ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΧΤΑ
ΑΚΑΔΗΜΑΙΚΑ
ΜΑΘΗΜΑΤΑ

Στατιστική

7ο Μάθημα: Ο Έλεγχος X^2

Γεώργιος Μενεξές
Τμήμα Γεωπονίας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

7ο Μάθημα

Ο Έλεγχος X^2

Πίνακες Συμπτώσεων ή Συνάφειας

Εισαγωγή

- Σε εφαρμογές, οι παρατηρούμενες συχνότητες κατατάσσονται σε *κ* κατηγορίες (ή κλάσεις) ως προς την κατηγορική μεταβλητή Α και ως προς *λ* κλάσεις της κατηγορικής μεταβλητής Β. Ο πίνακας συχνοτήτων που σχηματίζεται έχει *κ* γραμμές και *λ* στήλες. Οι διδιάστατοι αυτοί πίνακες ονομάζονται πίνακες συμπτώσεων ή συνάφειας.
- Σε κάθε παρατηρούμενη συχνότητα ενός πίνακα συνάφειας $\kappa \times \lambda$ αντιστοιχεί μια αναμενόμενη ή θεωρητική συχνότητα που υπολογίζεται σύμφωνα με μια υπόθεση.

Εισαγωγή (*συνέχεια*)

Βασική Ιδέα:

Εξετάζουμε τη συμφωνία μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων κάτω από την ισχύ της υπόθεσης.

Εισαγωγή (*συνέχεια*)

- Pearson's Chi Squared

Karl Pearson (27 March 1857 – 27 April 1936)

Παράδειγμα 1

- Έχουμε 5 δείγματα (μηδικής) F_1 απογόνων και θέλουμε να ελέγξουμε αν οι αναλογίες των κλάσεων γύρης είναι οι ίδιες στους αντίστοιχους πληθυσμούς (α=0,05)

	Κλάσεις Παραγωγής Γύρης				
F1	1η	2η	3η	4η	Σύνολο
1ο	11	6	9	14	40
2ο	7	6	7	9	29
3ο	14	5	7	11	37
4ο	11	4	7	20	42
5ο	22	2	12	16	52
Σύνολο	65	23	42	70	200

r_1

n

c_1

Στατιστικός Έλεγχος

- H_0: Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης είναι ίδιες και για τις πέντε ομάδες απογόνων (F_1 γενεές)
- H_1: Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης δεν είναι ίδιες και για τις πέντε ομάδες απογόνων (F_1 γενεές)

σε επίπεδο σημαντικότητας α=0,05

Προφίλ Γραμμών

- Κατασκευάζουμε τον πίνακα που περιέχει τα “προφίλ” των γραμμών:

Δείγματα F1 απογόνων $\dfrac{\sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha\ \kappa\varepsilon\lambda\iota o\acute{\upsilon}}{\sigma\acute{\upsilon}\nu o\lambda o\ \gamma\rho\alpha\mu\mu\acute{\eta}\varsigma}\times 100$ **Crosstabulation**

% within Δείγματα F1 απογόνων

		Κλάσεις Παραγωγής Γύρης				Total
		1	2	3	4	
Δείγματα F1 απογόνων	1	27.5%	15.0%	22.5%	35.0%	100.0%
	2	24.1%	20.7%	24.1%	31.0%	100.0%
	3	37.8%	13.5%	18.9%	29.7%	100.0%
	4	26.2%	9.5%	16.7%	47.6%	100.0%
	5	42.3%	3.8%	23.1%	30.8%	100.0%
Total		32.5%	11.5%	21.0%	35.0%	100.0%

$$\frac{11}{40}\times 100 = 27.5$$

$$\frac{16}{52}\times 100 = 30.8$$

Παρατήρηση

Κάθε γραμμή του πίνακα που περιέχει τα προφίλ γραμμών αποτελεί υλοποίηση μιας Πολυωνυμικής Κατανομής.

Πίνακας Συμπτώσεων Απολύτων Συχνοτήτων

$$\frac{11}{40}\times 100 = 27.5$$

	Κλάσεις Παραγωγής Γύρης				
F1	1η	2η	3η	4η	Σύνολο
1^o	11	6	9	14	40
2^o	7	6	7	9	29
3^o	14	5	7	11	37
4^o	11	4	7	20	42
5^o	22	2	12	16	52
Σύνολο	65	23	42	70	200

$$\frac{16}{52}\times 100 = 30.8$$

Αναμενόμενες Συχνότητες

- Υπολογίζουμε τις αναμενόμενες συχνότητες σε κάθε κελί του πίνακα:

$$\alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha = \frac{\sigma\acute{\upsilon}\nu o\lambda o\ \gamma\rho\alpha\mu\mu\acute{\eta}\varsigma \times \sigma\acute{\upsilon}\nu o\lambda o\ \sigma\tau\acute{\eta}\lambda\eta\varsigma}{\gamma\varepsilon\nu\iota\kappa\acute{o}\ \sigma\acute{\upsilon}\nu o\lambda o}$$

		Κλάσεις Παραγωγής Γύρης				
		1	2	3	4	Total
Δείγματα F1 απογόνων	1	13.0	4.6	8.4	14.0	40.0
	2	9.4	3.3	6.1	10.2	29.0
	3	12.0	4.3	7.8	13.0	37.0
	4	13.7	4.8	8.8	14.7	42.0
	5	16.9	6.0	10.9	18.2	52.0
Total		65.0	23.0	42.0	70.0	200.0

$$\frac{40 \times 65}{200} = 13.0$$

$$\frac{50 \times 72}{200} = 18.2$$

Πίνακας Συμπτώσεων Απολύτων Συχνοτήτων

$$\boxed{\frac{40 \times 65}{200} = 13.0}$$

	Κλάσεις Παραγωγής Γύρης				
F1	**1η**	**2η**	**3η**	**4η**	Σύνολο
1ο	**11**	**6**	**9**	**14**	**40**
2ο	**7**	**6**	**7**	**9**	**29**
3ο	**14**	**5**	**7**	**11**	**37**
4ο	**11**	**4**	**7**	**20**	**42**
5ο	**22**	**2**	**12**	**16**	**52**
Σύνολο	**65**	**23**	**42**	**70**	**200**

Βοηθητικός Πίνακας

Δείγματα F1 απογόνων * Κλάσεις Παραγωγής Γύρης Crosstabulation

			Κλάσεις Παραγωγής Γύρης				Total
			1	2	3	4	
Δείγματα F1 απογόνων	1	Count	11	6	9	14	40
		Expected Count	13.0	4.6	8.4	14.0	40.0
	2	Count	7	6	7	9	29
		Expected Count	9.4	3.3	6.1	10.2	29.0
	3	Count	14	5	7	11	37
		Expected Count	12.0	4.3	7.8	13.0	37.0
	4	Count	11	4	7	20	42
		Expected Count	13.7	4.8	8.8	14.7	42.0
	5	Count	22	2	12	16	52
		Expected Count	16.9	6.0	10.9	18.2	52.0
Total		Count	65	23	42	70	200
		Expected Count	65.0	23.0	42.0	70.0	200.0

Count: Συχνότητα

Expected Count: Αναμενόμενη Συχνότητα

Το Στατιστικό X^2

- Υπολογίζουμε το στατιστικό X^2:

$$X^2 = \sum \frac{(\pi\alpha\rho\alpha\tau\eta\rho o\acute{\upsilon}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha - \alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha)^2}{\alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha}$$

$$\Downarrow$$

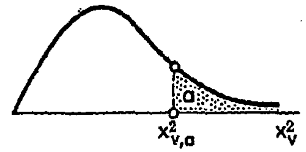
$$X^2 = \frac{(11-13{,}0)^2}{13{,}0} + \frac{(6-4{,}6)^2}{4{,}6} + \cdots + \frac{(16-18{,}2)^2}{18{,}2} = 12{,}125$$

Χρήση Πίνακα της X^2 Κατανομής

- Συγκρίνουμε το στατιστικό X^2 με την κρίσιμη τιμή της X^2 Κατανομής, με (5-1)(4-1)=12 β.ε., σε επίπεδο σημαντικότητας α=0,05:

