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

> **! IMPORTANT NOTES**
>
> - 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 second Midterm. 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.
> - Please keep in mind that the final exam is cumulative.

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

- Transforming Data

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

Random Variables and Distributions (Chapters 3 and 4 of *OpenIntro Statistics*)

- 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

Inferential Statistics (Chapter 5, and Parts of Chapters 6 and 7 from *Open-Intro Statistics*)

- 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 in a single proportion

    - 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

- Hypothesis Testing across multiple populations

    - Two-sample t-test
        * Satterthwaite approximation
    - Analysis of Variance (ANOVA)
        * SS_G, SS_E, MS_G, MS_E
        * F-distribution

Statistical Modeling (Parts of Chapters 8 and 9 of *OpenIntro Statistics*, and beyond)

- Pearson's correlation coefficient

- Statistical models

    - Response variable
    - Explanatory variable
    - Signal function

- Regression vs. Classification models/problems
  - Trends/Associations

- Simple Linear Regression

  - Ordinary Least Squares (OLS) estimators $\hat{\beta}_0$ and $\hat{\beta}_1$
  - Residuals
  - Fitted values
  - OLS regression line
  - Prediction using the OLS regression line
  - Extrapolation
  - Inference on the slope
    * Confidence intervals
    * Hypothesis testing

- Regression Diagnostics

  - Residuals plots and how to read them

Sampling (Chapter 1 of *OpenIntro Statistics*)

- Research Process

- Sampling Procedures

  - Simple Random Sampling
  - Stratified Sampling
  - Cluster Sampling
  - Convenience Sampling

- Bias (including non-response bias)

- Observational studies vs. experiments

  - Relationship to causal associations

- Longitudinal vs. cross-sectional studies

- Experimental Design

Programming

- General terminology

  - Code cells
  - Running / executing code
  - Expressions
  - Order of operations
  - Variable assignment/re-assignment
  - Modules
  - Different Types of Error

- Data Types

- Data Classes (lists, arrays, tables)

- Comparisons

- Conditional expressions

- Functions
    - Docstring
    - Return statement

- Visualization using Python

- Loops

- Random number generation, and sampling from a distribution

- Markdown Syntax

- Exploratory Data Analysis (EDA)