Introduction

Chapter 1: Introduction to Statistics

Chapter 2: Data Types and Distribution Shapes

Chapter 3: Descriptive Statistics in R

Chapter 4: Probability

Chapter 5: Discrete Random Variables

Chapter 6: Continuous Random Variables and Probability Distributions

Objectives for Exam 1

Introduction

This document outlines the objectives for Exam 1, covering the essential concepts from chapters 1 through 6.

Chapter 1: Introduction to Statistics

- Define and demonstrate knowledge of the **three branches of statistics**:
  - **Data Collection**: The process of gathering information.
  - **Descriptive Statistics**: Summarizing and organizing data.
  - **Inferential Statistics**: Drawing conclusions from data.
- Define and distinguish between a *population* and a *sample* including their respective symbols; population parameters by Greek letters, sample statistics are denoted by Latin letters.
- Determine whether a listing of objects refers to a **population or a sample**.
- Identify situations that exemplify **probability** or **inferential statistics**.

Chapter 2: Data Types and Distribution Shapes

- Identify data as **univariate, bivariate, or multivariate**.
- Recognize and classify variables as **categorical/qualitative** or **numerical/quantitative**.
- Describe the **shape of a distribution**:
  - **Peaks**: unimodal, bimodal, multimodal.
  - **Symmetry**: symmetric, right skewed, or left skewed.
  - **Outliers**: Identify and distinguish “real” outliers from the explicit points.
- **Interpret histograms** to describe shape and identification of outliers.

Chapter 3: Descriptive Statistics in R

- Given R output, identify the statistics: **mean, median, variance, standard deviation, and quartiles**.
- Understand and state the **formulas** for sample mean and sample variance:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

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

- Calculate **standard deviation from variance**:

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

- Calculate the **Interquartile Range (IQR)** and explain **quartiles** in non-mathematical terms.

$$\text{IQR} = Q_3 - Q_1$$

- Write down the **five-number summary** from R output and interpret **modified boxplots**.
- Using the five number summary identify inner and outer fences.

$$\text{IF}_\text{L} = Q_1 - 1.5 \times \text{IQR}, \ \ \text{IF}_\text{H} = Q_3 + 1.5 \times \text{IQR}$$

$$\text{OF}_\text{L} = Q_1 - 3 \times \text{IQR}, \ \ \text{OF}_\text{H} = Q_3 + 3 \times \text{IQR}$$

- Identify **explicit** points using the **1.5 IQR rule** and evaluate if they are **“real”**.
- Draw/complete a modified boxplot from the **five number summary** and **1.5 IQR rule**.
- Interpret the results of a modified boxplot or side-by-side boxplots.
- Decide on the appropriate **measures of location and spread** for given data.

Chapter 4: Probability

- Write down the **sample space** for experiments and determine **disjoint events**.
- Understand the **frequentist interpretation of probability**:

$$\lim_{n \to \infty} \frac{n(E)}{n} \approx P(E)$$

- State and check the axioms associated with a probability space Ω:

$$\text{For any event E} \subseteq \Omega, \quad 0 \leq P(E) \leq 1$$

$$P(\Omega) = 1$$

$$\text{For any event E} \subseteq \Omega, P(E) = \sum_{\omega \in \text{E}} P(\omega)$$

- Calculate a **probability** using:
    - The theoretical (or classical) approach (if equally likely can be assumed):

$$P(A) = \frac{\text{Number of outcomes in A}}{\text{Total number of outcomes}}$$

    - Empirical approach using a provided probability distribution table.
- Use **Venn diagrams** to visualize and calculate probabilities.
- Calculate probabilities using **probability rules**:
    - **General Addition Rule** for any events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

    - **Special Addition Rule** for **disjoint events** A and B

$$P(A \cup B) = P(A) + P(B)$$

    - **Complement Rule**

$$P(A') = 1 - P(A)$$

    - **Law of Partitions**: If $B_1, \ldots, B_n$ are exhaustive and mutually exclusive events then

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

    - **Law of Total Probability**: If $B_1, \ldots, B_n$ are exhaustive and mutually exclusive then

$$P(A) = \sum_{i=1}^{n} P(A \ \ B_i) P(B_i)$$

    - Calculate **conditional probabilities** using **probability rules**:

$$P(B\ A) = \frac{P(A \cap B)}{P(A)}$$

  - State and use the **general multiplication rule** to determine probabilities of intersections.
    - **General Multiplication Rule** for Two Events:

$$P(A \cap B) = P(A)P(B\ A) = P(B)P(A\ B)$$

    - **General Multiplication Rule** for Three Events:

$$P(A \cap B \cap C) = P(A)P(B\ A)P(C\ A \cap B) = P(B)P(A\ B)P(C\ A \cap B) = P(C)P(B\ C)P(A\ B \cap C) = ...$$

