1. Fundamentals

GHV Chapters 4-5

DATA 335 – Univerrsity of Calgary – Winter 2025

Statistical models and statistical inference

- A *statistical model* is a probability distribution.
- A statistical model is characterized by unknown and often unknowable numbers called *parameters*. They are our quantities of interest.
- Statistical models facilitate *statistical inference* – procedures for turning data into parameters estimates, avatars for their uncertainty.
  - Frequentist inference: point estimation, standard errors, confidence intervals, hypothesis tests
  - Bayesian inference: posterior distribution

Estimators for mean and variance

- Let $x_0, \ldots, x_{n-1}$ be a *random sample*[1] from the a model (distribution) F with mean μ and variance σ^2.
- The *sample mean*

$$\bar{x} = \frac{x_0 + \cdots + x_{n-1}}{n}$$

  estimates μ.
- The *sample variance*

$$s^2 = \frac{1}{n-1} \sum_{i<n} (x_i - \bar{x})$$

  estimates σ^2.

[1] independent and identically distributed

Estimators have distributions

- Since the x_i are random variables, the estimators $\bar{x}$ and s^2 are computed computed from them are, too.
- In partifcular, they have distributions.
- Distributions of random variables computed from random samples from other distributions are called *sampling distributions*.
- **(Demo)** Visualize sampling distributions with histograms