Τι είναι η στατιστική μέθοδος X^2

Η Στατιστική είναι η επιστήμη των πιθανοτήτων. Ο βαθμός τυχαιότητας ενός αποτελέσματος προσδιορίζεται από την σύγκριση των αποτελεσμάτων ενός πειράματος, με προγενέστερα αποτελέσματα που ήδη έχουν θεωρητική επαλήθευση. Αυτά τα σημεία αναφοράς είναι γνωστά ως «*πίνακες κατανομής συχνοτήτων σε επίπεδα σημαντικότητας*». Μια παραδοσιακή μέθοδος εξακρίβωσης ότι το αποτέλεσμα δεν ήταν τυχαίο, είναι η επίτευξη του μέσου όρου μεταξύ του απόλυτα σωστού (100%) και του εντελώς τυχαίου (50%). Δηλαδή οτιδήποτε δείχνει κάτω από 75% δηλώνει ότι το αποτέλεσμα ενδέχεται να είναι τυχαίο.

Με την μέθοδο Χί εις το Τετράγωνο (chi-square) μπορούμε να εξετάσουμε δεκάδες μεταβλητές ταυτόχρονα. Πιο συγκεκριμένα, ο στατιστικός έλεγχος με την μέθοδο X^2 είναι μια διαδικασία που μας επιτρέπει να αξιολογήσουμε αν και κατά πόσο υπάρχει αιτιώδης σχέση μεταξύ δύο ή περισσότερων μεταβλητών. Ειδικά όταν θέλουμε να γνωρίσουμε αν η υψηλή τιμή μιας μεταβλητής σχετίζεται με μια υψηλή ή χαμηλή τιμή μιας άλλης μεταβλητής. Έτσι χρησιμοποιούμε την μέθοδο X^2 για να βρούμε τις φυσιολογικές κατανομές και για να ανακαλύψουμε αν η ανεξαρτησία ή μη, μεταξύ δύο ή περισσότερων μεταβλητών, είναι στατιστικά σημαντική ή όχι.

Η στατιστική δοκιμασία X^2 για κάθε μία μεταβλητή, στην ουσία εξετάζει αν υπάρχει διαφορά μεταξύ των δεδομένων που έχουν συλλεχθεί κατά την διάρκεια της έρευνας (**πραγματικές συχνότητες**) και αυτών που θα περιμέναμε να εμφανιστούν αν ίσχυε η μηδενική υπόθεση (**αναμενόμενες συχνότητες**). Αυτό στην πράξη σημαίνει ότι, αν οι Πραγματικές συχνότητες είναι τυχαίες, θα πρέπει να πλησιάζουν αρκετά τις Αναμενόμενες συχνότητες. Με απλά λόγια, το X^2 αντανακλά το μέγεθος των διαφορών μεταξύ των **Π**ραγματικών συχνοτήτων και των **Α**ναμενόμενων συχνοτήτων. Όσο μεγαλύτερη είναι αυτή η διαφορά, τόσο πιθανότερο είναι να προκύψει στατιστικώς σημαντικό αποτέλεσμα.

Κοντολογίς, ο υπολογισμός του X^2 στηρίζεται στην μέτρηση της διαφοράς μεταξύ **Π**αρατηρούμενων Τιμών Συχνοτήτων και **Α**ναμενόμενων Τιμών Συχνοτήτων. Και με απλά λόγια η συνάρτηση είναι $\mathbf{X^2= (\Pi\text{-}A)^2 /A}$

Πως υπολογίζουμε το X^2

Ας υποθέσουμε ότι έχουμε 100 ομοφυλόφιλους άνδρες ασθενείς με AIDS που απεβίωσαν λόγω αυτής της ασθένειας. Και ας υποθέσουμε ότι έχουμε 100 περίπου συνομηλίκους (*άνδρες και γυναίκες*) από τον γενικό πληθυσμό, τους οποίους γνωρίζουμε προσωπικά και είμαστε βέβαιοι ότι δεν έχουν AIDS ούτε σεξουαλικές παρεκκλίσεις, είναι καλά στην υγεία τους, γι' αυτό θα τους χρησιμοποιήσουμε ως δείγμα ελέγχου στην έρευνά μας.