- **Independence**
  - Two events, A and B, are independent if the occurrence of one does not affect the probability of the other: $P(A\ B) = P(A), \quad (B\ A) = P(B)$
  - Special multiplication rule for independent events (Only use if you know for a fact they are independent.)

$$P(A \cap B) = P(A) \times P(B)$$

- **Bayes' Rule**:
- Baye's Rule for 2 Events:

$$P(A\ B) = \frac{P(B\ A)P(A)}{P(B\ A)P(A) + P(B\ A')P(A')}$$

- General Baye's Rule for n Events: If $A_1, \ldots, A_n$ are exhaustive and mutually exclusive events

$$P(A_i\ B) = \frac{P(B\ A_i)P(A_i)}{\sum_{i=1}^{n} P(B\ A_i)P(A_i)}$$

Chapter 5: Discrete Random Variables

- Recognize the properties of a **valid probability distribution** for discrete variables:
  - Each probability

$$p_X(x)$$

satisfies

$$0 \leq p_X(x) \leq 1$$

.
  - The sum of all probabilities $\sum p_X(x) = 1$.
- Calculate probabilities using a **probability mass function (pmf)**.
- Calculate the **mean of a discrete random variable (Expected value)**:

$$\mathrm{E}(X) = \mu_X = \sum x \cdot p(x)$$

- Calculate the **variance and standard deviation for a discrete random variable**:
  - Variance:

$$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}[(X - \mu_X)^2] = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

  - Standard deviation:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

Rules for Expected Value and Variance

- **LOTUS** For any real valued function $g(\cdot)$ and discrete random variable X

$$\mathrm{E}[g(X)] = \sum_x g(x)p_X(x)$$

- **Linearity of Expectation**: For any two random variables X and Y, and constants a and b,

$$\mathrm{E}(aX \pm bY) = a\mathrm{E}(X) \pm b\mathrm{E}(Y)$$

- **Variance of a Linear function**: For any random variable X and constants $a \neq 0$ and b,

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

  This shows that adding a constant b to a random variable does not change its variance, while multiplying by a scales the variance by a^2.

- **Variance of the Sum/Difference of Two Independent Random Variables**: If X and Y are independent,

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Named Distributions

- For a **Binomial distribution**, understand when it applies (BInS criteria) and how to calculate probabilities, expected values, and variances:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\text{E}(X) = np, \quad \sigma_X = \sqrt{np(1-p)}$$

- For a **Poisson distribution**, recognize when it applies and how to calculate probabilities, expected values, and variances:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$\text{E}(X) = \lambda, \;\; \sigma_X = \sqrt{\lambda}$$

Chapter 6: Continuous Random Variables and Probability Distributions

- Determine if a function is a **legitimate density function** and calculate the **normalization constant** if necessary.
  - **Legitimate Density Functions**: A function $f(x)$ is a legitimate density function if it satisfies two conditions:
    1. $f(x) \geq 0$ for all x.
    2. The total area under the curve of $f(x)$ over its entire range equals 1, i.e., $\int_{-\infty}^{\infty} f(x)dx = 1$.
  - **Normalization Constant**: The constant required to ensure the total area under the probability density function (pdf) equals 1.
- Calculate **probabilities** for a continuous random variable using the density function:

$$P(a < X < b) = \int_a^b f(x)dx$$

- Calculate and use the **cumulative distribution function (CDF)**:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- **Percentiles and Median**:
  - **Percentile**: Solve $F(y) = p$ to find the yth percentile.
  - **Median**:

$$\int_{-\infty}^{\tilde{\mu}} f(x)dx = 0.5$$

  - **Mean (Expected Value)**:

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Named Distributions

(Note: The distributions have been written in short hand notation. You need to realize where the pdf/cdf is 0 and where the cdf is 1.)

- **Normal Distribution**:
  - Use the z-table for calculating probabilities and percentiles.
  - Normal probability plots help determine if data follow a normal distribution. Deviations suggest skewness or a non-normal distribution.

- For a **Uniform Distribution**, understand when it applies and how to calculate probabilities, percentiles, expected values, and variances:
  - Probability Density Function (pdf):

$$f(x) = \frac{1}{b-a}, \quad \text{for } \ a \le x < b$$

  - CDF:

$$F(x) = \frac{x-a}{b-a}, \quad \text{for } \ a \le x < b$$

  - Mean and Standard Deviation:

$$E(X) = \frac{a+b}{2}, \quad \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

- For a **Exponential Distribution**, understand when it applies and how to calculate probabilities, percentiles, expected values, and variances:
  - pdf:

$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } \ x \ge 0$$

  - CDF:

$$F(x) = 1 - e^{-\lambda x}, \quad \text{for } \ x \ge 0$$

  - Mean and Standard Deviation:

$$E(X) = \frac{1}{\lambda}, \quad \sigma = \frac{1}{\lambda}$$