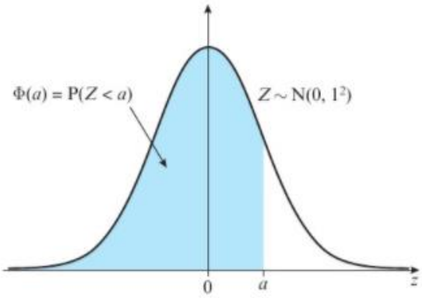
A2 Statistics – Chapter 3 – The Normal Distribution

$\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, where μ is the population mean, σ is the standard deviation and σ^2 is variance

- Continuous, used to find probabilities for **data distributed symmetrically about a mean value**
- The distribution has a bell-shaped curve with asymptotes at each end.
- The points of inflection are at $\mu + \sigma$ and $\mu - \sigma$
- As the distribution is symmetrical, mean = median = mode
- The total area under any normal curve is equal to 1
- The standardised normal distribution has parameters $\mu = 0$ and $\sigma = 1$ $\Longrightarrow$ $\boldsymbol{Z \sim N(0\,,1)}$



- A variable X can be standardised using the formula $\boldsymbol{z = \dfrac{x - \mu}{\sigma}}$
- Your calculator can give you cumulative probabilities for any normal distribution, including Z.
- Your calculator has an inverse normal function which gives the value of your variable when you input the area **of the lower tail only** as a probability, including values of z if you

Approximations

The **normal** distribution may be used to approximate the **binomial** distribution $B(n, p)$ if

- n is large, p is not close to 0 or 1 (so that the binomial is roughly symmetrical)

The approximating normal distribution has parameters $\mu = np,\ \sigma^2 = npq = np(1-p) \quad \Rightarrow \quad \boldsymbol{N(np, npq)}$

Continuity Corrections

When using a Normal distribution (continuous) to approximate the binomial distribution (discrete), you must apply **continuity corrections** – **do not forget these** and be very careful with inclusive/exclusive inequalities.

Examples:

Binomial	Normal
$P(X = 45)$	$P(44.5 < X < 45.5)$
$P(X < 45)$	$P(X < 44.5)$
$P(X \leq 45)$	$P(X < 45.5)$

Sample Means

For **samples of size n** from a normally distributed population $X \sim N(\mu\,,\sigma^2)$, the **distribution of sample means** is also normal with the same mean (μ), but the sample standard deviation (s) decreases as sample size is increased:

Population: $X \sim N(\mu\,,\sigma^2)$ $\rightarrow$ **Sample means**: $\bar{X} \sim N\left(\mu\,,\frac{\sigma^2}{n}\right)$

Standard deviation of sample means: $s = \frac{\sigma}{\sqrt{n}}$ (also known as the standard error of the mean)

Standardizing formula for sample means: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (where $\bar{x}$ is your measured sample mean)

You may have to calculate the standard deviation of a large sample from summary statistics using $s = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$

This can be used as a good estimate of the population standard deviation for (roughly) $n > 30$

Hypothesis Testing using Sample Means – Example

Test results for the module S1 are normally distributed with a mean of 65 and a standard deviation of 10.
After the introduction of a dynamic new teacher (Mr Ludlam), the results of a group of 8 students had a mean of 72.
Is there evidence that the results have significantly improved, at the 5% level of significance?

Solution

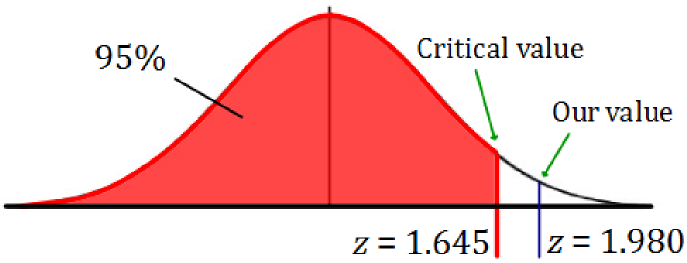
$\mathrm{H}_0 : \mu = 65$

$\mathrm{H}_1 : \mu > 65$ where μ is the population mean test result

Let X represent the mark an individual student achieves, so $X \sim N(65\,,10^2)$

Distribution of sample means: $\bar{X} \sim N\left(65\,,\frac{10^2}{8}\right)$ (sample size $n = 8$)

One-tailed test, 5% significance level (upper tail, so $p = 0.95$)



Test Statistic: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{72-65}{10/\sqrt{8}} = \mathbf{1.980}$

Critical Value, upper tail, 5% level: $\Phi^{-1}(0.95) = \mathbf{1.645}$

Since $1.980 > 1.645$, the result is significant, reject H_0

There is sufficient evidence **to suggest** that Mr Ludlam has had a positive effect.