 600 =$ 180	600
$\pi_i - \theta_i$	0	10	-10	
$(\pi_i - \theta_i)^2$	0	100	100	
$\frac{(\pi_i - \theta_i)^2}{\theta_i}$	0	0.48	0.56	1.04

Ελέγχουμε εάν ισχύει η ανισότητα του στατιστικού κριτηρίου:

$$\begin{aligned} \chi^2 &= \frac{(210 - 210)^2}{210} + \frac{(220 - 210)^2}{210} + \frac{(170 - 180)^2}{180} = 1.04 < \chi_{k-1, \alpha}^2 = \chi_{3-1, 0.05}^2 = \chi_{2, 0.05}^2 \\ &= 5.99 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς απορρίπτουμε την H_0 . Επομένως τα πειραματικά δεδομένα υποστηρίζουν ότι οι τρεις φυλές A, B, Γ της περιοχής βρίσκονται σε ποσοστό 35%, 35% και 30% αντίστοιχα.

3. Στη βιβλιογραφία αναφέρεται ότι τα ποσοστά των ομάδων αίματος A, B, AB και O σε έναν πληθυσμό είναι 0.41, 0.10, 0.04 και 0.45 αντίστοιχα. Μια ομάδα ερευνητών, προκειμένου να ελέγξει αν τα ποσοστά των ομάδων αίματος σε αυτόν τον πληθυσμό είναι πράγματι αυτά που αναφέρονται στην βιβλιογραφία, επέλεξε τυχαία 200 άτομα από αυτόν τον πληθυσμό και για καθένα κατέγραψε την ομάδα αίματός του. Στον πίνακα που ακολουθεί φαίνεται η συχνότητα κάθε ομάδας αίματος που παρατηρήθηκε στο δείγμα.

	Ομάδα αίματος			
	A	B	AB	O
Συχνότητα (παρατηρηθείσα)	89	18	12	81

α) Σε επίπεδο σημαντικότητας 5%, τα ποσοστά που παρατηρούνται στο δείγμα συμφωνούν ή όχι, με τα αντίστοιχα ποσοστά που αναφέρονται στη βιβλιογραφία; **β)** Με βάση το συμπέρασμά σας στο (α), μπορείτε να αποφασίσετε σε επίπεδο σημαντικότητας 1%, αν τα ποσοστά που παρατηρούνται στο δείγμα συμφωνούν ή όχι, με τα αντίστοιχα ποσοστά που αναφέρονται στη βιβλιογραφία; Εξηγήστε.

α) Πρόκειται για χ^2 στατιστικό έλεγχο καλής προσαρμογής. Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Τα πειραματικά δεδομένα συμφωνούν με τα αντίστοιχα της βιβλιογραφίας,

δηλαδή οι ομάδες αίματος A, B, AB, O στον πληθυσμό βρίσκονται σε ποσοστό

$$p_A = 0.41, \quad p_B = 0.10, \quad p_{AB} = 0.04, \quad p_O = 0.45 \text{ αντίστοιχα.}$$

H_1 : Τα πειραματικά δεδομένα δεν συμφωνούν με τα αντίστοιχα της βιβλιογραφίας,

δηλαδή οι ομάδες αίματος A, B, AB, O στον πληθυσμό δεν βρίσκονται σε ποσοστό

$$p_A = 0.41, \quad p_B = 0.10, \quad p_{AB} = 0.04, \quad p_O = 0.45 \text{ αντίστοιχα.}$$

Αφού υπολογίσουμε τις θεωρητικές συχνότητες $\theta_i = n \cdot p_i, i = 1,2,3,4$ με το στατιστικό κριτήριο που δίνεται παρακάτω ελέγχουμε εάν ισχύουν οι παραπάνω υποθέσεις:

Εάν

$$X^2 = \sum_{i=1}^k \frac{(\pi_i - \theta_i)^2}{\theta_i} > \chi_{k-1, \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_i οι παρατηρηθείσες συχνότητες και θ_i οι θεωρητικές. Θα πρέπει: $\theta_i \geq 5, i = 1, 2, 3$

Στον παρακάτω πίνακα παραθέτουμε τις παρατηρηθείσες και τις αναμενόμενες συχνότητες:

	Ομάδες αίματος				Άθροισμα
	A	B	AB	O	
Παρατηρηθείσες συχνότητες π_i	89	18	12	81	200
Αναμενόμενες ή θεωρητικές συχνότητες $\theta_i = n \cdot p_i$	$0.41 \cdot 200 =$ 82	$0.10 \cdot 200 =$ 20	$0.04 \cdot 200 =$ 8	$0.45 \cdot 200 =$ 90	200
$\pi_i - \theta_i$	7	-2	4	-9	
$(\pi_i - \theta_i)^2$	49	4	16	81	
$\frac{(\pi_i - \theta_i)^2}{\theta_i}$	0.598	0.2	2	0.9	3.698

Ελέγχουμε εάν ισχύει η ανισότητα του στατιστικού κριτηρίου:

$$\chi^2 = \frac{(89 - 82)^2}{82} + \frac{(18 - 20)^2}{20} + \frac{(12 - 8)^2}{8} + \frac{(81 - 90)^2}{90} = 3.698 < \chi_{k-1, \alpha}^2 = \chi_{4-1, 0.05}^2$$

$$= \chi_{3, 0.05}^2 = 7.81$$

Άρα σε επίπεδο σημαντικότητας 5%, δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς δεν μπορούμε να απορρίψουμε την H_0 . Επομένως δεν μπορούμε να ισχυριστούμε ότι οι ομάδες αίματος A, B, AB, O στον πληθυσμό βρίσκονται σε ποσοστό $p_A = 0.41$, $p_B = 0.10$, $p_{AB} = 0.04$ και $p_O = 0.45$ αντίστοιχα.

β) Σε αυτή την περίπτωση έχουμε τις ίδιες στατιστικές υποθέσεις και ο στατιστικός έλεγχος είναι ο ίδιος. Το μόνο που αλλάζει είναι το επίπεδο σημαντικότητας. Επομένως όταν $\alpha = 0.01$, το $\chi_{k-1, \alpha}^2 = \chi_{4-1, 0.01}^2 = \chi_{3, 0.01}^2 = 11.34$. Άρα το $3.698 < 11.34$ και συνεπώς και σε επίπεδο σημαντικότητας 1% αποδεχόμαστε την H_0 .

5. Η κατανομή του βάρους των νεογέννητων τριών φυλών χοίρων έδωσε τα παρακάτω αποτελέσματα:

Φυλές	Βάρος σε κιλά			
	≤ 3	4	5	≥ 6
A	5	15	12	18
B	3	7	12	8
Γ	2	8	6	4

Ελέγξτε εάν το βάρος των νεογέννητων χοιριδίων εξαρτάται από τη φυλή στην οποία ανήκουν ($\alpha=0.05$).

α) Πρόκειται για χ^2 στατιστικό έλεγχο ανεξαρτησίας ή ομοιογένειας.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Το βάρος των νεογέννητων των τριών φυλών χοίρων *δεν διαφέρει* στατιστικά (ομοιογένεια)
ή το βάρος των νεογέννητων χοιριδίων είναι *ανεξάρτητο* από τη φυλή. (ανεξαρτησία)

H_1 : Το βάρος των νεογέννητων των τριών φυλών χοίρων *διαφοροποιείται* ανάλογα με τη φυλή
ή το βάρος των νεογέννητων χοιριδίων *εξαρτάται* από τη φυλή στην οποία ανήκουν.

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$\chi^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_{ij} οι παρατηρηθείσες συχνότητες, θ_{ij} οι θεωρητικές, κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας. Για να εφαρμόσουμε το τεστ χ^2 , θα πρέπει για όλες τις θεωρητικές συχνότητες να ισχύει: $\theta_i \geq 5, \forall (i, j)$.

