Chapter 24

The normal distribution

Syllabus reference: 5.9

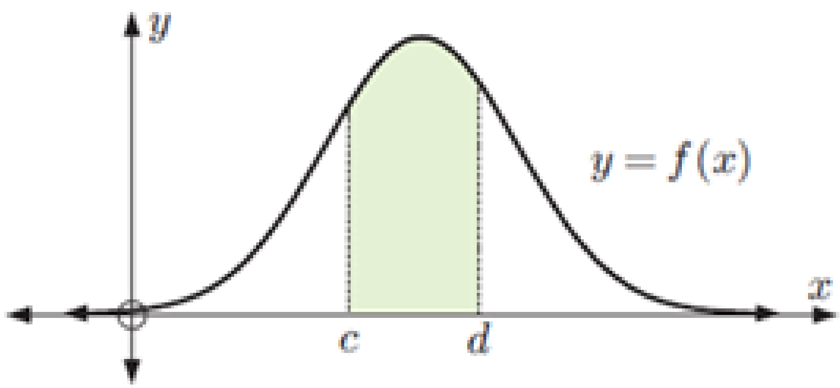
Contents:

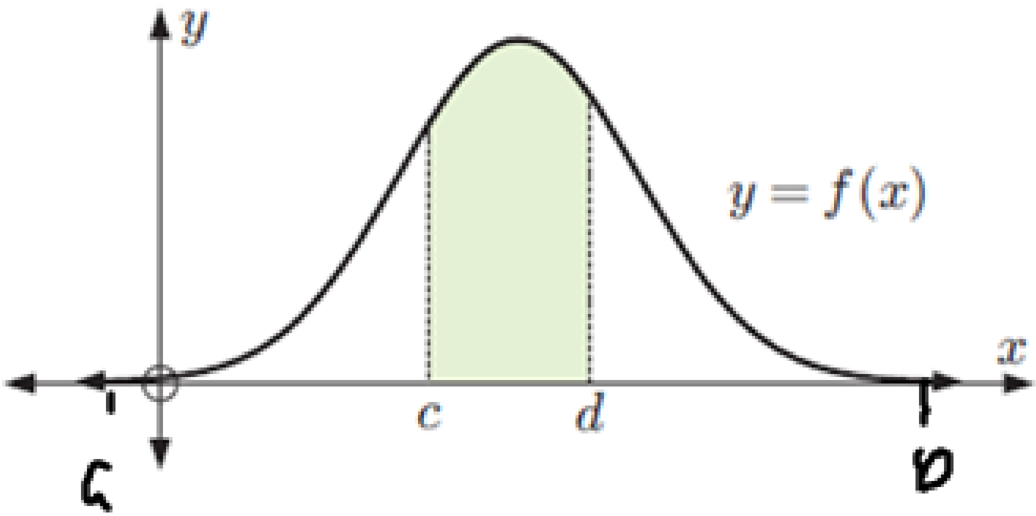
24A – The Normal Distribution

From Ch. 23: **discrete** random variables
- examined binomial probability distributions where X could take non-negative integer values, $x = 0, 1, 2, 3, \ldots$

For a **continuous** random variable X, x can be any real value.

Use a **probability density function** to define the probability distribution.
Probabilities are found be calculating areas under the probability density function curve.





For a continuous random variable X, the **probability density function** is a function $f(x)$ such that $f(x) \geqslant 0$ on its entire domain.

If the domain of the function is $a \leqslant x \leqslant b$, then $\int_a^b f(x)\,dx = 1$.

The probability that X lies in the interval $c \leqslant X \leqslant d$ is $\mathbf{P}(c \leqslant X \leqslant d) = \int_c^d f(x)\,dx$.

For a continuous variable X, the probability that X is *exactly* equal to a particular value is zero:

$$P(X = x) = 0 \text{ for all } x.$$

For example, the probability that an egg will weigh *exactly* 72.9 g is zero.

If you were to weigh an egg on scales that weigh to the nearest 0.1 g, a reading of 72.9 g means the weight lies somewhere between 72.85 g and 72.95 g. No matter how accurate your scales are, you can only ever know the weight of an egg within a range.

So, for a continuous variable we can only talk about the probability that an event lies in an **interval**.

A consequence of this is that $P(c \leqslant X \leqslant d) = P(c < X \leqslant d) = P(c \leqslant X < d) = P(c < X < d)$.

This would not be true if X was discrete.