Ανατρέχουμε στους Πίνακες της X^2 Κατανομής

ΠΙΝΑΚΑΣ 7: Κριτικές τιμές της κατανομής χι-τετράγωνο

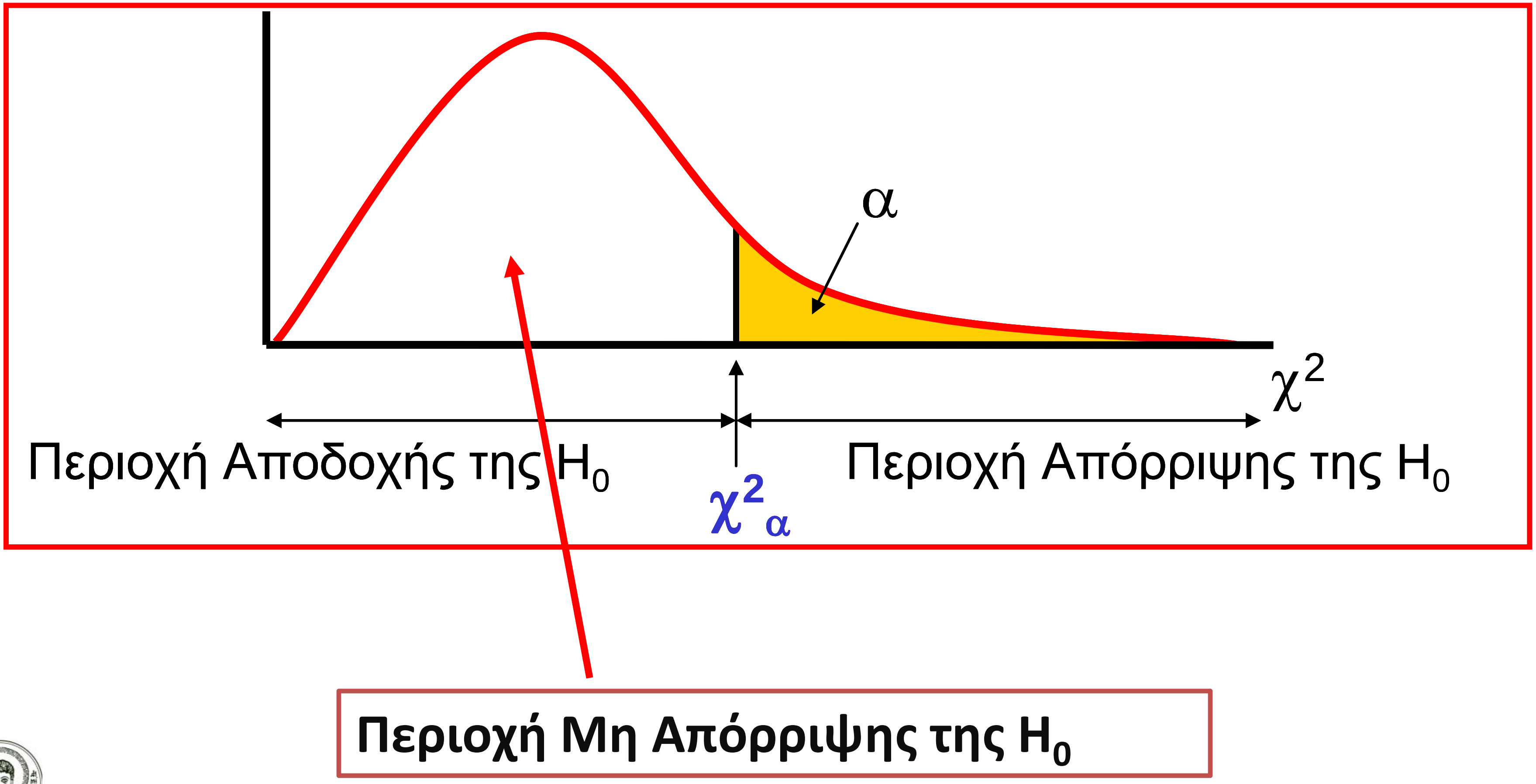


ν	α									
	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Σύγκριση

- $X^2(12)_{0,05}=21,03$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=12,125$ (δειγματικό)
- $X^2=12,125<21,03=X^2(12)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

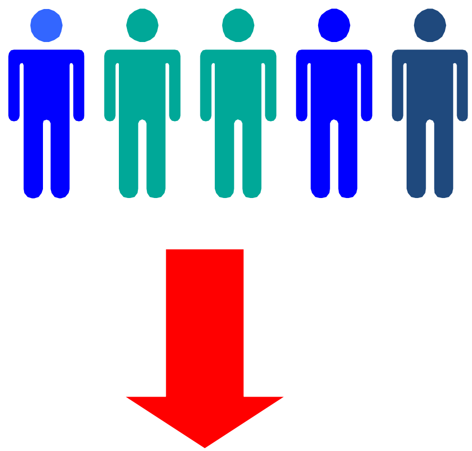
Περιοχή Απόρριψης της H_0



Απόφαση-Συμπέρασμα

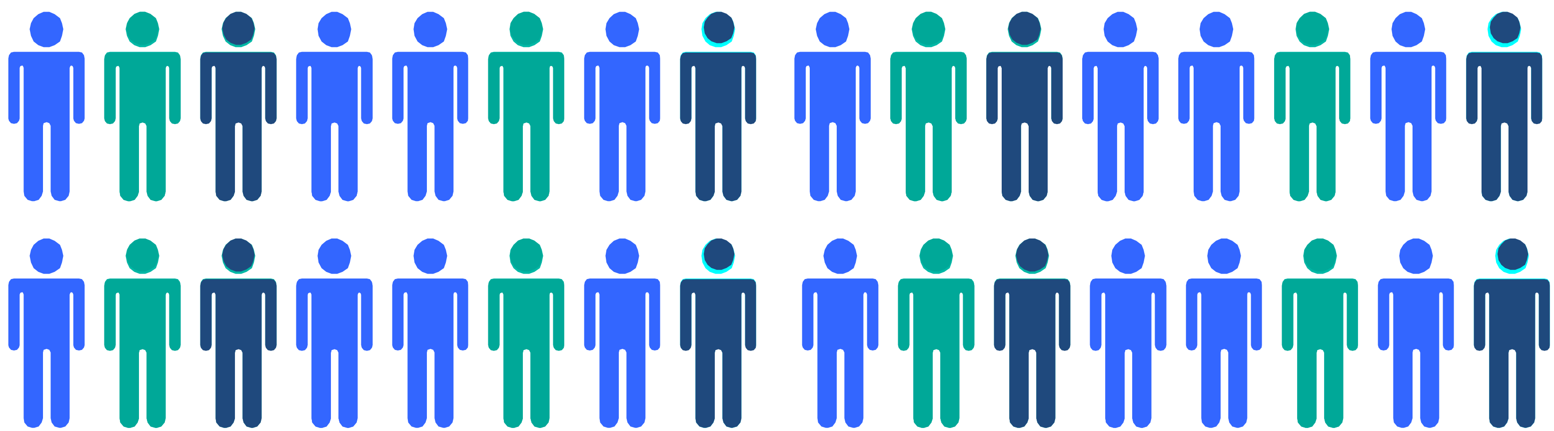
- Η μηδενική υπόθεση δεν απορρίπτεται σε ε.σ. α=0,05.
- Με βάση τα διαθέσιμα δεδομένα δεν μπορεί να απορριφθεί η μηδενική υπόθεση σε ε.σ. α=0,05.
- Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης και για τις πέντε ομάδες απογόνων (F_1 γενεές) δεν διαφέρουν στατιστικά σημαντικά σε ε.σ. α=0,05.
- Οι αναλογίες των τεσσάρων κλάσεων παραγωγής γύρης "είναι ίδιες" και για τις πέντε ομάδες απογόνων (F_1 γενεές) σε ε.σ. α=0,05.

Υπενθύμιση



Από το δείγμα…

…γενικεύουμε για τον αντίστοιχο Πληθυσμό



Αποτελέσματα του SPSS

X^2 δειγματικό

Βαθμοί Ελευθερίας

Παρατηρούμενη Στάθμη Σημαντικότητας *p*-value

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
				Sig.	99% Confidence Interval		Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound		Lower Bound	Upper Bound
Pearson Chi-Square	12.125[a]	12	.436	.442[b]	.429	.454			
Likelihood Ratio	12.430	12	.412	.445[b]	.432	.458			
Fisher's Exact Test	12.176			.427[b]	.414	.439			
Linear-by-Linear Association	.196[c]	1	.658	.663[b]	.650	.675	.332[b]	.320	.344
N of Valid Cases	200								

a. 4 cells (20.0%) have expected count less than 5. The minimum expected count is 3.34.

b. Based on 10000 sampled tables with starting seed 334431365.

c. The standardized statistic is -.443.

Αν $p<\alpha$, τότε απορρίπτεται η H_0

Αν $p\geq\alpha$, η H_0 δεν απορρίπτεται

Προϋποθέσεις Εφαρμογής του Ελέγχου

Τα συμπεράσματα του ελέγχου X^2 είναι *έγκυρα* μόνο όταν:

1. Τα δείγματα είναι τυχαία.
2. Οι παρατηρήσεις είναι ανεξάρτητες.
3. Όλες οι αναμενόμενες (θεωρητικές) συχνότητες είναι μεγαλύτερες από 1.
4. Το πολύ 20% από τις αναμενόμενες συχνότητες είναι μικρότερες από 5.

Προβλήματα και Λύσεις

Αν δεν ισχύουν οι προϋποθέσεις 3 και 4:

1. Κάνουμε σύμπτυξη γειτονικών κλάσεων (γραμμών ή/και στηλών). Δηλαδή ενώνουμε κατηγορίες γραμμών ή/και στηλών. Αυτό έχει ως συνέπεια την τροποποίηση των βαθμών ελευθερίας.
2. Υπολογίζουμε την παρατηρούμενη στάθμη σημαντικότητας (*p*-value) είτε με την Ακριβή Μέθοδο (*Exact Method*) είτε με τη μέθοδο προσομοίωσης *Monte-Carlo*.
3. Εφαρμόζουμε Permutation Tests.

Ειδικές Περιπτώσεις

1. Πίνακες συμπτώσεων $\kappa \times 2$ ή $2 \times \lambda$
2. Πίνακες Συμπτώσεων 2×2

Παραδείγματα Ειδικών Περιπτώσεων

Παράδειγμα 2

Φυλλοφόρα μοσχεύματα έξι ποικιλιών ελιάς που ριζοβόλησαν ή όχι μετά από 84 ημέρες κάτω από υδρονέφωση

	Μοσχεύματα		
Ποικιλία	Με ρίζες	Χωρίς ρίζες	Σύνολο
Καλαμών	85	75	160
Σεβιλλάνο	87	73	160
Χαλκιδικής	97	63	160
Αμφίσσης	109	51	160
Πατρών	109	51	160
Κοθρέικη	150	10	160
Σύνολο	637	323	960

Παράδειγμα 2 (*συνέχεια*)

- Να ελεγχθεί αν τα ποσοστά ριζοβολίας των 6 φυλλοφόρων μοσχευμάτων ελιάς είναι ίσα (δεν διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05).

Στατιστικός Έλεγχος

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_6$$

$$\mathrm{H}_1 : \textit{τουλάχιστον δύο ποσοστά διαφέρουν}$$

Παράδειγμα 2 (*συνέχεια*)

- Οι υπολογισμοί μπορούν να γίνουν όπως και στο Παράδειγμα 1.
- Επειδή ο πίνακας συμπτώσεων είναι $\kappa \times 2$ υπάρχει πιο εύκολος τρόπος υπολογισμού του στατιστικού X^2

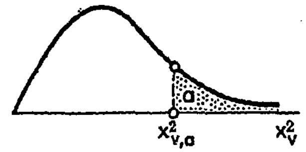
Υπολογίζουμε πρώτα την ποσότητα:

$$\boxed{\frac{85^2}{160} + \frac{87^2}{160} + \cdots + \frac{150^2}{160} = 440,406}$$

…και στη συνέχεια

$$\boxed{X^2 = \frac{440,406 - \left(\frac{637^2}{960}\right)}{\frac{637 \times 323}{960^2}} = 79,40}$$

ΠΙΝΑΚΑΣ 7: Κριτικές τιμές της κατανομής χι-τετράγωνο



ν	α									
	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Παράδειγμα 2 (*συνέχεια*)

- $X^2(5)_{0,05}=11,07$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=79,40$ (δειγματικό)
- $X^2=79,40>11,07=X^2(5)_{0,05}$
- Δειγματικό X^2 > Θεωρητικό X^2

Απόφαση-Συμπέρασμα

- Η μηδενική υπόθεση απορρίπτεται σε ε.σ. α=0,05.
- Τα ποσοστά ριζοβολίας των 6 φυλλοφόρων μοσχευμάτων ελιάς δεν είναι όλα ίσα μεταξύ τους. Τουλάχιστον σε δύο ποικιλίες τα αντίστοιχα ποσοστά διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05.

