IMBS3 - ISCID-CO, site de Dunkerque, 2015/2016

Mathématiques appliquées à la gestion - TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES

Fiche de Mathématiques 1 - Prérequis - Exercices de révision.

Un test statistique est un procédé d'inférence : son but est d'énoncer des propriétés de la population en s'appuyant sur un échantillon d'observations. À l'aide d'un test, on construit aussi des intervalles de confiance qui expriment le degré de certitude associé à une simulation. L'objectif du test est de répondre à des problèmes décisionnels dans un environnement incertain. Par exemple on peut se demander à partir d'un échantillon si la taille des hommes est différente de la taille des femmes, si un nouveau médicament est vraiment plus efficace ou encore si l'incidence du cancer du poumon est plus élevée actuellement qu'il y a vingt ans. On ne dispose en pratique que de l'échantillon soumis aux fluctuations pour répondre à ces questions. L'objet des tests statistiques est de distinguer ce qui est plausible de ce qui est trop peu vraisemblable. Ce cours-TD introduit les concepts nécessaires pour développer et appliquer les tests paramétriques et non paramétriques.

1 Prérequis

Il est indispensable dans ce cours de maîtriser les bases de proababilités nécessaires à la compréhension des méthodes d'analyse statistique ainsi que les notions de base pour l'estimation des paramètres et les tests d'hypothèses.
Il faudra donc connaître

- Le calcul des probabilités et des variables aléatoires réelles
- Quelques lois de probabilité
  - discrètes telles que les lois de Bernoulli, les lois binomiales, les lois multinomiales, et les lois de Poisson,
  - continues telles que les lois normales, les lois exponentielles, les lois gamma, les lois du χ^2, les lois béta, les lois de Fischer-Snédécor et les lois de Student.
- Les approximations telles que
  - l'approximation normale de la loi binomiale,
  - l'approximation normale d'une somme de variables indépendantes,
  - l'approximation de Poisson de la loi binomiale,
  - l'approximation normale du χ^2.
- Le principe de l'estimation ponctuelle et celui de l'estimation par intervalles de confiance.

2 Exercices de révision

Exercice 1 Dans cet exercice, les probabilités demandées sont calculées à 10^{-3} près.
On envisage l'installation d'une pompe à chaleur en relève de chaudière dans un hôtel "deux étoiles" en construction.

On se propose d'étudier si le contrat de maintenance forfaitaire annuel proposé par l'installateur, après la période de garantie d'un an, est plus avantageux que la facturation au prix réel des interventions ponctuelles.
Une étude statistique permet au constructeur d'affirmer que la probabilité de l'événement "la pompe à chaleur tombe en panne une fois pendant un mois donné" est $0,125$.

Dans un but de simplification, on admet que, pendant un mois donné, la pompe à chaleur ne peut tomber en panne qu'au plus une fois et que les pannes éventuelles survenues deux mois d'une même année sont indépendantes. On note X la variable aléatoire qui, à chaque année (de douze mois), associe le nombre de pannes survenues à la pompe.

1. Expliquer pourquoi X suit une loi binomiale. Donner les paramètres de cette loi.

2. Calculer la probabilité des événements suivants

   (a) il n'y a pas de panne dans l'année,

   (b) il y a au plus deux pannes dans l'année.

3. Calculer l'espérance mathématique, notée $E(X)$, de la variable aléatoire X. Que représente $E(X)$?

4. Les résultats d'une étude statistique menée auprès de nombreux utilisateurs de ce modèle de pompe à chaleur n'ayant souscrit de contrat de maintenance annuel permettent d'admettre que le coût d'une intervention est de 320 euros. Soit Y la variable aléatoire qui à chaque année associe le montant total en euros des frais de réparation de la pompe à chaleur.

   (a) Écrire une relation entre les variables Y et X.

   (b) Déterminer l'espérance mathématique, notée $E(Y)$, de la variable Y. Que représente $E(Y)$?

   (c) Le contrat de maintenance forfaitaire annuel de la pompe à chaleur est proposé par l'installateur au prix de 685 euros TTC.
   Quelle est la solution de maintenance la plus intéressante sur une longue période?

5. On approche la loi binomiale du 1. par une loi de Poisson de paramètre $\lambda = np$ où n et p sont les paramètres de cette loi binomiale.
   En utilisant la loi de Poisson, déterminer les probabilités respectives de deux événements du (a) et du (b) de la question 2.