THE NORMAL PROBABILITY DENSITY FUNCTION

If X is **normally distributed** then its **probability density function** is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for} \quad -\infty < x < \infty$$

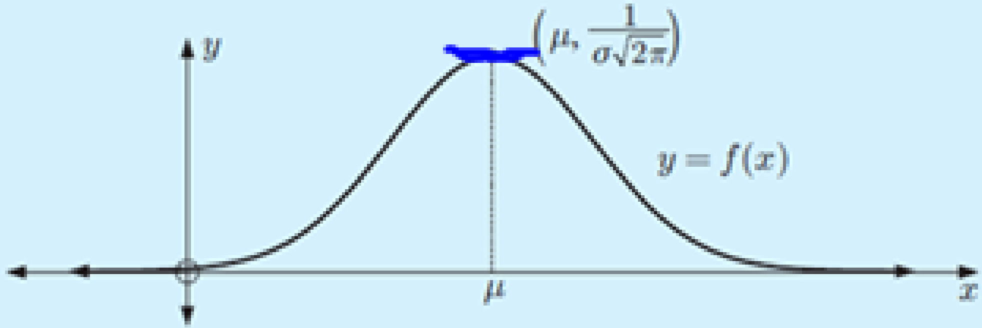
where μ is the mean and σ^2 is the variance of the distribution.

We write $\quad X \sim \mathrm{N}(\mu, \sigma^2)$.

$f(x) = ae^{-bx^2}$

↑ Chain Rule

- The curve $y = f(x)$, which is called a **normal curve**, is symmetrical about the vertical line $x = \mu$.
- As $x \to \pm\infty$ the normal curve approaches its asymptote, the x-axis.
- $f(x) > 0$ for all x.
- Since the total probability must be 1, $\int_{-\infty}^{\infty} f(x)\,dx = 1$.



- More scores are distributed closer to the mean than further away. This results in the typical **bell shape**.

THE GEOMETRIC SIGNIFICANCE OF μ AND σ

Differentiating $\quad f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

we obtain $\quad f'(x) = \dfrac{-1}{\sigma^2\sqrt{2\pi}} \left(\dfrac{x-\mu}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\therefore \quad f'(x) = 0$ only when $x = \mu$, and this corresponds to the point on the graph when $f(x)$ is a maximum.

Differentiating again, we obtain $\quad f''(x) = \dfrac{-1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[\dfrac{1}{\sigma} - \dfrac{(x-\mu)^2}{\sigma^3}\right]$

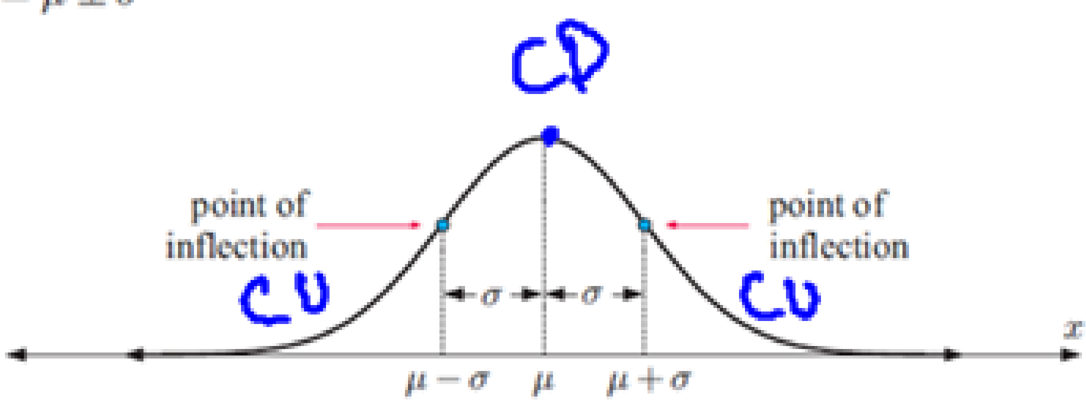
$\therefore \quad f''(x) = 0$ when $\dfrac{(x-\mu)^2}{\sigma^3} = \dfrac{1}{\sigma}$

