ACTL2131 2.1 - Data Visualisation and Descriptive Statistics

Tadhg Xu-Glassop

2025T1

Population v.s. Sample

Population

The **population** represents the entire group we are interested in. Analysing the entire population might be unfeasible however.

Sample

A **sample** represents a subgroup of the population; it is our window into the population that we can use to learn more about it.

For example, say we are interested in the height of Australian university students. Then, the population of our study is all Australian university students, and a sample could be this class.

Descriptive Statistics - Sample Mean

Intuitively, the expected value (mean) of the population is very insightful for **location** of the population! Given a sample, how should we estimate it?

Descriptive Statistics - Sample Mean

Intuitively, the expected value (mean) of the population is very insightful for **location** of the population! Given a sample, how should we estimate it?

Sample Mean

Let $X_1, X_2, \ldots, X_n$ be IID observations from a population with mean μ. Then, we define the sample mean as

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

and $\mathbb{E}[\bar{X}_n] = \mu$.

The sample mean is an intuitive and unbiased estimator for the population mean.

Descriptive Statistics - Sample Variance

Further, the variance of the population is very insightful for the **scale** / **spread** of the population! Given a sample, how should we estimate it?

Descriptive Statistics - Sample Variance

Further, the variance of the population is very insightful for the **scale** / **spread** of the population! Given a sample, how should we estimate it?

Sample Variance

Let $X_1, X_2, \ldots, X_n$ be IID observations from a population with variance σ^2. The sample variance is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2.$$

We have that $\mathbb{E}[s^2] = \sigma^2$.

The sample variance *looks* like the population variance with a small correction (dividing by $n-1$ instead of n). It is unbiased.