LECTURE 17: MIDTERM GUIDANCE

OCT 23 2025

WHAT THE MIDTERM IS **NOT**

- It is NOT a test of your
    - Speed of question answering or problem solving
    - Proficieny in complex computations
    - Skills in implementation/coding (R)

WHAT WE **ARE** ASSESSING

- Understanding of the fundamentals of statistical methods
  - The key concepts, such as Distributions, Hypothesis Testing, and Simple Linear Regression ….
- Your ability to provide solutions to real problems
  - Such as estimating key statistics, applying hypothesis testing, fitting sample data with a curve (with linear regression)

SYLLABUS: FOR REVISION

- **Skip Lectures 1-4** (basic of central tendencies, dispersion, probability)
  - That was essentially revision before starting on statistical methods for data science
- **Lectures 5-13 cover everything for this Midterm**
- **Topics**
  - Distributions
    - The CDF and the PDF
    - Kernel methods
    - Bootstrapping
  - Hypothesis Testing
    - Fundamentals, Error types
    - Types of tests: Single population/Two populations (Univariate/Bivariate), Paired/ Independent
    - Problem solving
      - We will provide lookup tables (subset) for t-stat and Z-stat (if required)
  - Simple Linear Regression (only upto Lecture 13)

FORMULAE: EXPECT YOU TO KNOW

- Standard Deviation
- Standard Error (SE)
- T-stat and Z-stat
    - In terms of your statistic of interest, and the SE
- Confidence interval
- Simple Linear Regression: the basic equation for the model

FORMULAE THAT WE ARE **NOT** ASSUMING YOU HAVE IN READY MEMORY, OR KNOW

- Integrals etc, for distributions
- Formulae for various kernels
  - Those will be supplied if required
- Other statistics such as SSR, SSE, TSS etc that we only briefly covered for the coefficient of determination R^2
  - Though keep in mind the key takeaway was:

$$R^2 = \text{Corr}\left(y, \hat{y}\right)^2 = \frac{\text{Var}\left(\hat{y}\right)}{\text{Var}\left(y\right)} = \frac{\text{Cov}\left(y, \hat{y}\right)^2}{\text{Var}\left(y\right)\text{Var}\left(\hat{y}\right)}$$

- Slightly more involved formulae, for instance the formula for the slope in simple linear regression β_1
- For standard deviation across two populations with different variances (Welch test)

FORMAT

- ONE hour
- Questions
  - 10 short questions (multiple choice)
  - 5 problem solving questions