Οι θεωρητικές συχνότητες υπολογίζονται από τη σχέση:

$$\theta_{ij} = \frac{(\text{άθροισμα } i \text{ γραμμής}) \times (\text{άθροισμα } j \text{ στήλης})}{\text{συνολικό άθροισμα}} \quad \forall (i, j).$$

Στη συνέχεια αφού υπολογίσουμε τα αθροίσματα των γραμμών και των στηλών του πίνακα συνάφειας υπολογίζουμε τις θεωρητικές συχνότητες:

$$\theta_{11} = \frac{50 \cdot 10}{100} = 5 \quad \theta_{12} = \frac{50 \cdot 30}{100} = 15 \quad \theta_{13} = \frac{50 \cdot 30}{100} = 15 \quad \theta_{14} = \frac{50 \cdot 30}{100} = 15$$

$$\theta_{21} = \frac{30 \cdot 10}{100} = 3 \quad \theta_{22} = \frac{30 \cdot 30}{100} = 9 \quad \theta_{23} = \frac{30 \cdot 30}{100} = 9 \quad \theta_{24} = \frac{30 \cdot 30}{100} = 9$$

$$\theta_{31} = \frac{20 \cdot 10}{100} = 2 \quad \theta_{32} = \frac{20 \cdot 30}{100} = 6 \quad \theta_{33} = \frac{20 \cdot 30}{100} = 6 \quad \theta_{34} = \frac{20 \cdot 30}{100} = 6$$

Φυλές	Βάρος σε κιλά				Άθροισμα
	≤3	4	5	≥6	
A	5 (5)	15 (15)	12 (15)	18 (15)	50
B	3 (3)	7 (9)	12 (9)	8 (9)	30
Γ	2 (2)	8 (6)	6 (6)	4 (6)	20
Άθροισμα	10	30	30	30	100

Καταρχάς παρατηρούμε ότι δεν ισχύουν οι προϋποθέσεις $\theta_{ij} \geq 5$, για όλα τα (i, j) . Σε αυτές τις περιπτώσεις ενοποιούμε δύο διπλανές κατηγορίες, έτσι ώστε στη συνέχεια να ισχύει: $\theta_{ij} \geq 5$, $\forall (i, j)$. Σε αυτήν την περίπτωση ενώνουμε τις δύο πρώτες στήλες του πίνακα συνάφειας και έχουμε:

Φυλές	≤4	5	≥6	Άθροισμα
A	20 (20)	12 (15)	18 (15)	50
B	10 (12)	12 (9)	8 (9)	30
Γ	10 (8)	6 (6)	4 (6)	20
Άθροισμα	40	30	30	100

καθώς:

$$\theta_{11} = \frac{50 \cdot 40}{100} = 20 \quad \theta_{12} = \frac{50 \cdot 30}{100} = 15 \quad \theta_{13} = \frac{50 \cdot 30}{100} = 15$$

$$\theta_{21} = \frac{30 \cdot 40}{100} = 12 \quad \theta_{22} = \frac{30 \cdot 30}{100} = 9 \quad \theta_{23} = \frac{30 \cdot 30}{100} = 9$$

$$\theta_{31} = \frac{20 \cdot 40}{100} = 8 \quad \theta_{32} = \frac{20 \cdot 30}{100} = 6 \quad \theta_{33} = \frac{20 \cdot 30}{100} = 6$$

Με τη δοκιμασία χ^2 ελέγχουμε εάν υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στις παρατηρηθείσες και τις θεωρητικές συχνότητες. Το στατιστικό κριτήριο που χρησιμοποιούμε είναι:

Εάν

$$\chi^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας.

Αντικαθιστώντας έχουμε:

$$\begin{aligned} \chi^2 &= \frac{(20 - 20)^2}{20} + \frac{(12 - 15)^2}{15} + \dots + \frac{(4 - 6)^2}{6} = 3.81 < \chi^2_{(k-1)(l-1),\alpha} \\ &= \chi^2_{(3-1)(3-1),0.05} = \chi^2_{4,0.05} = 9.49 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς δεν μπορούμε να απορρίψουμε την H_0 . Επομένως το βάρος των νεογέννητων των τριών φυλών χοίρων δεν διαφοροποιείται σημαντικά ανάλογα με τη φυλή ή θα μπορούσαμε να πούμε ότι είναι ανεξάρτητο της φυλής.

6. Στον παρακάτω πίνακα δίνεται η συχνότητα προτίμησης από το καταναλωτικό κοινό τεσσάρων ποικιλιών μήλων. 450 άτομα και των δύο φύλων δοκίμασαν τις ποικιλίες μήλων επιλέγοντας μόνο μία.

