Robert Davidson

ST1112: Statistics

70% Exam
30% Continuous Assessment (3 parts)

Contents

1 Inferential Statistics

The ultimate goal in statistical inference is to estimate population paramaters (like the mean μ) based on sample statistics (like the sample mean $\bar{X}$).

1.1 Probability vs Statistics

- **Probability** deals with known underlying processes: one starts with a model (like proportion of red vs. green jelly beans in a jar) and computes probability of specific outcomes
- **Statistics** works in reverse: one observes outcomes (sample data) and attempts to infer the underlying process or population paramaters (e.g. proportion of red jellybeans)

1.2 Definitions and Concepts

Definition 1.1: Population

A **population** is the complete set of items (or individuals) of interest.

Definition 1.2: Sample

A **sample** is a subset of that populaton, intended to represent the population

For example the sample mean $\bar{X}$ is an estimate of the population mean μ.

Concept 1.1: Sampling Variation

When we take multiple samples from the same population, each sample's mean $\bar{X}$ will be different. This is variability is called **sampling variation**.

Larger sample sizes tend to reduce this variation, that is as n gros, the sample mean $\bar{X}$ becomes a better estimate of the population mean μ.

Concept 1.2: Sampling Distributions

The sample mean itself is a **random variable** becuase different samples yield different mean values.

The distribution of all possible sample means (of a given sample size n) is called the **sampling distribution** of the sample mean ($\bar{X}$).

Definition 1.3: Expected Value of the Sample Mean

$$E(\bar{X}) = \mu$$

This means if you averaged all possible sample means, you would get the population mean μ.

Definition 1.4: Standard Error of the Mean

$$SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

where σ is the population standard deviation and n is the sample size.

This value is valled the **standard error** of the mean and measures how much the sample mean $\bar{X}$ fluctuates around the population mean μ.