

Modélisation statistique

Semaine 1

Dr. Léo Belzile
HEC Montréal

Bienvenue



Léo Belzile

- + Professor adjoint de Sciences de la décision
- + Doctorat de l'EPFL (Suisse)
- + Je suis spécialisé dans l'analyse des valeurs extrêmes (inondations, vagues de chaleur, etc.)
- + Je suis originaire de la Gaspésie
- + J'aime faire de la randonnée dans les Alpes
- + Attentes pour le cours: participation active en classe et rétroaction!

Et vous? Faisons connaissance!

- + Origine
- + Formation précédente et parcours académique
- + Ville/région/pays actuel
- + Quelles sont vos attentes par rapport au cours?

Sessions de groupe Zoom

Pour compléter des activités en petits groupes

- + Plus d'options comme participants (vous pouvez partager votre écran).
- + Si besoin, demandez de l'aide (je reçois une notification).
- + Je peux vous écrire (fenêtre contextuelle)
- + Vous perdez temporairement accès à l'historique de clavardage de la réunion.
- + Ne quittez pas la salle sans que je vous ai encouragé à le faire svp!
- + Votre microphone est fermé par défaut quand vous revenez.
- + Lorsque j'annonce la fin des sessions de groupe, il y a une latence d'une minute (quittez pour revenir immédiatement).

Pratique des sessions de groupe dans Zoom

Activité brise-glace: présentez-vous à vos camarades de classe en donnant cinq faits à propos de vous, par exemple.

- + Je vis en Tunisie
- + Je grimpe en salle
- + Je possède un labrador appelée Marguerite
- + J'ai travaillé dans une vie précédente comme photographe
- + J'ai une formation d'ingénieur

Récapitulatif de la semaine 1

- + Procédures de test
- + Théorème central limite
- + Graphiques
- + Analyse exploratoire des données

Séance de réponses aux questions

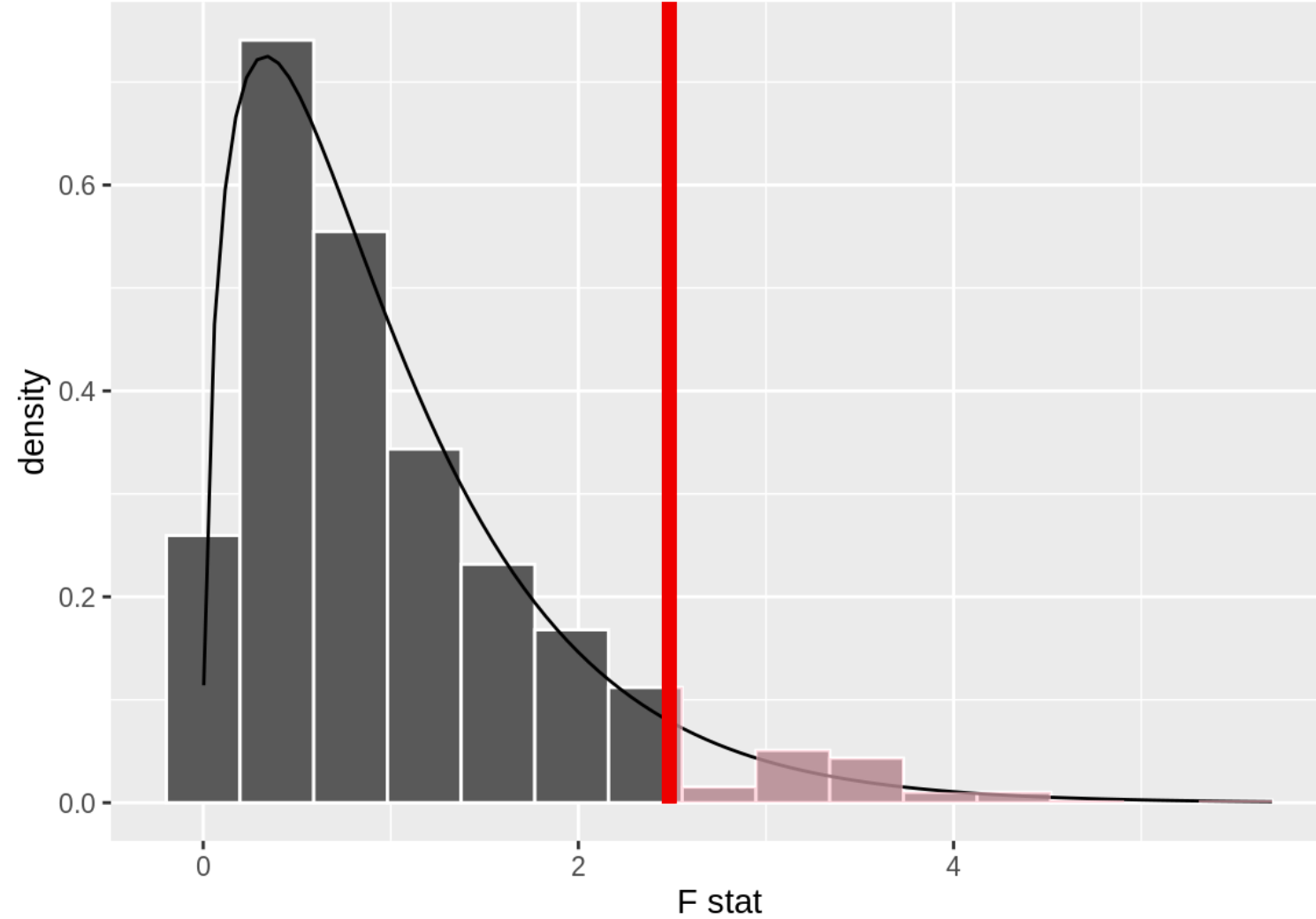
- + Levez la main pour poser une question.
 - + Quand je vous nomme, ouvrez votre micro.
- + ou posez directement vos questions dans le clavardage.

