

MSI : examen final (1h30)



Documents autorisés : calculatrice, feuille A4 manuscrite recto/verso. Les exercices sont indépendants.

QCM (5 points)

Les réponses sont à renseigner sur la feuille jointe au sujet et à remettre dans la copie.

Plusieurs réponses peuvent être vraies pour chaque question. Une question est dite juste si et seulement si toutes les affirmations vraies sont cochées et aucune affirmation fausse n'est cochée.

- 1 question juste : +1/2 de points
- 1 question fausse : -1/2 de points
- 1 question sans réponse : 0 point

Exemples :

- 8 questions justes et 2 sans réponse = 4 points
- 10 questions justes = 5 points
- 5 questions justes, 5 questions fausses = 0 points

Question 1 :

La théorie de l'estimation travaille avec des variables aléatoires. Le mode de tirage considéré pour obtenir un échantillon est

- A. L'échantillonnage aléatoire simple.
- B. La méthode des quotas.
- C. L'échantillonnage systématique.
- D. Aucune de ces affirmations n'est vraie.

Question 2 :

- A. L'échantillonnage systématique nécessite d'avoir une liste de l'ensemble de la population.
- B. Des méthodes d'enquêtes non aléatoires sont souvent utilisées dans les sondages politiques.
- C. Dans un échantillonnage stratifié, les strates sont obligatoirement sans individu en commun.

- D. L'échantillonnage par grappe nécessite d'avoir une liste de l'ensemble de la population.
- E. Aucune de ces réponses n'est vraie.

Question 3 :

Des policiers de Cork font des contrôles sur des individus qui leur paraissent suspects à la sortie d'un pub.

- A. La méthode d'échantillonnage est une méthode aléatoire simple.
- B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié.
- C. La méthode d'échantillonnage est une méthode aléatoire systématique.
- D. La méthode d'échantillonnage est une méthode des quotas.
- E. Aucune de ces affirmations n'est vraie.

Question 4 :

Pour mesurer le temps passé par les clients dans son magasin, un commerçant de Dublin chronomètre un client sur dix, dans l'ordre d'entrée au magasin.

- A. La méthode d'échantillonnage est une méthode aléatoire simple.
- B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié.
- C. La méthode d'échantillonnage est une méthode aléatoire systématique.
- D. La méthode d'échantillonnage est une méthode des quotas.
- E. Aucune de ces affirmations n'est vraie.

Question 5 :

Pour une enquête sur des élèves de première année du cycle junior, dans le comté du Tipperary, on tire au hasard un établissement secondaire sur 10 du comté et on interroge tous les élèves des établissements ainsi sélectionnés.

- A. La méthode d'échantillonnage est une méthode aléatoire simple.
- B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié.
- C. La méthode d'échantillonnage est une méthode aléatoire par grappe.
- D. La méthode d'échantillonnage est une méthode d'échantillonnage à plusieurs degrés.
- E. Aucune de ces affirmations n'est vraie.

Question 6 :

- A. Un échantillon aléatoire de 1 000 Irlandais est aussi représentatif des Irlandais qu'un échantillon aléatoire de 1 000 Anglais l'est des Anglais.
- B. Il existe des méthodes d'échantillonnage non aléatoires utilisées en statistique.
- C. Doubler la taille de l'échantillon permet de réduire de moitié les fluctuations.
- D. A taille égale, les résultats des échantillons pris dans une population plus homogène seront plus sujets aux fluctuations d'échantillonnage.
- E. Aucune de ces affirmations n'est vraie.

Question 7 :

On souhaite effectuer des mesures de consommation de Guinness dans les différents comtés du Pays en donnant plus d'importance aux comtés pour lesquels la variabilité est plus grande.

- A. La méthode d'échantillonnage est une méthode aléatoire simple.
- B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié avec allocation proportionnelle.
- C. La méthode d'échantillonnage est une méthode aléatoire systématique.
- D. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié avec allocation optimale.
- E. Aucune de ces affirmations n'est vraie.

