



Δημοκρίτειο Πανεπιστήμιο Θράκης

Μεταπτυχιακό Πρόγραμμα:  
Διοίκηση Τουριστικών Επιχειρήσεων  
και  
Οργανισμών για Στελέχη  
(Executive MBA in Tourism)

by

Dr. Efstathios Dimitriadis

Mathematic

Ph.D in Applied Statistics

M.Sc in Statistics and Demography

M.Sc in Quality Assurance

# $\chi^2$ Έλεγχος Ανεξαρτησίας

Ο  $\chi^2$  έλεγχος ανεξαρτησίας, μπορεί να χρησιμοποιηθεί για τον έλεγχο της υπόθεσης που αφορά τη σχέση δύο ποιοτικών μεταβλητών (ονομαστικής ή/ και ιεραρχικής κλίμακας) στον πληθυσμό.

- ▣ Η **Μηδενική Υπόθεση ( $H_0$ )** δηλώνει ότι δεν υπάρχει συσχέτιση μεταξύ των δύο διασταυρούμενων μεταβλητών στον πληθυσμό. **Στατιστικά Ανεξάρτητες.**
- ▣ Η **Εναλλακτική Υπόθεση ( $H_1$ )** προτείνει ότι οι δύο μεταβλητές σχετίζονται στον πληθυσμό και επομένως οι μεταβλητές **Εξαρτώνται Στατιστικά.**

# $\chi^2$ Έλεγχος Ανεξαρτησίας

- ▣ Ο έλεγχος ανεξαρτησίας  $\chi^2$  προβάλλει τα δεδομένα ως ένα δείγμα στο οποίο κάθε άτομο ταξινομείται σε δύο διαφορετικές μεταβλητές..
- ▣ Τα δεδομένα παρουσιάζονται συνήθως σε ένα πίνακα με τις κατηγορίες της μίας μεταβλητής να καθορίζουν τις **σειρές** και τις κατηγορίες της δεύτερης μεταβλητής να καθορίζουν τις **στήλες**.
- ▣ Τα δεδομένα, που ονομάζονται **παρατηρούμενες συχνότητες**, δείχνουν απλά πόσα άτομα από το δείγμα βρίσκονται σε κάθε κελί της μήτρας.
- ▣ Η μηδενική υπόθεση χρησιμοποιείται για την κατασκευή μιας εξιδανικευμένης κατανομής δείγματος των **αναμενόμενων συχνοτήτων** που περιγράφει τον τρόπο εμφάνισης του δείγματος εάν τα δεδομένα ήταν σε απόλυτη συμφωνία με την μηδενική υπόθεση.

## *Αναμενόμενη Συχνότητα -Expected Frequency*

Για τον έλεγχο ανεξαρτησίας, η *αναμενόμενη συχνότητα (expected frequency)* για κάθε κελί του πίνακα δίνεται

από τον τύπο: 
$$e_{ij} = \frac{(i.\text{row total}) * (j.\text{column total})}{N}$$

Για να αποκτήσετε τις αναμενόμενες συχνότητες για οποιοδήποτε κελί σε οποιαδήποτε διασταύρωση στην οποία οι δύο μεταβλητές θεωρούνται ανεξάρτητες, *πολλαπλασιάστε* τα *σύνολα γραμμών και στηλών* αυτού του κελιού και *διαιρέστε* το αποτέλεσμα με τον *συνολικό αριθμό* περιπτώσεων στον πίνακα.

# Pearson's Chi-Square

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad \begin{array}{l} \text{Όπου: } f_{ij} = \text{παρατηρούμενες συχνότητες} \\ e_{ij} = \text{αναμενόμενες συχνότητες} \end{array}$$

d.f = (r-1)(k-1)     r = πλήθος γραμμών  
k = πλήθος στηλών

Για 2\*2 πίνακα με  $N < 40$ , αν υπάρχουν  $e_{ij} < 10$  ή 5, προτείνεται η διόρθωση του Yates, αν και δεν είναι ευρέως αποδεκτή.

$$\chi_{yates}^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(|f_{ij} - e_{ij}| - 0,5)^2}{e_{ij}}$$

Για μεγάλο πίνακα, όλες οι αναμενόμενες συχνότητες  $> 1$  και όχι περισσότερο από το 20% του συνόλου των κελιών να έχουν αναμενόμενες συχνότητες  $< 5$ .