Tests

L'inférence statistique consiste à tirer des conclusions valables sur une population à partir d'un échantillon représentatif de cette dernière.

- + Choisir le modèle (souvent implicite)
- + Écrire nos hypothèses
- + Choisir et calculer la statistique de test
- + Et si \mathcal{H}_0 ? Déterminer ce qu'on aurait pu avoir dans un monde alternatif comme valeur de statistique de test
 - + à l'aide d'arguments asymptotiques (théorème central limite)
 - + distribution exacte (tests nonparamétriques avec petites taille d'échantillons)
 - + à l'aide de simulations du modèle nul (tests de permutation et autoamorçage)

Simulation-Based and Theoretical F Null Distributions



Souvent, on dérive la loi nulle en faisant des hypothèse sur la loi des données.

Graphiques

Une variable:

- + continue: -
- + catégorielle: -

Two variables

- + continue²: -
- + continue × catégorielle: -
- + catégorielle²: -

Amélioration des graphiques



Sondage Wooclap

- + Allez à l'adresse: <https://www.wooclap.com/MODSTATS1> pour démarrer le sondage

Démonstration de SAS onDemand

Atelier

Télécharger le fichier **.zip** qui contient 30 graphiques

- + Quelle histoire le graphique raconte
- + Quelles variables et coordonnées?
- + Types de variables et choix de géométrie
- + Est-ce que le graphique est complet?
- + Soulignez les bons points et les points à améliorer

Devoir 1

Je vous demanderai de choisir deux graphiques en ligne et répéter cet exercice (un mauvais et un bon)

Sources possibles pour trouver de bons graphiques

- + fil Twitter de:
 - + BBC [[@BBCNewsGraphics](#)]
 - + Washington Post [[@PostGraphics](#)]
 - + New York Times [[@nytgraphics](#)]
 - + etc.
- + journaux et magazines
- + gouvernements, par exemple [US Census](#)

| Discussion

Faites une analyse exploratoire des données **assurance** en petits groupes

Questions

–

Résumé

–