

# Travaux Dirigés de Statistiques Inférentielles

Ecole Supérieure d'Informatique (ESI)

Deuxième année

## Estimation ponctuelle et estimation par intervalle de confiance

### Exercice 1

Afin de mieux gérer les demandes de crédits de ses clients, un directeur d'agence bancaire réalise une étude relative à la durée de traitement des dossiers, supposée suivre une distribution normale.

Un échantillon, non exhaustif de 30 dossiers a donné :

Durée (mn)	[0;10]	[10;20]	[20;30]	[30;40]	[40;50]	[50;60]
Effectif	3	6	10	7	3	1

1. Calculer la moyenne et l'écart-type des durées de traitement des dossiers de cet échantillon.
2. En déduire les estimations ponctuelles de la moyenne  $m$  et de l'écart-type  $\sigma$  de la population des dossiers.
3. Donner une estimation de  $m$  par intervalle de confiance au seuil de risque 5%.

### Exercice 2

Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés; les observations sur 100 employés tirés au sort sont les suivantes:

taux de cholestérol en cg:(centre classe)	effectif d'employés
120	9
160	22
200	25
240	21
280	16
320	7

1. Calculer la moyenne  $m_e$  et l'écart-type  $\sigma_e$  sur l'échantillon.
2. Estimer la moyenne et l'écart-type pour le taux de cholestérol dans toute l'entreprise.
3. Déterminer un intervalle de confiance pour la moyenne.

### Exercice 3

Un échantillon de 10000 personnes sur une population étant donné, on sait que le taux moyen de personnes à soigner pour un problème de cholestérol élevé est de 7,5%. Donner un intervalle dans lequel on soit "sûr" à 95%, de trouver le nombre exact de personnes à soigner sur les 10000.

### Exercice 4

Un vol Bobo Dioulasso - Ouagadougou est assuré par un Airbus de 150 places ; pour ce vol des estimations ont montré que la probabilité pour qu'une personne confirme son billet est  $p = 0.75$ . La compagnie vend  $n$  billets,  $n > 150$ . Soit  $X$  la variable aléatoire "nombre de personnes parmi les  $n$  possibles, ayant confirmé leur réservation pour ce vol".

1. Quelle est la loi exacte suivie par  $X$  ?
2. Quel est le nombre maximum de places que la compagnie peut vendre pour que, à au moins 95%, elle soit sûre que tout le monde puisse monter dans l'avion, c'est-à-dire  $n$  tel que :  $P[X > 150] \leq 0.05$ ?
3. Reprendre le même exercice avec un avion de capacité de 300 places; faites varier le paramètre  $p = 0, 5; p = 0, 8$ .

## Tests statistiques

### Exercice 1

Un échantillon de 112 malades atteints d'un cancer du colon a été comparé à 185 témoins non malades quant à leur consommation moyenne de caféine. Pour les malades, cette consommation moyenne est égale à 147.2 mg/jour (l'écart-type échantillonnal est de 101.8mg/j) et pour les témoins, elle vaut 132.9 mg/j (l'écart-type échantillonnal est de 115.7 mg/j).

1. Tester, avec un risque de première espèce  $\alpha = 5\%$ , si la consommation moyenne de caféine diffère entre les malades et les non malades.
2. Peut-on inférer une association entre la consommation de caféine et le cancer du colon ?

### Exercice 2

Pour un sondage électoral, on constitue deux échantillons d'électeurs de tailles 300 et 200 respectivement dans 2 circonscriptions A et B. Cela met en évidence des intentions de vote de 56% et 48% pour un candidat donné. Tester, au seuil de 5% les hypothèses suivantes:

1. Il y a une différence entre les circonscriptions.

2. Le candidat est préféré dans la circonscription A.

### Exercice 3

Une entreprise de logiciels a développé un nouvel algorithme de tri qu'elle prétend plus rapide que l'algorithme standard actuel. Pour tester cette affirmation, les temps d'exécution des deux algorithmes ont été mesurés sur 30 jeux de données différents. Les mesures de temps (en millisecondes) sont supposées suivre une distribution normale. Les temps d'exécution (en millisecondes) des deux algorithmes sur les 30 jeux de données sont les suivants :

Algorithme standard	102	105	98	110	115	97	101	103	99	100	104
Nouvel algorithme	95	96	92	97	100	94	93	91	92	94	95

1. Formulez les hypothèses nulles et alternatives pour tester si le nouvel algorithme est significativement plus rapide que l'algorithme standard.
2. Calculez la différence des temps d'exécution pour chaque jeu de données.
3. Effectuez un test t bilatéral pour les différences de temps d'exécution. Utilisez un niveau de signification de 5 %.
4. Interprétez les résultats du test.

### Exercice 4

On se demande s'il y a un lien entre la réussite ou l'échec à un examen et le fait que l'étudiant assiste ou non au cours de révision. Les résultats pour 225 étudiants choisis au hasard sont indiqués ci-dessous. Choisir un seuil critique de 1%. Peut-on dire que ceux qui assistent au cours de révision sont plus nombreux à réussir l'examen ?

Examen			
Révision	Réussi	Échoué	Total
Oui	79	21	100
Non	76	49	125
Total	155	70	225