Παράδειγμα 3

Φυλλοφόρα μοσχεύματα δύο ποικιλιών ελιάς που ριζοβόλησαν ή όχι μετά από 84 ημέρες κάτω από υδρονέφωση

	Μοσχεύματα		
Ποικιλία	**Με ρίζες**	**Χωρίς ρίζες**	**Σύνολο**
Χαλκιδικής	**100**	**60**	**160**
Αμφίσσης	**109**	**51**	**160**
Σύνολο	**209**	**111**	**320**

Χρήσιμοι Συμβολισμοί

	Μοσχεύματα		
Ποικιλία	**Με ρίζες**	**Χωρίς ρίζες**	**Σύνολο**
Χαλκιδικής	n_{11}	n_{12}	$n_{1.}$
Αμφίσσης	n_{21}	n_{22}	$n_{2.}$
Σύνολο	$n_{.1}$	$n_{.2}$	$n_{..}$

Παράδειγμα 3 (*συνέχεια*)

- Να ελεγχθεί αν τα ποσοστά ριζοβολίας των 2 φυλλοφόρων μοσχευμάτων ελιάς είναι ίσα (δεν διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας *α*=0,05).

Στατιστικός Έλεγχος

$$\mathrm{H}_0 : p_1 = p_2$$

$$\mathrm{H}_1 : \textit{τα δύο ποσοστά διαφέρουν } (p_1 \neq p_2)$$

Παράδειγμα 3 (*συνέχεια*)

- Υπολογίζουμε τις αναμενόμενες συχνότητες όπως και στο Παράδειγμα 1

Στο κελί (1,1) η E_{11}= $\dfrac{160\times 209}{320}=104,5$

Στο κελί (2,1) η E_{21}= $\dfrac{160\times 209}{320}=104,5$

Στο κελί (1,2) η E_{12}= $\dfrac{160\times 111}{320}=55,5$

Στο κελί (2,2) η E_{22}= $\dfrac{160\times 111}{320}=55,5$

Παράδειγμα 3 (*συνέχεια*)

Διόρθωση Συνέχειας του *Yates*

$$X^2 = \frac{\left(\left|100-104{,}5\right|-0{,}5\right)^2}{104{,}5} + \frac{\left(\left|60-55{,}5\right|-0{,}5\right)^2}{55{,}5} + \frac{\left(\left|109-104{,}5\right|-0{,}5\right)^2}{104{,}5} + \frac{\left(\left|51-55{,}5\right|-0{,}5\right)^2}{55{,}5} =$$
$$= 0{,}88$$

Γενική Σχέση

$$X^2 = \sum \frac{\left(\left|O-E\right|-\frac{1}{2}\right)^2}{E}$$

O: Παρατηρούμενη Συχνότητα

E: Αναμενόμενη-Θεωρητική Συχνότητα

Παράδειγμα 3 (*συνέχεια*)

Β' Τρόπος

$$X^2 = \frac{\left(\left| (100 \times 51) - (109 \times 60) \right| - \frac{320}{2} \right)^2 \times 320}{209 \times 111 \times 160 \times 160} = 0,88$$

Γενική Σχέση

$$X^2 = \frac{\left(\left| n_{11} n_{22} - n_{12} n_{21} \right| - \frac{n_{..}}{2} \right)^2 n_{..}}{n_{1.} n_{2.} n_{.1} n_{.2}}$$

Παράδειγμα 3 (*συνέχεια*)

- $X^2(1)_{0,05}=3,84$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=0,88$ (δειγματικό)
- $X^2=0,88<3,84=X^2(1)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Παράδειγμα 3 (*συνέχεια*)

- Τα δύο ποσοστά ριζοβολίας μπορούν να συγκριθούν και με το z-τεστ.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} \quad \Longleftrightarrow \quad X^2 = \frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}$$

Παράδειγμα 3 (*συνέχεια*)

- Ο Στατιστικός Έλεγχος z (z-test)

  H_0: p_1-p_2=0

  H_1: p_1-$p_2 \neq 0$

  ή

  H_0: p_1=p_2

  H_1: $p_1 \neq p_2$

  Σε ε.σ. α

Παράδειγμα 3 (*συνέχεια*)

- Απορριπτική Περιοχή:

$$R = \left\{ |z| > z_{a/2} \right\}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{s}$$

$$s \approx \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Σύμφωνα με τη Μηδενική Υπόθεση τα δύο ποσοστά είναι ίσα και επομένως μπορούμε να συγχωνεύσουμε τα δύο δείγματα σε ένα και να υπολογίσουμε ένα κοινό

p (και *q*=1-*p*)

Τυπικό Σφάλμα της Διαφοράς των Δύο Ποσοστών

Παράδειγμα 3 (*συνέχεια*)

- Άλλη εκτίμηση του Τυπικού Σφάλματος της διαφοράς των δύο ποσοστών:

$$\boxed{\begin{aligned} s^2_{\hat{p}_1-\hat{p}_2} &= \frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2} \Rightarrow \\ s_{\hat{p}_1-\hat{p}_2} &= \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} \end{aligned}}$$

Απόφαση-Συμπέρασμα

- Η μηδενική υπόθεση δεν απορρίπτεται σε ε.σ. α=0,05.
- Με βάση τα διαθέσιμα δεδομένα δεν μπορεί να απορριφθεί η μηδενική υπόθεση σε ε.σ. α=0,05.
- Τα ποσοστά ριζοβολίας των 2 φυλλοφόρων μοσχευμάτων ελιάς "είναι ίσα" μεταξύ τους. Στις δύο ποικιλίες τα αντίστοιχα ποσοστά δεν διαφέρουν στατιστικά σημαντικά σε επίπεδο σημαντικότητας α=0,05.

Παρατηρήσεις

- Η διόρθωση συνέχειας του *Yates* γίνεται διότι προσπαθούμε να προσεγγίσουμε μια συνεχή κατανομή (X^2) μέσω μιας διακριτής κατανομής (Διωνυμική).
- Η διόρθωση του *Yates* δεν συνιστάται από όλους τους στατιστικούς.
- Η σύγκριση ποσοστών με τον έλεγχο X^2 για να είναι ικανοποιητική θα πρέπει όλα τα περιθώρια αθροίσματα-σύνολα του 2×2 πίνακα συμπτώσεων να είναι μεγαλύτερα από 15.

Παρατηρήσεις (*συνέχεια*)

- Σε κάθε περίπτωση, όλες οι τεχνικές και πιθανοθεωρητικές προϋποθέσεις εφαρμογής του ελέγχου X^2 πρέπει να ελέγχονται.
- Όλοι οι περιορισμοί παύουν να έχουν σημαντική επίδραση στα αποτελέσματα του ελέγχου όταν η παρατηρούμενη στάθμη σημαντικότητας (*p*-value) υπολογίζεται είτε με την Ακριβή Μέθοδο (*Exact Method*) είτε με τη μέθοδο προσομοίωσης *Monte-Carlo*.
- Στην ειδική περίπτωση των 2×2 πινάκων όπου τα δείγματα είναι σχετικά μικρά εφαρμόζεται ο Ακριβής Έλεγχος του *Fisher* (*Fisher's Exact Test*).

Ειδικές Εφαρμογές του Ελέγχου X^2

Άλλες Εφαρμογές του Ελέγχου

- Έλεγχοι Ανεξαρτησίας ή Συνάφειας
- Έλεγχοι Ομοιογένειας
- Έλεγχοι Καλής Προσαρμογής

Έλεγχοι Ανεξαρτησίας ή Συνάφειας

- Σε πρόσφατη έρευνα ρωτήθηκαν κάποια άτομα σχετικά με το ποιο θεωρούν ως το σημαντικότερο πρόβλημα σήμερα.
- Οι απαντήσεις στην ερώτηση αυτή διασταυρώθηκαν με τις διάφορες ηλικιακές κατηγορίες των ατόμων που συμμετείχαν στην έρευνα.
- Τα αποτελέσματα δίνονται στον πίνακα που ακολουθεί:

Πίνακας Συμπτώσεων

Κλάσεις Ηλικιών	Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία
18-24	24	4	23	82	12
25-34	42	2	20	146	13
35-44	54	0	12	136	18
45-54	40	1	6	121	9
55-64	37	1	7	68	15
65+	20	5	3	61	13

Πηγή: Μπεχράκης (1999)

Ερευνητικό Ερώτημα

- Υπάρχει σχέση (συσχέτιση, συνάφεια) μεταξύ της ηλικίας των ερωτώμενων και της άποψή τους σχετικά με το ποιο είναι το σημαντικότερο πρόβλημα (α=0,05);

Στατιστικός Έλεγχος Υπόθεσης:

H_0: Δεν υπάρχει σχέση μεταξύ των δύο χαρακτηριστικών (ηλικία, πρόβλημα)

H_1: Υπάρχει σχέση

Προφίλ Γραμμών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% within Κλάσεις Ηλικιών

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	
Κλάσεις Ηλικιών	18-24 ετών	16.6%	2.8%	15.9%	56.6%	8.3%	100.0%
	25-34	18.8%	.9%	9.0%	65.5%	5.8%	100.0%
	35-44	24.5%		5.5%	61.8%	8.2%	100.0%
	45-54	22.6%	.6%	3.4%	68.4%	5.1%	100.0%
	55-64	28.9%	.8%	5.5%	53.1%	11.7%	100.0%
	65+	19.6%	4.9%	2.9%	59.8%	12.7%	100.0%
Total		21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Προφίλ Στηλών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% within Πρόβλημα

