Eléments de statistique

Examen écrit du 11 janvier 2016 - Théorie

Consignes

1. Répondez aux deux questions sur des feuilles différentes, rendez au moins une feuille pour chaque question, et indiquez votre nom, prénom et matricule sur chaque feuille.
2. **Ni notes de cours ou personnelles, ni GSM, ni calculatrice ne peuvent être utilisés !**

Question 1 : estimation

1.a Construction d'un intervalle de confiance pour un paramètre θ à partir d'un échantillon i.i.d. de taille n et d'une statistique $\mathcal{T}$ estimant la valeur de θ :
  - expliquer la notion d'intervalle de probabilité $1 - \alpha$ pour la statistique $\mathcal{T}$, puis à l'aide d'un graphique et de formules montrez comment on en déduit l'intervalle de confiance pour le paramètre de population θ pour un niveau de risque α donné.
  - Donner l'interprétation de la notion d'intervalle de confiance construit pour un niveau α, et expliquer l'influence de la valeur de n et de α sur sa largeur.

1.b Supposons que la variable parente soit gaussienne : construire, en justifiant les différentes étapes, l'intervalle de confiance pour la moyenne μ à partir de la statistique m_x obtenue d'un échantillon i.i.d. de taille n, respectivement dans le cas où la variance σ^2 de la variable parente est connue, et dans le cas où cette variance est inconnue.

Question de Théorie 2 : tests d'hypothèses

2.a Méthode de Neyman-Pearson
  - Poser le problème adressé et énoncer le théorème de Neyman-Pearson.
  - Définir la notion de statistique exhaustive, en donner en une interprétation intuitive, et en expliquer l'intérêt pour la construction de tests d'hypothèses et pour les problèmes d'estimation.
  - Dans le cas où la variable parente suite une loi Gaussienne de variance connue, et que les deux hypothèses correspondent donc respectivement aux valeurs μ_0 et μ_1 du paramètre μ, indiquez la statistique exhaustive naturelle et expliquez à quoi finalement le test d'hypothèse se réduit (supposez que $\mu_0 < \mu_1$).
  - Montrez sur un graphique la région de rejet de ce test d'hypothèses, en mettant en évidence α ; expliquez comment la puissance du test dépend de la taille de l'effet étudié.

2.b Expliquer la nature du paradoxe de Simpson, sur base d'un exemple numérique concret. Expliquer ensuite la notion de "facteur de confusion" en utilisant un graphe de dépendances. Quels enseignements peut-on tirer de cette analyse d'un point de vue pratique ? Comment peut-on en principe se prémunir de l'existence d'un facteur de confusion ?