$$\therefore \quad (x-\mu)^2 = \sigma^2$$

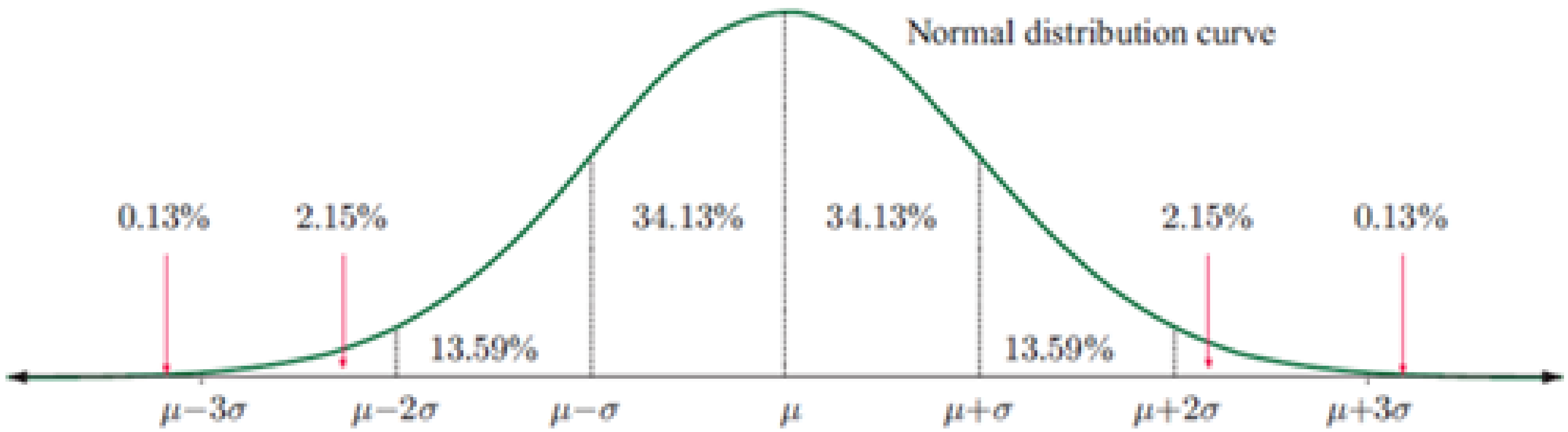
$$\therefore \quad x - \mu = \pm\sigma$$

$$\therefore \quad x = \mu \pm \sigma$$

So, the points of inflection are at $x = \mu + \sigma$ and $x = \mu - \sigma$.



For a normal distribution with mean μ and standard deviation σ, the percentage breakdown of where the random variable could lie is shown below.

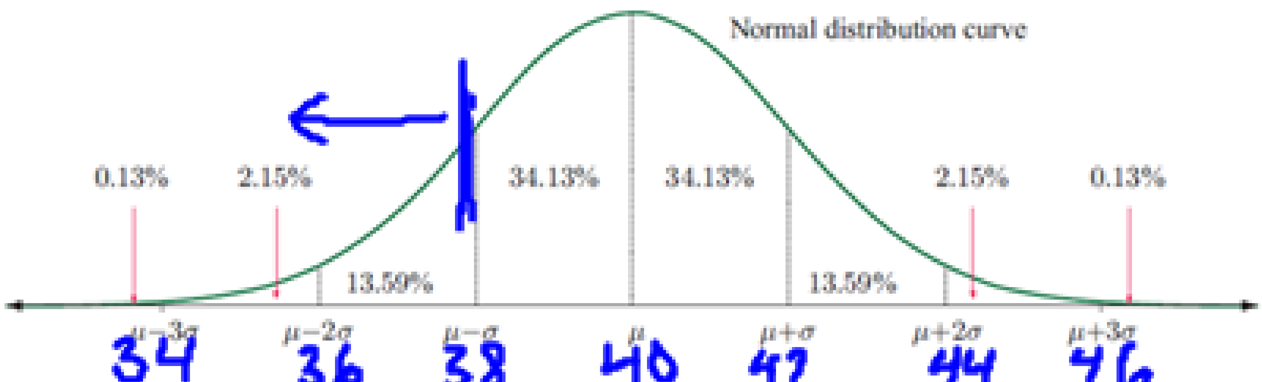


Notice that:

- $\approx 68.26\%$ of values lie between $\mu - \sigma$ and $\mu + \sigma$
- $\approx 95.44\%$ of values lie between $\mu - 2\sigma$ and $\mu + 2\sigma$
- $\approx 99.74\%$ of values lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Example:

In an Oreo factory, the masses of cookies are normally distributed with a mean of 40 g and a standard deviation of 2 g.



(a) Find the percentage of cookies with a mass between 36 g and 42 g.

13.59% + 34.13% + 34.13% = 81.85%

(b) Find the probability that the mass of a cookie is between 38 g and 46 g.

