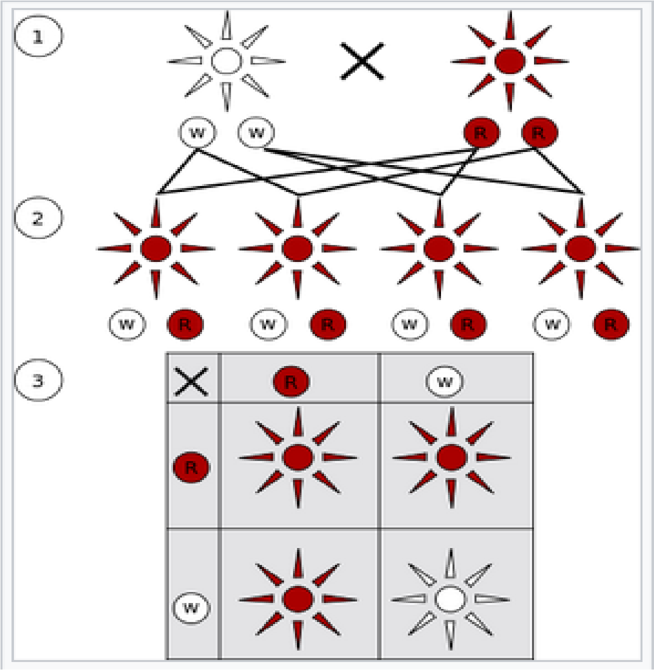
- Κριτήριο χ^2 του Pearson
  - ✓ έλεγχος καλής προσαρμογής
  - ✓ έλεγχος ομοιογένειας
  - ✓ έλεγχος ανεξαρτησίας (πίνακες συνάφειας)

- X^2 έλεγχος
- καλής προσαρμογής



εάν τα δεδομένα ακολουθούν (υπακούουν) συγκεκριμένο θεωρητικό μοντέλο:

αναλογία κόκκινων λευκών 3:1 από διασταύρωση ετεροζυγωτών

- ομοιογένειας

  εάν το υπό μελέτη χαρακτηριστικό σε επαναλήψεις του πειράματος σε δύο ή περισσότερους πληθυσμούς έχει την ίδια συμπεριφορά

- Ανεξαρτησίας

  εάν δύο χαρακτηριστικά που μελετώνται είναι ανεξάρτητα

Σε κάθε περίπτωση έχουμε μια "πολυωνυμική" κατανομή με κ δυνατά αποτελέσματα $A_1, A_2, \ldots, A_k$

π_i πλήθος (από τις n παρατηρήσεις του δείγματος) που εμφανίζεται η A_i (παρατηρούμενες τιμές) $i=1,\ldots,k$

$\pi_1 + \pi_2 + \ldots + \pi_k = n$

Μας ενδιαφέρει να εξετάσουμε εάν η υπόθεση $P(\text{εμφάνισης } A_i) = p_i$ είναι αληθής

H_0: $P(A_1) = p_1$ κ' $P(A_2) = p_2$ ... κ' $P(A_k) = p_k$

H_1: $P(A_1) \neq p_1$ ή $P(A_2) \neq p_2$ ... ή $P(A_k) \neq p_k$

$p_1 + p_2 + \ldots + p_k = 1$ $\quad$ $1:1$

ισχύει ότι $\pi_i \sim Binomial(n, p_i)$ (κάτω από H_0)

οπότε $E(\pi_i) = np_i$ $\quad$ $Var(\pi_i) = np_i q_i$ $\quad i=1,\ldots,k$

Ονομάζουμε θεωρητικές τιμές τις $\theta_i = np_i$ και όταν ισχύει η H_0 θα πρέπει να μη διαφέρουν πολύ από τις παρατηρούμενες τιμές π_i.

- Εάν $n >>$ και $np_i > 5$ τότε από Κεντρικό Οριακό

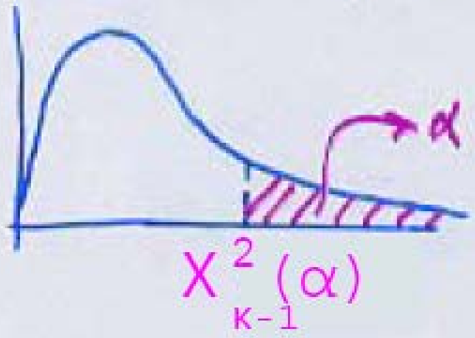
$$\frac{\pi_i - np_i}{\sqrt{np_i q_i}} \sim N(0,1) \Rightarrow \left(\frac{\pi_i - \theta_i}{\theta_i q_i}\right)^2 \sim X_1^2$$

όπου $q_i = 1 - p_i$

• χρησιμοποιούμε την ποσότητα

$$X^2 = \sum_{i=1}^{k-1} \left(\frac{\pi_i - np_i}{np_i q_i}\right)^2 \underset{\text{αποδεικνύεται}}{\overset{\text{περίπου}}{=\!=}} \sum_{i=1}^{k} \frac{(\pi_i-\theta_i)^2}{\theta_i} \sim X^2_{k-1}$$

για τον έλεγχο και απορρίπτουμε την H_0 σε επίπεδο σημαντικότητας α, εαν $X^2 > X^2_{k-1}(\alpha)$



• Σε περίπτωση που τα p_i δεν είναι γνωστά αλλά πρέπει να τα εκτιμήσουμε (εκτιμώνται s άγνωστους παραμέτρους) τότε $X^2 \sim X^2_{k-1-s}$

και απορρίπτουμε την H_0 σε ε.σ. α,

εαν $X^2 > X^2_{k-1-s}(\alpha)$

X^2 έλεγχος: καλής προσαρμογής σε θεωρητικό μοντέλο

H_0: The sample comes from a population having a 9 : 3 : 3 : 1 ratio of yellow-smooth to yellow-wrinkled to green-smooth to green-wrinkled seeds.

H_1: The sample comes from a population not having a 9 : 3 : 3 : 1 ratio of the above four seed phenotypes.

	Yellow smooth	*Yellow wrinkled*	*Green smooth*	*Green wrinkled*	n
f_i	152	39	53	6	250

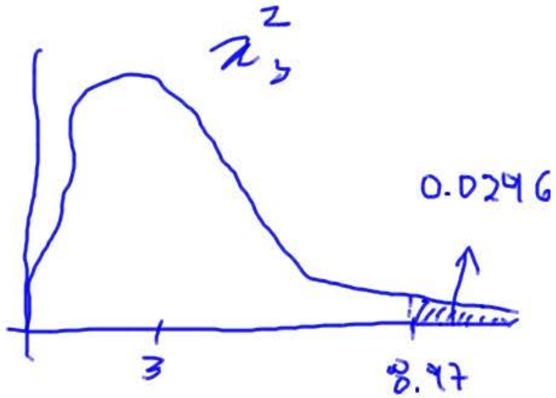
H_0: $p_1=9/(9+3+3+1)$, $p_2=3/16$, $p_3=3/16$, $p_4=1/16$

H_1: κάποιο διαφέρει

H_0: $p_1=0.5625$, $p_2=0.1875$, $p_3=0.1875$, $p_4=0.0625$

H_1: κάποιο διαφέρει από την αντίστοιχη τιμή

	YS	YW	GS	GW	Σύνολο
π_i	152	39	53	6	250
p_i	9/16	3/16	3/16	1/16	1
$\theta_i = np_i$	140.63	46.87	46.87	15.63	250
$\frac{(\pi_i-\theta_i)^2}{\theta_i}$	0.92	1.32	0.8	5.92	8.97



$$X^2 = \sum_{i=1}^{k} \frac{(\pi_i-\theta_i)^2}{\theta_i} \left(= \sum_{i=1}^{k} \frac{\pi_i^2}{\theta_i} - n \right) \sim \chi^2_{k-1}$$ εδώ η τιμή 8.97 προέρχεται χ^2_3

Απορρίπτω την H_0 σε ε.σ. α όταν $X^2 > \chi^2_{k-1}(\alpha)$

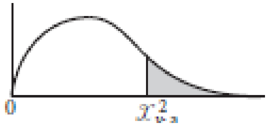
$X^2 = 8.97$

$\chi^2_3(0.1) = 6.25$ απορ. (H_1) $\quad \chi^2_3(0.05) = 7.81$ απορ. (H_1) $\quad \chi^2_3(0.025) = 9.34$ αποδέχομαι (H_0)

Οπότε $0.025 < \text{p-value} < 0.05$ $\quad$ p-value $= P(\chi^2_3 > 8.97) = 1 - \text{CDF.CHISQ}(8.97, 3)$ $= 0.0296$

```
> 1-pchisq(8.97,df=3)
[1] 0.02969242
```

Πίνακας: Τιμών $\mathcal{X}^2_{v;\,a}$ της $\mathcal{X}^2$ κατανομής για τις οποίες $P(X^2 > \mathcal{X}^2_{v;\,a}) = a$



Για $v > 100$, $\mathcal{X}^2_{v;\,a} = \frac{1}{2}\left(z_a + \sqrt{2v-1}\right)^2$

β.ε.	a=0,995	a=0,990	a=0,975	a=0,950	a=0,900
1	0,0000393	0,0001571	0,0009821	0,0039321	0,0157908
2	0,0100251	0,0201007	0,0506356	0,102587	0,210720
3	0,0717212	0,114832	0,215795	0,351846	0,584375
4	0,206990	0,297110	0,484419	0,710721	1,063623
5	0,411740	0,554300	0,831211	1,145476	1,61031
6	0,675727	0,872085	1,237347	1,63539	2,20413
7	0,989265	1,239043	1,68987	2,16735	2,83311
8	1,344419	1,646482	2,17973	2,73264	3,48954
9	1,734926	2,087912	2,70039	3,32511	4,16816
10	2,15585	2,55821	3,24697	3,94030	4,86518
11	2,60321	3,05347	3,81575	4,57481	5,57779
12	3,07382	3,57056	4,40379	5,22603	6,30380
13	3,56503	4,10691	5,00874	5,89186	7,04150
14	4,07468	4,66043	5,62872	6,57063	7,78953
15	4,60094	5,22935	6,26214	7,26094	8,54675
16	5,14224	5,81221	6,90766	7,96164	9,31223
17	5,69724	6,40776	7,56418	8,67176	10,0852
18	6,26481	7,01491	8,23075	9,39046	10,8649
19	6,84398	7,63273	8,90655	10,1170	11,6509
20	7,43386	8,26040	9,59083	10,8508	12,4426
21	8,03366	8,89720	10,28293	11,5913	13,2396
22	8,64272	9,54249	10,9823	12,3380	14,0415
23	9,26042	10,19567	11,6885	13,0905	14,8479
24	9,88623	10,8564	12,4011	13,8484	15,6587
25	10,5197	11,5240	13,1197	14,6114	16,4734
26	11,1603	12,1981	13,8439	15,3791	17,2919
27	11,8076	12,8786	14,5733	16,1513	18,1138
28	12,4613	13,5648	15,3079	16,9279	18,9392
29	13,1211	14,2565	16,0471	17,7083	19,7677

a=0,10	a=0,05	a=0,025	a=0,010	a=0,005	β.ε.
2,70554	3,84146	5,02389	6,63490	7,87944	1
4,60517	5,99147	7,37776	9,21034	10,5966	2
6,25139	7,81473	9,34840	11,3449	12,8381	3
7,77944	9,48773	11,1433	13,2767	14,8602	4
9,23635	11,0705	12,8325	15,0863	16,7496	5
10,6446	12,5916	14,4494	16,8119	18,5476	6
12,0170	14,0671	16,0128	18,4753	20,2777	7
13,3616	15,5073	17,5346	20,0902	21,9550	8
14,6837	16,9190	19,0228	21,6660	23,5893	9
15,9871	18,3070	20,4831	23,2093	25,1882	10
17,2750	19,6751	21,9200	24,7250	26,7569	11
18,5494	21,0261	23,3367	26,2170	28,2995	12
19,8119	22,3621	24,7356	27,6883	29,8194	13
21,0642	23,6848	26,1190	29,1413	31,3193	14
22,3072	24,9958	27,4884	30,5779	32,8013	15
23,5418	26,2962	28,8454	31,9999	34,2672	16
24,7690	27,5871	30,1910	33,4087	35,7185	17
25,9894	28,8693	31,5264	34,8053	37,1564	18
27,2036	30,1435	32,8523	36,1908	38,5822	19
28,4120	31,4104	34,1696	37,5662	39,9968	20
29,6151	32,6705	35,4789	38,9321	41,4010	21
30,8133	33,9244	36,7807	40,2894	42,7956	22
32,0069	35,1725	38,0757	41,6384	44,1813	23
33,1963	36,4151	39,3641	42,9798	45,5585	24
34,3816	37,6525	40,6465	44,3141	46,9278	25
35,5631	38,8852	41,9232	45,6417	48,2899	26
36,7412	40,1133	43,1944	46,9630	49,6449	27
37,9159	41,3372	44,4607	48,2782	50,9933	28
39,0875	42,5569	45,7222	49,5879	52,3356	29
40,2560	43,7729	46,9792	50,8922	53,6720	30
51,8050	55,7585	59,3417	63,6907	66,7659	40

Παραδείγματα

- Ρίχνουμε ένα ζάρι 120 φορές και παίρνουμε

Αποτέλεσμα (A_i)	1	2	3	4	5	6
Αρ. Εμφανίσ. (n_i)	11	21	29	30	19	10

Να ελέγξετε εαν το ζάρι είναι αμερόληπτο (δίκαιο) σε επίπεδο σημαντικότητας $\alpha = 0.05$

- Η υπόθεση που θέλουμε να ελέγξουμε

$H_0: \; p_1 = p_2 = \dots = p_6 = \frac{1}{6}$

(δίκαιο νόμισμα)

Θεωρητικές (αναμενόμενες) τιμές $\theta_i = np_i = 120 \cdot \frac{1}{6} = 20$

A_i	1	2	3	4	5	6	Σύνολο
n_i	11	21	29	30	19	10	120
θ_i	20	20	20	20	20	20	120
$\frac{(n_i-\theta_i)^2}{\theta_i}$	4.05	0.05	4.05	5	0.05	5	18.20

$X^2 = 18.2$ από πίνακες $X^2_{6-1}(0.05) = X^2_5(0.05) = 11.07$

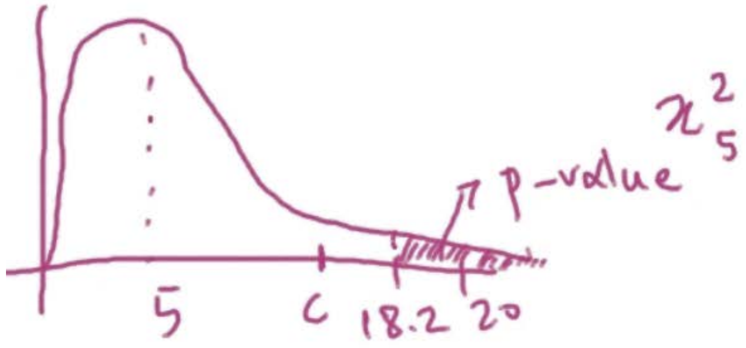
όμως $18.2 > 11.07$ άρα σε ε.σ. 0.05 απορρίπτεται η μηδενική υπόθεση

```
> 1-pchisq(18.2,df=5)
[1] 0.002705916
```

επειδή p-value =1-CDF.CHISQ(18.2,5)=0.0027

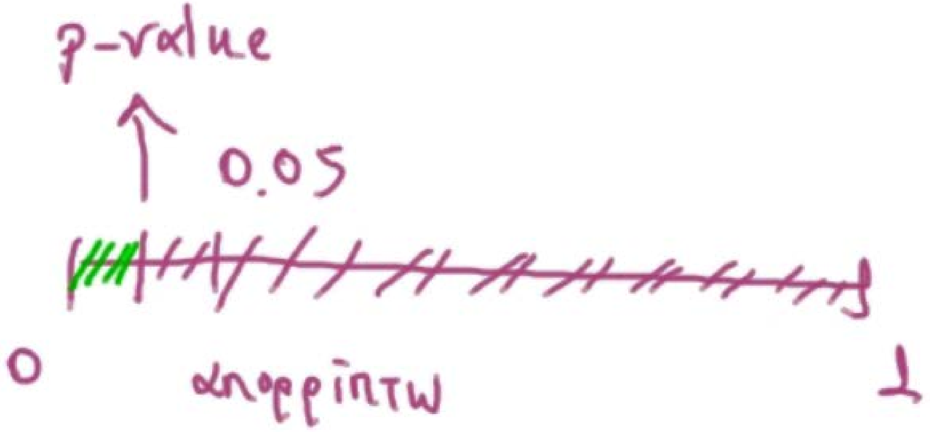
η Ho απορρίπτεται για κάθε επίπεδο σημαντικότητας >0.0027

p-value για το ζάρι



$c < 18.2$ απορρίπτω

$c > 18.2$ αποδέχομαι



$P(\chi^2_5 > 18.2) = 1 - \text{CDF.CHISQ}(18.2, 5)$

$= \text{p-value} < 0.05$