The normal distribution

Modelling continuous data with a normal distribution, using tables, finding probabilities, working back to the mean or standard deviation, and approximating the binomial with a continuity correction.

The normal distribution

The normal curve

Many continuous quantities, such as heights and masses, follow a symmetric bell-shaped curve, written $X \sim \mathrm{N}(\mu, \sigma^2)$. The mean μ sets its centre, and the standard deviation σ sets how spread out it is.

- The total area under the curve is 1, and an area gives a probability.
- For a continuous variable, $\mathrm{P}(X = a) = 0$, so $\leq$ and $<$ give the same result.
- About 68% of values lie within σ of the mean, and about 95% within 2σ.

Standardising

Any normal variable can be converted to the standard normal $Z \sim \mathrm{N}(0, 1)$, which is what the tables give.

$Z = \dfrac{X - \mu}{\sigma}$, and the tables give $\Phi(z) = \mathrm{P}(Z < z)$.

Finding probabilities

Sketch the curve, shade the area you want, and use the symmetry:

Area wanted	Calculation (for $z > 0$)
$\mathrm{P}(Z < z)$	$\Phi(z)$
$\mathrm{P}(Z > z)$	$1 - \Phi(z)$
$\mathrm{P}(Z < -z)$	$1 - \Phi(z)$
$\mathrm{P}(Z > -z)$	$\Phi(z)$
$\mathrm{P}(-z < Z < z)$	$2\Phi(z) - 1$

TRY IT

Note Academy

$X \sim \mathrm{N}(50, 16)$. Find $\mathrm{P}(X > 55)$ and $\mathrm{P}(45 < X < 55)$.

$\sigma = 4$, so $z = \dfrac{55 - 50}{4} = 1.25$ and $\Phi(1.25) = 0.8944$.

$\mathrm{P}(X > 55) = 1 - 0.8944 = 0.1056$

$\mathrm{P}(45 < X < 55) = 2(0.8944) - 1 = 0.7888$

NOTE

The bracket in $\mathrm{N}(50, 16)$ gives the **variance**. Take the square root before standardising.

Going backwards from a probability

Given a probability, use the tables in reverse to find z, and then x.

TRY IT

$X \sim \mathrm{N}(50, 16)$. Find a where $\mathrm{P}(X < a) = 0.9$.

$\Phi(z) = 0.9$ gives $z = 1.282$. So $a = 50 + 1.282 \times 4 = 55.1$.

A probability below 0.5 to the left of a means a is below the mean, so its z is negative. Check this with a sketch.

Finding μ or σ

TRY IT

$X \sim \mathrm{N}(70, \sigma^2)$ and $\mathrm{P}(X > 80) = 0.1$. Find σ.

$\mathrm{P}(X < 80) = 0.9$, so $\dfrac{80 - 70}{\sigma} = 1.282$ and $\sigma = 7.80$.

Two unknowns need two equations.

TRY IT

$\mathrm{P}(X < 30) = 0.1$ and $\mathrm{P}(X > 60) = 0.2$. Find μ and σ.

Note Academy

$\frac{30-\mu}{\sigma} = -1.282$ $\frac{60-\mu}{\sigma} = 0.842$

Subtracting gives $30 = 2.124\sigma$, so $\sigma = 14.1$, and then $\mu = 30 + 1.282 \times 14.1 = 48.1$.

Normal approximation to the binomial

A binomial with a large n has a bell-shaped distribution, so it can be approximated by a normal curve.

> **If $np > 5$ and $n(1-p) > 5$, then $\mathrm{B}(n, p)$ is approximately $\mathrm{N}(np, np(1-p))$.**

The continuity correction

The binomial is discrete and the normal is continuous, so each whole number is treated as the interval reaching 0.5 either side of it.

Binomial	Normal
$X = 45$	$44.5 < Y < 45.5$
$X \geq 45$	$Y > 44.5$
$X > 45$	$Y > 45.5$
$X \leq 45$	$Y < 45.5$
$X < 45$	$Y < 44.5$

> **TRY IT**
>
> $X \sim \mathrm{B}(100, 0.4)$. Use a normal approximation to find $\mathrm{P}(X \geq 45)$.

$np = 40$ and $np(1-p) = 24$, both of which allow this approximation, so use $Y \sim \mathrm{N}(40, 24)$.

$\mathrm{P}(Y > 44.5)$: $z = \frac{44.5-40}{\sqrt{24}} = 0.919$

$1 - \Phi(0.919) = 1 - 0.8209 = 0.179$

> **NOTE**
>
> Show that you have checked the condition, and write the numbers for np and $n(1-p)$. Each is commonly worth a mark.

Note Academy