6. On considère que, pour un événement, l'approximation d'une loi binomiale par une loi de Poisson est justifiée lorsque l'erreur relative $\dfrac{p - p'}{p}$ est, en valeur absolue, inférieure à 10% (p étant la probabilité de cet événement mesurée avec la loi de Poisson). Pour chacun des deux événements précédents, déterminer si l'approximation de la loi binomiale du 1. par la loi de Poisson du 5. est justifiée.

Exercice 2 Un laboratoire veut fabriquer des pilules se composant de deux substances A et B. Pour chaque pilule de la fabrication, on considère les masses a et b respectivement des 2 substances A et B qui la constituent. On désigne par X et Y respectivement les variables aléatoires qui associent à chaque pilule la masse a et la masse b des substances de cette pilule. On suppose que ces variables sont indépendantes et suivent des lois normales de moyennes respectives $m_X = 8,55$ mg et $m_Y = 5,20$ mg et de même écart-type $\sigma_X = \sigma_Y = 0,05$ mg.

1. Déterminer les probabilités $p(8,45 \leq X \leq 8,70)$ et $p(5,07 \leq Y \leq 5,33)$

2. Les normes imposées pour la fabrication sont les suivantes : $8,45 \leq a \leq 8,70$ et $5,07 \leq b \leq 5,33$.

   (a) Calculer le pourcentage de pilules qui seront hors normes à la sortie de la chaîne de fabrication.

   (b) En déduire que le procédé de fabrication ne peut être retenu si on veut que le pourcentage de pilules défectueuses ne dépasse pas 3%. On modifie alors la fabrication de la substance B. La moyenne de Y ne change pas mais son écart-type est modifié. Trouver la valeur minimum de ce nouvel écart-type pour que le pourcentage de pièces défecteuses soit inférieur à 3%.

3. (a) Déterminer la moyenne et l'écart-type de la variable aléatoire S qui associe à chaque pilule sa masse totale, les variables X et Y gardant leurs caractéristiques de la question 1.

   (b) On admet que S est encore une variable aléatoire normale dont les paramètres sont ceux calculés précédemment. Calculer $p(13,6 \leq S \leq 13,8)$.

4. On assure le conditionnement des pilules par boîtes de 100 unités. Une boîte est constituée à partir d'un tirage au hasard dans un stock assez grand pour qu'on puisse estimer que les tirages successifs se font avec remise. On désigne par Z la variable aléatoire qui, à chaque boîte associe le nombre de pilules hors normes au sens de la question 2.(a). On pourra prendre pour probabilité p d'une pilule hors-norme $p = 0,01$.

   (a) Dans ces conditions, montrer que Z est une variable binomiale dont on précisera les paramètres.

   (b) Dire pourquoi on peut approcher cette variable par une loi de Poisson. En utilisant cette loi, donner une valeur approximative de $p(Z \geq 5)$.

5. On désigne par U la variable aléatoire qui à chaque boîte associe le nombre de pilules dont la masse totale est supérieure à $13,8$. Là aussi, on peut supposer que U est une variable binomiale de paramètres n et p.

(a) Calculer p.

(b) Dire pourquoi on peut approcher U par une variable normale. À l'aide de cette approximation, donner une valeur approchée de $p(U \in \{70, 71, \ldots, 85\})$.

Exercice 3 (Les quatre questions de cet exercice sont indépendantes.)

Dans un groupe d'assurances, on s'intéresse aux sinistres susceptibles de survenir, une année donnée, aux véhicules de la flotte d'une importante entreprise de maintenance de chauffage collectif.
Dans cet exercice, sauf mention contraire, les résultats approchés sont à arrondir à 10^{-3}.

1. Étude du nombre de sinistres par véhicule.
   Soit X la variable aléatoire qui, à tout véhicule tiré au hasard dans un des parcs de la flotte, associe le nombre de sinistres survenant pendant l'année considérée. On admet que X suit la loi de Poisson de paramètre $0,28$.

   (a) Calculer la probabilité de l'évènement A : "un véhicule tiré au hasard dans le parc n'a aucun sinistre pendant l'année considérée".

   (b) Calculer la probabilité de l'évènement B : "un véhicule tiré au hasard dans le parc a, au plus, deux sinistres pendant l'année considérée".

2. Étude du nombre de sinistres dans une équipe de 15 conducteurs.
   On note E l'évènement : "un conducteur tiré au hasard dans l'ensemble des conducteurs de l'entreprise n'a pas de sinistre pendant l'année considérée". On suppose que la probabilité de l'évènement E est $0,6$. On tire au hasard 15 conducteurs dans l'effectif des conducteurs de l'entreprise. Cet effectif est assez important pour que l'on puisse assimiler ce prélèvement à un tirage avec remise de 15 conducteurs. On considère la variable aléatoire Y qui, à tout prélèvement de 15 conducteurs, associe le nombre de conducteurs n'ayant pas de sinistre pendant l'année considérée.