Ας υποθέσουμε επίσης ότι στα ωροσκόπια των 100 ομοφυλόφιλων ανδρών που ασθένησαν και απεβίωσαν από AIDS βρήκαμε να ξεχωρίζει η γενεθλιακή Αφροδίτη σε Αντίθεση με Πλούτωνα 10 φορές, ενώ στο δείγμα ελέγχου 4 φορές. Τα νούμερα αυτά μπορεί να φαίνονται μικρά αλλά δεν είναι. Και εξηγούμαι: στην παραδοσιακή αστρολογία εξετάζουμε συνολικά 640 πιθανούς γεωκεντρικούς συνδυασμούς, εκ των οποίων οι 222 συνδυασμοί είναι όψεις. Πιο συγκεκριμένα, έχουμε ΠΕΝΤΕ κύριες όψεις (*Σύνοδο, Αντίθεση, Τετράγωνο, Τρίγωνο, Εξάγωνο*), ΕΠΙ ΔΥΟ ωριαία σημεία (*Ωροσκόπο, Μεσουράνημα*), ΕΠΙ ΔΕΚΑ πλανήτες (*Ήλιο, Σελήνη, Ερμή, Αφροδίτη, Άρη, Δία, Κρόνο, Ουρανό, Ποσειδώνα, Πλούτωνα*) =**222** πιθανούς γεωκεντρικούς συνδυασμούς όψεων. Τόσο είναι το πλήθος των όψεων, αλλά σε ένα ωροσκόπιο ποτέ δεν θα συμβούν ταυτόχρονα όλοι οι συνδυασμοί. Καταλαβαίνετε όμως ότι, αφού έχουμε πάντα 222 πιθανούς συνδυασμούς όψεων, σε ένα τόσο μικρό δείγμα έρευνας -*μονάχα 100 περιπτώσεων*-, αν μια όψη επαναληφθεί 10 φορές, θα πρέπει να την αντιμετωπίζουμε ως σημαντική. Αλλά χάρη παραδείγματος τετραπλασιάζω την αναλογία κι έτσι θα έχουμε 16 προς 40.

Ο σκοπός μας είναι να μάθουμε αν η γενεθλιακή Αφροδίτη σε Αντίθεση με Πλούτωνα είναι όψη στατιστικά σημαντική, ή αν αυτές οι δύο μεταβλητές είναι ανεξάρτητες. Αν οι πιθανότητες δείξουν πως είναι ανεξάρτητες, τότε θα πρέπει να είναι περίπου ίδιο το πλήθος των ατόμων σε κάθε κελί του παρακάτω πίνακα. Εάν στην μέτρηση φανεί ότι ο Πραγματικός αριθμός είναι πολύ διαφορετικός, τότε σχετίζονται οι δύο μεταβλητές.

Για να τα συγκεκριμενοποιήσουμε όλα αυτά, θα χρειαστεί να τα τοποθετήσουμε σε ένα πίνακα:

	Με αυτή την όψη	**Χωρίς αυτή την όψη**	**Σύνολο**
Aids	40	60	100
Normal	16	84	100
Σύνολο	56	144	200

Στον πίνακα βλέπουμε ότι η ολική πιθανότητα να έχει κάποιος AIDS είναι 100 / 200 =**0,5**
Η ολική πιθανότητα να έχει γενεθλιακή Αφροδίτη Αντίθεση Πλούτωνα είναι 56 / 200 =**0,28**
Η διπλή πιθανότητα να έχει κάποιος AIDS και την ανωτέρω όψη, είναι 0,5x0,28 =**0,14**
Συνεπώς η Αναμενόμενη τιμή είναι 0,14x200 =**28**

Ένας πιο γρήγορος και εύκολος τρόπος υπολογισμού των Αναμενόμενων τιμών για κάθε κελί του πίνακα, είναι να πολλαπλασιάσουμε την συνολική οριζόντια γραμμή με το σύνολο της κάθετης στήλης, και στην συνέχεια τα διαιρούμε με το τελικό σύνολο. Έτσι, η Αναμενόμενη τιμή για το πρώτο κελί AIDS είναι (100x56) / 200 =**28**, για το δεύτερο κελί AIDS είναι (100x144) /200 =**72,** για το πρώτο κελί Normal είναι (56x100) /200 =**28**, και για το δεύτερο κελί Normal είναι (144x100) /200 =**72** (*τα νούμερα δείχνουν να επαναλαμβάνονται επειδή εξετάζουμε ακριβώς 100 άτομα από κάθε ομάδα*).

Όλα αυτά θα τα αντιληφθούμε καλύτερα αν τα τοποθετήσουμε εκ νέου μέσα στον πίνακα, ως εξής:

	Με αυτή την όψη	**Χωρίς αυτή την όψη**	**Σύνολο**
Aids	40 (28)	60 (72)	100
Normal	16 (28)	84 (72)	100
Σύνολο	56	144	200

Ο παρακάτω τύπος περιγράφει μια ακόμη λειτουργία που οφείλουμε να εκτελέσουμε σε κάθε κελί, για να βρούμε έναν νέο αριθμό. Όταν όλοι οι νέοι αριθμοί αθροιστούν, το αποτέλεσμα που προκύπτει θα μας φανερώσει την αντίστροφη πιθανότητα X^2 ως εξής:

$$\mathbf{X^2} = (40-28)^2 /28 + (60-72)^2 /72 + (16-28)^2 /28 + (84-72)^2 /72 = 5,14+2+5,14+2 = \mathbf{14,28}$$

Σε γενικές γραμμές, όσο πιο μεγάλος προκύπτει ο τελικός αριθμός της αντίστροφης πιθανότητας X^2 οι Αναμενόμενες τιμές είναι πολύ διαφορετικές από τις Πραγματικές τιμές, και συνεπώς, τόσο πιθανότερο είναι οι μεταβλητές που εξετάσαμε να μην είναι τυχαίες, αλλά να συνδέονται.

Το X^2 έχει βαθμούς ελευθερίας, ο τύπος της οποίας είναι (αριθμός γραμμών -1) **x** (αριθμός στηλών -1) που στο παράδειγμα το οποίο σας έδειξα είναι (2-1) x (2-1) = 1x1 = **1** βαθμός ελευθερίας.

Έτσι κοιτάμε τον ακόλουθο πίνακα. ΕΝΑ ΑΠΟΤΕΛΕΣΜΑ ΑΠΟ ΤΗΝ ΔΕΥΤΕΡΗ ΣΤΗΛΗ 0,1 ΚΑΙ ΠΕΡΑ, ΑΡΧΙΖΕΙ ΝΑ ΘΕΩΡΕΙΤΑΙ ΟΛΟΕΝΑ ΠΙΟ ΔΥΣΚΟΛΑ ΑΜΦΙΣΒΗΤΗΣΙΜΟ. Οπότε με αντίστροφη πιθανότητα X^2 =14,28 και ΕΝΑ βαθμό ελευθερίας, βλέπουμε την τελευταία στήλη **0,001** σε επίπεδο σημαντικότητας, που σημαίνει ότι μπορούμε να απορρίψουμε την μηδενική υπόθεση. Με άλλα λόγια, δείχνει να είναι <u>στατιστικά σημαντική</u> η γενεθλιακή όψη Αφροδίτης Αντίθεση Πλούτωνα στους ομοφυλόφιλους άνδρες που απεβίωσαν από AIDS, και μπορούμε να την υιοθετήσουμε ως δείκτη αστρολογικής προδιάθεσης για τέτοιες περιπτώσεις.

Πίνακας κατανομής συχνοτήτων με επίπεδα σημαντικότητας:

Chi Square Distribution Table

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	32.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	32.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Table from Ronald J. Wonnacott and Thomas H. Wonnacott, *Statistics: Discovering Its Power*, New York: John Wiley and Sons, 1982, p.352.

Τέλος, για όσους ενδιαφέρει, σε ένα άλλο πείραμα που έκανα μεταγενέστερα, με πολύ μεγαλύτερο δείγμα, εξετάζοντας μόνο τον γενικό πληθυσμό (*ετεροφυλόφιλους άνδρες και γυναίκες*), εμφανίστηκε η γενεθλιακή όψη Ερμής Τετράγωνο Πλούτωνα ως στατιστικά σημαντική στην προδιάθεση για AIDS.

