

Macro complémentaire de rééchantillonnage

Olivier MARTIN omartin@inapg.inra.fr

La macro Rééchantillonnage permet de mettre en oeuvre les techniques Jackknife et Bootstrap en procédant automatiquement au rééchantillonnage et aux calculs des estimations. Une option permet également de générer, dans le cas du Bootstrap, les histogrammes de distribution empirique des estimations obtenues sur les échantillons artificiels.

Utilisation de la macro Rééchantillonnage *

A l'ouverture du fichier **Rééchantillonnage.xls** une fenêtre d'accueil rappelle la procédure d'exécution de la macro et propose l'ouverture d'une feuille de calcul vierge ou d'une feuille contenant l'un des 4 exemples suivants:

1. Régression linéaire
2. Coefficient de corrélation
3. Rapport
4. Comparaison de moyenne

Le bouton OK exécute l'option choisie.

Les exemples illustrent la mise en forme des données à respecter pour faciliter l'exécution de la procédure.

La feuille de calcul ouverte contient 2 boutons :

- MENU : permet de revenir à la boîte de dialogue initiale.
- EXECUTER : permet de lancer la macro (ou Ctrl+maj+r)

A l'exécution, s'affiche une interface permettant de saisir la plage de données contenant l'échantillon (avec ou sans intitulé en première ligne) et la plage de formule contenant les statistiques calculées sur l'échantillon (avec ou sans intitulé en première colonne).

La méthode sélectionnée par défaut est le Jackknife.

La sélection de la méthode Bootstrap implique le choix du nombre d'échantillons artificiels à générer (200 par défaut) et offre l'option de construction d'un histogramme de distribution empirique par paramètre statistique traité.

Le bouton OK lance la procédure.

Le bouton ANNULER quitte la procédure.

Les plages de données et de formules sont des plages uniques de cellules (les plages multiples ne sont pas valides).

La plage de formule doit être sous forme d'une colonne de formules calculées à partir de l'échantillon avec, éventuellement, la colonne des intitulés à gauche.

Des contrôles interviennent préalablement à l'exécution de la procédure et informent l'utilisateur sur les corrections à envisager.

Enfin une procédure de détection comptabilise les données non numériques et les données manquantes pouvant apparaître dans l'échantillon.

Les résultats sont exhibés dans un nouveau classeur Excel comprenant une feuille Echantillon contenant une copie de la feuille initiale, une feuille Rééchantillonnage Bootstrap ou Jackknife contenant les résultats numériques et de façon optionnelle, une feuille graphique par paramètre statistique rééchantillonné (chacune de ces feuilles porte le nom du paramètre statistique dont elle contient l'histogramme des fréquences).

Sur ces graphiques apparaissent les seuils des 2,5ème et 97,5ème percentiles de la distribution.

Paramètres

La feuille Rééchantillonnage comprend pour chaque paramètre :

- dans le cas du **Bootstrap** :
 - * l'estimation classique,
 - * la série des estimations obtenues sur chaque échantillon artificiel généré,
 - * les estimations Bootstrap de moyenne et d'écart type, ainsi que les bornes des intervalles de confiance 95% pour les méthodes 'Standard' et 'Percentile'.
- dans le cas du **Jackknife** :
 - * l'estimation classique,
 - * la série des estimations obtenues sur chaque échantillon artificiel généré,
 - * les estimations Jackknife de moyenne et écart type, ainsi que les bornes de l'intervalle de confiance 95% (sous conjecture de Tukey).