		Πρόβλημα					
		Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	Total
Κλάσεις Ηλικιών	18-24 ετών	11.1%	30.8%	32.4%	13.4%	15.0%	14.6%
	25-34	19.4%	15.4%	28.2%	23.8%	16.3%	22.4%
	35-44	24.9%		16.9%	22.1%	22.5%	22.1%
	45-54	18.4%	7.7%	8.5%	19.7%	11.3%	17.8%
	55-64	17.1%	7.7%	9.9%	11.1%	18.8%	12.9%
	65+	9.2%	38.5%	4.2%	9.9%	16.3%	10.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Πίνακας Αντιστοιχιών

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

% of Total

		Πρόβλημα					Total
		Τζόγος	Στεγαστικό	Σπουδές	Οικον ομικό	Υ γεία	
Κλάσεις Ηλικιών	18-24 ετών	2.4%	.4%	2.3%	8.2%	1.2%	14.6%
	25-34	4.2%	.2%	2.0%	14.7%	1.3%	22.4%
	35-44	5.4%		1.2%	13.7%	1.8%	22.1%
	45-54	4.0%	.1%	.6%	12.2%	.9%	17.8%
	55-64	3.7%	.1%	.7%	6.8%	1.5%	12.9%
	65+	2.0%	.5%	.3%	6.1%	1.3%	10.3%
Total		21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Αποτελέσματα Ελέγχου

Sig.→*p*-value

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
					99% Confidence Interval			99% Confidence Interval	
				Sig.	Lower Bound	Upper Bound	Sig.	Lower Bound	Upper Bound
Pearson Chi-Square	59.480[a]	20	.000	.000[b]	.000	.000			
Likelihood Ratio	55.136	20	.000	.000[b]	.000	.000			
Fisher's Exact Test	52.065			.000[b]	.000	.000			
Linear-by-Linear Association	.362[c]	1	.548	.563[b]	.550	.575	.291[b]	.279	.303
N of Valid Cases	995								

a. 6 cells (20.0%) have expected count less than 5. The minimum expected count is 1.33.

b. Based on 10000 sampled tables with starting seed 272886377.

c. The standardized statistic is -.601.

Αφού *p*<0,05 η Μηδενική Υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας *α*=0,05

Συμπέρασμα

- Υπάρχει στατιστικά σημαντική συσχέτιση, σε επίπεδο σημαντικότητας $α$=0,05, μεταξύ της ηλικίας των ερωτώμενων και της άποψής τους σχετικά με το ποιο είναι το σημαντικότερο πρόβλημα.

Ποια είναι η ένταση-βαθμός της συνάφειας;

Symmetric Measures

		Value	Approx. Sig.	Monte Carlo Sig.		
					99% Confidence Interval	
				Sig.	Lower Bound	Upper Bound
Nominal by Nominal	Phi	.244	.000	.000[c]	.000	.000
	Cramer's V	.122	.000	.000[c]	.000	.000
N of Valid Cases		995				

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on 10000 sampled tables with starting seed 1090229469.

Η τιμή του δείκτη συνάφειας *V* του Cramer μαρτυρά ασθενούς εντάσεως συσχέτιση μεταξύ των δύο χαρακτηριστικών που διασταυρώνονται

Ο δείκτης *V* του *Cramer*

$$V = \sqrt{\frac{X^2}{Np}}$$

όπου $p = min(k-1, l-1)$.

k **είναι ο αριθμός των γραμμών και *l* ο αριθμός των στηλών του πίνακα συμπτώσεων**

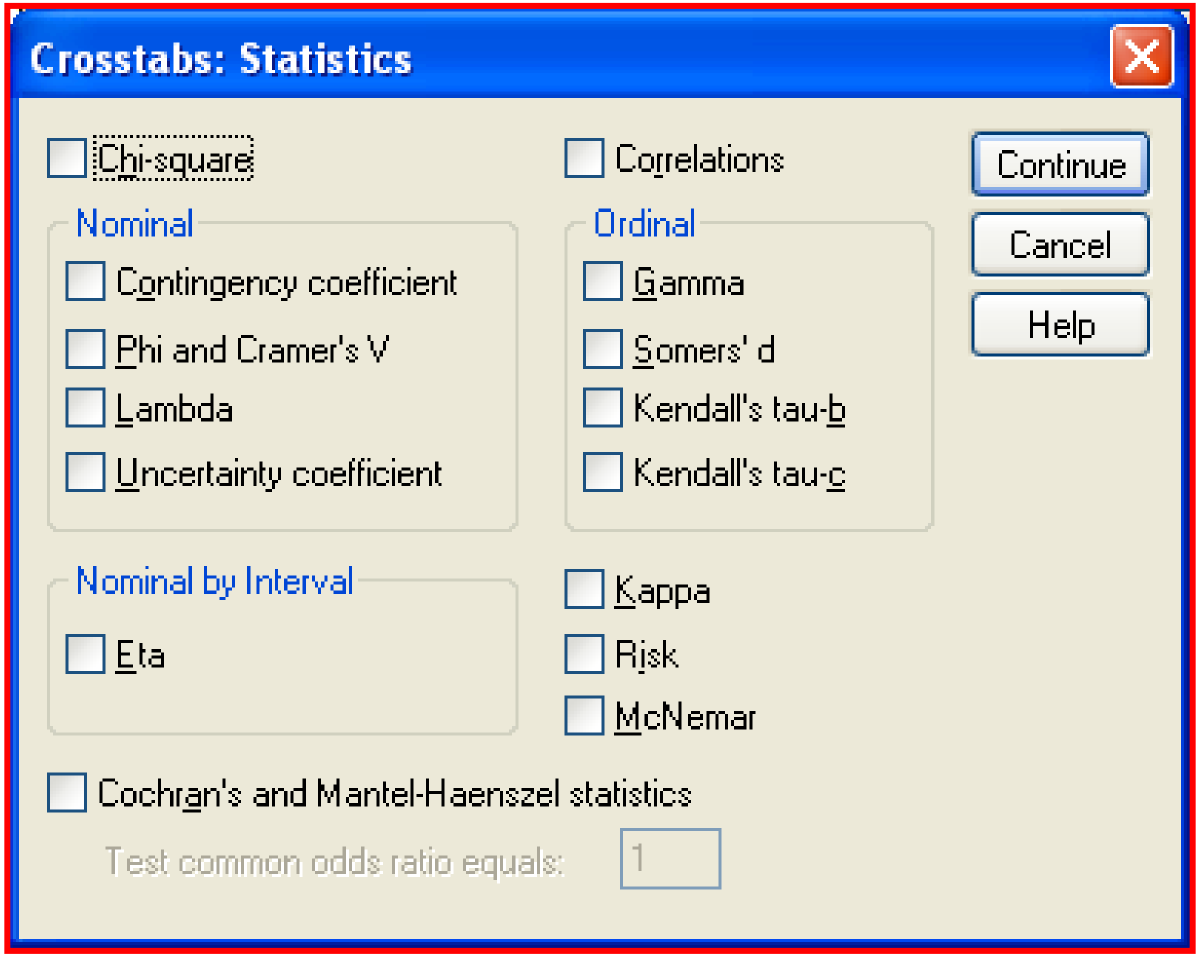
Παίρνει τιμές στο διάστημα [0, 1]

Ο δείκτης V του *Cramer (συνέχεια)*

- 0-0,10 ➔ Κλινικά, Βιολογικά, Πρακτικά Ασήμαντη Συνάφεια
- 0,10-0,20 ➔ Ασθενής
- 0,20-0,40 ➔ Μέτρια
- 0,40-0,60 ➔ Ισχυρή
- 0,60-0,80 ➔ Πολύ Ισχυρή
- 0,80-1,0 ➔ Πάρα πολύ Ισχυρή έως Απόλυτη

Οι Νόρμες αυτές είναι γενικές. Το τι είναι βιολογικά σημαντικό εξαρτάται κάθε φορά από το πεδίο έρευνας και τα χαρακτηριστικά που εξετάζονται.

Άλλοι Δείκτες Συνάφειας-Συσχέτισης στο SPSS



Που οφείλεται κυρίως η συνάφεια των δύο χαρακτηριστικών;

Κλάσεις Ηλικιών * Πρόβλημα Crosstabulation

			Πρόβλημα					Total
			Τζόγος	Στεγαστικό	Σπουδές	Οικον ομικό	Υ γεία	
Κλάσεις Ηλικιών	18-24 ετών	% within Κλάσεις Ηλικιών	16.6%	2.8%	15.9%	56.6%	8.3%	100.0%
		Adjusted Residual	-1.7	1.7	4.4	-1.4	.1	
	25-34	% within Κλάσεις Ηλικιών	18.8%	.9%	9.0%	65.5%	5.8%	100.0%
		Adjusted Residual	-1.2	-.6	1.2	1.3	-1.4	
	35-44	% within Κλάσεις Ηλικιών	24.5%	.0%	5.5%	61.8%	8.2%	100.0%
		Adjusted Residual	1.1	-1.9	-1.1	.0	.1	
	45-54	% within Κλάσεις Ηλικιών	22.6%	.6%	3.4%	68.4%	5.1%	100.0%
		Adjusted Residual	.3	-1.0	-2.1	2.0	-1.6	
	55-64	% within Κλάσεις Ηλικιών	28.9%	.8%	5.5%	53.1%	11.7%	100.0%
		Adjusted Residual	2.1	-.6	-.8	-2.1	1.6	
	65+	% within Κλάσεις Ηλικιών	19.6%	4.9%	2.9%	59.8%	12.7%	100.0%
		Adjusted Residual	-.6	3.4	-1.7	-.4	1.8	
Total		% within Κλάσεις Ηλικιών	21.8%	1.3%	7.1%	61.7%	8.0%	100.0%

Adjusted Residual:

Διορθωμένο Τυποποιημένο Υπόλοιπο

Διορθωμένα Τυποποιημένα Υπόλοιπα-ΔΤΥ

- Κελιά με όπου τα αντίστοιχα Διορθωμένα Τυποποιημένα Υπόλοιπα είναι σε απόλυτη τιμή μεγαλύτερη του 1,96≈2 συνεισφέρουν στατιστικά σημαντικά, σε ε.σ. α=0,05, στη σημαντικότητα του στατιστικού X^2 και σε αυτά τα κελιά οφείλεται, κυρίως, η συνάφεια ή η αλληλεπίδραση των δύο μεταβλητών.
- Η τιμή 1,96 αντιστοιχεί στην κρίσιμη τιμή της Τυποποιημένης Κανονικής Κατανομής για α/2=0,025. Αν θέσουμε το α=0,01, τότε η τιμή σύγκρισης των τυποποιημένων υπολοίπων είναι 2,58, ενώ αυτή μειώνεται σε 1,64 για α=0,10.

