

## STT 7340 Sondages : modèles et techniques (4CR) 3-0-0-6

### BUT GENERAL DU COURS

Initier les étudiants à des méthodes d'analyse statistique pour les données récoltées dans un plan d'échantillonnage complexe, semblable à celles disponibles dans des centres de données de recherche ou à l'Institut de la Statistique du Québec. Modifier une analyse statistique pour tenir compte que les données ont été récoltées avec un plan de sondage complexe.

### OBJECTIFS SPECIFIQUES

À la fin du cours, l'étudiant devra être capable de :

- calibrer les poids d'une enquête à partir de données administratives
- utiliser l'échantillonnage à deux phases pour traiter non-réponse;
- appliquer les méthodes de linéarisation, du bootstrap et du jackknife pour estimer les variances de statistiques non-linéaires;
- utiliser des équations d'estimation pour estimer les paramètres d'un modèle statistique à partir de données d'enquête;
- analyser des tableaux de fréquences construits à l'aide d'unités recueillies dans un plan d'échantillonnage complexe;
- ajuster des modèles de régression et de régression logistique à des données recueillies à l'aide d'un plan complexe;
- comprendre la construction et l'utilisation des poids bootstrap pour estimer les variances dans des plans complexes;
- comprendre la documentation des enquêtes répertoriées dans Equinox, un portail réalisé pour favoriser l'accès aux données de recherche, voir

<http://equinox.uwo.ca.ezproxy.bibl.ulaval.ca/ViewEntryScreen.asp>

CONTENU (les 4 premiers points couvrent en partie les chapitres 7 à 11 du livre de S. Lohr)

#### 1-Définition d'un plan de sondage complexe :

- Révision des notions d'échantillonnage à plusieurs degrés avec des probabilités de sélection variables ;
- Formules de variance générales pour des probabilités de sélection quelconques;
- Calage des poids d'échantillonnage et incorporation de l'information auxiliaire ;
- Modification aux poids pour le traitement de la non-réponse totale : échantillonnage à deux phases ;
- Effet de plan ;

#### 2-Estimation de la variance de statistiques non-linéaires dans des plans complexes

- Méthode de linéarisation ;
- Méthode des groupes aléatoires ;
- Méthode du jackknife ;
- Méthode du bootstrap.

3-Ajustement de modèles statistiques à des échantillons récoltés à l'aide d'un plan de sondage complexe en utilisant une pseudo-vraisemblance

4-Analyse de tableaux de fréquences construits à partir de données récoltées dans des plans complexes

- Correction de Rao-Scott de la statistique du chi-deux pour tester l'indépendance ;
- Modèles log-linéaires

5-Modèle de régression pour des données recueillies avec un plan complexe

- Faut-il utiliser les poids ?
- Modèle mixte et échantillonnage par grappes
- Régression logistique

#### REFERENCES

Livre suggéré: *Sampling: Design and Analysis* par Sharon S. Lohr publié par Duxbury Press.

Autres références:

- *Notes de Cours*
- *Analysis of Survey Data* édité par R. L. Chambers et C.J. Skinner publié par Wiley en 2003
- Articles de la littérature statistique distribués durant le cours

#### ÉVALUATION

Devoir (3-4 durant la session) = 40%, Examen (mercredi 26 octobre & 21 décembre)= 60%.

#### HORAIRE

lundi 15h30-17h30, mercredi 8h30-10h30, Salle 1610 (Vachon) et 2540 (Pouliot)  
Août 2008