

MAT-8185

Techniques avancées de programmation statistique: SAS

Devoir 2 et 3 — jeudi 4 mai 2017

[ 8 mai 2017 ]

### Instructions

- Le présent devoir compte pour 15% + 15% de la note globale.
- Le travail doit être remis par courriel à [larribe.fabrice@uqam.ca](mailto:larribe.fabrice@uqam.ca) au plus tard à 9h00 le mercredi 10 mai.
- Le code SAS doit être propre (formaté et commenté).
- Vous devez envoyer le code SAS, le contenu du journal, et une sortie de la procédure Print pour chaque data lu (seulement les 10 premières observations pour chaque base de données), dans un format autre que Microsoft Word (pdf par exemple).
- Vous devez mettre des titres pour que la sortie soit explicite.

### Question 1

Vous allez faire une estimation de l'erreur standard de la moyenne par la méthode bootstrap. À cette fin, suivez ce cheminement:

- a. Générez dans un Data nommé [Echantillon](#) de 250 observations d'une loi  $N(100, 25)$  dans une variable nommée X;
- b. Dans un unique data, générez 1000 échantillons aléatoires avec remise de taille 250 de la variable X.
- c. Vous utiliserez ensuite Proc Means pour calculer la moyenne de chacun de ces 1000 échantillons, et pour calculer ensuite l'erreur standard de la moyenne.
- d. Vérifiez que l'erreur standard trouvée est proche de la vraie.

Ressource: [Bootstrap](#)

## Question 2

Vous allez faire une estimation de l'erreur standard de la moyenne par la méthode du Jackknife. À cette fin, suivez ce cheminement:

- a. Générez dans un Data nommé [Echantillon](#) 250 observations d'une loi  $N(100, 25)$  dans une variable X;
- b. Dans un unique data, générez 250 échantillons de taille  $n - 1$  de la variable X, tel que l'échantillon  $i$  corresponde à l'échantillon original sans l'observation  $i$ .
- c. Vous utiliserez ensuite Proc Means pour calculer la moyenne de chacun de ces 100 échantillons, et pour calculer ensuite l'écart-type de la moyenne.
- d. Vérifiez que l'erreur standard trouvée est proche de la vraie.

Ressource: [Jackknife](#)

### Question 3

Calcul de risque de première espèce et de puissance dans le contexte d'un test  $t$ . Vous **devez** suivre les étapes suivantes:

- Générez 1000 échantillons de taille  $n$  d'une  $N(0, 1)$ ,  $n = 5, 10, 15, \dots, 500$  dans un unique **Data** (le data aura donc  $1000 \cdot (5 + 10 + 15 + \dots) = 252\,500\,000$  observations), dans une unique variable **X**.
- Utilisez un **Proc Means** pour calculer la **p-value** d'un test  $t$  de l'hypothèse  $H_1 : \mu \neq 0$  pour chacun des 1 010 101 échantillons. De la procédure **Means**, obtenez les p-values dans un nouveau Data, qui contiendra donc 1 010 101 observations.
- Calculez pour chacune des tailles d'échantillons (il y en a 101) la proportion des erreurs de première espèces faites.
- Représentez dans un graphique en abscisse la taille des échantillons, et en ordonnée la proportion des erreurs de première espèce.

Puissance.

- Utilisez la même stratégie, mais en générant cette fois 1000 échantillons de taille variable ( $n = 5, 10, \dots, 500$ ) d'une  $N(0.2, 0)$ , et représentez la puissance en fonction de la taille d'échantillon.

*∴ fin ∴*