ΔΤΥ (*συνέχεια*)

- Έχουν μέση τιμή μηδέν και διακύμανση ίση με τη μονάδα και η ασυμπτωτική τους συμπεριφορά προσεγγίζει καλύτερα την Τυποποιημένη Κανονική Κατανομή. Υπολογιστικά τα διορθωμένα τυποποιημένα υπόλοιπα δίνονται από τη σχέση:

$$\boxed{\frac{f_{ij} - \dfrac{f_{i+} f_{+j}}{N}}{\sqrt{\dfrac{f_{i+} f_{+j}}{N}\left(1-\dfrac{f_{i+}}{N}\right)\left(1-\dfrac{f_{+j}}{N}\right)}}}$$

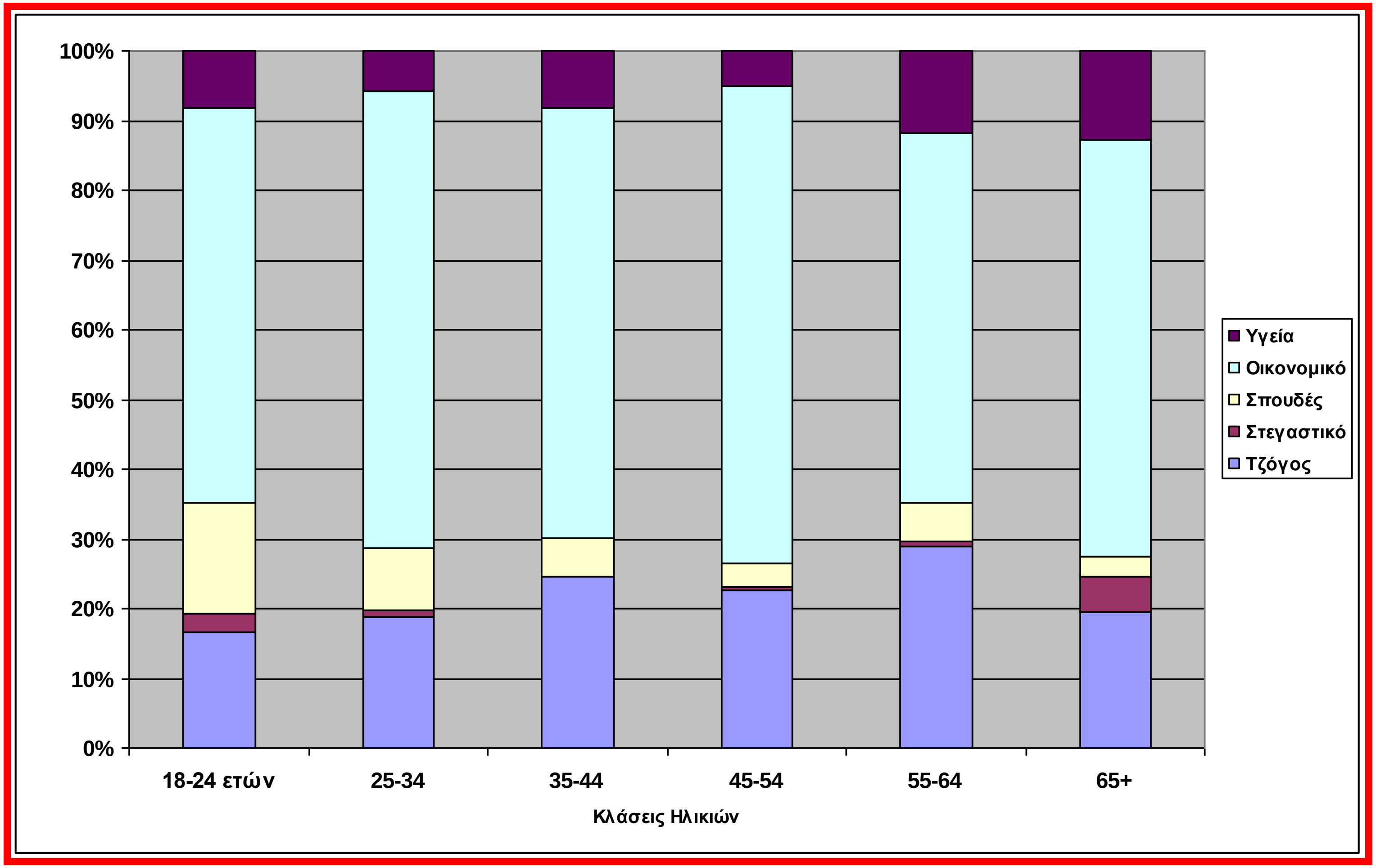
όπου f_{i+} και f_{+j} είναι οι περιθώριες συχνότητες της γραμμής i και της στήλης j αντίστοιχα του πίνακα συμπτώσεων **F** με γενικό στοιχείο f_{ij} $\left(i=1,\dots k,\ j=1,\dots,l\right)$.

ΔΤΥ (*συνέχεια*)

Το πρόσημο των ΔΤΥ έχει την ακόλουθη φυσική ερμηνεία:

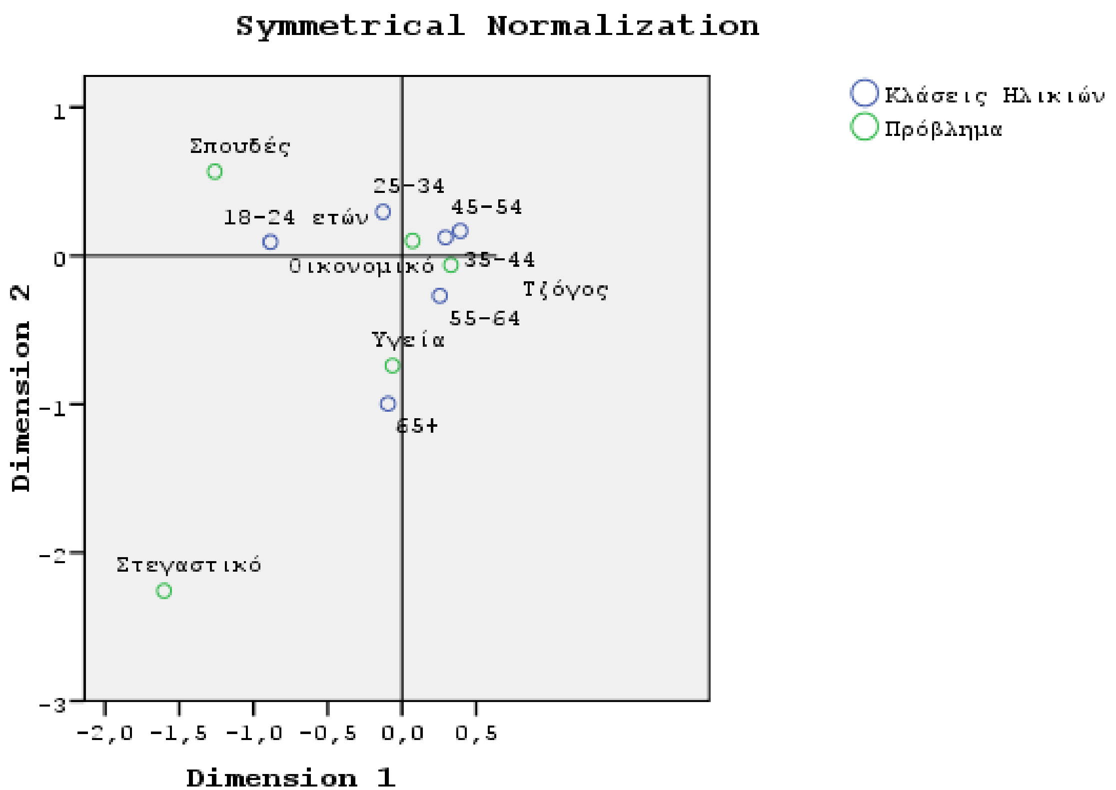
- Αν σε κάποιο κελί το αντίστοιχο ΔΤΥ είναι σε απόλυτη τιμή μεγαλύτερο του 2 και έχει αρνητικό πρόσημο, αυτό σημαίνει ότι στο συγκεκριμένο κελί υπάρχουν στατιστικά σημαντικά λιγότερες παρατηρήσεις (f_{ij}) σε σύγκριση με αυτές που αναμένονται κάτω από την υπόθεση της ανεξαρτησίας των δύο μεταβλητών.
- Αν σε κάποιο κελί το αντίστοιχο ΔΤΥ είναι σε απόλυτη τιμή μεγαλύτερο του 2 και έχει θετικό πρόσημο, τότε στο συγκεκριμένο κελί υπάρχουν στατιστικά σημαντικά περισσότερες παρατηρήσεις (f_{ij}) σε σχέση με το αν οι δύο μεταβλητές ήταν ανεξάρτητες.

Γραφική Απεικόνιση της Συσχέτισης (1)



Γραφική Απεικόνιση της Συσχέτισης (2)

Παραγοντικό Επίπεδο της Correspondence Analysis



Έλεγχοι Ομοιογένειας

- Όπως στα Παραδείγματα 1 και 2.
- Η διαφορά μεταξύ των ελέγχων ανεξαρτησίας και ελέγχων ομοιογένειας έγκειται στο γεγονός ότι στους δεύτερους τα σύνολα-αθροίσματα των γραμμών του αντίστοιχου πίνακα είναι προκαθορισμένα και όχι τυχαίες μεταβλητές όπως στους πρώτους.

