Chapter 3 - The Normal Distribution

Personal notes:

BF MATHS

3.1 - Intro to normal distribution

<u>Starter</u> (Knowledge Check)

1) The probability that a one-month old Labrador puppy weighs under 2kg is 0.735. Two puppies are chosen at random from different litters. Find:
   a) P(both weigh under 2kg)
   b) P(exactly one weighs under 2kg)
   (S1 Chp 5)

2) $X \sim B(20, 0.4)$. Find
   a) $P(X = 6)$
   b) $P(X \geq 8)$
   c) $P(3 \leq X \leq 10)$
   (S1 Chp 6)

3) The probability that a plate made using a particular production process is faulty is given as 0.16. A sample of 20 plates is taken. Find:
   a) the probability that exactly two plates are faulty
   b) the probability that no more than three plates are faulty
   (S1 Chp 6)

<u>What is a normal distribution?</u>

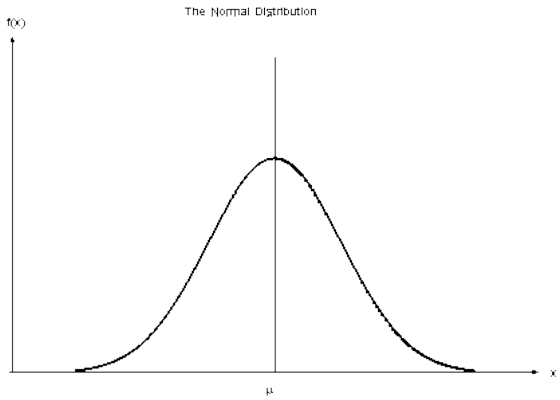
Normal distribution is a ______________ *probability* distribution that can be used to model many naturally occurring characteristics that behave in a way where values are more likely to group around a central value than to take extreme values.

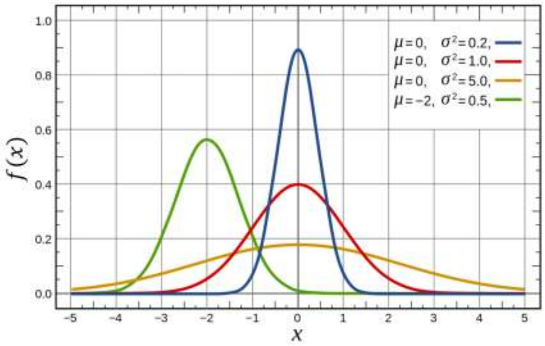
Examples of continuous variables:

- Heights of people within a given population
- Weights of tigers in a jungle
- Errors in scientific measurements
- Size variations in manufactured objects

3.1 - Intro to normal distribution

When put all the values in a graph, it shows a **bell-shaped curve.**



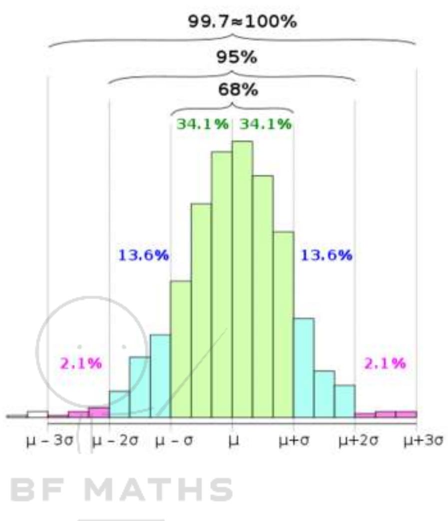


We can set the mean μ and the standard deviation σ of the Normal Distribution. If a random variable X is normally distributed, then we write $X \sim N(\mu, \sigma^2)$

<u>Other Key Information</u>

- For a normal distribution to be used, the variable has to be ________________.
- With a discrete variable (e.g. outcomes of a dice), all the probabilities had to add up to 1. Similarly, for a continuous variable, the
  ____.
- If I were to find P(170 < X < 190)**,** I would find the ______________________________
- How about P(X = 200)? Since the variable is continuous, the probability that someone is exactly 200cm tall is **infinitely small**, we normally round it to **0.**
- The curve has points of inflexion one standard deviation from the mean, i.e. $\mu \pm \sigma$. (Recall this is the point where the curve changes concavity, from decreasing gradient -> increasing gradient, or vice versa)
- Normal distribution is symmetrical
  -> Mean = Mode = Median