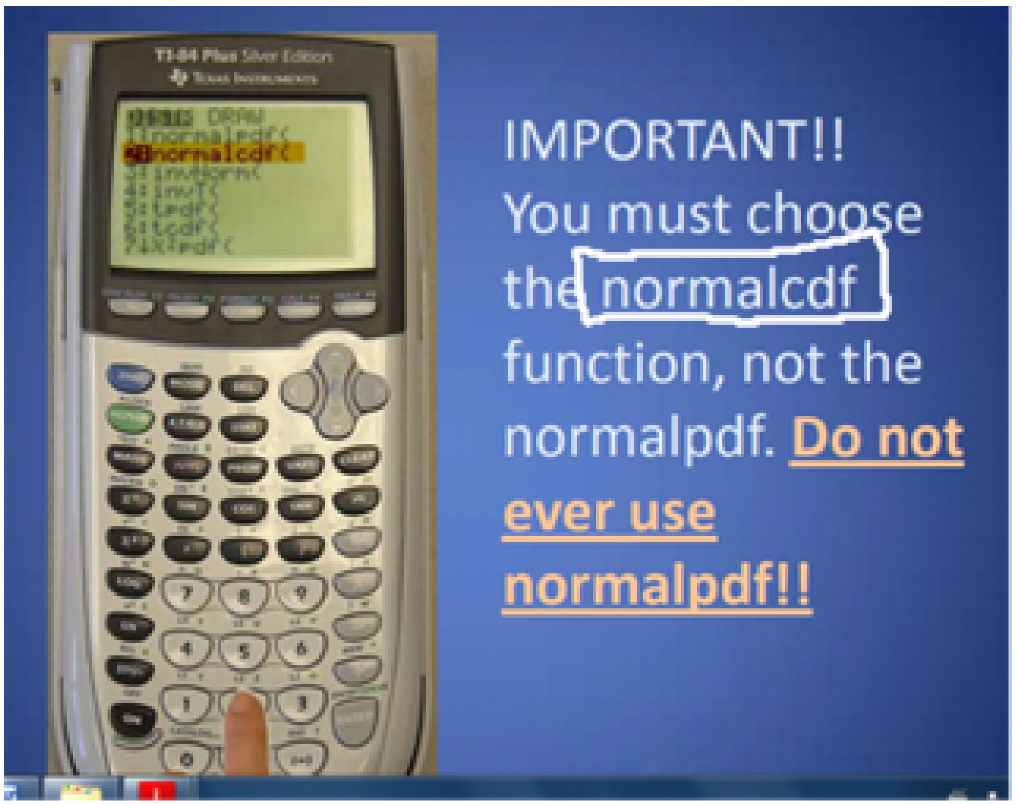
0.6826 + 0.1359 + 0.0215 = 0.84

(c) Find the value of k such that approximately 16% of the cookies are below k g.

k = 38g

24 B – Probabilities Using a Calculator

Need to use the calculator to find probabilities that are not just regions of width of σ from the mean.



Example: ← is modelled by a Normal dis. function

Suppose $X \sim N(50, 3^2)$. (mean ← 50; variance: ~~st~~ st.dev = 3)

This means X is normally distributed with mean 50 and standard deviation 3.

Find:

(a) $P(40 \le X \le 65)$

(b) $P(X \ge 35)$

$P(X \geq 35) = 1 - P(X \leq 35)$

```
lower:-1E99
upper:35
μ:50
σ:3
Paste
```

```
normalcdf(-1E99▸
         2.87105E-7
1-Ans
       .9999997129
```

(c) $P(X \le 42)$

24 C – The Standard Normal Distribution (*Z*-Distribution)

Every normal *X*-distribution can be transformed into the standard normal distribution, also called the *Z*-distribution using the transformation

$$z = \frac{x - \mu}{\sigma}$$

No matter what the values of μ and σ are of the *X*-distribution, the *Z*-distribution will always have mean 0 and standard deviation 1.

modelled by a Norm. dist

$$Z \sim N(0, 1)$$

St.dev.

mean

Read pg. 640 for the *Z*-transformation…

The **probability density function** for the *z*-distribution is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