Question 8 :

- A. Une statistique qui conserve l'information sera qualifiée d'exhaustive.
- B. Les statistiques exhaustives ne conservent pas l'information.
- C. Le principe de factorisation nous donne un moyen de reconnaître si une statistique est exhaustive.
- D. Pour les lois de type exponentiel, il n'existe pas de statistiques exhaustives.
- E. Aucune de ces affirmations n'est vraie.

Question 9 :

- A. La borne de Cramer Rao est la borne maximale de la variance, égale à l'information de Fisher.
- B. Un estimateur qui atteint la borne de Cramer Rao est qualifié d'efficace.
- C. Les estimateurs du maximum de vraisemblance ne sont jamais efficaces.
- D. Aucune de ces affirmations n'est vraie.

Question 10 :

Dans la théorie des tests statistiques :

- A. La puissance est la probabilité de rejeter H_0 sachant qu'en vérité, H_0 était vraie.
- B. La puissance est la probabilité d'opter pour H_1 alors que H_0 est la vérité.
- C. La puissance est un concept à minimiser pour améliorer le test.
- D. La puissance est le complémentaire de l'erreur de première espèce alpha.
- E. Aucune de ces affirmations n'est vraie.

EXERCICE 1

NOM :

Prénom :

Groupe :

Les réponses sont à renseigner dans la feuille jointe exclusivement :

Mettre une croix quand l'affirmation est vraie.

A

B

C

D

E

N° de question

1

2

3

4

5

6

7

8

9

10

Cadre réservé au correcteur

Nombre de questions justes =

Nombre de questions fausses =

Nombre de questions sans réponse =

d'où la note sur 5 points :

Problème : le rugby irlandais (15 points)

Les trois parties sont indépendantes.

Le XV du Trèfle, l'équipe nationale irlandaise de rugby, s'illustre depuis 1875 et un premier test match contre l'Angleterre disputé à Dublin. Malgré une défaite inaugurale 7-0 contre le XV de la Rose, l'équipe irlandaise s'est affirmée au fil du temps comme l'une des sélections nationales les plus fortes du monde, comme en atteste d'ailleurs aujourd'hui sa troisième position au classement de l'IRB (<http://www.worldrugby.org/rankings>).

Cet exercice se penche sur les performances de l'équipe irlandaise au sein du tournoi des 6 Nations, compétition européenne existant sous sa forme actuelle depuis 2000, et regroupant les sélections anglaise, irlandaise, galloise, écossaise, française et italienne. Le tableau ci-dessous résume les performances du XV du Trèfle sur 15 années de compétition, soit 75 matches :

Tableau 1 - Performances irlandaises au Tournoi des VI Nations, depuis 2000

Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Classement	3	2	3	2	2	3	2	2	4	1	2	3	3	5	1	-
Nombre Victoires	3	4	3	4	4	3	4	4	2	5	3	3	2	1	4	49
Nombre essais	17	11	16	10	17	12	12	17	9	12	11	10	13	5	16	188

PARTIE 1 : estimation

Question 1 (2 points)

Soit X l'évènement : « L'Irlande gagne un match du Tournoi des 6 Nations. » On suppose que X suit une loi de Bernoulli de paramètre p . Cette loi est caractérisée par la fonction de masse suivante :

$$P(X = x) = p^x(1 - p)^{1-x} \mathbb{1}_{\{0,1\}}(x)$$

Estimez la valeur de son unique paramètre avec la méthode du maximum de vraisemblance (faire toute la démonstration puis l'application numérique grâce aux données du problème).

Question 2 (1 point)

On suppose que les résultats des différents matchs sont indépendants. Soit Y la variable aléatoire représentant le nombre de victoires irlandaises pour chaque tournoi des 6 Nations. En sachant que chaque équipe doit disputer 5 matchs par tournoi, et en vous aidant de la réponse à la question précédente, identifier la nature du phénomène (quelle loi ? valeur des paramètres ?).

Question 3 (1 point)

La Figure 1 générée avec le logiciel R ci-dessous reprend les informations du Tableau 1 relatives au nombre d'essais inscrits par les irlandais à chaque édition du Tournoi. La courbe orange constitue l'ajustement de la série par la loi Normale.