Έλεγχοι Ομοιογένειας (*συνέχεια*)

- Ρωτήθηκαν 300 κληρικοί, 250 εκπαιδευτικοί και 300 Ιδ. υπάλληλοι σχετικά με το εάν καταναλώνουν αλκοόλ.

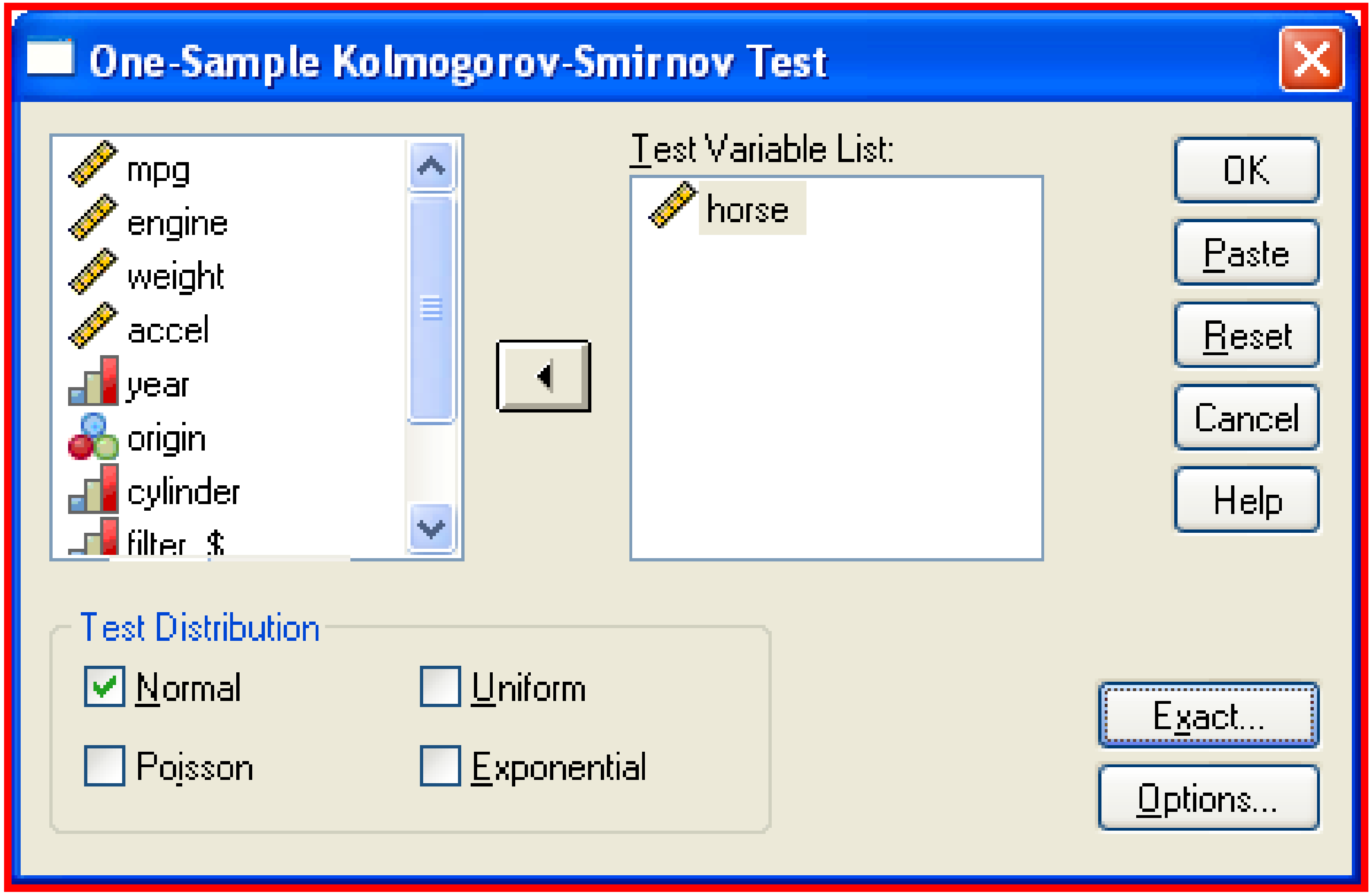
	ΝΑΙ	ΟΧΙ	Σύνολο
Κληρικοί	32	268	300
Εκπαιδευτικοί	51	199	250
Ιδ. Υπάλληλοι	67	233	300
Σύνολο	150	700	850

Τα σύνολα γραμμών προκαθορισμένα π.χ. λόγω Στρωματοποιημένης Τυχαίας Δειγματοληψίας

Έλεγχοι Καλής Προσαρμογής

Το πρόβλημα ελέγχου της καλής προσαρμογής μιας θεωρητικής κατανομής σε μια δειγματική ανάγεται στη λήψη απόφασης σχετικά με το εάν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των τιμών του πληθυσμού και του δείγματος.

Έλεγχοι Καλής Προσαρμογής στο SPSS (1)



Αποτελέσματα

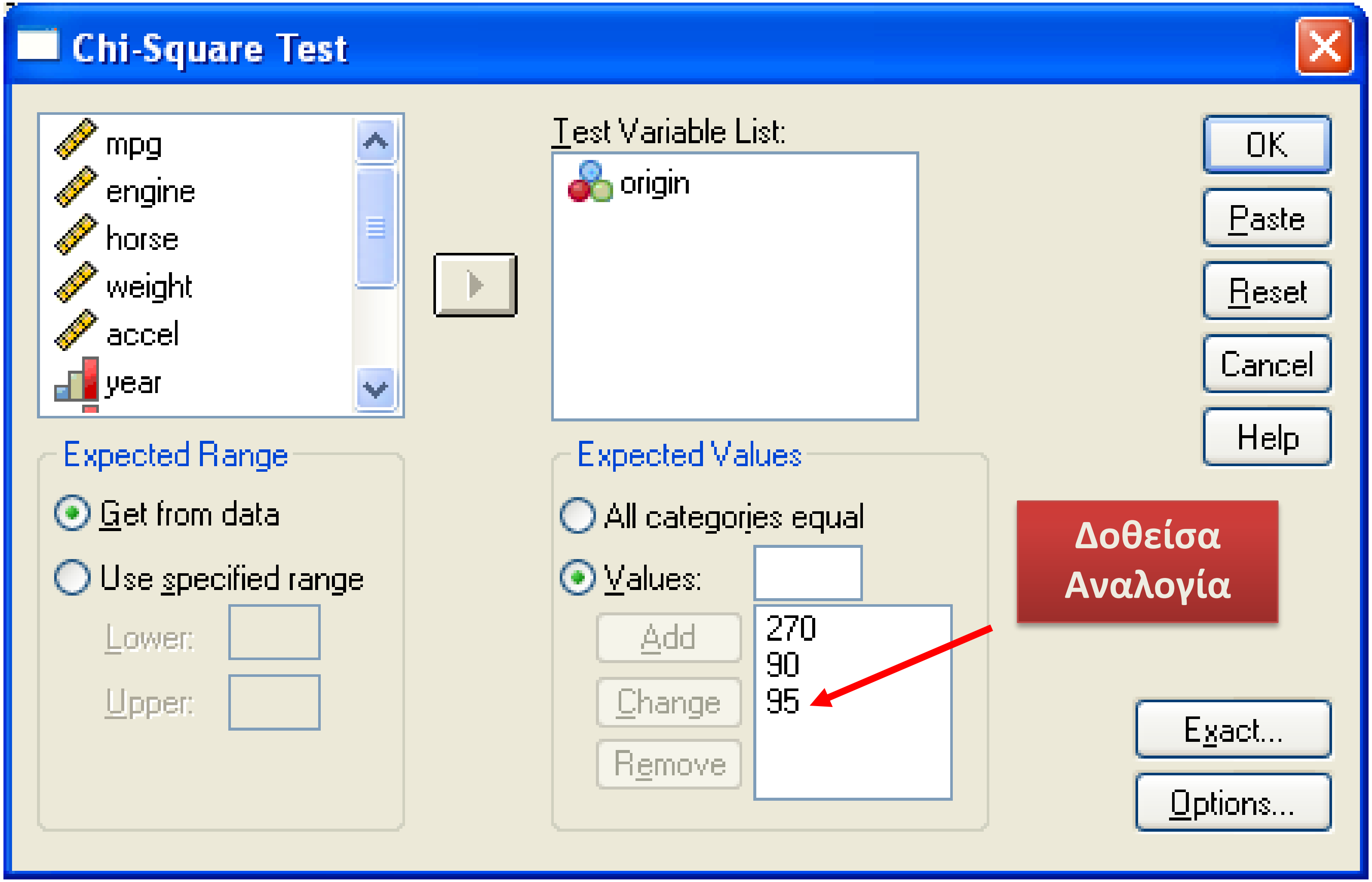
One-Sample Kolmogorov-Smirnov Test

			Horsepower
N			400
Normal Parameters[a,b]	Mean		104.83
	Std. Deviation		38.522
Most Extreme Differences	Absolute		.160
	Positive		.160
	Negative		-.082
Kolmogorov-Smirnov Z			3.198
Asymp. Sig. (2-tailed)			.000
Monte Carlo Sig. (2-tailed)	Sig.		.000[c]
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Test distribution is Normal.

Αφού *p*<0,05 η Μηδενική Υπόθεση Απορρίπτεται, τα δεδομένα ΔΕΝ προσαρμόζονται στην Κανονική Κατανομή

Έλεγχοι Καλής Προσαρμογής στο SPSS (2)



Αποτελέσματα

Test Statistics

			Country of Origin
Chi-Square[a]			1.665
df			2
Asymp. Sig.			.435
Monte Carlo Sig.	Sig.		.447[b]
	99% Confidence Interval	Lower Bound	.434
		Upper Bound	.460

Αφού *p*>0,05 η Μηδενική Υπόθεση ΔΕΝ Απορρίπτεται, τα δεδομένα προσαρμόζονται στην δοθείσα αναλογία

Παράδειγμα 4

Σύγκριση μιας δεδομένης κατανομής συχνότητας με μια θεωρητική/

Η πιθανότητα ένα νεογέννητο ζώο να είναι αρσενικό ή θηλυκό είναι ½, εφόσον τουλάχιστον δεν υπάρχουν κληρονομικές ανωμαλίες. Στον πίνακα που ακολουθεί δίνεται η συχνότητα των τοκετών 5 νεογνών ινδικών χοιριδίων που έδωσαν από 0 έως 5 αρσενικά. Να ελεγχθεί η μηδενική υπόθεση ότι η κατανομή του αριθμού των αρσενικών είναι η Διωνυμική με παράμετρο $p=½$ ($α=0,05$).

