

Statistiques Appliquées  
Contrôle Continu N°2  
TD N°5  
Gwenn PARENT

**Questions de cours : (3 points)**

1. Rappelez les propriétés des estimateurs MCO ?
2. Donnez leur formule algébrique pour un ajustement simple.
3. Pourquoi ces estimateurs sont-ils des variables aléatoires ?

**Problème : Evaluation du prix moyen des cigarettes, du pain et du riz sur un échantillon de 26 pays. (17 points)**

”Tabac et pauvreté forment à eux deux un cercle vicieux. Dans la plupart des pays, le tabagisme est généralement plus répandu parmi les plus pauvres. C’est pourquoi les dépenses de tabac représentent une part importante du revenu des familles défavorisées. Or, l’argent qui passe dans le tabac ne peut être dépensé pour des besoins essentiels comme l’alimentation, le logement, l’éducation et les soins de santé. Le tabac peut en outre aggraver la pauvreté des fumeurs et de leurs familles du fait que ces derniers sont beaucoup plus susceptibles de tomber malade et de mourir prématurément d’un cancer, d’une crise cardiaque, d’une maladie respiratoire ou d’autres maladies liées au tabagisme, privant leurs familles d’un revenu très précieux et leur imposant des dépenses supplémentaires pour les soins de santé. Par ailleurs, même si l’industrie du tabac emploie des milliers de personnes, la grande majorité d’entre elles gagnent très peu tandis que les grandes compagnies de tabac engrangent d’énormes bénéfices.” Organisation Mondiale de la Santé (2000).

**Toutes les questions du problème sont indépendantes, vous pouvez les traiter séparément (sauf les questions 6 et 11 qui sont des applications numériques des questions précédentes).**

Nous souhaitons étudier le prix moyen des cigarettes, du pain et du riz dans le monde. Nous disposons pour cela des prix (calculés en temps de travail pour un ouvrier) d’un paquet d’une marque présente dans tous les pays sélectionnés, d’un paquet d’une marque locale, ainsi que d’un kg de pain et d’un kg de riz pour un échantillon de 26 pays (13 pays développés et 13 pays en développement). (cf. le tableau ci-joint). Nous faisons l’hypothèse que les temps moyens qu’un ouvrier de chaque pays doit travailler pour se payer ces différents biens suivent une loi normale d’espérance  $m$  et de variance  $\sigma^2$ . Nous adopterons par la suite les notations suivantes :

$$X_1, X_2, \dots, X_N \approx \mathcal{N}(m_1, \sigma_1^2) \text{ pour les Marlboro}$$

$$Y_1, Y_2, \dots, Y_N \approx \mathcal{N}(m_2, \sigma_2^2) \text{ pour les marques locales}$$

$$Z_1, Z_2, \dots, Z_N \approx \mathcal{N}(m_3, \sigma_3^2) \text{ pour le pain}$$

$$T_1, T_2, \dots, T_N \approx \mathcal{N}(m_4, \sigma_4^2) \text{ pour le riz}$$

1. En reprenant les notations définies ci-dessus, indiquez à quoi correspondent les chiffres 25, 77 puis 15, 67 du tableau 1. Expliquez quelle est la différence entre un paramètre, et son estimateur. (1,5 points)
2. Qu'observez-vous à première vue en comparant le temps de travail moyen nécessaire à un ouvrier pour l'achat d'un paquet de Marlboro et d'un kg de riz ? Quelle analyse succincte pouvez-vous faire du tableau 2 ? (1 point)
3. Calculez l'estimateur du maximum de vraisemblance du paramètre  $m_1$ . Comment s'appelle l'estimateur que vous trouvez ? (3 points)

Indication : Fonction de densité de la loi normale  $\mathcal{N}(m, \sigma^2)$  :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

4. Vous disposez dans le tableau 1 de deux indicateurs statistiques, rappelez comment ces indicateurs ont été trouvés (donnez les formules de calcul) ? Quel est le lien entre écart-type et variance ? Quel est l'intérêt de l'écart-type par rapport à la variance ? Quels autres outils statistiques connaissez-vous, qui permettent d'apporter des indications supplémentaires sur la véritable moyenne  $m_4$  du temps nécessaire à l'achat d'un kg de riz dans le monde ? (2 points)
  5. Vous vous interrogez sur la significativité de la différence entre le prix moyen d'un kg de riz, et celui d'un paquet de Marlboro. Vous souhaitez donc connaître (avec une confiance de  $\alpha\%$ ) une borne supérieure de la moyenne du temps nécessaire à l'achat d'un kg de riz. Comment procédez-vous ? (1,5 points)
  6. Application numérique avec  $\alpha = 80\%$ . (0,5 point)
  7. Afin de pouvoir comparer cette borne supérieure de confiance pour  $m_4$ , avec le prix d'un paquet de Marlboro, calculez une borne inférieure de confiance (toujours à 80% de confiance) pour  $m_1$ . Que concluez-vous sur la différence de prix moyens entre ces deux biens ? (1,5 points)
  8. La variance du prix moyen d'un kg de pain est inconnue, donnez un encadrement à 90% de confiance de la moyenne  $m_3$ . Indication : variance empirique corrigée :  $S^2 = 634$  (2 points)
- Nous souhaitons désormais prévoir le prix d'un kg de riz au Kenya.
9. Quelles hypothèses doit-on poser pour étudier cette question à partir des éléments dont nous disposons ? Ces hypothèses vous paraissent-elles justifiées, expliquez. (1 point)
  10. Nous considérons que les hypothèses nécessaires pour l'étude de cette question sont vérifiées malgré tout, donnez un encadrement à 90% de confiance du prix **prévu** d'un kg de riz au Kenya :  $T_k$ . (2 points)
  11. Procédez également à l'application numérique à partir des informations disponibles dans le tableau 2 (nous considérerons ici que le Kenya est un pays en développement) ? Expliquez d'où provient la différence avec la prévision de la question précédente. (1 point)