Παράδειγμα με πολλές μεταβλητές X^2

Έστω ότι έχουμε 6.418 σεισμούς από 4R και άνω, που έχουν συμβεί εντός 50 ετών στον ελλαδικό χώρο, και θέλουμε να δούμε αν στα ωροσκόπιά τους, οι όψεις μεταξύ Άρη–Ουρανού είναι στατιστικά σημαντικές ή όχι. Για να το διερευνήσουμε παίρνουμε άλλα 6.418 εντελώς τυχαία ελληνικά ωροσκόπια, σε άλλες χρονιές, ώρες και ημερομηνίες, διεσπαρμένα κατά μισό αιώνα πίσω. Αυτό είναι το δείγμα ελέγχου.

Ταξινόμηση Πραγματικών Τιμών

Όψεις Άρη - Ουρανού	Με Σύνοδο	Χωρίς Σύνοδο	Με Αντίθεση	Χωρίς Αντίθεση	Με Τετράγωνο	Χωρίς Τετράγωνο	Με Τρίγωνο	Χωρίς Τρίγωνο	Με Εξάγωνο	Χωρίς Εξάγωνο	Με Χιασμό	Χωρίς Χιασμό	Με 45°	Χωρίς 45°	Με 135°	Χωρίς 135°	Σύνολο
Σεισμοί	37	6381	35	6383	52	6366	70	6348	64	6354	40	6378	26	6392	21	6397	**6418**
Δείγματα	43	6375	30	6388	52	6366	126	6292	57	6361	45	6373	38	6380	47	6371	**6418**
Σύνολο	**80**	**12756**	**65**	**12771**	**104**	**12732**	**196**	**12640**	**121**	**12715**	**85**	**12751**	**64**	**12772**	**68**	**12768**	**12836**

Για να υπολογίσουμε εύκολα και γρήγορα τις Αναμενόμενες τιμές για κάθε κελί αυτού του πίνακα, πολλαπλασιάζουμε την συνολική οριζόντια γραμμή με το σύνολο της κάθετης στήλης, και στην συνέχεια την διαιρούμε με το γενικό σύνολο. Ας πάρουμε πρώτα τους σεισμούς: για την Σύνοδο έχουμε (6418x80) / 12836 =**40** χωρίς όψη Συνόδου έχουμε (6418x12756) /12836 =**6378** Ας πάρουμε τώρα τα δείγματα: με όψη Συνόδου έχουμε (6418x80) / 12836 =**40** χωρίς όψη Συνόδου έχουμε (6418x12756) /12836 =**6378** ΟΙ ΑΝΑΜΕΝΟΜΕΝΕΣ ΤΙΜΕΣ ΠΑΝΤΑ ΕΜΦΑΝΙΖΟΝΤΑΙ ΙΔΙΕΣ ΟΤΑΝ ΕΧΟΥΜΕ ΙΣΑΡΙΘΜΟ ΠΛΗΘΟΣ ΔΕΙΓΜΑΤΟΣ

Επαναλαμβάνουμε την ίδια διαδικασία για όλες τις υπόλοιπες όψεις και μη όψεις, τόσο για τους σεισμούς, όσο και για τα δείγματα, και τα αποτελέσματα που βρίσκουμε είναι οι Αναμενόμενες τιμές τις οποίες τοποθετούμε μέσα σε ένα νέο πίνακα για να έχουμε καθαρή εικόνα, ως εξής:

Ταξινόμηση Αναμενόμενων Τιμών

Όψεις Άρη - Ουρανού	Με Σύνοδο	Χωρίς Σύνοδο	Με Αντίθεση	Χωρίς Αντίθεση	Με Τετράγωνο	Χωρίς Τετράγωνο	Με Τρίγωνο	Χωρίς Τρίγωνο	Με Εξάγωνο	Χωρίς Εξάγωνο	Με Χιασμό	Χωρίς Χιασμό	Με 45°	Χωρίς 45°	Με 135°	Χωρίς 135°	Σύνολο
Σεισμοί	40	6378	32,5	6385,5	52	6366	98	6320	60,5	6357,5	42,5	6375,5	32	6386	34	6384	**6418**
Δείγματα	40	6378	32,5	6385,5	52	6366	98	6320	60,5	6357,5	42,5	6375,5	32	6386	34	6384	**6418**
Σύνολο	**80**	**12756**	**65**	**12771**	**104**	**12732**	**196**	**12640**	**121**	**12715**	**85**	**12751**	**64**	**12772**	**68**	**12768**	**12836**

Τώρα που έχουμε βρει όλες τις Πραγματικές τιμές και όλες τις Αναμενόμενες τιμές, εκτελούμε τον παρακάτω τύπο συγκρίνοντας τα 32 κελιά για να βρούμε τις τελικές τιμές, ως εξής:

37-40)^2 /40=**0,225** + (6381-6378)^2 /6378=**0,0014** + (35-32,5)^2 /32,5=**0,19** + (6383-6385,5)^2 /6385,5=**0,0009** + (52-52)^2 /52=**0** + (6366-6366)^2 /6366=**0** + (70-98)^2 /98=**8** + (6348-6320)^2 /6320=**0,124** + (64-60,5)^2 /60,5=**0,202** + (6354-6357,5)^2 /6357,5=**0,0019** + (40-42,5)^2 /42,5=**0,147** + (6378-6375,5)^2 /6375,5=**0,00098** + (26-32)^2 /32=**1,125** + (6392-6386)^2 /6386=**0,0056** + (21-34)^2 /34=**4,97** + (6397-6384)^2 /6384=**0,026** + (43-40)^2 /40=**0,225** + (6375-6378)^2 /6378=**0,0014** + (30-32,5)^2 /32,5=**0,19** + (6388-6385,5)^2 /6385,5=**0,00097** + (52-52)^2 /52=**0** + (6366-6366)^2 /6366=**0** + (126-98)^2 /98=**8** + (6292-6320)^2 /6320=**0,124** + (57-60,5)^2 /60,5=**0,202** + (6361-6357,5)^2 /6357,5=**0,0019** + (45-42,5)^2 /42,5=**0,147** + (6373-6375,5)^2 /6375,5=**0,00098** + (38-32)^2 /32=**1,125** + (6380-6386)^2 /6386=**0,0056** + (47-34)^2 /34=**4,97** + (6371-6384)^2 /6384 =**0,026**

Άρα το αντίστροφο X^2 ισούται με το άθροισμα όλων αυτών, που είναι ο αριθμός **30**

Στο παράδειγμα που κάναμε οι βαθμοί ελευθερίας είναι (αριθμός γραμμών -1) **x** (αριθμός στηλών -1) δηλαδή (2-1) x (16-1) = 1x15 = **15** βαθμοί ελευθερίας. Έτσι κοιτάζουμε ξανά τον πίνακα σημαντικότητας με αντίστροφη πιθανότητα X^2 =30 και ΔΕΚΑΠΕΝΤΕ βαθμούς ελευθερίας, οπότε βλέπουμε την 5η στήλη **0,10** που σημαίνει ότι μπορούμε να απορρίψουμε την μηδενική υπόθεση, αφού η στατιστική επιβεβαιώνει ότι είναι σημαντικές οι όψεις του Άρη με τον Ουρανό κατά την εκδήλωση σεισμών. Αλλά όχι όλες. Μόνο αυτές που έδωσαν υψηλές Αναμενόμενες τιμές, όπως το Τρίγωνο (120 μοίρες) και το Ενάμισι Τετράγωνο (135 μοίρες).

Διόρθωση YATES

Την διόρθωση Γέιτς (*Yates' correction*) πρέπει να την κάνουμε όταν το δείγμα μας είναι μικρό, γύρω στις 50 περιπτώσεις, και έχουμε μόνο έναν βαθμό ελευθερίας. Η διόρθωση Γέιτς υπολογίζεται σύμφωνα με τον τύπο $X^2 = ([Π-Α]-0,5)^2 /Α$

Με απλά λόγια σημαίνει ότι, η διαφορά μεταξύ **Π**ραγματικής συχνότητας και **Α**ναμενόμενης συχνότητας για κάθε κελί μειώνεται κατά 0,5 προτού υψωθεί εις το Τετράγωνο. Αυτό έχει ως αποτέλεσμα μικρότερη τιμή του X^2 και κατά συνέπεια, μικρότερες πιθανότητες σφάλματος.

Περισσότερα μπορείτε να δείτε στο μεγάλο αρχείο της ανάλυσης των αποτελεσμάτων (My_Results.xls)