**Κανόνας Απόρριψης:**  $\chi^2 > \chi_{\alpha, d.f}^2$

# Παράδειγμα 1<sup>ο</sup>

Ερευνητές σε μια κοινότητα της Καλιφόρνια ζήτησαν από ένα δείγμα 175 ιδιοκτητών αυτοκινήτων να επιλέξουν το αγαπημένο τους από τρία δημοφιλή περιοδικά αυτοκινήτων. Από τους 111 κατόχους εισαγόμενων αυτοκινήτων στο δείγμα, 54 επιλέγουν το Αυτοκίνητο και Οδηγός, 25 το Τάσεις Αυτοκινήτου και 32 το 4-τροχοί. Από τους 64 ιδιοκτήτες εγχώριων αυτοκινήτων στο δείγμα, 19 επιλέγουν το Αυτοκίνητο και Οδηγός, 22 το Τάσεις Αυτοκινήτου και 23 το 4-τροχοί. Σε επίπεδο σημαντικότητας 5%, η κατοχή εισαγόμενου ή εγχώριου αυτοκινήτου είναι ανεξάρτητη από τις προτιμήσεις του περιοδικού;

## 1<sup>ο</sup> βήμα: Δημιουργία Πίνακα Συχνοτήτων

	Αυτοκίνητο και Οδηγός	Τάσεις Αυτοκινήτου	4-τροχοί	Σύνολο
Εισαγόμενο	54	25	32	<b>111</b>
Εγχώριο	19	22	23	<b>64</b>
<b>Σύνολο</b>	<b>73</b>	<b>47</b>	<b>55</b>	<b>175</b>

## 2<sup>ο</sup> βήμα: Υπολογισμός των αναμενόμενων συχνοτήτων

$$e_{11} = \frac{111 \cdot 73}{175} = 46,3 \quad e_{12} = \frac{111 \cdot 47}{175} = 29,8 \quad e_{13} = \frac{111 \cdot 55}{175} = 34,9$$

$$e_{21} = \frac{64 \cdot 73}{175} = 26,7 \quad e_{22} = \frac{64 \cdot 47}{175} = 17,2 \quad e_{23} = \frac{64 \cdot 55}{175} = 20,1$$

	Αυτοκίνητο και Οδηγός	Τάσεις Αυτοκινήτου	4-τροχοί	Σύνολο
Εισαγόμενο	54 46,3	25 29,8	32 34,9	<b>111</b> <b>111</b>
Εγχώριο	19 26,7	22 17,2	23 20,1	<b>64</b> <b>64</b>
<b>Σύνολο</b>	<b>73</b>	<b>47</b>	<b>55</b>	<b>175</b>

### 3<sup>ο</sup> βήμα: Υπολογισμός του $\chi^2$

$$\chi^2 = \frac{(54 - 46,3)^2}{46,3} + \frac{(25 - 29,8)^2}{29,8} + \frac{(32 - 34,9)^2}{34,9} + \frac{(19 - 26,7)^2}{26,7} + \frac{(22 - 17,2)^2}{17,2} + \frac{(23 - 20,1)^2}{20,1} = 6,2732$$

### 4<sup>ο</sup> βήμα: Υπολογισμός της κριτικής τιμής

Η κριτική τιμή (για  $\alpha=5\%$  και 2 d.f ) είναι : 5,991

Αν  $\chi^2 > 5,991$  απορρίπτεται η  $H_0$ .

### 5<sup>ο</sup> βήμα: Συμπεράσματα

Επειδή  $\chi^2 = 6,2732 > 5,991 = \chi^2_{0.05,2}$  η μηδενική υπόθεση απορρίπτεται και συνεπώς η προτίμηση στο περιοδικό και η προέλευση του αυτοκινήτου (εγχώριο ή εισαγόμενο) δεν είναι μεταβλητές ανεξάρτητες.



## Παράδειγμα 2°

Στην προσπάθεια να διαπιστωθεί αν οι μεταβλητές “φύλο” και “Κατηγορία εργαζόμενου” είναι μεταβλητές εξαρτημένες πραγματοποιήθηκε έρευνα σε δείγμα 474 εργαζομένων σε επιχείρηση. Η κατανομή των εργαζομένων ανά φύλο και κατηγορία παρουσιάζονται στον πίνακα.

