Chapter 3

The Normal Distribution

Chapter Overview

1:: Characteristics of the Normal Distribution

What shape is it? What parameters does it have?

2:: Finding probabilities on a standard normal curve.

"Given that IQ is distributed as $X \sim N(100, 15^2)$, determine the probability that a randomly chosen person has an IQ above 130."

3:: Finding unknown means/standard deviations.

In Wales, 30% of people have a height above 1.6m. Given the mean height is 1.4m and heights are normally distributed, determine the standard deviation of heights.

:: Binomial → Normal Approximations

How would I approximate $X \sim B(10, 0.4)$ using a Normal distribution? Under what conditions can we make such an approximation?

5:: Hypothesis Testing

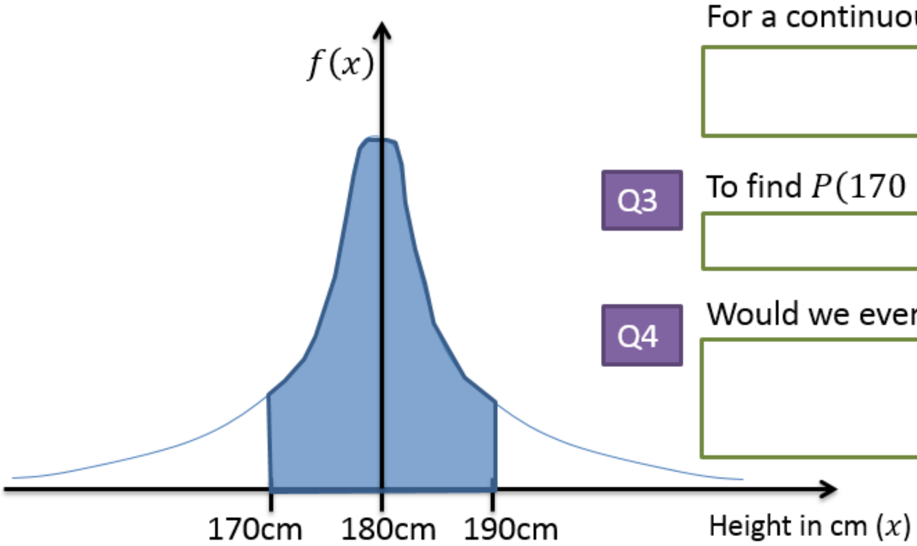
Specification

4.2	Understand and use the Normal distribution as a model; find probabilities using the Normal distribution	The notation $X \sim N(\mu, \sigma^2)$ may be used. Knowledge of the shape and the symmetry of the distribution is required. Knowledge of the probability density function is not required. Derivation of the mean, variance and cumulative distribution function is not required. Questions may involve the solution of simultaneous equations. Students will be expected to use their calculator to find probabilities connected with the normal distribution.
	Link to histograms, mean, standard deviation, points of inflection	Students should know that the points of inflection on the normal curve are at $x = \mu \pm \sigma$. The derivation of this result is not expected.
	and the binomial distribution.	Students should know that when n is large and p is close to 0.5 the distribution $B(n, p)$ can be approximated by $N(np, np[1 - p])$ The application of a continuity correction is expected.

What does it look like?

The following shows what the probability distribution might look like for a random variable X, if X is the height of a randomly chosen person.

Normal Distribution Q & A



Q1 For a Normal Distribution to be used, the variable has to be:

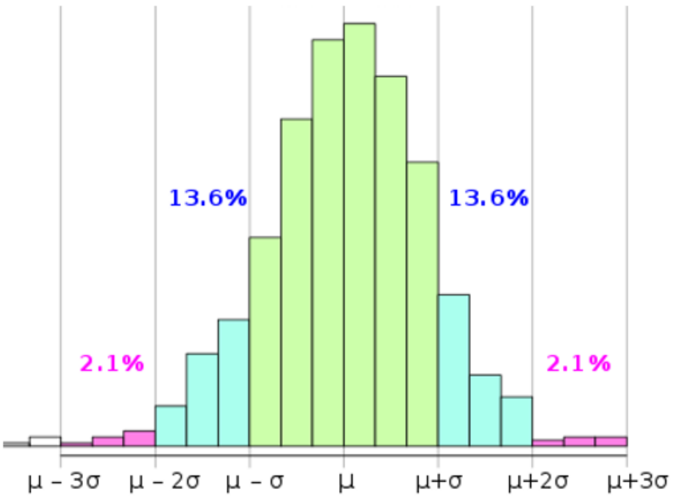
Q2 With a discrete variable, all the probabilities had to add up to 1.
For a continuous variable, similarly:

Q3 To find $P(170 < X < 190)$, we could:

Q4 Would we ever want to find $P(X = 200)$ say?

The 68-95-99.7 rule

<u>You need to memorise this!</u>



The histogram above is for a quantity which is approximately normally distributed.

$\approx 68\%$

$\approx 95\%$

$\approx 99.7\%$

For practical purposes we consider all data to lie within $\mu \pm 5\sigma$

Examples

[Textbook] The diameters of a rivet produced by a particular machine, X mm, is modelled as $X \sim N(8, 0.2^2)$. Find:

a) $P(X > 8)$

b) $P(7.8 < X < 8.2)$

Test Your Understanding

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) $P(X > 115)$

Getting normal values from your calculator

IQ is distributed using $X \sim N(100, 15^2)$. Find

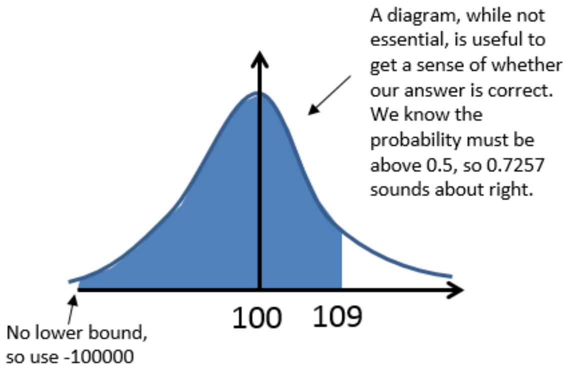
(a) $P(X < 109)$

(b) $P(X \geq 93)$

(c) $P(110 < X < 120)$

(d) $P(X < 80 \text{ or } X > 106)$

1. Press MODE.
2. Choose DISTRIBUTION (option 7)
3. Choose Normal CD (i.e. "Cumulative Distribution")
4. Since the lower value is effectively $-\infty$, use any value at least 5σ below the mean (-100000 will do!). Press = after each value.
5. Put the upper value as 109.
6. Set $\sigma = 15$ and $\mu = 100$
7. You should obtain $P(X < 109) = 0.7257\ (4dp)$



Getting normal values from your calculator

IQ is distributed using $X \sim N(100, 15^2)$. Find
(a) $P(X < 109)$
(b) $P(X \geq 93)$
(c) $P(110 < X < 120)$
(d) $P(X < 80 \; or \; X > 106)$

Test Your Understanding

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

Inverse Normal Distribution

We now know how to use a calculator to value of the variable to obtain a probability. But we might want to do the reverse: given a probability of being in a region, how do we find the value of the boundary?

[Textbook] $X \sim N(20, 3^2)$. 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. $P(16 < X < a) = 0.3$

a

1. MODE → Distributions
2. Choose 'Inverse Normal'.
3. Put the area as 0.75 (this is the area up to the a value to determine). Put $\mu = 20$ and $\sigma = 3$.
4. You should get **22.0235**.

b

c

DRAW A SKETCH!

Further Example

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

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

Test Your Understanding

$X \sim N(80, 7^2)$. Using your calculator,

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

Standard Normal Distribution

✎ Z is the

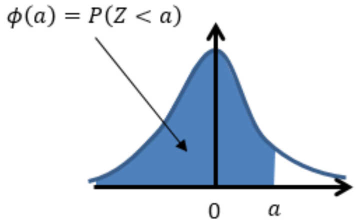
If again we use IQ distributed as
$X \sim N(100, 15^2)$ then:

IQ	Z
100	
130	
85	
165	
62.5	

$$Z = \frac{X - \mu}{\sigma}$$

and $Z \sim N(0, 1^2)$. Z is known as a **standard** normal distribution.

Standard Normal Distribution



✐ $\Phi(a) = P(Z < a)$ is the cumulative distribution for the standard normal distribution. The values of $\Phi(a)$ can be found in a z-table.

This is a traditional z-table in the old A Level syllabus (but also found elsewhere). You no longer get given this and are expected to use your calculator.

This is from the new formula booklet. This is sometimes known as a 'reverse z-table', because you're looking up the z-value for a probability. Beware: p here it the probability of **exceeding** z rather than being up to z. Let's use it…

THE NORMAL DISTRIBUTION FUNCTION

The function tabulated below is $\Phi(z)$, defined as $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.02	0.9783
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.04	0.9793
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.06	0.9803
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.08	0.9812
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.10	0.9821
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.12	0.9830
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.14	0.9838
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.16	0.9846
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.18	0.9854
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.20	0.9861

Percentage Points of The Normal Distribution

The values z in the table are those which a random variable $Z \sim N(0, 1)$ exceeds with probability p; that is, $P(Z > z) = 1 - \Phi(z) = p$.