	Ποικιλίες μήλων			
Φύλο	A	B	Γ	Δ
Άνδρες	32	44	72	52
Γυναίκες	78	56	63	53

α) Σε επίπεδο σημαντικότητας 5% ελέγξτε αν η προτίμηση της ποικιλίας μήλων εξαρτάται από το φύλο.

β) Ελέγξτε αν το ποσοστό των γυναικών που προτιμούν την ποικιλία A είναι μεγαλύτερο από αυτό που προτιμούν την ποικιλία B ($\alpha = 0.05$).

γ) Ελέγξτε σε επίπεδο σημαντικότητας 5% εάν το ποσοστό των ανδρών που προτιμούν την ποικιλία A είναι μικρότερο του 20%.

α) Πρόκειται για χ^2 στατιστικό έλεγχο ανεξαρτησίας ή ομοιογένειας.

Διατυπώνουμε τις στατιστικές υποθέσεις:

H_0 : Η προτίμηση της ποικιλίας μήλων δεν διαφοροποιείται ανάλογα με το φύλο (ομοιογένεια)
ή η προτίμηση της ποικιλίας μήλων είναι ανεξάρτητη του φύλου (ανεξαρτησία)

H_1 : Η προτίμηση της ποικιλίας μήλων διαφοροποιείται ανάλογα με το φύλο
ή η προτίμηση της ποικιλίας μήλων είναι εξαρτάται από το φύλο.

Το στατιστικό κριτήριο για τον έλεγχο των παραπάνω υποθέσεων είναι:

Εάν

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου π_{ij} οι παρατηρηθείσες συχνότητες, θ_{ij} οι θεωρητικές, κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας. Επιπλέον θα πρέπει: $\theta_{ij} \geq 5, \forall (i, j)$.

Οι θεωρητικές συχνότητες υπολογίζονται από τη σχέση:

$$\theta_{ij} = \frac{(\text{άθροισμα } i \text{ γραμμής}) \times (\text{άθροισμα } j \text{ στήλης})}{\text{συνολικό άθροισμα}} \quad \forall (i, j).$$

Στη συνέχεια αφού υπολογίσουμε τα αθροίσματα των γραμμών και των στηλών του πίνακα συνάφειας υπολογίζουμε τις θεωρητικές συχνότητες:

$$\theta_{11} = \frac{200 \cdot 110}{450} = 48.9 \quad \theta_{12} = \frac{200 \cdot 100}{450} = 44.4 \quad \theta_{13} = \frac{200 \cdot 135}{450} = 60 \quad \theta_{14} = \frac{200 \cdot 105}{450} = 46.7$$

$$\theta_{21} = \frac{250 \cdot 110}{450} = 61.1 \quad \theta_{22} = \frac{250 \cdot 100}{450} = 55.6 \quad \theta_{23} = \frac{250 \cdot 135}{450} = 75 \quad \theta_{24} = \frac{250 \cdot 105}{450} = 58.3$$

	Ποικιλίες μήλων				
Φύλο	A	B	Γ	Δ	Άθροισμα
Άνδρες	32 (48.9)	44 (44.4)	72 (60)	52 (46.7)	200
Γυναίκες	78 (61.1)	56 (55.6)	63 (75)	53 (58.3)	250
Άθροισμα	110	100	135	105	450

Καθώς όλα τα $\theta_{ij} \geq 5$, ελέγχουμε εάν ισχύει το στατιστικό κριτήριο:

Εάν

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} > \chi_{(\kappa-1)(\lambda-1), \alpha}^2 \quad \text{απορρίπτουμε την } H_0,$$

όπου κ ο αριθμός των γραμμών και λ ο αριθμός των στηλών του πίνακα συνάφειας.