<u>The 68-95-99.7 rule</u>



- ≈ 68% of the data lies within **one** s.d. of the mean (i.e. $\mu \pm \sigma$)
- ≈ 95% of the data lies within **two** s.d. of the mean (i.e. $\mu \pm 2\sigma$)
- ≈ 99.7% of the data lies within **three** s.d. of the mean (i.e. $\mu \pm 3\sigma$)

*Although a normal random variable could take any value, but in practice any observations more than 5 standard deviations from the mean would have probabilities close to 0.

3.1 - Intro to normal distribution

Example

The diameters of a bottle lid produced by a particular machine, X mm, is modelled as $X \sim N(5, 0.1^2)$. Find:

a) $P(X > 5)$
b) $P(4.9 < X < 5.1)$
c) $P(4.7 < X < 5.3)$
d) $P(X > 5.2)$

Practice

IQ ("Intelligence Quotient") for a given population is, by definition, distributed using $X \sim N(100, 15^2)$. Find:

a) $P(70 < X < 130)$
b) $P(X > 115)$

Practice

The distribution of incomes, in £000s per year, of employees of a bank is shown on the right.

State, with reasons, why the normal distribution is not a suitable model for this data.

3.2 - Probabilities for normal distributions

Notes

- We can find probabilities for a normal distribution using the **normal cumulative distribution** function on your calculator.

Example

Given $X \sim N(35, 3^2)$, find:

a) $P(X < 33)$
b) $P(X \geq 21)$
c) $P(31.9 < X < 39.7)$
d) $P(X < 26 \text{ or } X > 38)$

Practice

Given $X \sim N(40, 3^2)$, find

a) $P(X > 45)$
b) $P(X \leq 38)$
c) $P(41 \leq X \leq 44)$

Practice

The criteria for joining Mensa is an IQ of at least 131.
Assuming that IQ has the distribution $X \sim N\left(100, 15^2\right)$ for a population, determine:

a) What percentage of people are eligible to join Mensa.
b) If 30 adults are randomly chosen, the probability that at least 3 of them will be eligible to join. (Hint: Binomial distribution?)

3.2 - Probabilities for normal distributions

Practice

The masses, X grams, of a large population of squirrels are modelled as a normal distribution with $X \sim N(480, 1600)$.

a) Find the probability that a squirrel chosen at random has a mass greater than 490g. **(1 mark)**

A naturalist takes a random sample of 30 squirrels from the population.

b) Find the probability that at least 15 of the squirrels have a mass between 470g and 490g. **(4 marks)**

Practice

The diameters of bolts, Dmm, made by a particular machine are modelled as $D \sim N(13, 0.1^2)$.

a) Find the probability that a bolt, chosen at random, has a diameter less than 12.8mm. **(1 mark)**

Bolts are considered to be 'perfect' if the diameter lies between 12.9mm and 13.1mm. A random sample of 40 bolts is taken.

b) Find the probability that more than 25 of the bolts are 'perfect'. **(4 marks)**

3.3 - Inverse normal distributions

Notes

- When given the values of μ and σ, we can find out the probability of any given critiera.
  e.g. $X \sim N(60, 5^2)$, we are able to work out $P(X < 45)$, or $P(X > 50)$

- We can also use "Inverse" function in the calculator to work backwards.

Example

Given $X \sim N(20, 3^2)$. Find, correct to two decimal places, the values of *a* such that:

a) $P(X < a) = 0.75$
b) $P(X > a) = 0.4$
c) $P(16 < X < a) = 0.3$

Example

Satellite dishes made using a particular manufacturing process have a diameter, *D* cm, which can be modelled using a normal distribution, $D \sim N(50, 2.5^2)$.

a) Given that 70% of the dishes are less than x cm, find x.
b) Find the interquartile range of the dish diameters.

3.3 - Inverse normal distributions

Practice

Given $X \sim N(80, 7^2)$, find the value of:

a) *a* such that $P(X > a) = 0.65$
b) *b* such that $P(75 < X < b) = 0.4$
c) *c* such that $P(c < X < 76) = 0.2$
d) Interquartile range of X.

Practice

If the IQ of a population is distributed using $X \sim N(100, 15^2)$.

a) Determine the IQ corresponding to top 30% of the population.
b) Determine the interquartile range of IQ.

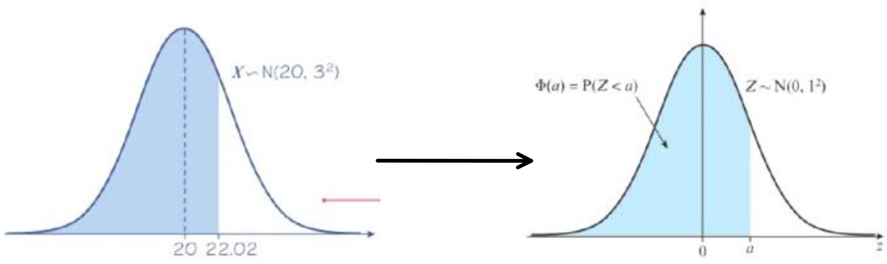
BF MATHS

3.4 - Standardise normal distribution

Notes

- It is usually quite hard to compare two normal distributions with different mean (μ) and standard deviation (σ). Standardising normal distribution allows easier comparision and easier calculation of probabilities.

- When a normal distribution is standardised,



- To standardise a normal distribution, we can code X with the formula:

Example

The random variable $X \sim N(50, 4^2)$. Write in terms of $\phi(z)$ for some value z:

a) $P(X < 53)$

b) $P(X \geq 55)$

3.4 - Standardise normal distribution

<u>Example</u>

The systolic blood pressure of an adult population, S mmHg, is modelled as a normal distribution with mean 127 and standard deviation 16. A medical research wants to study adults with blood pressures higher than the 95th percentile. Find the minimum blood pressure for an adult included in her study.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

<u>Example</u>

a) Determine $P(Z > -1.3)$
b) Determine $P(-2 < Z < 1)$
c) Determine the value of *a* such that $P(Z > a) = 0.7$
d) Determine the value of *a* such that $P(-a < Z < a) = 0.6$

BF MATHS

3.4 - Standardise normal distribution

<u>Example/Practice</u>

IQ is distributed with mean 100 and standard deviation 15. Using the percentage points table, determine the IQ corresponding to the

a) top 10% of people - $P(Z > z) = 0.1$
b) bottom 20% of people - $P(Z < z) = 0.2$
c) top 5% of people
d) Bottom 15% of people

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

<u>Practice</u>

If $X \sim N(100, 15^2)$, determine, in terms of Φ:

a) $P(X > 115)$
b) $P(77.5 < X < 112)$

<u>Practice</u>

Find the a such that:

a) $P(-a < Z < a) = 0.2$
b) $P(0 < Z < a) = 0.35$

3.5 - Finding μ and σ

Notes

- We can use standardised normal distribution to work out unknown mean (μ) or unknown standard deviation (σ).

Example

The random variable $X \sim N(\mu, 3^2)$. Given that $P(X > 20) = 0.2$, find the value of μ.

Example

A machine makes metal sheets with width, X cm, modelled as a normal distribution such that $X \sim N(50, \sigma^2)$.

a) Given that $P(X < 46) = 0.2119$, find the value of σ.
b) Find the 90th percentile of the widths.

BF MATHS

3.5 - Finding μ and σ

Example

The random variable $X \sim N(\mu, \sigma^2)$.

Given that $P(X > 35) = 0.025$ and $P(X < 15) = 0.1469$, find the value of μ and σ.

Practice (Ex. 3E)

1 The random variable $X \sim N(\mu, 5^2)$ and $P(X < 18) = 0.9032$.
Find the value of μ.

2 The random variable $X \sim N(11, \sigma^2)$ and $P(X > 20) = 0.01$.
Find the value of σ.

Exam Practice

Edexcel S1 May 2013 (R)

The time taken to fly from London to Berlin has a normal distribution with mean 100 minutes and standard deviation d minutes.

Given that 15% of the flights from London to Berlin take longer than 115 minutes,

(*b*) find the value of the standard deviation d.

(4)

BF MATHS

3.5 - Finding μ and σ

Exam Practice

Edexcel S1 Jan 2011

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.

(*c*) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$, find the value of μ and the value of σ. **(6)**

Exam Practice

Edexcel S1 Jan 2002

5. The duration of the pregnancy of a certain breed of cow is normally distributed with mean μ days and standard deviation σ days. Only 2.5% of all pregnancies are shorter than 235 days and 15% are longer than 286 days.

(*a*) Show that $\mu - 235 = 1.96\sigma$. **(2)**

(*b*) Obtain a second equation in μ and σ. **(3)**

(*c*) Find the value of μ and the value of σ. **(4)**

(*d*) Find the values between which the middle 68.3% of pregnancies lie. **(2)**

3.6 - Approximating binomial distribution

Notes

- Random variables with a binomial distribution are known to be discrete. This means that there are a countable number of outcomes that can occur in a binomial distribution. E.g. Binomial variable can only take a value of 3 or 4, but not any values in between.
- **If n is large (greater than 50) and p is close to 0.5**, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $Y \sim N(\mu, \sigma^2)$, where
  $\mu =$

  $\sigma =$

- Why does it only work when n is large?

  Because binomail involves

- Why does it only work when p is close to 0.5?

  Because

Binomial -> Normal Distribution

$X \sim B(30, 0.2) = Y \sim N($

$X \sim B(50, 0.4) = Y \sim N($

$X \sim B(10, 0.7) = Y \sim N($

Continuity Correction

- As mentioned above, binomial random variables only take discrete data whilst normal random variables could take any values (continuous). Hence, when we approximate binomial using normal distribution, we have to apply **continuity correction**.

E.g.

$P(X \leq 7) \approx$

$P(X < 20) =$

$P(X > 5) =$

$P(1 \leq X \leq 10) \approx$

$P(3 < X < 6) =$

$P(X = 3) \approx$

**

3.6 - Approximating binomial distribution

<u>Example</u>

The binomial random variable $X \sim B(150, 0.48)$ is approximated by the normal random variable $Y \sim N(72, 6.12^2)$. Use this approximation to find:

a) $P(X \leq 70)$
b) $P(80 \leq X \leq 90)$

<u>Example</u>

For a particular type of flower bulbs, 55% will produce yellow flowers. A random sample of 80 bulbs is planted.

a) Calculate the actual probability that there are exactly 50 yellow flowers.
b) Use a normal approximation to find a estimate that there are exactly 50 yellow flowers.
c) Hence determine the percentage error of the normal approximation for 50 yellow flowers.

3.6 - Approximating binomial distribution

Practice

Given the random variable $X \sim B(150, 0.45)$. Use a suitable approximation to estimate

a) $P(X \leq 60)$

b) $P(X > 75)$

Exam Practice

Edexcel S2 Jan 2004 Q3

The discrete random variable X is distributed $\mathrm{B}(n, p)$.

(*a*) Write down the value of p that will give the most accurate estimate when approximating the binomial distribution by a normal distribution. **(1)**

(*b*) Give a reason to support your value. **(1)**

(*c*) Given that $n = 200$ and $p = 0.48$, find $\mathrm{P}(90 \leq X < 105)$. **(7)**

BF MATHS

3.7 - Hypothesis testing with normal distribution

Understanding the concept

- What is the purpose of hypothesis testing?

  Ans: To decide whether the difference between population parameter and the sample statistic is due to *chance*.
  E.g. - The mean recorded of rotten oranges in 1000 is 3, you random selected 50 oranges and found one is rotten. Is that by chance or the mean recorded was actually wrong?

- Hence, when we perform hypothesis testing on the mean of a normally distributed random variable, we need to look at the *mean of a sample* taken from the whole population.

Imagine there are 10 children, one each of age between 0 to 9.

