

Robert Davidson
ST1112: Statistics

70% Exam
30% Continuous Assessment (3 parts)

Contents

1	Inferential Statistics	3
1.1	Probability vs Statistics	3
1.2	Definitions and Concepts	3
	Definition 1.1	3
	Definition 1.2	3
	Definition 1.3	3
	Concept 1.1	3
	Concept 1.2	3
	Definition 1.4	3
	Definition 1.5	4
	Definition 1.6	4
	Definition 1.7	4
	Definition 1.8	4

1 Inferential Statistics

The ultimate goal in statistical inference is to estimate population parameters (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 parameters (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 population, intended to represent the population

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

Definition 1.3: Population Mean (μ)

μ is the population mean (also called the expected value or average). It represents the central tendency of a population distribution.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

where N is the population size and x_i are the individual values in the population.

Concept 1.1: Sampling Variation

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

Larger sample sizes tend to reduce this variation, that is as n grows, 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** because 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 1.4: 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 1.5: 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 called the **standard error** of the mean and measures how much the sample mean \bar{X} fluctuates around the population mean μ .

Definition 1.6: Central Limit Theorem

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

where \bar{X} is the sample mean, μ is the population mean, and σ is the population standard deviation.

The **Central Limit Theorem** states that the sampling distribution of the sample mean \bar{X} (the distribution of all sample means) approaches a normal distribution as the sample size n increases, **regardless of the shape of the population distribution**.

This means that for large enough sample sizes, we can use the normal distribution to make inferences about the population mean μ .

Practically, many apply the rule of thumb $n \geq 30$ to treat \bar{X} as normally distributed.

Definition 1.7: Bessel's Correction

Definition 1.8: Unbiased Estimators

We say a statistic T is an **unbiased estimator** of a population parameter θ , if $E(T) = \theta$.

For example, the sample mean \bar{X} is an unbiased estimator of the population mean μ because $E(\bar{X}) = \mu$.

The sample standard deviation s (using Bessel's correction, dividing by $n - 1$) is an unbiased estimator of the population standard deviation σ .