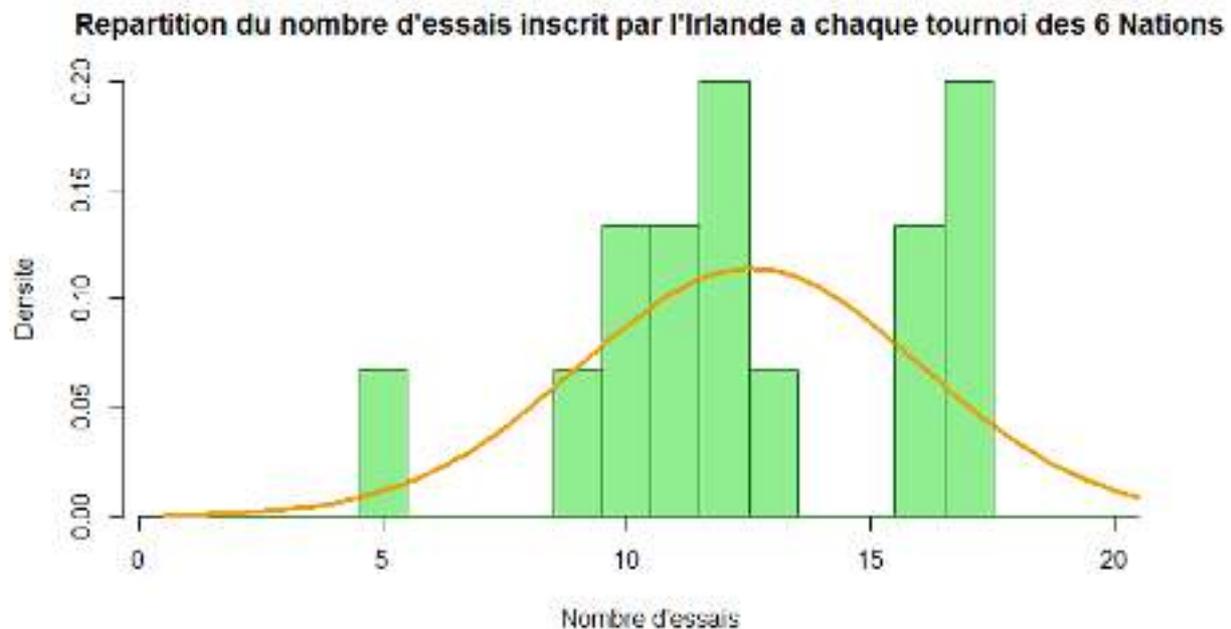


Figure 1 - Répartition du nombre d'essai inscrits par l'Irlande, depuis 2000

Donnez les formules des estimateurs du maximum de vraisemblance des paramètres de la loi Normale (sans démonstration). A l'aide des informations contenues dans le tableau ci-dessus, déterminez la valeur de ces estimateurs.

Question 4 (1 point)

Un test de Shapiro est effectué sur cette série. Comme en TP, le test se fonde sur l'hypothèse nulle selon laquelle la distribution suit une loi normale. Le test effectué sous R retourne une *p-value* de 0.17. Que pouvez-vous en conclure, au niveau de confiance 90% ? Commentez par rapport à l'allure de la figure 1.

PARTIE 2 : intervalles de confiance et tests d'hypothèses

Question 5 (2 points)

On suppose à présent que le nombre d'essais inscrits par tournoi par l'équipe d'Irlande suit une loi normale $N(m, \sigma)$. A partir de notre échantillon des 15 tournois de 2000 à 2014, donnez un intervalle de confiance pour le paramètre m au niveau de confiance 90%.

Question 6 (2 points)

Peut-on contredire l'hypothèse selon laquelle le nombre moyen d'essais marqués dans un tournoi par l'Irlande vaut 12 ? On décrira bien toutes les étapes du test statistique, de la construction à la décision. Le risque est fixé à 5%.