Αντικαθιστώντας έχουμε:

$$\begin{aligned} \chi^2 &= \frac{(32 - 48.9)^2}{48.9} + \frac{(44 - 44.4)^2}{44.4} + \dots + \frac{(53 - 58.3)^2}{58.3} = 15.93 > \chi_{(k-1)(l-1),\alpha}^2 \\ &= \chi_{(2-1)(4-1),0.05}^2 = \chi_{3,0.05}^2 = 7.81 \end{aligned}$$

Άρα σε επίπεδο σημαντικότητας 5%, ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς απορρίπτουμε την H_0 . Επομένως η προτίμηση της ποικιλίας μήλων *διαφοροποιείται* ανάλογα με το φύλο (είναι *διαφορετική* ανάμεσα σε άνδρες και γυναίκες) ή αλλιώς θα μπορούσαμε να πούμε ότι η προτίμηση της ποικιλίας μήλων είναι *εξαρτάται* από το φύλο.

β) Παίρνουμε τις απαντήσεις της δειγματοληψίας που αφορούν μόνο τις γυναίκες ($n = 250$).

Θέλουμε να συγκρίνουμε τα ποσοστά των γυναικών που προτιμούν τις ποικιλίες Α και Β.

Διατυπώνουμε τις στατιστικές υποθέσεις:

$H_0: p_A = p_B$ (αρχική ή μηδενική υπόθεση)

$H_1: p_A > p_B$ (εναλλακτική υπόθεση)

Πρόκειται για στατιστικό έλεγχο υποθέσεων για τη σύγκριση των άγνωστων ποσοστών δύο πληθυσμών (τα ποσοστά των γυναικών που προτιμούν τις ποικιλίες Α και Β).

Από τα πειραματικά δεδομένα έχουμε:

$n = 250$ το μέγεθος δείγματος, $\hat{p}_A = 78/250 = 0.312$ το ποσοστό των γυναικών που προτιμούν την ποικιλία Α και $\hat{p}_B = 56/250 = 0.224$ το ποσοστό των γυναικών που προτιμούν την ποικιλία Β.

Καθώς ισχύουν οι προϋποθέσεις: $n_i \hat{p}_i \geq 5$ και $n_i(1 - \hat{p}_i) \geq 5$, $i = 1, 2$ η απορριπτική περιοχή της H_0 είναι: Εάν

$$\frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} > Z_\alpha \quad \text{απορρίπτουμε την } H_0 \quad (\text{στατιστικό κριτήριο})$$

Αντικαθιστώντας και σε επίπεδο σημαντικότητας $\alpha = 0.05$, οπότε $z_\alpha = z_{0.05} = 1.645$ και

$$\hat{p} = \frac{78+56}{250+250} = \frac{134}{500} = 0.268 \quad \text{έχουμε:}$$

$$\frac{0.312 - 0.224}{\sqrt{0.268(1 - 0.268)\left(\frac{1}{250} + \frac{1}{250}\right)}} = 2.2 > z_\alpha = z_{0.05} = 1.645$$

Επομένως ισχύει η ανισότητα του στατιστικού κριτηρίου και συνεπώς η H_0 απορρίπτεται σε επίπεδο σημαντικότητας 5%. Άρα βασιζόμενοι στη συγκεκριμένη δειγματοληπτική έρευνα μπορούμε να ισχυριστούμε ότι το ποσοστό των γυναικών που προτιμούν την ποικιλία Α είναι μεγαλύτερο από αυτό που προτιμούν την ποικιλία Β.

γ) Πρόκειται για στατιστικό έλεγχο υποθέσεων για το άγνωστο ποσοστό ενός πληθυσμού (ποσοστό των ανδρών που προτιμούν την ποικιλία Α). Παίρνουμε τις απαντήσεις της δειγματοληπτικής έρευνας που αφορούν μόνο τους άνδρες ($n = 200$) και διατυπώνουμε τις στατιστικές υποθέσεις:

$H_0: p_0 = 0.20$ (αρχική ή μηδενική υπόθεση) (Θέτουμε κάτι γενικό ή κάτι που ισχύει από παλιά)

$H_1: p_A < 0.20$ (εναλλακτική υπόθεση) (Θέτουμε το ερώτημα του ερευνητή)

Από τα πειραματικά δεδομένα έχουμε $n=200$ το μέγεθος του δείγματος και $\hat{p}_A = 32/200 = 0.16$ το ποσοστό στο δείγμα. Ισχύουν οι προϋποθέσεις: $np_0 = 200 \cdot 0.20 = 40 \geq 5$ και $n(1 - p_0) = 200 \cdot (1 - 0.20) = 160 \geq 5$, συνεπώς η απορριπτική περιοχή της H_0 δίνεται από τη σχέση:

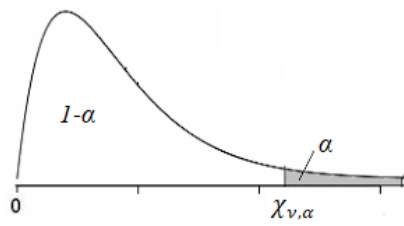
Εάν $\frac{\hat{p}_A - p_0}{\sqrt{p_0(1-p_0)/n}} < -Z_\alpha$ απορρίπτουμε την H_0 (στατιστικό κριτήριο)

Αντικαθιστώντας και σε επίπεδο σημαντικότητας $\alpha = 0.05$, οπότε $z_\alpha = z_{0.05} = 1.645$ έχουμε:

$$\frac{0.16 - 0.2}{\sqrt{0.2(1 - 0.2)/200}} = \frac{-0.04}{0.02828} = -1.41 > -z_\alpha = -z_{0.05} = -1.645$$

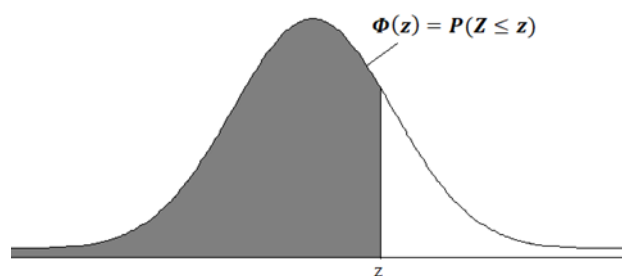
Άρα δεν ισχύει η ανισότητα του στατιστικού κριτηρίου και επομένως η H_0 δεν μπορεί να απορριφθεί σε επίπεδο σημαντικότητας 5%. Συνεπώς βασιζόμενοι στη συγκεκριμένη δειγματοληπτική έρευνα δεν μπορούμε να ισχυριστούμε ότι το ποσοστό των ανδρών που προτιμούν την ποικιλία Α είναι μικρότερο του 20%.

Στατιστικός πίνακας της χ^2 κατανομής



v	α				
	0.05	0.025	0.010	0.005	0.001
1	3.84	5.02	6.63	7.88	10.83
2	5.99	7.38	9.21	10.60	13.82
3	7.81	9.35	11.34	12.84	16.27
4	9.49	11.14	13.28	14.86	18.47
5	11.07	12.83	15.09	16.75	20.52
6	12.59	14.45	16.81	18.55	22.46
7	14.07	16.01	18.48	20.28	24.32
8	15.51	17.53	20.09	21.95	26.12
9	16.92	19.02	21.67	23.59	27.88
10	18.31	20.48	23.21	25.19	29.59
11	19.68	21.92	24.72	26.76	31.26
12	21.03	23.34	26.22	28.30	32.91
13	22.36	24.74	27.69	29.82	34.53
14	23.68	26.12	29.14	31.32	36.12
15	25.00	27.49	30.58	32.80	37.70
16	26.30	28.85	32.00	34.27	39.25
17	27.59	30.19	33.41	35.72	40.79
18	28.87	31.53	34.81	37.16	42.31
19	30.14	32.85	36.19	38.58	43.82
20	31.41	34.17	37.57	40.00	45.31
21	32.67	35.48	38.93	41.40	46.80
22	33.92	36.78	40.29	42.80	48.27
23	35.17	38.08	41.64	44.18	49.73
24	36.42	39.36	42.98	45.56	51.18
25	37.65	40.65	44.31	46.93	52.62
26	38.89	41.92	45.64	48.29	54.05
27	40.11	43.19	46.96	49.64	55.48
28	41.34	44.46	48.28	50.99	56.89
29	42.56	45.72	49.59	52.34	58.30
30	43.77	46.98	50.89	53.67	59.70

Πίνακας Τυπικής Κανονικής κατανομής



Παράδειγμα: $\Phi(0.82)=0.7939$, $\Phi(1.28)=0.8997$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998