

Eléments de statistique

Examen du 21 janvier 2017

Question 1

La quantité de cannabis absorbée mensuellement par un consommateur régulier suit une loi de moyenne $\mu_{\mathcal{X}} = 149g$ et d'écart-type $\sigma_{\mathcal{X}} = 40,3g$. On tire un échantillon i.i.d. de 15 consommateurs réguliers de cannabis.

- Sous l'hypothèse d'une loi normale pour la variable \mathcal{X} , estimer la probabilité que la moyenne de l'échantillon soit comprise entre 134g et 164g? (**4 points**)
- Sans faire d'hypothèse sur la distribution de la variable \mathcal{X} , que peut-on dire sur la probabilité que la moyenne de l'échantillon soit comprise entre 134g et 164g? (**4 points**)
- La probabilité calculée au point b) devait-elle forcément être plus petite que celle calculée au point a)? (**2 points**)

Supposons désormais que les consommateurs achètent des paquets de 25g et que le nombre théorique de sachets achetés suit la loi indiquée à la table 1. On tire toujours un échantillon i.i.d. de 15 consommateurs réguliers de cannabis.

TABLE 1 – Quantité achetée mensuellement par un consommateur de cannabis.

| Nombre de sachets de 25g | Probabilité |
|--------------------------|-------------|
| 3 | 5% |
| 4 | 15% |
| 5 | 20% |
| 6 | 25% |
| 7 | 18% |
| 8 | 10% |
| 9 | 5% |
| 10 | 2% |

- Que valent l'espérance et la variance de la moyenne d'échantillon? (**2 points**)
- Que valent les espérances de s_x^2 et de s_{n-1}^2 ? (**2 points**)
- Peut-on utiliser les formules du formulaire pour calculer les variances de s_x^2 et de s_{n-1}^2 ? Justifier. (**1 point**)
- Déterminez le rapport entre les variances de s_x^2 et de s_{n-1}^2 . (**1 point**)
- Quel est l'intérêt de corriger la variance? Quel est le désavantage? (**2 points**)
Suggestion : repartez de vos réponses aux sous-questions e) et g).
- Si vous deviez calculer la probabilité que la moyenne de l'échantillon soit comprise entre 134g et 164g dans les conditions qui nous occupent, pensez-vous qu'elle serait forcément plus grande que celle calculée au point b)? (**2 points**)

Question 2

On étudie la vitesse de consommation du carburant de la fusée à propulsion à propergol solide Ariane 5. Selon les ingénieurs qui ont conçu les propulseurs d'appoints (EAP), la consommation de poudre moyenne devrait être de 2 tonnes par seconde. Entre 2010 et 2012, Ariane 5 a effectué 18 lancements et on a respectivement observé les consommations suivantes : 1,99 tonnes/sec, 2,12 tonnes/sec, 1,79 tonnes/sec, 2,11 tonnes/sec, 1,83 tonnes/sec, 2,3 tonnes/sec, 2,12 tonnes/sec, 2,09 tonnes/sec, 2,23 tonnes/sec, 1,93 tonnes/sec, 2,28 tonnes/sec, 1,99 tonnes/sec, 1,94 tonnes/sec, 1,87 tonnes/sec, 1,79 tonnes/sec, 1,91 tonnes/sec, 2,02 tonnes/sec et 2,05 tonnes/sec.