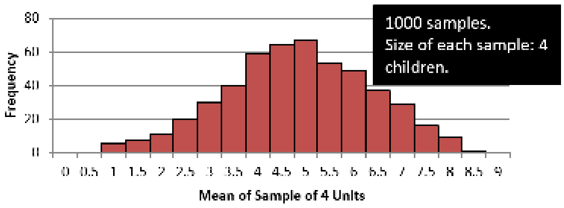
0	1	2	3	4	5	6	7	8	9

This is our population. The **population mean** (μ) = 4.5
Suppose we took a sample of 4 children, I work out the mean age and record it. I then took another sample of 4 children, do the same by working out the mean and record it…

					$\bar{x}$
Sample 1:	1	3	7	8	4.75
Sample 2:	6	2	0	9	4.25
	…				

Sample mean $\bar{x}$	Tally
4.00	
4.25	|
4.50	
4.75	|
5.00	

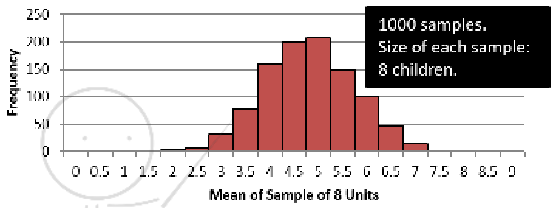
When I took 1000 samples, $\bar{X}$ naturally becomes *normally distributed*.



Mean
Now, if I work out the mean of the sample means, what is the answer close to?

Hence, the mean of $\bar{X}$ =

Standard deviation (how spread out the data is)
Instead of taking samples of 4 children, I took samples of 8 children. Then each mean I calculated would be more likely to centralise in the middle (as it is less likely to pick all the smaller ages, so less probability to get tiny mean values like 1.5).



Compared to samples of 4 children, the data is ____.

Hence, the standard deviation of $\bar{X}$ depends on the ________________________.

3.7 - Hypothesis testing with normal distribution

Notes

- For a random **sample** of size **n** taken from a random variable $X \sim \mathrm{N}(\mu, \sigma)$, the sample mean, $\bar{X}$, is normally distributed with

Example

Test the hypothesis at the stated level of significance:

$H_0: \mu = 15$, $H_1: \mu > 15$, $n = 40$, $\bar{x} = 16.5$, $\sigma = 3.5$ at 1% level.

Practice

Test the hypothesis at the stated level of significance:

$H_0: \mu = 100$, $H_1: \mu < 100$, $n = 36$, $\bar{x} = 98.5$, $\sigma = 5.0$ at 5% level.

Example

A certain company sells fruit juice in cartons. The amount of juice in a carton has a normal distribution with a standard deviation of 3ml.

The company claims that the mean amount of juice per carton, μ, is 60ml. A trading inspector has received complaints that the company is overstating the mean amount of juice per carton and wishes to investigate this complaint. The trading inspector takes a random sample of 16 cartons and finds that the mean amount of juice per carton is 59.1ml.

Using a 5% level of significance, and stating your hypotheses clearly, test whether or not there is evidence to uphold this complaint.

3.7 - Hypothesis testing with normal distribution

Notes

- Recall the *z-value* of a normally distributed random variable $X \sim N(\mu, \sigma^2)$ is

- The *z-value* of the sample mean of a normally distributed random variable $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ is

Example

A machine products bolts of diameter D where D has a normal distribution with mean 0.580 cm and standard deviation 0.015 cm. The machine is serviced and after the service a random sample of 50 bolts from the next production run is taken to see if the mean diameter of the bolts has changed from 0.580 cm. The distribution of the diameters of bolts after the service is still normal with a standard deviation of 0.015 cm.

a) Find, at the 1% level, the critical region for this test, stating your hypotheses clearly.

The mean diameter of the sample of 50 bolts is calculated to be 0.587 cm.

b) Comment on this observation in light of the critical region.

3.7 - Hypothesis testing with normal distribution

Practice

The random variable X has a normal distribution with mean μ and standard deviation 2. A random sample of 25 observations is taken and the sample mean $\bar{X}$ is calculated in order to test the null hypothesis $\mu = 7$ against the alternative hypothesis $\mu > 7$ using a 5% level of significance. Find the critical region for $\bar{X}$.

Exam Practice

Edexcel S3 June 2011 Q7a

Roastie's Coffee is sold in packets with a stated weight of 250 g. A supermarket manager claims that the mean weight of the packets is less than the stated weight. She weighs a random sample of 90 packets from their stock and finds that their weights have a mean of 248 g and a standard deviation of 5.4 g.

(*a*) Using a 5% level of significance, test whether or not the manager's claim is justified. State your hypotheses clearly.

(5)