Significance tests use data from samples to test hypotheses.

You will use data on successful applications for courses in higher education to answer questions about proportions, for example, to test whether equal proportions of male and female applicants are accepted.

Information sheet A Number of applicants

The simulated data given below give the total number of applicants for courses of higher education at a sample of universities and colleges. Actual data can be accessed from the UCAS website www.ucas.co.uk.

Gender

	2000		2010	
Gender	**Applicants**	**Accepts**	**Applicants**	**Accepts**
Male	35,117	5,273	45,455	7,062
Female	37,785	5,521	54,030	8,353

Age

	Number accepted	
Age	**2000**	**2010**
18 and under	5,821	7,614
19	2,727	4,028
20	859	1,433
21	401	653
22	248	405
23	126	256
24	86	164
25 - 29	253	476
30 - 39	201	284
40 and over	72	102
Total	**10,794**	**15,415**

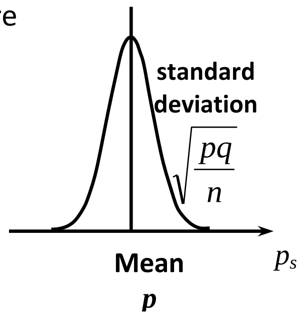
Domicile

	2000		2010	
Domicile	**Applications**	**Accepts**	**Applications**	**Accepts**
UK	67,699	10,220	90,540	14,053
EU(not UK)	1,856	191	3,297	407
Non EU	3,347	383	5,648	955
Total	**72,902**	**10,794**	**99,485**	**15,415**

Information sheet B Testing a proportion

Distribution of a sample proportion

If large samples of size n are taken from a population in which there are a proportion p with a certain attribute, then the distribution of sample proportions, p_s, is approximately normal with mean p and standard deviation $\sqrt{\frac{pq}{n}}$ where $q = 1 - p$.



Think about…

Why is it important that samples are large?

Summary of method for testing a proportion

To test whether the proportion of a population has a value p:

Null Hypothesis H_0 : population proportion, p = value suggested

Alternative Hypothesis H_1 : $p \neq$ value suggested (two-tail test)
or $p <$ value suggested or $p >$ value suggested (one-tail test)

Test statistic $z = \dfrac{p_s - p}{\sqrt{\frac{pq}{n}}}$ where p_s is the proportion in a sample of size n

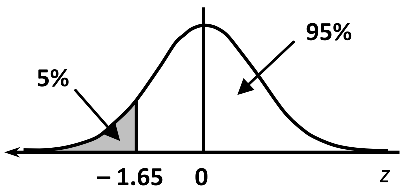
and $q = 1 - p$

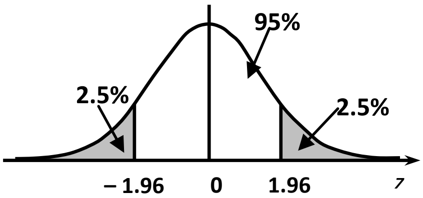
> **Think about…**
> Explain the formula for the test statistic.

Compare the test statistic with critical values of z.

Critical values:

Test type	Significance level	Critical values
one-tail test	5%	1.65 or −1.65
	1%	2.33 or −2.33
two-tail test	5%	± 1.96
	1%	± 2.58





If the test statistic is **in the critical region** (that is, a tail of the distribution), **reject the null hypothesis in favour of the alternative**.

If the test statistic is **not in the critical region**, **accept the null hypothesis**.

Testing a proportion: Example

In 2010, a newspaper article said that the proportion of people accepted on higher education courses who were over 20 years old had reached 16%.

Using 2010 data to test this percentage:

H_0 : population proportion, p = 0.16

H_1 : p < 0.16 (one-tail test)

Think about…

Why is a one-tail test used here rather than a two-tail test?

Test statistic $z = \dfrac{p_s - p}{\sqrt{\dfrac{pq}{n}}}$ where p = 0.16 and q = 1 – 0.16 = 0.84

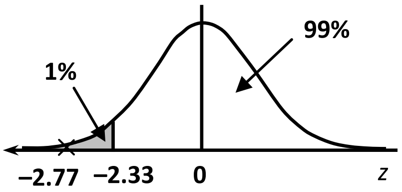
From the 2010 data, $p_s = \dfrac{2340}{15415}$ = 0.1518 and n = 15 415

So $z = \dfrac{0.1518 - 0.16}{\sqrt{\dfrac{0.16 \times 0.84}{15415}}} = -2.77$

For a one-tail 1% significance test, the critical value is –2.33.