Question 7 (1 point)

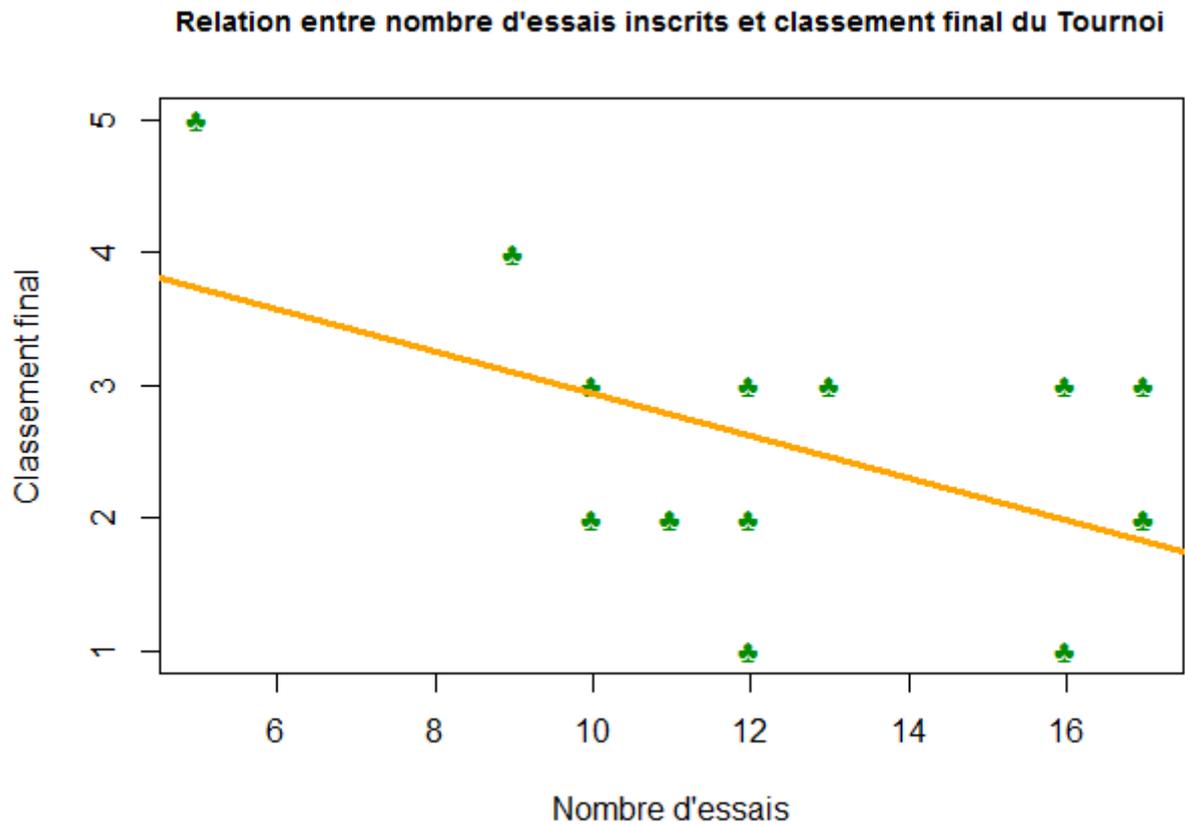
Répondez à la question 6 à l'aide de la question 5. Expliquez et illustrez ainsi le lien entre intervalles de confiance et test.

Question 8 (2 points)

En suppose à présent qu'en vérité, le nombre moyen d'essais marqués par l'Irlande ne vaut pas 12 mais 12.5. Quelle est l'erreur de deuxième espèce du test construit à la question 6 ? Que vaut donc la puissance du test ? Comment expliquer les résultats obtenus ?

PARTIE III : régression linéaire

La figure suivante représente le classement de l'équipe d'Irlande dans le tournoi en fonction du nombre d'essais inscrits. Elle a été produite à partir du tableau 1 (mêmes données, même problème)



La droite en rouge a été construite grâce à la fonction de R $lm(\text{Classement} \sim \text{Nombre d'essais})$.

Question 9 (1 point)

Comment appelle-t-on cette droite ? Quelle est la variable explicative ? Quelle est la variable expliquée ?

Question 10 (1 point)

Représentez sur la figure les résidus empiriques et expliquez en détails ce que sont ces résidus (on rendra la figure avec la copie).

Question 11 (1 point)

Quel est le coefficient de détermination R^2 ? Que représente ce coefficient ?