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

Examples

[Textbook] The random variable $X \sim N(50, 4^2)$. Write in terms of $\Phi(z)$ for some value of z.
(a) $P(X < 53)$ (b) $P(X \geq 55)$

a

b

[Textbook] 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.

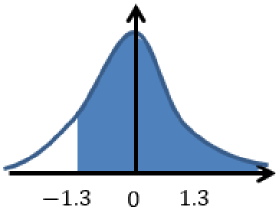
p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

Further Examples

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

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

a



Test Your Understanding

1 IQ is distributed with mean 100 and standard deviation 15. Using an appropriate table, determine the IQ corresponding to the
(a) top 10% of people.
(b) bottom 20% 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

a

b

2 If $X \sim N(100, 15^2)$, determine, in terms of Φ:
(a) $P(X > 115)$
(b) $P(77.5 < X < 112)$

a

b

3 Find the a such that:
(a) $P(-a < Z < a) = 0.2$
(b) $P(0 < Z < a) = 0.35$

a

b

Missing μ and σ

In the last section, you may have thought, "what's the point of standardising to Z when I can just use the DISTRIBUTION mode on my calculator?"
Fair point, but both forward and reverse normal lookups on the calculator **required you to specify μ and σ.**

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

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

[Textbook] 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.

When both are missing

If both μ and σ are missing, we end up with simultaneous equations which we must solve.

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 $\mathrm{P}(Y < 160) = 0.99$ and $\mathrm{P}(Y > 152) = 0.90$, find the value of μ and the value of σ.
(6)

Test Your Understanding

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)

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)

Approximating a Binomial Distribution

If we're going to use a normal distribution to approximate a Binomial distribution, it makes sense that we set the mean and standard deviation of the normal distribution to match that of the original binomial distribution:

$$\mu = \boxed{}$$
$$\sigma = \boxed{}$$

✎ If n is large and p close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where

$$\mu =$$
$$\boldsymbol{\sigma} =$$

Quickfire Questions:

$X \sim B(10{,}0.2) \quad \rightarrow \quad Y \sim \boxed{}$

$X \sim B(20{,}0.5) \quad \rightarrow \quad Y \sim \boxed{}$

$X \sim B(6, 0.3) \quad \rightarrow \quad Y \sim \boxed{}$

Continuity Corrections

Examples

Discrete	Continuous
$P(X \leq 7)$	
$P(X < 10)$	
$P(X > 9)$	
$P(1 \leq X \leq 10)$	
$P(3 < X < 6)$	
$P(3 \leq X < 6)$	
$P(3 < X \leq 6)$	
$P(X = 3)$	

Full Example

[Textbook - Edited] 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 flowers.

(b) Use a normal approximation to find a estimate that there are exactly 50 flowers.

(c) Hence determine the percentage error of the normal approximation for 50 flowers.

Test Your Understanding

Edexcel S2 Jan 2004 Q3

The discrete random variable X is distributed $B(n, p)$.

(*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 $P(90 \leq X < 105)$. **(7)**

Hypothesis Testing on the Sample Mean

Examples

[Textbook] 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.

Finding the critical region

[Textbook] 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.

Test Your Understanding

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)

Conditional Probabilities

This is not in the textbook. But given the recent Chapter 2 on Conditional Probabilities and the fact that the type of question below occurred frequently in S1 papers, it seems worthwhile to cover!

Edexcel S1 May 2014(R) Q4

The time, in minutes, taken to fly from London to Malaga has a normal distribution with mean 150 minutes and standard deviation 10 minutes.

The time, X minutes, taken to fly from London to another city has a normal distribution with mean μ minutes.

Given that $\mathrm{P}(X < \mu - 15) = 0.35$

(*c*) find $\mathrm{P}(X > \mu + 15 \mid X > \mu - 15)$. **(3)**

Test Your Understanding

Edexcel S1 Jan 2013 Q4a,c

The length of time, L hours, that a phone will work before it needs charging is normally distributed with a mean of 100 hours and a standard deviation of 15 hours.

(*a*) Find $P(L > 127)$.

(3)

Alice is about to go on a 6 hour journey. Given that it is 127 hours since Alice last charged her phone,

(*c*) find the probability that her phone will not need charging before her journey is completed.

(4)

One Last Toughie…

IQ is distributed as $X \sim N(100, 15^2)$.
A person is considered to be 'pretty smart' if their IQ is at least 130.
Determine the median IQ of 'pretty smart' people.