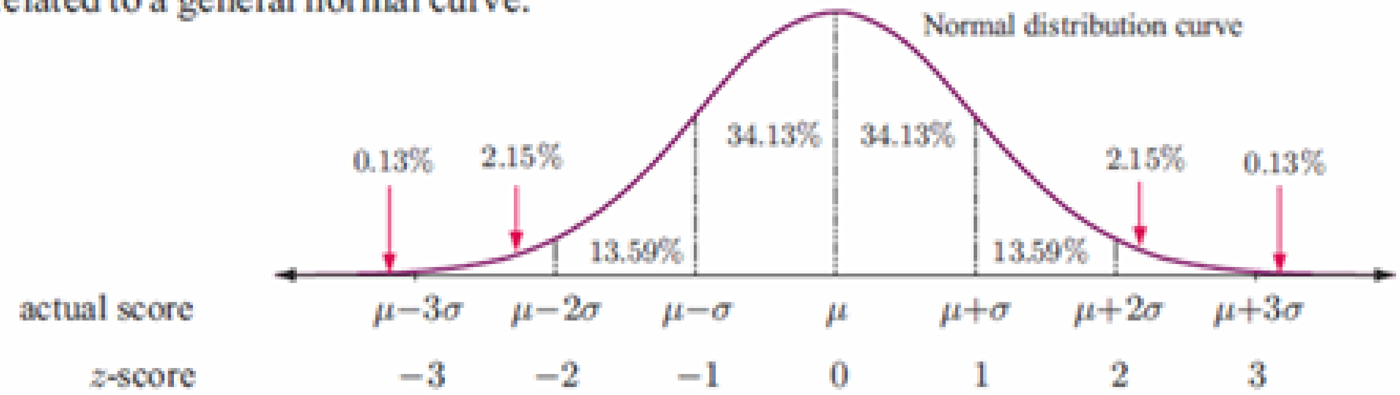
If x is an observation from a normal distribution with mean μ and standard deviation σ, the ***z*-score** of x is the number of standard deviations x is from the mean.

$$z = \frac{x - \mu}{\sigma}$$

For example, if $z = 1.34$, then x is 1.34 standard deviations to the right of the mean.
If $z = -0.31$, then x is 0.31 standard deviations to the left of the mean.

Pg. 641

This diagram shows how the z-score is related to a general normal curve:



This means that for $X \sim \text{N}(\mu, \sigma^2)$ and $Z \sim \text{N}(0, 1^2)$ and for any $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$, with corresponding z-values $z_1 = \frac{x_1 - \mu}{\sigma}$ and $z_2 = \frac{x_2 - \mu}{\sigma}$, then:

- $\text{P}(X \geqslant x_1) = \text{P}(Z \geqslant z_1)$

- $P(X \leqslant x_1) = P(Z \leqslant z_1)$

- $P(x_1 \leqslant X \leqslant x_2) = P(z_1 \leqslant Z \leqslant z_2)$

z-scores are useful when comparing two populations with different means and standard deviations (as long as both distributions are approximately normal).

$$z = \frac{x - \mu}{\sigma}$$

Example:

Pg. 642

1 The table shows Emma's midyear exam results. The exam results for each subject are normally distributed with the mean μ and standard deviation σ shown in the table.

a Find the z-score for each of Emma's subjects.

b Arrange Emma's subjects from 'best' to 'worst' in terms of the z-scores.

Subject	*Emma's score*	μ	σ
English	48	40	4.4
Mandarin	81	60	9
Geography	84	55	18
Biology	68	50	20
Maths	84	50	15

English

$$z = \frac{48-40}{4.4} = 1.82$$

Mand

$$z = \frac{81-60}{9} = 2.33$$

Geo

$$z = \frac{84-55}{18} = 1.61$$

Bio

$$z = \frac{68-50}{20} = 0.9$$

math

$$z = \frac{84-50}{15} = 2.27$$

Pg. 643

5 If Z is the standard normal distribution, find the following probabilities using technology. In each case sketch the regions.

a $P(-0.86 \leqslant Z \leqslant 0.32)$

d $P(Z \leqslant -0.53)$

g $P(Z > 4)$

24D – Quantiles or *k*-values

When finding quantiles, we are given a probability and are asked to calculate the corresponding measurement.

For example, the masses of Oreo cookies are normally distributed with a mean of 40 g and a standard deviation of 2 g.

Find the value of k such that approximately 16% of the cookies are below k g.

We need to find k such that $P(X \leq k) = 0.16$.

Use the **inverse normal function** on the calculator:

Pg. 645

3 Suppose $X \sim N(38.7, 8.2^2)$. Illustrate with a sketch and find k such that:

a $P(X \leqslant k) = 0.9$

b $P(X \geqslant k) = 0.8$

** If we are trying to find an unknown mean μ or standard deviation σ, then we **always** need to convert to z-scores.

Pg. 646

1 The IQs of students at school are normally distributed with a standard deviation of 15. If 20% of students have an IQ higher than 125, find the mean IQ of students at school.

Pg. 647

5 **a** Find the mean and standard deviation of a normally distributed random variable X, given that $P(X \geqslant 80) = 0.1$ and $P(X \leqslant 30) = 0.15$.

b In a Mathematics examination it was found that 10% of the students scored at least 80, and no more than 15% scored under 30. Assuming the scores are normally distributed, what proportion of students scored more than 50?