		Κατηγορία εργαζόμενου			Σύνολο
		Κατώτερο στέλεχος	Μεσαίο στέλεχος	Ανώτερο στέλεχος	
Φύλο	Άνδρας	157	27	74	258
	Γυναίκα	206	0	10	216
Σύνολο		363	27	84	474

Πραγματοποιήθηκε ο  $\chi^2$  έλεγχος ανεξαρτησίας και τα αποτελέσματα εμφανίζονται στους επόμενους πίνακες.

		Κατηγορία Εργαζόμενου			
		Κατώτερο Στέλεχος	Μεσαίο Στέλεχος	Ανώτερο Στέλεχος	Σύνολο
Φύλο	Άνδρας	157	27	74	258
		197,6	14,7	45,7	258,0
		60,9%	10,5%	28,6%	100%
	Γυναίκα	206	0	10	216
		165,4	12,3	38,3	216,0
		95,4%	0%	4,6%	100%
Σύνολο		43,3%	100%	88,1%	54,4%
		33,1%	5,7%	15,6%	54,4%
		206	0	10	216
		165,4	12,3	38,3	216,0
	95,4%	0%	4,6%	100%	
	56,7%	0%	11,9%	45,6%	
	43,5%	0%	2,1%	45,6%	
	363	27	84	474	
	363,0	27,0	84,0	474,0	
	76,6%	5,7%	17,7%	100%	
	100%	100%	100%	100%	
	76,6%	5,7%	17,7%	100%	

- Συχνότητα
- Αναμενόμενη Συχνότητα
- Ποσοστό γραμμής
- Ποσοστό στήλης
- Συνολικό ποσοστό

Από τον πίνακα που ακολουθεί, προκύπτει ότι οι μεταβλητές φύλο και θέση εργασίας δεν είναι ανεξάρτητες καθώς η τιμή του Pearson Chi- square (79,277) είναι στατιστικά σημαντική (sig.=0,000<0,05).

<b>Chi-Square Tests</b>			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	79,277 <sup>a</sup>	2	,000
Likelihood Ratio	95,463	2	,000
Linear-by-Linear Association	67,463	1	,000
N of Valid Cases	474		
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12,30.			

Στον πίνακα που ακολουθεί υπάρχουν τα μέτρα της έντασης της σχέσης των δύο εξαρτημένων μεταβλητών. Οι τιμές και των δύο μέτρων (Phi, Cramer's V) κυμαίνονται μεταξύ 0 και 1. Στο συγκεκριμένο παράδειγμα η τιμή 0,409 δηλώνει ασθενή εξάρτηση μεταξύ των δύο μεταβλητών.

<b>Symmetric Measures</b>			
		Value	Approximate Significance
Nominal by Nominal	Phi	,409	,000
	Cramer's V	,409	,000
N of Valid Cases		474	

# Μέτρα Σχέσης

## a. Για δύο Ονομαστικής κλίμακας μεταβλητές

1. **Contingency coefficient**: Χρησιμοποιείται μόνο για πίνακες  $5*5$  ή μεγαλύτερους
2. **Phi**: Χρησιμοποιείται για  $2*2$  πίνακες
3. **Cramer's V**: Χρησιμοποιείται για πίνακες μεγαλύτερους των  $2*2$  και μικρότερους των  $5*5$ . Είναι το δημοφιλέστερο μέτρο σχέσης.
4. **Lambda**: Μέτρο κατεύθυνσης
5. **Uncertainty coefficient**: Μέτρο κατεύθυνσης

*!!! Όλοι οι δείκτες κυμαίνονται μεταξύ 0 (καμία σχέση) και 1 (τέλεια σχέση).*

b. Για δύο Ιεραρχικής κλίμακας μεταβλητές

1. Gamma
2. Somers' d
3. Kendall's tau –b
4. Kendall's tau –c

*!!! Όλοι οι δείκτες κυμαίνονται μεταξύ -1 (τέλεια αρνητική σχέση) και +1 (τέλεια θετική σχέση), ενώ το 0 δείχνει καμία σχέση.*

c. Μία Διαστημικής κλίμακας μεταβλητή (εξαρτημένη) και μία Ονομαστικής κλίμακας μεταβλητή (ανεξάρτητη).

