Statistics & Sampling Distributions

Population and samples : a (statistical) **population** is the complete set of all possible measurements or values, corresponding to the entire collection of units, for which inferences are to be made from taking a **sample** - the set of measurements or values that are actually collected from a population.

Simple random sample : every item in the population is equally likely to be in the sample, independently of which other members of the population are chosen.

Parameter : a quantity that describes an aspect of a population, eg. the population mean, μ, or variance, σ^2.

Statistic : a quantity calculated from the sample, e.g. the sample mean, $\bar{x}$, or variance, s^2.

Sampling distributions : the value of a statistic will in general vary from sample to sample, in which case it will have its own probability distribution, called its **sampling distribution**. A statistic used to estimate the value of a *parameter* θ in a distribution is called an **estimator** (the random variable) or an **estimate** (the value).

If $\hat{\theta}$ is an estimator of θ , the mean of its sampling distribution, $\mathbf{E}[\hat{\theta}]$, is called the *sampling mean*. The variance, $\mathrm{Var}(\hat{\theta})$, is called the *sampling variance*.

$\sqrt{\mathrm{Var}(\hat{\theta}}$) is called the *standard error* of $\hat{\theta}$. If $\mathbf{E}[\hat{\theta}] = \theta$, then $\hat{\theta}$ is an unbiased estimator of θ e.g. $\bar{X}$ is an unbiased estimator for μ and has sampling variance $\frac{\sigma^2}{n}$ where Var(X_i) $= \sigma^2$, $(i = 1, 2, \ldots, n)$.

Corrected sum of squares

$$S_{xx} = \sum (x_i - \bar{x})^2 \equiv \sum x_i^2 - n\bar{x}^2 \equiv \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

has expectation $(n-1)\sigma^2$ so that dividing S_{xx} by $(n-1)$ will give an unbiased estimator of σ^2, denoted s^2.

Normal and Chi-squared distributions

If $X_1, X_2, \ldots X_n$ are independently and identically $\sim$ N(μ, σ^2), then $\sum \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$, a Chi-squared distribution with n **degrees of freedom**.

Also $\bar{X} \sim$ N $\left(\mu, \frac{\sigma^2}{n}\right)$ independently of $\frac{S_{xx}}{\sigma^2} \sim \chi^2_{(n-1)}$.