Final Exam Topics
PSTAT 5A: Spring 2023, with Ethan P. Marzban

> **❗ IMPORTANT NOTE**
>
> This list is not meant to be comprehensive: just because a topic/concept does not appear below does not mean it isn't fair game for the final exam Conversely, just because something *does* show up below doesn't mean it is guaranteed to appear on the exam. Your best resource for reviewing are the notes you have (hopefully) been creating each lecture.

Descriptive Statistics (Chapter 2 of *OpenIntro Statistics*)

- Structure of data and data representation

    - Data matrix / observational units / variables
    - Data as a *set* / data aggregation

- Classification of variables

- Appropriate visualizations for numerical and categorical data

    - Barplots
    - Histograms
    - Boxplots

- Appropriate visualizations of the relationship between variables

    - Numerical vs. numerical (scatterplot)
    - Numerical vs. categorical (side-by-side boxplot)
    - Categorical vs. categorical (contingency table)

- Numerical summaries of data

    - Measures of central tendency (mean, median)
    - Measures of spread (range, variance, standard deviation, IQR)
    - Five-number summary

- Transformations (HW01)

Probability (Chapter 3 of *OpenIntro Statistics*)

- Basics of probability
    - Experiment
    - Outcomes / outcome space / different representations of outcome spaces
    - Events
    - Probability as a function
    - Two approaches to probability (long-run frequency and classical)

- – Equally likely outcomes

- Set operations
  - – Union / intersection / complement
  - – DeMorgan's Laws
  - – Venn Diagrams

- Probability rules
  - – Probability of the empty set
  - – Complement Rule
  - – Addition Rule
  - – Axioms of probability

- Counting
  - – Fundamental principle of counting
  - – Slot diagrams
  - – Sampling with/without order
  - – $n!, (n)_k , \binom{n}{k}$

- Conditional probability
  - – Definition of $\mathbb{P}(E \mid F)$
  - – Independence of events (HW03)

- Random Variables
  - – State space
  - – Expected value
  - – Variance/Standard Deviation
  - – Discrete vs. continuous random variables
    - * Probability Mass Function (PMF)
    - * Probability Density Function (PDF)
    - * Cumulative Distribution Function (CDF)
  - – Binomial Distribution
  - – Uniform Distribution
  - – Normal Distribution
    - * Standardization

Inferential Statistics (Chapters 5, 6, and 7 of *OpenIntro*)

- Population; population parameter
- Sample statistic; point estimator
- Sampling distribution of a point estimator
- Central Limit Theorem for Proportions
  - – Success-Failure Conditions
  - – Substitution Approximation
- Sampling distribution of $\overline{X}$

- Central Limit Theorem for the Sample Mean
- $t-$distribution with k degrees of freedom
- Confidence Intervals
  - Confidence Level
  - Relationship between the width of a confidence interval and the confidence level
  - Percentiles; reading the $z-$ and $t-$tables
- Hypothesis Testing
  - Null and Alternative Hypotheses
  - Type I and Type II errors
  - Level of significance α
  - Two-sided, lower-tailed, and upper-tailed tests
  - Test statistic
  - Critical value
  - p-value
- Testing across multiple populations
  - Two-sample t-test
  - Analysis of Variance (ANOVA)

Regression and Correlation (Chapter 8 of *OpenIntro*)

- Pearson's correlation coefficient
- Ordinary Least Squares Regression
  - Response vs. explanatory variables
  - OLS regression line
  - Residual
  - Fitted Values
  - Residual Sum of Squares (RSS)
  - Inference on the slope of the true linear fit

Sampling Techniques/Experimental Design (1.3 of *OpenIntro* and More)

- Types of Sampling

  - Simple Random Sample
  - Culster Sampling
  - Convenience Sampling

- Longitudinal vs. Cross-Sectional studies

- Observational study vs. experiment

- Treatment vs. control group

Programming (All Labs)

- General terminology

    - Code cells
    - Running / executing code
    - Expressions
    - Order of operations
    - Variable assignment/re-assignment
    - Modules
        * Importing modules with a nickname
    - Different Types of Error

- Data Types

- Data Classes (lists, arrays, tables)

- Comparisons

- Conditional expressions

- Functions

    - Docstring
    - Return statement
    - Default values

- Loops / iteration

- Data Visualization

- Know the modules we have been using extensively (e.g. `datascience`, `numpy`, `scipy.stats`, `matplotlib.pyplot`, etc.)

- Simulations

- Quantiles

- QQ-plots

- Importing Data

- Markdown Syntax

- LaTeX