Travaux Pratiques 03 : Intervalle de fluctuation et prise de décision

Exercice 1 : Une histoire de sucreries

Une production de M&M's est composée de six couleurs, supposées être réparties de manière égale : bleu, vert, jaune, rouge, orange, marron.

Jonathan ouvre un sachet contenant 16 M&M's, et affirme être lésé : il ne contient aucun bonbon bleu !!

Couleur	Effectif	Fréquence	Fréquence théorique
Bleu	0	0	1/6 ≃ 0,17
Vert	2	2/16 = 0,125	1/6
Jaune	3	3/16 ≃ 0,19	1/6
Rouge	4	4/16 = 0,25	1/6
Orange	5	5/16 ≃ 0,31	1/6
Marron	2	2/16 = 0,125	1/6

1. Compléter le tableau.
2. On s'intéresse à la proportion de M&M's bleus dans un sachet acheté au hasard en magasin. Il n'y a dans le sachet de l'exercice aucun M&M's bleu… Commenter cette situation.
   Le sachet contient en théorie 1/6ème de bonbons bleus. Mais les bonbons étant distribués au hasard dans les sachets, cette situation pourrait être possible.
3. On décide de **simuler** la production de 1000 sachets à l'aide du tableur :
   - Ouvrir la feuille « TP03 – Echantillonnage – MMS »
   - Compléter la cellule B4 avec la formule =ent(alea()*6)+1 ; nous conviendrons que les M&M's bleus sont représentés par le chiffre 1. Quelle autre simulation cette formule permet + elle ?
     Le lancer d'un dé à six faces, équilibré.
   - Simuler alors la production de 1000 sachets contenant 16 M&M's.
   - Compléter la cellule S4 avec la formule =NB.SI($B4:$P4;1). Que permet – elle ? Recopier la formule.
     Elle permet de compter le nombre de M&M's bleus présents dans le sachet.
   - Quel rang occupe le premier sachet qui ne contient aucun bonbon bleu ?
     Le $n^{ème}$ rang (n en fonction de la feuille de calcul obtenue)
   - Déterminer en W4 la moyenne de bonbons bleus par sachets
   - Déterminer les quartiles de la série en colonne W (formule pour W6 : =QUARTILE($S$4:$S$1003;1)).
   - ***Pour les plus avancés :*** déterminer dans les cellules prévues à cet effet le nombre de sachets contenant k bonbons bleus, pour k allant de 0 à 16. Puis leurs fréquences.
     Formule pour V12 : =NB.SI($S$4:$S$1003;U12)
4. Déterminer par le calcul ***l'intervalle de fluctuation***, au niveau de confiance 95%, de la proportion de bonbons bleus présent dans un sachet de n = 16 bonbons. Quelle explication pouvez – vous alors donner à Jonathan ?
   $I =]\frac{1}{6} - \frac{1}{\sqrt{16}} ; \frac{1}{6} + \frac{1}{\sqrt{16}}[=] -0,08 ; 0,42 [$. Obtenir 0 bonbons bleus dans le sachet, même si ce n'est pas le plus probable, est possible. Sur 1000 sachets, on peut estimer que 95% d'entre eux contiennent entre 0 et 40% de bonbons bleus.

Historique des couleurs de M&M's : 1941 – Marron ; 1960 – Rouge, Vert, Jaune ; 1964 – Orange ; 1976 – disparition du rouge ; 1987 – retour du Rouge ; 1995 – Bleu

Exercice 2 : Une histoire judiciaire : l'affaire Rodrigo Partida

Arrêté au Texas en Novembre 1976, Rodrigo Partida est condamné à huit ans de prison pour cambriolage d'une résidence et tentative de viol.
Il attaque ce jugement au motif que la désignation des jurés du comté est discriminante à l'égard des américains d'origine mexicaine : 79,1% de la population du comté est d'origine mexicaine, or sur les 870 personnes convoqués pour être jurés sur une période de référence, il n'y eût que 339 personnes d'origine mexicaine !

Compléter d'abord ce petit comparatif :

Population	***Jury***
N = ?	n = 870
p = 0,791	f = 0,390

La question, dès lors, est la suivante : est – il vraisemblable que le hasard puisse désigner un tel jury ?

Partie 1 : Simulation

Ouvrir la feuille "TP Echantillonnage – Partida"

1. Simuler la désignation d'un juré de ce comté sur la cellule A1 à l'aide de la formule : =ENT(ALEA() + 0,791)
Le résultat 1 signifie que le juré est "mexicain", le résultat 0 signifie qu'il ne l'est pas.
2. Simuler le tirage des 870 jurés, dans les cellules A1:A870.
3. Dans la cellule A871 entrer la formule =SOMME(A1:A870). Qu'obtient-on ? Appuyer plusieurs fois sur la touche F9 ; indiquer autour de quel nombre le total de jurés d'origine mexicaine semble osciller.
4. Dans la cellule A872, calculer la fréquence d'américains d'origine mexicaine dans le jury.
5. Constituer de la sorte 100 jurys.
6. Représenter avec un nuage de points les données de la ligne 872.
7. En utilisant la touche F9, estimer « à l'œil » une fourchette approximative dans laquelle se situe le nuage de points. Déterminer ensuite l'intervalle de fluctuation expérimental en X879 et Y879.
8. Arrivez- vous à obtenir un jury contenant une proportion de 39% d'américains d'origine mexicaine ? Comment expliquez-vous cette situation ?
Il semble impossible d'obtenir un tel jury, qui ne se situe pas dans l'intervalle de fluctuation.

Partie 2 : Intervalle de fluctuation

Calculer les bornes de l'intervalle de fluctuation, au niveau de risque 5%, et conclure.

$$I =]0{,}791 - \frac{1}{\sqrt{870}} ; 0{,}791 + \frac{1}{\sqrt{870}}[=]0{,}757 ; 0{,}825[$$

Extrait de l'attendu de la cour suprême dans l'affaire Partida

« Si les jurés étaient tirés au hasard dans l'ensemble de la population, le nombre d'américains mexicains dans l'échantillon pourrait alors être modélisé par une distribution binomiale… Etant donné que 79,1 % de la population est mexico américaine, le nombre attendu d'américains mexicains parmi les 870 personnes convoquées en tant que grands jurés pendant la période de 11 ans est approximativement 688. Le nombre observé est 339. Bien sûr, dans n'importe quel tirage considéré, une certaine fluctuation par rapport au nombre attendu est prévisible. Le point essentiel cependant, est que le modèle statistique montre que les résultats d'un tirage au sort tombent vraisemblablement dans le voisinage de la valeur attendue… La mesure des fluctuations prévues par rapport à la valeur attendue est l'écart type, défini pour la distribution binomiale comme la racine carrée de la taille de l'échantillon (ici 870) fois la probabilité de sélectionner un américain mexicain (ici 0,791) fois la probabilité de sélectionner un non américain mexicain (ici 0,209)… Ainsi, dans ce cas, l'écart type est approximativement de 12. En règle générale pour de si grands échantillons, si la différence entre la valeur attendue et le nombre observé est plus grand que deux ou trois écarts types, alors l'hypothèse que le tirage du jury était au hasard serait suspect à un spécialiste des sciences humaines. Les données sur 11 années reflètent ici une différence d'environ 29 écarts types. Un calcul détaillé révèle qu'un éloignement aussi important de la valeur attendue se produirait avec moins d'une chance sur 10^{140}. »