1. Eta:

*!!!! Ο δείκτης κυμαίνεται μεταξύ 0 (καμία σχέση) και 1 (τέλεια σχέση)*

## Παράδειγμα 3<sup>ο</sup>.

Ερώτηση:

Υπάρχει συσχέτιση μεταξύ της προτίμησης των μαθητών για διαδικτυακή ή πρόσωπο με πρόσωπο εκπαίδευση και του επιπέδου εκπαίδευσής τους;

### Επίπεδο Εκπαίδευσης

- Προπτυχιακό
- Μεταπτυχιακό

### Μέθοδος Εκπαίδευσης

- Πρόσωπο με πρόσωπο
- Διαδικτυακή

Ο πίνακας παρουσιάζει τα αποτελέσματα από έρευνα σε 200 φοιτητές.

	Προπτυχιακό	Μεταπτυχιακό	Σύνολο
Διαδικτυακή	40	70	<b>110</b>
Πρόσωπο με πρόσωπο	80	10	<b>90</b>
Σύνολο	<b>120</b>	<b>80</b>	<b>200</b>

# Έλεγχος Καλής Προσαρμογής

## Goodness of Fit test

$H_0$ : Ο πληθυσμός *ακολουθεί* μια πολυωνυμική κατανομή με καθορισμένες πιθανότητες για κάθε μια από τις  $k$  κατηγορίες

$H_1$ : Ο πληθυσμός *δεν ακολουθεί* μια πολυωνυμική κατανομή με καθορισμένες πιθανότητες για κάθε μια από τις  $k$  κατηγορίες

Στατιστικός έλεγχος: 
$$\chi^2 = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i}$$

Όπου:  $f_i$  = παρατηρούμενες συχνότητες  
 $e_i$  = αναμενόμενες συχνότητες  
 $k$  = πλήθος κατηγοριών

Κανόνας Απόρριψης:  
Απορρίπτεται η  $H_0$   
όταν:  $\chi^2 > \chi_{\alpha, d.f}^2$

**!!!! Ο στατιστικός έλεγχος ακολουθεί την  $\chi^2$  κατανομή με  $k-1$  βαθμούς ελευθερίας υπό την προϋπόθεση ότι οι αναμενόμενες συχνότητες για όλες τις κατηγορίες είναι 5 ή μεγαλύτερες**



## Παράδειγμα

Τα προηγούμενα χρόνια το μερίδιο αγοράς συγκεκριμένης κατηγορίας προϊόντων είχε σταθεροποιηθεί στο 30% για την εταιρεία Α, 50% για την Β και 20% για την Γ. Πρόσφατα η εταιρεία Γ ανέπτυξε ένα νέο και προηγμένο προϊόν με σκοπό να βελτιώσει το δικό της μερίδιο αγοράς. Η Scott Marketing Research προσλήφθηκε από την εταιρεία Γ προκειμένου να καθορίσει αν το νέο προϊόν άλλαξε τις ισορροπίες στο μερίδιο αγοράς.

Η εταιρεία Scott Marketing Research χρησιμοποίησε ένα δείγμα 200 καταναλωτών και τους ρώτησε για την προτίμησή τους μεταξύ των τριών εταιρειών. Οι απαντήσεις, σχετικά με την προτίμηση, παρουσιάζονται στον επόμενο πίνακα. Ο έλεγχος καλής προσαρμογής πραγματοποιήθηκε σε επίπεδο σημαντικότητας  $\alpha=5\%$ .

Παρατηρούμενες Συχνότητες		
Εταιρεία Α	Εταιρεία Β	Εταιρεία Γ
48	98	54

## Λύση

$$H_0: P_A=30\% \quad P_B=50\% \quad P_\Gamma=20\%$$

$H_1$ : Η αναλογία στον πληθυσμό δεν είναι

$$P_A=30\% \quad P_B=50\% \quad P_\Gamma=20\%$$

Υπολογίζονται πρώτα οι *αναμενόμενες συχνότητες* πολλαπλασιάζοντας το σταθερό ποσοστό προτίμησης των τριών εταιρειών με το συνολικό πλήθος των ερωτηθέντων. Στη συνέχεια εφαρμόζουμε τον τύπο προκειμένου να υπολογίσουμε την τιμή του  $\chi^2$ .

Αναμενόμενες Συχνότητες		
Εταιρεία Α	Εταιρεία Β	Εταιρεία Γ
$200 \cdot 0,3 = 60$	$200 \cdot 0,5 = 100$	$200 \cdot 0,2 = 40$

## Παρατηρούμενες Συχνότητες- Αναμενόμενες Συχνότητες

Εταιρεία Α	Εταιρεία Β	Εταιρεία Γ
48	98	54
$200 * 0,3 = 60$	$200 * 0,5 = 100$	$200 * 0,2 = 40$
$48 - 60 = -12$	$98 - 100 = -2$	$54 - 40 = 14$
$(-12)^2 = 144$	$(-2)^2 = 4$	$14^2 = 196$
$144 / 60 = 2,4$	$4 / 100 = 0,04$	$196 / 40 = 4,9$
$\chi^2 = 2,4 + 0,04 + 4,9 = 7,34$		

$$\chi_{0.05,2}^2 = 5,991 \quad \chi^2 = 7,34 > \chi_{0.05,2}^2 = 5,991$$

Η μηδενική υπόθεση απορρίπτεται και συνεπώς το νέο προϊόν άλλαξε τις ισορροπίες στο μερίδιο αγοράς. Ειδικότερα, το νέο προϊόν βελτίωσε τη θέση της εταιρείας Γ.

## Άσκηση

Τα προηγούμενα χρόνια στις εξετάσεις του μαθήματος «Εισαγωγή στη Στατιστική» το 30% αποτύγχανε (0-4,9), το 40% περνούσε με βαθμολογία 5,0-6,4, το 20% με βαθμολογία 6,5-8,5 και το 10% με βαθμολογία 8,6-10,0. Στις εξετάσεις του μαθήματος στην εξεταστική περίοδο του Φεβρουαρίου προσήλθαν 300 φοιτητές και οι επιδόσεις τους παρουσιάζονται στον επόμενο πίνακα. Να πραγματοποιηθεί έλεγχος καλής προσαρμογής σε επίπεδο σημαντικότητας  $\alpha=5\%$ .

Παρατηρούμενες Συχνότητες			
0-4,9	5,0-6,4	6,5-8,5	8,6-10,0
100	140	50	10

## Λύση

$$H_0: P_A=30\% \quad P_B=40\% \quad P_\Gamma=20\% \quad P_\Delta=10\%$$

$H_1$ : Η αναλογία στον πληθυσμό δεν είναι

$$P_A=30\% \quad P_B=40\% \quad P_\Gamma=20\% \quad P_\Delta=10\%$$

Υπολογίζονται πρώτα οι *αναμενόμενες συχνότητες* πολλαπλασιάζοντας το σταθερό ποσοστό προτίμησης των τριών εταιρειών με το συνολικό πλήθος των ερωτηθέντων. Στη συνέχεια εφαρμόζουμε τον τύπο προκειμένου να υπολογίσουμε την τιμή του  $\chi^2$ .

Αναμενόμενες Συχνότητες			
0-4,9	5,0-6,4	6,5-8,5	8,6-10,0
$300 \cdot 0,3 = 90$	$300 \cdot 0,4 = 120$	$300 \cdot 0,2 = 60$	$300 \cdot 0,1 = 30$

## Παρατηρούμενες Συχνότητες- Αναμενόμενες Συχνότητες

0-4,9	5,0-6,4	6,5-8,5	8,6-10,0
100	140	50	10
$300 \cdot 0,3 = 90$	$300 \cdot 0,4 = 120$	$300 \cdot 0,2 = 60$	$300 \cdot 0,1 = 30$
$100 - 90 = 10$	$140 - 120 = 20$	$50 - 60 = -10$	$10 - 30 = -20$
$(10)^2 = 100$	$(20)^2 = 400$	$(-10)^2 = 100$	$(-20)^2 = 400$
$100/90 = 1,11$	$400/120 = 3,33$	$100/60 = 1,67$	$400/30 = 13,33$
$\chi^2 = 1,00 + 3,33 + 1,67 + 13,33 = 19,44$			

$$\chi_{0.05,2}^2 = 5,991 \quad \chi^2 = 7,34 > \chi_{0.05,2}^2 = 5,991$$

Η μηδενική υπόθεση απορρίπτεται και συνεπώς το νέο προϊόν άλλαξε τις ισορροπίες στο μερίδιο αγοράς. Ειδικότερα, το νέο προϊόν βελτίωσε τη θέση της εταιρείας Γ.

7,815