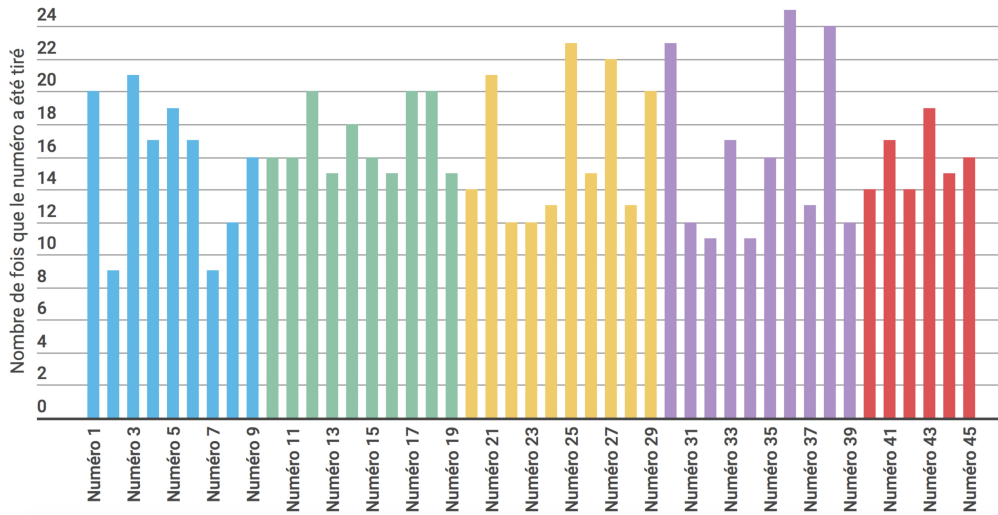
- a) Réalisez un test bilatéral pour vérifier l'hypothèse que les EAP ont une consommation moyenne de 2 tonnes par seconde, au seuil de signification $\alpha = 1\%$.
(4 points)
- b) Réalisez un graphique de l'évolution du risque β en fonction de la valeur théorique (inconnue) de la consommation moyenne de poudre. Pour ce faire, calculez la valeur du risque β pour les différentes valeurs possibles suivantes : 1,85 tonnes/sec, 1,9 tonnes/sec, 1,95 tonnes/sec, 2 tonnes/sec, 2,05 tonnes/sec, 2,1 tonnes/sec et 2,15 tonnes/sec et interpolez ces points sur le graphique. Pour faciliter les choses, on supposera que l'écart-type de la consommation de poudre est connu, et vaut dans chacun des cas $\sigma = 0,1566$ tonnes/seconde. (6 points)
- c) Si on avait travaillé avec un échantillon plus grand, la courbe d'évolution du risque β en fonction de la valeur théorique (inconnue) de la consommation moyenne de poudre aurait-elle été identique à celle obtenue au point précédent ? Si non, aurait-elle été située en dessus ou au dessous de celle-ci ? Expliquez. (2 points)

Lors de la plupart des lancements, Ariane 5 transporte deux satellites, mais il arrive parfois qu'un seul satellite soit embarqué (et jamais aucun). Sur les 18 lancements effectués entre 2010 et 2012, 34 satellites ont été transportés.

- d) Si vous deviez utiliser la méthode du maximum de vraisemblance pour estimer la proportion de lancements d'Ariane 5 lors desquels un seul satellite est embarqué, qu'obtiendriez-vous ? Justifiez. (2 points)
- e) Utilisez l'approche bayésienne pour estimer la proportion de lancements d'Ariane 5 lors desquels un seul satellite est embarqué. Vous supposerez que cette proportion suit a priori une loi $\beta(3, 27)$. (4 points)
- f) Si vous deviez utiliser l'approche bayésienne pour construire un intervalle de confiance à 95% de la proportion de lancements d'Ariane 5 lors desquels un seul satellite est embarqué, à quel moment du raisonnement auriez-vous besoin d'une aide logicielle pour continuer ? Quelle équation poserait problème ? (2 points)

Question 3

La loterie nationale belge a publié les statistiques des numéros de lotto qui sont sortis le plus souvent en 2016 :



Pour que les calculs ne soient pas trop longs et que le graphique ne prenne pas trop de temps, on s'intéresse uniquement aux numéros de 1 à 15.

- Au vu de ces statistiques, peut-on conclure au seuil $\alpha = 5\%$ que tous les numéros ont la même probabilité d'être tirés ? (**6 points**)
- Réalisez un Boxplot du nombre de tirages par numéro. Y a-t-il des données aberrantes ? (**5 points**)

On tire un échantillon i.i.d de 18 personnes ayant gagné plus d'un million au lotto, depuis que celui-ci est passé à 45 numéros le 1 octobre 2011. La moyenne de cet échantillon est de 3 279 654 €, avec un écart-type de 529 567 €.

- Donnez l'intervalle de confiance à 80% du gain moyen obtenu par les personnes ayant gagné plus d'un million au lotto. (**4 points**)

Un site internet prétend que le lotto truque les résultats du mois de décembre, pour qu'il y ait plus de gagnants pendant les fêtes de fin d'année. En décembre 2016, 28 269 230 bulletins ont été remplis et 7 068 007 d'entre-eux furent gagnants. Sur les autres mois de 2016, 291 036 295 ont été remplis et 72 756 963 d'entre-eux furent gagnants. En décembre 2015, 26 625 621 bulletins ont été remplis et 6 656 906 d'entre-eux furent gagnants. Sur les autres mois de 2015, 277 390 417 ont été remplis et 69 347 148 d'entre-eux furent gagnants.

- Effectuez un test d'indépendance au seuil $\alpha = 2\%$ pour voir s'il y a indépendance entre le fait de jouer en décembre et de gagner. (**5 points**)