Παράδειγμα 4 (*συνέχεια*)

Αριθμός Αρρένων Ινδικών Χοιριδίων σε Τοκετούς των Πέντε Νεογνών

Αριθμός Αρρένων	Αριθμός Τοκετών	Θεωρητική Πιθανότητα	Αναμενόμενη Συχνότητα
0	1	0,03125	2,72
1	12	0,15625	13,59
2	27	0,31250	27,19
3	32	0,31250	27,19
4	12	0,15625	13,59
5	3	0,03125	2,72
Σύνολο	87	1,0000	87

???

Παράδειγμα 4 (*συνέχεια*)

- Θεωρητικές Πιθανότητες:

  Υπολογίζονται με βάση τη Διωνυμική Κατανομή για *n*=5 και *p*=1/2.

$$P(X=0)=\binom{5}{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{5}=0,03125$$

$$P(X=1=\binom{5}{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{4}=0,15625$$

$$\kappa.\lambda\pi.$$

Παράδειγμα 4 (*συνέχεια*)

- Αναμενόμενες Συχνότητες:

  Πολλαπλασιάζοντας τις θεωρητικές συχνότητες (πιθανότητες) με το σύνολο των τοκετών (87 συνολικά τοκετοί).

  Για παράδειγμα:

  $2,72=0,03125\times 87$

  $13,59=0,15625\times 87$

Παράδειγμα 4 (*συνέχεια*)

Στατιστικός Έλεγχος

- H_0: Δεν υπάρχει διαφορά μεταξύ δειγματικών και θεωρητικών τιμών.
- H_0: Τα δεδομένα προσαρμόζονται στη Διωνυμική Κατανομή
- H_0: Η κατανομή του αριθμού των αρσενικών είναι η Διωνυμική με παράμετρο p=½.
- H_1: όχι η H_0, δηλ. υπάρχει διαφορά μεταξύ δειγματικών και θεωρητικών τιμών, τα δεδομένα δεν προσαρμόζονται στη Διωνυμική Κατανομή, η κατανομή του αριθμού των αρσενικών δεν είναι η Διωνυμική με παράμετρο p=½.

Παράδειγμα 4 (*συνέχεια*)

- Υπολογίζουμε το στατιστικό X^2:

$$\boxed{X^2 = \frac{\left(1-2,72\right)^2}{2,72} + \frac{\left(12-13,59\right)^2}{13,59} + \cdots + \frac{\left(3-2,72\right)^2}{2,72} = 2,34}$$

- Συγκρίνουμε το στατιστικό X^2=2,34 με την κρίσιμη της X^2 Κατανομής με 6-1=5 β.ε. σε επίπεδο σημαντικότητας α=0,05.

Παράδειγμα 4 (*συνέχεια*)

- $X^2(5)_{0,05}=11,07$ (θεωρητικό, κρίσιμη τιμή)
- $X^2=2,34$ (δειγματικό)
- $X^2=2,34<11,07=X^2(5)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Η μηδενική Υπόθεση παραμένει

Παράδειγμα 4 (*συνέχεια*)

Απόφαση-Συμπέρασμα:

- Με βάση τα διαθέσιμα δεδομένα η μηδενική υπόθεση δεν μπορεί να απορριφθεί σε ε.σ. α=0,05.
- Δεν ανιχνεύθηκαν στατιστικά σημαντικές διαφορές μεταξύ θεωρητικών και δειγματικών τιμών.
- Δεν έχουμε λόγους να αμφιβάλλουμε ότι η κατανομή του αριθμού των αρρένων είναι η Διωνυμική (σε ε.σ. α=0,05).

Παράδειγμα 5

Κατανομή των Αποδόσεων σε χλγ. Χλωρού Βάρους 168 Μεμονωμένων Φυτών Μηδικής

Πραγματικά Όρια Κλάσεων	Αριθμός Φυτών	Θεωρητική Πιθανότητα	Αναμενόμενη Συχνότητα
<0,55	0	0,0154	2,59
0,55-1,05	5	0,0247	4,15
1,05-1,55	8	0,0500	8,40
1,55-2,05	20	0,0861	14,46
2,05-2,55	20	0,1253	21,05
2,55-3,05	27	0,1547	25,99
3,05-3,55	30	0,1617	27,17
3,55-4,05	20	0,1432	24,06
4,05-4,55	13	0,1075	18,06
4,55-5,05	11	0,0684	11,49
5,05-5,55	7	0,0368	6,18
5,55-6,05	7	0,0168	2,82
>6,05	0	0,0094	1,58
Σύνολο	168	1,0000	168,00

Παράδειγμα 5 (*συνέχεια*)

Να ελεγχθεί η υπόθεση ότι η απόδοση των φυτών κατανέμεται κανονικά (σε ε.σ. α=0,05).

Δίνονται:

$$\boxed{\begin{aligned} \bar{X} &= 3{,}18 \\ s &= 1{,}22 \end{aligned}}$$

Στην πράξη υπολογίζονται από το δείγμα

Παράδειγμα 5 (*συνέχεια*)

Στατιστικός Έλεγχος:

- H_0: Η απόδοση των φυτών κατανέμεται κανονικά (ακολουθεί την Κανονική Κατανομή)
- H_1: Η απόδοση των φυτών δεν κατανέμεται κανονικά (δεν ακολουθεί την Κανονική Κατανομή)

σε ε.σ. α=0,05

Παράδειγμα 5 (*συνέχεια*)

- Θεωρητικές Πιθανότητες:

  Υπολογίζονται με τη βοήθεια της Τυποποιημένης Κανονικής Κατανομής.

  Για παράδειγμα:

$$P(Y \leq 0,55) = P\left(\frac{Y-3,18}{1,22} \leq \frac{0,55-3,18}{1,22}\right) = P(z \leq -2,16) =$$

$$= 0,5000 - 0,4846 = 0,0154$$

$$P(0,55 \leq Y \leq 1,55) = P\left(\frac{0,55-3,18}{1,22} \leq \frac{Y-3,18}{1,22} \leq \frac{1,55-3,18}{1,22}\right) =$$

$$= P(-2,16 \leq z \leq -1,75) = 0,4846 - 0,4599 = 0,0247$$

Παράδειγμα 5 (*συνέχεια*)

Πίνακας

Πιθανοτήτων $P(0<Z<z)$ *για την κανονική κατανομή* $N(0, 1)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Παράδειγμα 5 (*συνέχεια*)

- Προσοχή χρειάζεται στην κλάση 3,05-3,55 η οποία περιέχει το δειγματικό μέσο όρο. Στην περίπτωση αυτή προσθέτουμε το εμβαδό από 3,05 έως 3,18 και από 3,18 έως 3,55.

Παράδειγμα 5 (*συνέχεια*)

- Αναμενόμενες Συχνότητες:

  Πολλαπλασιάζοντας τις θεωρητικές συχνότητες (πιθανότητες) με το σύνολο των φυτών του δείγματος (168 συνολικά φυτά).

  Για παράδειγμα:

  $2{,}59=168\times0{,}0154$

  $4{,}15=168\times0{,}0247$

Παράδειγμα 5 (*συνέχεια*)

- Οι βαθμοί ελευθερίας είναι όσες οι κλάσεις μείον ένα, μείον το πλήθος των παραμέτρων που εκτιμήσαμε από το δείγμα (δηλ. 2, ο μέσος όρος και η τυπική απόκλιση). Συνεπώς, β.ε.=13-1-2=10.

Πόσοι είναι οι βαθμοί ελευθερίας στην περίπτωση που δίνονται οι παράμετροι *μ* και *σ*;

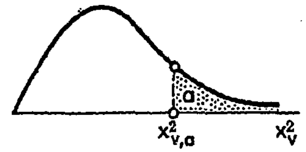
Παράδειγμα 5 (*συνέχεια*)

- Υπολογίζουμε το στατιστικό X^2:

$$\boxed{X^2 = \frac{(0-2,59)^2}{2,59} + \frac{(5-4,15)^2}{4,15} + \cdots + \frac{(0-1,58)^2}{1,58} = 15,30}$$

- Συγκρίνουμε το στατιστικό X^2=15,30 με την κρίσιμη της X^2 Κατανομής με 10 β.ε. σε επίπεδο σημαντικότητας *α*=0,05.

ΠΙΝΑΚΑΣ 7: Κριτικές τιμές της κατανομής χι-τετράγωνο



ν	α									
	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Παράδειγμα 5 (*συνέχεια*)

- $X^2(10)_{0,05}$=18,31 (θεωρητικό, κρίσιμη τιμή)
- X^2=15,30 (δειγματικό)
- $X^2=15,30<18,31=X^2(10)_{0,05}$
- Δειγματικό X^2 < Θεωρητικό X^2

Η μηδενική Υπόθεση παραμένει

Παράδειγμα 5 (*συνέχεια*)

Απόφαση-Συμπέρασμα:

- Με βάση τα διαθέσιμα δεδομένα η μηδενική υπόθεση δεν μπορεί να απορριφθεί σε ε.σ. α=0,05.
- Δεν ανιχνεύθηκαν στατιστικά σημαντικές διαφορές μεταξύ θεωρητικών και δειγματικών τιμών.
- Δεν έχουμε λόγους να αμφιβάλλουμε ότι η κατανομή του χλωρού βάρους είναι η Κανονική (σε ε.σ. α=0,05).

Παρατηρήσεις

- Σε κάθε περίπτωση, θα πρέπει να γίνεται έλεγχος σχετικά με το αν οι αναμενόμενες συχνότητες είναι όλες μεγαλύτερες από 1 και το πολύ το 20% από αυτές είναι μικρότερες από 5 (**περιορισμός του *Cochran***), αλλιώς πρέπει να γίνει συγχώνευση γειτονικών κλάσεων.
- Στο Παράδειγμα 5 προστέθηκαν οι κλάσεις <0,55 και >6,05 ώστε το άθροισμα των θεωρητικών πιθανοτήτων να ίσο με τη μονάδα και συνεπώς το άθροισμα των αναμενόμενων συχνοτήτων ίσο με 168.