   (a) Justifier que la variable aléatoire Y suit une loi binomiale et déterminer ses paramètres.

   (b) Calculer la probabilité que, dans un tel prélèvement, 10 conducteurs n'aient pas de sinistre pendant l'année considérée.

3. Étude du coût des sinistres.
   Dans ce qui suit, on s'intéresse au coût d'une certaine catégorie de sinistres survenus dans l'entreprise pendant l'année considérée. On considère la variable aléatoire C qui, à chaque sinistre tiré au hasard parmi les sinistres de cette catégorie, associe son coût en euros. On suppose que C suit la loi normale de moyenne 1200 et d'écart type 200. Calculer la probabilité qu'un sinistre tiré au hasard parmi les sinistres de ce type coûte entre 1000 euros et 1500 euros.

Exercice 4 La société PALOI envisage de réaliser une nouvelle construction. Cet ouvrage étant réalisé pour la première fois, l'entreprise ne dispose pas de données statistiques suffisantes sur la durée des tâches, et pourtant il serait important avant de signer le contrat indiquant la durée de réalisation et les pénalités éventuelles, de pouvoir mesurer le risque. Vous êtes chargés de résoudre ce problème.
À chaque responsable vous posez ces trois questions sur la durée des tâches :

- À combien estimez-vous la durée minimale de l'opération (estimation optimiste O_i)?
- À combien estimez-vous la durée maximale de l'opération (estimation pessimiste p_i)?
- À combien estimez-vous la durée la plus probable de l'opération (estimation normale n_i)?

Puis vous calculez pour chaque tâche T_i :

- l'espérance de la durée $E(T_i) = \dfrac{O_i + 4n_i + p_i}{6}$,
- la variance de la durée $V(T_i) = \dfrac{(p_i - O_i)^2}{36}$.

Au bout d'un certain nombre d'opérations on peut considérer que les variables aléatoires, durées des tâches sur le chemin critique, sont des variables aléatoires normales indépendantes.

Tâche	Durée normale	Durée la plus pessimiste	Durée la plus optimiste	Tâches antérieures
A	8	11	5	–
B	12	14	10	A
C	9	16	8	A
D	16	20	12	A,B,C
E	8	10	6	A

1. Établir le chemin critique en prenant pour durée de la tâche i, $E(T_i)$.
2. (a) Déterminer l'espérance de la variable aléatoire X durée du chemin critique, ainsi que la variance et son écart-type.
   (b) Calculer la probabilité de réaliser l'ouvrage en moins de 31 jours, en moins de 36 jours, en moins de 40 jours.
3. Le chef d'entreprise doit signer un contrat avec une durée maximale de 33 jours. Quel risque accepte t-il?
4. Lorsqu'il est réalisé dans un délai de 33 jours, le travail laisse un bénéfice de 200000 euros. Sachant que tout dépassement de délai implique le versement d'une pénalité journalière de 40000 euros, l'entrepreneur prend un risque financier en signant le contrat avec un délai de 33 jours.
   Apprécier le risque qu'il perde de l'argent sur ce chantier.

Exercice 5 Le département de contrôle de qualité d'une compagnie automobile examine l'efficacité de carburant d'un certain modèle de voiture. Les consommations d'essence (en litres) pour 12 voitures sur 100 kilomètres sont :

14,60 - 11,21 - 11,56 - 11,37 - 13,68 - 15,07
11,06 - 16,58 - 13,37 - 15,98 - 12,07 - 13,22

1. Supposons que les observations forment un échantillon simple de variables $\mathcal{N}(\mu, \sigma^2)$. Donner une estimation et un intervalle de confiance à 95% pour μ.
2. Expliquer comment votre intervalle trouvé dans 1. changerait si on avait $\sigma^2 = 3,72$ connue.

Exercice 6 Le second tour d'une élection présidentielle oppose le candidat A au candidat B. Pour évaluer la proportion p d'électeurs de la population souhaitant voter pour le candidat A plutôt que pour le candidat B, on tire au sort un échantillon de N individus dans une population de grande taille et on demande à chacun des individus pour lequel des deux candidats il a l'intention de voter. On associe à chaque individu sondé une variable aléatoire X_i pour $i = 1, \ldots, N$ telle que :

$$X_i = \begin{cases} 1 & \text{si l'individu } i \text{ a l'intention de voter pour A} \\ 0 & \text{sinon} \end{cases}$$

