

# Statistical modelling

Week 1

**Dr. Léo Belzile**  
**HEC Montréal**

# Welcome



**Léo Belzile**

- + Assistant professor of Decision Sciences
- + PhD from EPFL (Switzerland)
- + I study extreme value analysis (floods, heatwaves, etc.)
- + I come from the Gaspé Peninsula
- + I enjoy hiking in the Alps
- + Expectations: active learning, participation in class, feedback!

## Let's meet

- + Where are you from?
- + Academic background?
- + What are your expectations for this course?

# Zoom breakout rooms

A breakout room is used for sessions in small groups.

- + Different settings (you can share content/screen).
- + You can ask me for help; I will get a notification.
- + I can send everyone messages (pop-up)
- + You can't access the history of the chat from the main meeting while in the breakout rooms.
- + Please don't leave the breakout room until I instruct you to!
- + Your microphone will be muted when you come back.
- + When I end the breakout sessions, there is a one minute delay before going back (none if you leave them immediately).

# Zoom practice

We will practice Breakout rooms, which are small group sessions.

**Icebreaking activity:** introduce yourself to classmates and give five facts about yourself, e.g.,

- + I live in Tunisia
- + I am a mountain climber
- + I own a labrador named Daisy
- + I used to work as a photographer
- + I have an engineering background

# Recap of week 1

- + Testing procedure
- + Central limit theorem
- + Graphics
- + Exploratory data analysis

## Question and answer (Q&A) period

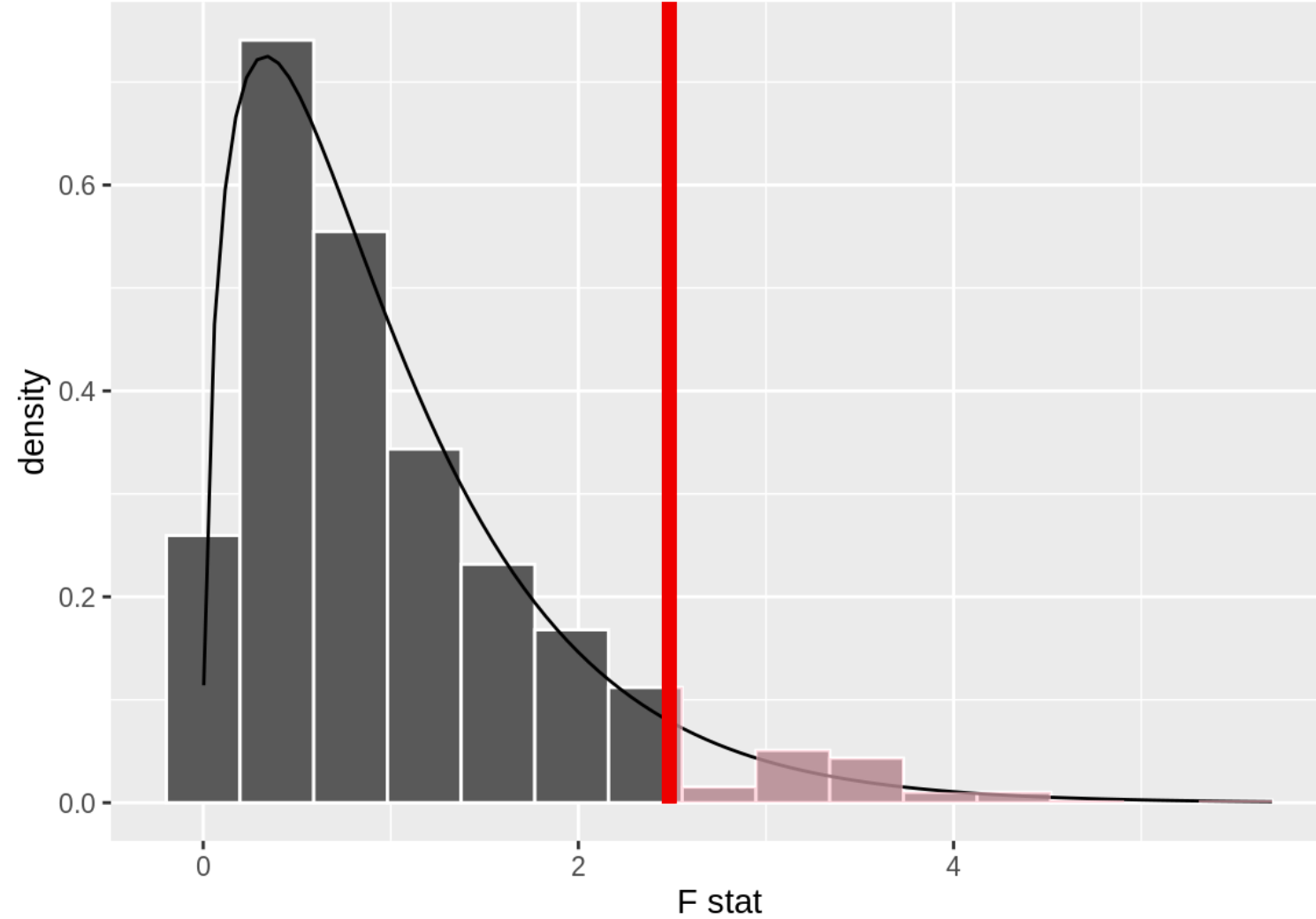
- + Raise your hand to ask a question.
  - + When asked, unmute yourself
- + or post question on the chat.

# Testing

Statistical inference is about drawing conclusion for population based on a representative sample.

- + Postulate a model (often implicitly)
- + Fix hypotheses
- + Choose and compute the test statistic
- + What if  $\mathcal{H}_0$ ? Figure out expected behaviour of test
  - + using asymptotic arguments (central limit theorem)
  - + exact distribution (nonparametric tests, small samples)
  - + using simulations from the null model (e.g., permutation, bootstrap)

## Simulation-Based and Theoretical F Null Distributions



*Often, the null hypothesis is derived by making assumptions about the underlying distribution of the data.*

# Graphics

Single variable:

- + continuous: -
- + categorical: -

Two variables:

- + continuous <sup>2</sup>: -
- + continuous  $\times$  categorical: -
- + categorical <sup>2</sup>: -



# Improving graphs



# Wooclap poll

- + Go to <https://www.wooclap.com/STATMOD1> to begin the poll

## **SAS onDemand demo**

# Workshop

Download the `.zip` folder containing 30 graphs.

- + Does the graph tell a story (i.e., is it standalone?)
- + What are the variables and the mapping?
- + Variable type: is the choice of geometry adequate?
- + Is the graph complete?
- + Highlight good features and improvement points

# Assignment 1

Pick a graph and repeat this commentary exercise

Good sources include

- + the Twitter handles:
  - + BBC [[@BBCNewsGraphics](#)]
  - + Washington Post [[@PostGraphics](#)]
  - + New York Times [[@nytgraphics](#)]
  - + etc.
- + newspaper and magazine
- + data providers, e.g., [US Census](#)

# |Discussion

Perform and exploratory data analysis of the `insurance` dataset in small groups.

## Questions

-

## Summary

-