Μηχανισμός του Ελέγχου Ανεξαρτησίας και Ομοιογένειας

Υπενθύμιση

Ορισμός: **Έστω δύο ενδεχόμενα A και B με $P(A) \neq 0$ και $P(B) \neq 0$, λέμε ότι το ενδεχόμενο A είναι ανεξάρτητο του B όταν:**

$$\mathbf{P(A/B)=P(A)}$$

Επειδή όμως: $$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

έχουμε: $$\mathbf{P(A \cap B) = P(A) \cdot P(B)}$$

Η τελευταία σχέση, είναι και η πιο γνωστή σχέση που συνδέει δύο ανεξάρτητα ενδεχόμενα, αυτή δε γενικεύεται για ένα πεπερασμένο σύνολο ανεξάρτητων ενδεχομένων.

Αν για τα ενδεχόμενα $A_1, A_2, \ldots, A_n$ με πιθανότητες διάφορες του μηδενός, $P(A_1) \neq 0, P(A_2) \neq 0, \ldots, P(A_n) \neq 0$ ισχύει η:

$$\mathbf{P(A_1 \cap A_2 \cap A_3 ... \cap A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot ... \cdot P(A_n)}$$

τα ενδεχόμενα αυτά είναι ανεξάρτητα.

Δηλαδή για να διαπιστωθεί ότι ένα πλήθος ενδεχομένων είναι ανεξάρτητα μεταξύ τους, αρκεί να διαπιστώσουμε ότι η πιθανότητα της τομής τους είναι ίση με το γινόμενο των πιθανοτήτων τους

Μηχανισμός του Ελέγχου

- Η μηδενική υπόθεση H_0 είναι (βλ. Παράδειγμα 1):

$$\boxed{H_0 : p_{1j} = p_{2j} = p_{3j} = p_{4j} = p_{5j}, \quad j = 1, \ldots 4}$$

Η αναλογία (%) για την *j* κλάση γύρης στο πρώτο δείγμα μηδικής

Μηχανισμός του Ελέγχου (*συνέχεια*)

- Αν η μηδενική υπόθεση είναι αληθής τότε τα δύο χαρακτηριστικά, Δείγματα Μηδικής (Α) και Κλάσεις Γύρης (Β), είναι ανεξάρτητα.
- Αν η μηδενική υπόθεση είναι αληθής τότε η πιθανότητα P(A∩B)=P(A).P(B)

$$\mathrm{P}(\mathrm{A} = \Delta\varepsilon\acute{\iota}\gamma\mu\alpha 1) = \frac{r_1}{n}$$

$$\mathrm{P}(\mathrm{B} = \mathrm{K}\lambda\acute{\alpha}\sigma\eta 1) = \frac{c_1}{n}$$

$$\mathrm{P}(\Delta 1 \cap \mathrm{K}1) = \frac{r_1}{n} \times \frac{c_1}{n}$$

***r*₁: Σύνολο πρώτης γραμμής**

c_1: Σύνολο πρώτης στήλης

***n*: Γενικό Σύνολο**

$$\alpha\nu\alpha\mu\varepsilon\nu\acute{o}\mu\varepsilon\nu\eta\ \sigma\upsilon\chi\nu\acute{o}\tau\eta\tau\alpha = n \times \left(\frac{r_1}{n} \times \frac{c_1}{n}\right) = \frac{r_1 \times c_1}{n}$$

Χρήσιμοι Συμβολισμοί

Πίνακας Συμπτώσεων ή Συνάφειας

Κλάσεις της μεταβλητής X	Κλάσεις της μεταβλητής Y						Άθροισμα ή Περιθώρια Κατανομή της X
	1	2	…	j	…	l	
1	f_{11}	f_{12}	…	f_{1j}	…	f_{1l}	f_{1+}
2	f_{21}	f_{22}	…	f_{2j}	…	f_{2l}	f_{2+}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	f_{i1}	f_{i2}	…	f_{ij}	…	f_{il}	f_{i+}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	f_{k1}	f_{k2}	…	f_{kj}	…	f_{kl}	f_{k+}
Άθροισμα ή Περιθώρια Κατανομή της Y	f_{+1}	f_{+2}	…	f_{+j}	…	f_{+l}	N

Προφίλ Γραμμών

Κλάσεις της μεταβλητής X	Κλάσεις της μεταβλητής Y						Άθροισμα
	1	2	…	j	…	l	
1	$\frac{f_{11}}{f_{1+}}$	$\frac{f_{12}}{f_{1+}}$	…	$\frac{f_{1j}}{f_{1+}}$	…	$\frac{f_{1l}}{f_{1+}}$	1
2	$\frac{f_{21}}{f_{2+}}$	$\frac{f_{22}}{f_{2+}}$	…	$\frac{f_{2j}}{f_{2+}}$	…	$\frac{f_{2l}}{f_{2+}}$	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	$\frac{f_{i1}}{f_{i+}}$	$\frac{f_{i2}}{f_{i+}}$	…	$\frac{f_{ij}}{f_{i+}}$	…	$\frac{f_{il}}{f_{i+}}$	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	$\frac{f_{k1}}{f_{k+}}$	$\frac{f_{k2}}{f_{k+}}$	…	$\frac{f_{kj}}{f_{k+}}$	…	$\frac{f_{kl}}{f_{k+}}$	1
Μέσο προφίλ γραμμών ή Κέντρο βάρους γραμμών	$\frac{f_{+1}}{N}$	$\frac{f_{+2}}{N}$	…	$\frac{f_{+j}}{N}$	…	$\frac{f_{+l}}{N}$	

Προφίλ Στηλών

Κλάσεις της μεταβλητής X	Κλάσεις της μεταβλητής Y 1	2	…	j	…	l	Μέσο προφίλ στηλών ή Κέντρο βάρους στηλών
1	$\frac{f_{11}}{f_{+1}}$	$\frac{f_{12}}{f_{+2}}$	…	$\frac{f_{1j}}{f_{+j}}$	…	$\frac{f_{1l}}{f_{+l}}$	$\frac{f_{1+}}{N}$
2	$\frac{f_{21}}{f_{+1}}$	$\frac{f_{22}}{f_{+2}}$	…	$\frac{f_{2j}}{f_{+j}}$	…	$\frac{f_{2l}}{f_{+l}}$	$\frac{f_{2+}}{N}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	$\frac{f_{i1}}{f_{+1}}$	$\frac{f_{i2}}{f_{+2}}$	…	$\frac{f_{ij}}{f_{+j}}$	…	$\frac{f_{il}}{f_{+l}}$	$\frac{f_{i+}}{N}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	$\frac{f_{k1}}{f_{+1}}$	$\frac{f_{k2}}{f_{+2}}$	…	$\frac{f_{kj}}{f_{+j}}$	…	$\frac{f_{kl}}{f_{+l}}$	$\frac{f_{k+}}{N}$
Άθροισμα	1	1	…	1	…	1	

Μάζες Γραμμών και Στηλών

	Κλάσεις της μεταβλητής Y						
Κλάσεις της μεταβλητής X	1	2	…	j	…	l	Μάζες Γραμμών ή Σχετική Κατανομή της X ή Κέντρο Βάρους Στηλών
1	f_{11}	f_{12}	…	f_{1j}	…	f_{1l}	$\frac{f_{1+}}{N}=r_1$
2	f_{21}	f_{22}	…	f_{2j}	…	f_{2l}	$\frac{f_{2+}}{N}=r_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	f_{i1}	f_{i2}	…	f_{ij}	…	f_{il}	$\frac{f_{i+}}{N}=r_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	f_{k1}	f_{k2}	…	f_{kj}	…	f_{kl}	$\frac{f_{k+}}{N}=r_k$
Μάζες Στηλών ή Σχετική Κατανομή της Y ή Κέντρο Βάρους Γραμμών	$\frac{f_{+1}}{N}=c_1$	$\frac{f_{+2}}{N}=c_2$	…	$\frac{f_{+j}}{N}=c_j$	…	$\frac{f_{+l}}{N}=c_l$	Άθροισμα=1

Πίνακας Αντιστοιχιών

Κλάσεις της μεταβλητής X	Κλάσεις της μεταβλητής Y					
	1	2	…	j	…	l
1	f_{11}/N	f_{12}/N	…	f_{1j}/N	…	f_{1l}/N
2	f_{21}/N	f_{22}/N	…	f_{2j}/N	…	f_{2l}/N
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	f_{i1}/N	f_{i2}/N	…	f_{ij}/N	…	f_{il}/N
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	f_{k1}/N	f_{k2}/N	…	f_{kj}/N	…	f_{kl}/N

Βιβλιογραφία

- Φωτιάδης, Ν. (1995). *Εισαγωγή στη Στατιστική για Βιολογικές Επιστήμες*. Θεσσαλονίκη: University Studio Press.
- Κολυβά, Φ. και Μπόρα-Σέντα, Ε. (1995). *Στατιστική: Θεωρία-Εφαρμογές*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.
- Παπαδημητρίου, Γ. (2001). *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής.
- Μενεξές, Γ. (2007). Πειραματικοί Σχεδιασμοί στην Ανάλυση Δεδομένων. Διδακτορική Διατριβή που υποβλήθηκε στο Τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας.

- Κολυβά, Φ. και Μπόρα-Σέντα, Ε. (1995). *Στατιστική: Θεωρία-Εφαρμογές*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.

- Μπεχράκης, Θ. (1999). *Πολυδιάστατη Ανάλυση Δεδομένων: Μέθοδοι και Εφαρμογές*, Αθήνα: Εκδόσεις ΝΕΑ ΣΥΝΟΡΑ-Α. Α. ΛΙΒΑΝΗΣ.

Σημείωμα Αναφοράς

«Στατιστική. Ο Έλεγχος X^2». Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: https://opencourses.auth.gr/courses/OCRS484/

Σημείωμα Αδειοδότησης

Τέλος Ενότητας

Επεξεργασία: Μαρία Αλεμπάκη

Θεσσαλονίκη, Εαρινό Εξάμηνο 2014-2015

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης