

Option Statistique :

Test d'ajustement d'une loi empirique à une loi de probabilité théorique

I. Un exemple : En lançant successivement 60 fois un dé, un joueur obtient les résultats suivants :

i (face)	1	2	3	4	5	6
O_i (effectifs observés)	15	7	7	11	6	14

Si le dé n'est pas truqué, sa loi théorique est la **loi uniforme**, c'est-à-dire que chaque face a la même probabilité $\frac{1}{6}$ d'apparaître, donc les effectifs théoriques pour 60 lancers sont de 10 pour chaque face :

i (face)	1	2	3	4	5	6
T_i (effectifs théoriques)	10	10	10	10	10	10

Pour mesurer l'écart entre la loi empirique obtenue et la loi théorique supposée, on calcule la quantité

$$\sum_{i=1}^r \frac{(O_i - T_i)^2}{T_i} \quad (1)$$

appelée Khi-deux de l'échantillon. Comme l'on connaît l'effectif total, l'un des 6 effectifs observés peut se déduire des 5 autres : on dit alors que c'est un **khi-deux à 5 degrés de liberté**. Cette quantité est aléatoire puisque si l'on répète la série de 60 lancers on obtiendra un autre Khi-deux. On peut montrer que lorsque la taille (ici $n = 60$) de l'échantillon est assez grande, sa loi de probabilité a une densité $f(x)$ que l'on peut assimiler à celle d'une v.a. notée $\chi^2(5)$ (voir l'exercice I ci-dessous). On désigne par $x(\alpha)$ la valeur du Khi-deux telle que

$$\int_{x(\alpha)}^{+\infty} f(x)dx = \alpha, \quad (2)$$

c'est-à-dire la valeur telle que $P(\chi^2 \geq x(\alpha)) = \alpha$.

Pour tester la qualité de l'ajustement de la loi empirique du dé à une loi uniforme on procède à un **test du Khi-deux** de la façon suivante :

- on choisit l'hypothèse nulle H_0 selon laquelle "le dé est normal" que l'on teste contre l'hypothèse alternative H_1 selon laquelle "le dé est truqué".
- on choisit un seuil α , risque de rejet à tort de l'hypothèse H_0 , et on en déduit l'intervalle d'acceptation $I(\alpha) := [0, x(\alpha)]$ définie par la relation (2). Sous l'hypothèse nulle, $P(\chi^2 \notin I(\alpha)) = \alpha$.
- On effectue une série de n lancers (n suffisamment grand) puis on calcule le Khi-deux selon la formule (1).
- On conclut : si la valeur trouvée n'appartient pas à $I(\alpha)$, c'est-à-dire si le Khi-deux est plus grand que $x(\alpha)$, on rejette l'hypothèse H_0 et le dé est déclaré truqué. Sinon, l'hypothèse H_0 n'est pas rejetée et on en déduit que les variations d'effectifs observées ne permettent pas de mettre en doute le fait que le dé soit normal.

Remarque importante : Lorsque les loi empiriques et théoriques comportent r classes et que les effectifs théoriques sont déterminés à l'aide de p relations (parmi lesquelles on compte généralement l'effectif total, mais parfois aussi l'espérance et la variance empiriques), alors le **nombre de degrés de liberté de la loi du Khi-deux** est $r - p$, c'est-à-dire qu'il convient de calculer l'intervalle $I(\alpha) = [0, x(\alpha)]$ au moyen de la densité de $\chi^2(r - p)$.

II. Exercice 1 : L'exercice consiste à tracer la densité d'une loi du $\chi^2(5)$.

1. Dans la colonne A d'une feuille de calcul, que l'on intitulera x , entrer une série de valeurs pour le Khi-deux équiréparties entre 0 et 15, de pas 0, 1.
2. Dans la colonne B, que l'on intitulera $P(X \geq x)$, calculer au moyen de la fonction $\text{LOI.KHIDEUX}(x; 5)$ la quantité α telle que $x = x(\alpha)$.
3. Dans la colonne C, que l'on intitulera Densité de Khi-deux(5), calculer la densité de la loi $\chi^2(5)$ par différence de deux valeurs consécutives de la colonne B (par exemple en C3, calculer $B2 - B3$).
4. Tracer le graphe de cette densité comme fonction des valeurs de la colonne A.
5. Indiquer entre quelles valeurs de x de la colonne A se situe la valeur $x(\alpha)$ correspondant à $\alpha = 0,05$. Puis, dans une cellule vide, calculer précisément $x(\alpha)$ en utilisant la fonction $\text{KHIDEUX.INVERSE}(\alpha; 5)$. En déduire l'intervalle d'acceptation $I(\alpha)$ au seuil $\alpha = 0,05$. Même question pour le seuil $\alpha = 0,01$.
6. Pour l'impression de cette feuille de calcul, placer votre figure en haut à droite des trois colonnes A, B et C. N'imprimer que la partie haute de ces trois colonnes, celle qui tient sur la première page. Sous la figure, écrire (sur trois colonnes) les résultats de la question précédente:

Pour $\alpha = 0,05$ $x(\alpha) = \dots\dots\dots$

Pour $\alpha = 0,01$ $x(\alpha) = \dots\dots\dots$

III. Exercice 2 : L'exercice consiste à simuler un jeu de dé puis à tester la qualité du dé.

1. Dans la colonne A d'une nouvelle feuille de calcul, que l'on intitulera *Tirages*, simuler, à l'aide de la fonction $\text{ALEA}()$, $n = 60$ nombres aléatoires de loi uniforme sur l'intervalle $[0; 6]$. En colonne B, que l'on intitulera *Classes*, entrer les entiers 1, 2, ..., 6.
2. Pour simuler les résultats de 60 tirages d'un dé, on va regrouper les nombres aléatoires précédents en 6 classes $[0; 1[$, ..., $[5; 6[$ au moyen de la fonction histogramme de l'utilitaire d'analyse (vérifier que la somme des fréquences fait bien 60). Tracer l'histogramme.
3. On se propose de tester la qualité du dé à l'aide d'un test d'ajustement à la loi uniforme. Pour cela, ajouter dans une colonne immédiatement à droite de la colonne *fréquences*, que l'on intitulera *ecarts*, les quantités $\sum_{i=1}^r \frac{(O_i - T_i)^2}{T_i}$ calculées pour chaque classe, puis la somme des 6 nombres obtenus. Peut-on considérer, au seuil $\alpha = 0,05$ que le dé simulé est "normal"? Qu'en serait-il au seuil $\alpha = 0,1$?
4. Si vous en avez le temps, recommencer avec un échantillon de 600 nombres aléatoires permettant de simuler 600 lancés de dé. Comparer.

IV. Exercice 3 : Afin de tester le réglage d'une machine qui effectue l'emballage d'un produit en paquets de 50 grammes, on a prélevé et pesé 1000 sachets et on a obtenu les résultats suivants: le poids moyen est de 50,8 et l'écart type de 3,72 et l'on a:

poids	moins de 45	[45 ; 47[	[47 ; 49[	[49 ; 51[	[51 ; 53[	[53 ; 55[	[55 ; 57[	plus de 57
0_i (effectifs observés)	68	103	152	206	175	156	85	55

On souhaite tester l'hypothèse selon laquelle le poids des sachets produits par cette machine suit une loi normale $\mathcal{N}(50,8; 3,72)$.

1. Sur une nouvelle feuille de calcul, saisir en colonne A, intitulée *poids*, la liste des 7 classes du tableau précédent et en colonne B les effectifs observés correspondant.
2. En colonne C, on se propose de calculer les effectifs théoriques mais on va tout d'abord effectuer des calculs préliminaires en colonnes D et E que l'on intitulera respectivement i et $N(i)$. Pour i , saisir les nombres 45, 47, ..., 57 qui sont les extrémités droites des intervalles de classes (noter qu'il manque une ligne à cette colonne). En colonne E, on calculera, au moyen de la fonction LOI.NORMAL , la fonction de répartition de la loi théorique $F(i) = P(X \leq i)$ pour une v.a. X qui suit une loi $\mathcal{N}(50,8; 3,72)$. On calcule alors en colonne C les effectifs théoriques de la façon suivante: $T_i = 1000(F(i) - F(i-1))$, sauf en première ligne ($i = 45$) où $T_i = 1000(F(i))$ et en dernière ligne ($i > 57$) où $T_i = 1000(1 - F(i))$.

3. Représenter effectifs observés et effectifs théoriques sur un même graphique.
4. Insérer une nouvelle colonne D intitulée *Ecart*s, et y calculer les quantités $\sum_{i=1}^r \frac{(O_i - T_i)^2}{T_i}$ pour chaque classe, puis la somme des nombres obtenus. Quel est le nombre de degrés de liberté du χ^2 suivi approximativement par cette somme?
5. Peut-on considérer, au seuil $\alpha = 0,05$, que le poids des sachets suit une loi normale $\mathcal{N}(50,8 ; 3,72)$?

V. Exercice 4 : Dans le but de contrôler la qualité d'une machine qui fabrique des dizaines de milliers de pièces identiques (dont quelques unes sont défectueuses), on a prélevé 100 échantillons de 100 pièces et on a obtenu les résultats suivants :

Nb pièces défaut.	0	1	2	3	4	5	6	7	8	9 et plus
Nb observations	3	6	11	23	19	17	10	6	5	0

La moyenne des observations étant égale à 4 et le nombre de pièces fabriquées étant très grand, on peut assimiler l'échantillonnage à un tirage avec remise et donc penser que la v.a. qui mesure le nombre de pièces défectueuses dans un lot de 100 suit une loi binomiale $\mathcal{B}(100 ; 0,04)$. On se propose de tester cette hypothèse.

1. Saisir sur deux colonnes d'une nouvelle feuille de calcul les données ci-dessus en ayant soin de regrouper en une seule classe les deux classes les plus basses et les trois les plus hautes. On gardera ainsi 7 classes.
2. Calculer les effectifs théoriques pour chacune de ces classes.
3. Calculer le Khi-deux de cet échantillon et décider si, au seuil de 5%, il convient de rejeter ou non l'hypothèse envisagée (le Khi-deux a, ici encore, 5 degrés de liberté. Expliquer pourquoi).