1. Quelle est la loi suivie par les variables aléatoires $X_1, X_2, \ldots, X_N$?
2. On considère l'estimateur (MV) de la probabilité de vote pour le candidat A défini par $\hat{p} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} X_i$. Montrer que l'estimateur $\hat{p}$ est sans biais et convergent.
3. Montrer que l'estimateur $\hat{p}$ est efficace au sens de la borne FDCR (Frechet-Darnois-Cramer-Rao).
4. Quelle est la loi asymptotique de l'estimateur $\hat{p}$?
5. On a compté 98 électeurs (parmi les 200 interrogés) déclarant voter pour le candidat A. Proposer une estimation ponctuelle du paramètre p.
6. Calculer un intervalle de confiance de niveau 95%, puis de niveau 98% du paramètre p en indiquant comment on construit l'intervalle. Comparer les deux intervalles obtenus.
7. On suppose que la vraie probabilité p que le candidat soit élu est égale à 52%. À partir de quelle taille d'échantillon les instituts de sondage donneraient gagnant le candidat A avec une probabilité de 95%?
   Remarque : on prévoit que le candidat A sera élu si la fréquence empirique $\hat{p}$ dans l'échantillon excède 50%.

Exercice 7 Dans un certain pays, deux études ont été réalisées pour évaluer le taux d'analphabétisme. Dans les deux cas, un échantillon a été prélevé au hasard, avec replacement, dans la population. Dans le premier échantillon, d'effectif 500, on dénombre 21 analphabètes. Dans le second échantillon, d'effectif 1000, on dénombre 57 analphabètes.

1. Calculer les proportions d'analphabètes observées dans les deux échantillons.
2. En utilisant l'approximation normale de la distribution d'échantillonnage exacte de la proportion d'analphabètes au sein d'un échantillon, déterminer la demi-largeur des intervalles de confiance pour le taux d'analphabétisme au sein de la population fournis par chacun des deux échantillons, aux niveaux de confiance de 95% et 99%.
3. Comparer les résultats précédents. Comment varie la largeur en fonction de la taille de l'échantillon, en fonction du niveau de confiance?
4. Représenter graphiquement sur un même axe les deux intervalles de confiance au niveau de confiance de 95%. Commenter les graphiques obtenus.
5. On décide de tenir compte des données des deux échantillons pour déterminer un nouvel intervalle de confiance (on fusionne les deux échantillons pour en obtenir un seul d'effectif 1500). Calculer la proportion d'analphabètes observée dans le nouvel échantillon.
6. En utilisant l'approximation normale de la distribution d'échantillonnage exacte de la proportion d'analphabètes au sein d'un échantillon, déterminer la demi-largeur des intervalles de confiance pour le taux d'analphabétisme au sein de la population fourni par le nouvel échantillon, aux niveaux de confiance de 95% et 99%.
7. Quel est l'avantage de fusionner les deux échantillons ?
8. Quelle taille minimale doit avoir un échantillon pour obtenir une demi-largeur inférieure ou égale à $0,5\%$ au niveau de confiance de 95% (en prenant la proportion observée dans l'échantillon fusionné comme estimation préliminaire du taux d'analphabétisme au sein de la population)? Arrondir le résultat à la centaine supérieure.
9. Quelle taille minimale doit avoir un échantillon pour obtenir une demi-largeur inférieure ou égale à $0,5\%$ au niveau de confiance de 99% (en prenant la proportion observée dans l'échantillon fusionné comme estimation préliminaire du taux d'analphabétisme au sein de la population)? Arrondir le résultat à la centaine supérieure.

Exercice 8 On veut estimer par intervalle de confiance la vitesse moyenne des automobiles dans un certain virage d'une route à grand trafic. Pour cela on a enregistré à l'aide d'un radar les vitesses $X_1(\omega) = x_1, \ldots, X_{400}(\omega) = x_{400}$ de 400 automobiles en une période de temps de 2 heures avec des conditions de circulation homogènes (météo, visibilité, densité de trafic,...). On a obtenu les statistiques suivantes :

$$\sum_{i=1}^{400} x_i = 35200 \text{ km/h}, \sum_{i=1}^{400} x_i^2 = 3107600 \text{ (km/h)}^2.$$

L'homogénéité des conditions de trafic permet de supposer que les variables aléatoires $X_1, \ldots,$
X_{400} dont on a ainsi observé une réalisation sont indépendantes et de même loi. Proposer un intervalle de confiance au niveau 98% pour la vitesse moyenne $E(X)$ en indiquant clairement quels résultats du cours légitiment les approximations faites. Les données numériques ci-dessus ont été "arrangées" pour permettre de faire facilement tous les calculs à la main si on ne dispose pas d'une calculatrice.