2°) Prise de décision et test du χ^2

Problème posé :

On veut déterminer à partir des effectifs (pas des fréquences) si le modèle théorique choisi et les valeurs réelles observées sont en adéquation

Autrement dit : on veut vérifier la validité du modèle théorique

On dispose donc de deux tableaux que l'on veut "confronter"

→ 1 tableau de k valeurs théoriques attendues notées E_i (E pour expected)

→ 1 tableau de k valeurs réelles observées notées O_i sur un échantillon donné.

Remarque : En supposant qu'on connaisse le total, on a ici $k-1$ degrés de liberté.

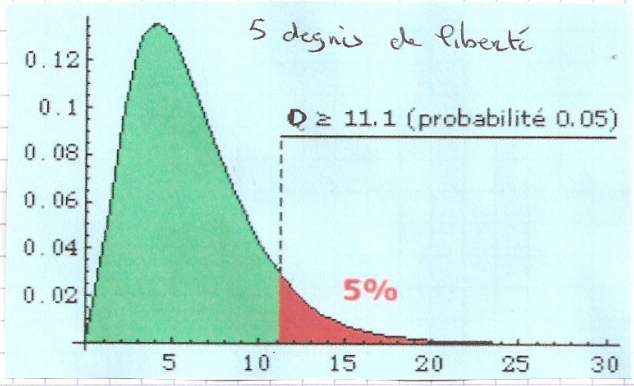
On considère la variable aléatoire D définie par

$$D = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Chaque échantillon observé donnera une valeur de D.

On admet alors que D suit approximativement une loi du χ^2 à $k-1$ degré de libertés.

Exemple pour 5 degrés de liberté



On a $p(D \geq 11,1) = 0,05$

Cela signifie que pour 95% des échantillons, la valeur de D sera inférieure à 11

Copyright : ©Programmaths 2020

On peut ainsi établir un test appelé le test du χ^2

Principe du test du χ^2

- On pose une hypothèse de départ notée H_0 (aussi appelée hypothèse nulle)

En règle générale on prend

H_0 : "le modèle théorique est valide"

- On fixe le risque d'erreur α : l'erreur de rejeter à tort l'hypothèse H_0 alors qu'elle est vraie

  Remarque : ce sera souvent 5% (par habitude)

- On calcule la valeur de $D = \sum \frac{(O_i - E_i)^2}{E_i}$

- On détermine le nombre de degrés de liberté (ddl)

- On lit dans la table de χ^2 la valeur correspondante : notée $\chi^2_{ddl,\alpha}$

si $D > \chi^2_{ddl,\alpha}$ Alors on rejette l'hypothèse H_0 au risque α

si $D \leq \chi^2_{ddl,\alpha}$ Alors on accepte l'hypothèse H_0

extrait de la table du χ^2

α	0,999	0,995	0,99	0,98	0,95	0,9	0,8	0,2	0,1	0,05	0,02	0,01	0,005	0,001
ddl														
1	0,0000	0,0000	0,0002	0,0006	0,0039	0,0158	0,0642	1,6424	2,7055	3,8415	5,4119	6,6349	7,8794	10,8276
2	0,0020	0,0100	0,0201	0,0404	0,1026	0,2107	0,4463	3,2189	4,6052	5,9915	7,8240	9,2103	10,5966	13,8155
3	0,0243	0,0717	0,1148	0,1848	0,3518	0,5844	1,0052	4,6416	6,2514	7,8147	9,8374	11,3449	12,8382	16,2662
4	0,0908	0,2070	0,2971	0,4294	0,7107	1,0636	1,6488	5,9886	7,7794	9,4877	11,6678	13,2767	14,8603	18,4668
5	0,2102	0,4117	0,5543	0,7519	1,1455	1,6103	2,3425	7,2893	9,2364	11,0705	13,3882	15,0863	16,7496	20,5150
6	0,3811	0,6757	0,8721	1,1344	1,6354	2,2041	3,0701	8,5581	10,6446	12,5916	15,0332	16,8119	18,5476	22,4577
7	0,5985	0,9893	1,2390	1,5643	2,1673	2,8331	3,8223	9,8032	12,0170	14,0671	16,6224	18,4753	20,2777	24,3219
8	0,8571	1,3444	1,6465	2,0325	2,7326	3,4895	4,5936	11,0301	13,3616	15,5073	18,1682	20,0902	21,9550	26,1245
9	1,1519	1,7349	2,0879	2,5324	3,3251	4,1682	5,3801	12,2421	14,6837	16,9190	19,6790	21,6660	23,5894	27,8772

Exemple 1 : On lance 600 fois un dé cubique. On obtient les résultats suivants :

face	1	2	3	4	5	6	Total
effectif	89	131	93	92	104	91	600

Le dé est-il parfaitement équilibré ?

- H_0 : le dé est équilibré.

  sous cette hypothèse l'effectif théorique est de 100 pour chaque face

- On fixe le risque d'erreur à 5% ($\alpha = 0{,}05$)

- $D = \frac{(89-100)^2}{100} + \frac{(131-100)^2}{100} + \frac{(93-100)^2}{100} + \frac{(92-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(91-100)^2}{100}$

  $\simeq 12{,}92$

- on a ici 5 degrés de liberté. donc

  $\chi^2_{5;0,05} = 11{,}07$

  $D > \chi^2_{5;0,05}$ donc on rejette l'hypothèse H_0.

Au risque de 5%, les écarts ne sont pas dus à la simple fluctuation d'échantillonnage. Il semble que le dé soit truqué.

Copyright : ©Programmaths 2020

Exemple 2 : On cherche à savoir à partir du génotype si la population des loups du parc de Yellostone est soumise à l'une des forces évolutives.
On vérifie alors la validité du modèle d'Handy-Weinberg à l'aide du tableau suivant

génotype	A//A	A//a	a//a	Total
Nombre de loups calculé avec H-W	48	287	430	765
Nombre de loups observés	31	321	413	765

- H_0 : le modèle H-W est valide : la population des loups de Yellowstone obéit à l'équilibre d'Handy-Weinberg, elle n'est soumise à aucune force évolutive
- On fixe le risque d'erreur à 5 % ($\alpha = 0,05$)
- $D = \frac{(31-48)^2}{48} + \frac{(321-287)^2}{287} + \frac{(413-430)^2}{430} = 10,72$
- Ici il y a 3 valeurs et on connait le total donc il y a 2 degrés de liberté.
- $\chi^2_{2;0,05} = 5,99$

$D > \chi^2_{2;0,05}$ donc on rejette H_0 au risque de 5 %.