The test statistic is in the critical region (less than the critical value).

The result is **significant** at the 1% level.



So reject the null hypothesis and accept the alternative.

The test has provided strong evidence that the proportion of people accepted on higher education courses who were over 20 years old had not reached 16%.

Think about…

Explain the reasoning behind this conclusion.

Information sheet C Testing the difference between proportions

Summary of method for testing the difference between proportions

To test the difference between proportions:

Null hypothesis,	$H_0 : p_A = p_B$	$(p_A - p_B = 0)$
Alternative hypothesis,	$H_1 : p_A \neq p_B$	(two-tail test)
	$p_A < p_B$	(one-tail test)
	$p_A > p_B$	(one-tail test)

The test statistic is $z = \dfrac{p_{SA} - p_{SB}}{\sqrt{pq\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)}}$

where p_{SA} and p_{SB} are the proportions from samples of size n_A and n_B.

p, the best estimate of the population proportion, is calculated from:

$$p = \frac{\text{Total number of items with attribute}}{\text{Total number of items in samples}} \quad \text{and } q = 1 - p$$

Think about…

Explain the formula for the test statistic.

Compare the test statistic with critical values of z.

Critical values:

Test type	**Significance level**	**Critical values**
one-tail test	5%	1.65 or −1.65
	1%	2.33 or −2.33
two-tail test	5%	± 1.96
	1%	± 2.58

If the test statistic is **in the critical region,**
reject the null hypothesis in favour of the alternative.

If the test statistic is **not in the critical region**, **accept the null hypothesis**.

Testing the difference between proportions: Example

Using 2010 data to test whether the proportion of males that were accepted is equal to the proportion of females that were accepted:

$H_0 : p_M = p_F \quad (p_M - p_F = 0)$
$H_1 : p_M \neq p_F \quad$ (two-tail test)

The test statistic is $z = \dfrac{p_{SM} - p_{SF}}{\sqrt{pq\left(\dfrac{1}{n_M} + \dfrac{1}{n_F}\right)}}$

In the 2010 sample, 7062 out of 45 455 applications from males were accepted, and 8353 out of 54 030 applications from females were accepted.

So $\quad p_{SM} = \dfrac{7062}{45\,455} = 0.155\,362, \quad p_{SF} = \dfrac{8353}{54\,030} = 0.154\,599,$

$\quad n_M = 45\,455$ and $n_F = 54\,030$

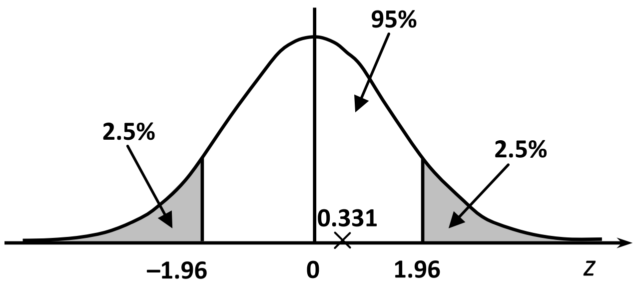
p, the best estimate of the population proportion, is calculated from

$$p = \frac{7062 + 8353}{45\,455 + 54\,030} = 0.154\,948 \quad \text{and} \quad q = 1 - 0.154\,948 = 0.845\,052$$

Using these values, the test statistic is given by:

$$z = \frac{0.155\,362 - 0.154\,599}{\sqrt{0.154\,948 \times 0.845\,052\left(\dfrac{1}{45\,455} + \dfrac{1}{54\,030}\right)}} = 0.331$$

For a two-tail test at the 5% level, the critical values of z are ± 1.96, so this value of z is not significant at the 5% level.



There is no significant difference between the proportion of males that were accepted and the proportion of females that were accepted.

Think about…

Explain the reasoning behind this conclusion.

Try this

1 Consider the data on Information sheet A.
Write a list of hypotheses you think could be tested using these data.

2 Choose some of the hypotheses you have listed in question 1.
Carry out significance tests on these hypotheses.
At least one of your tests should be of a proportion.
At least one of your tests should be of the difference between proportions.

Reflect on your work

- What are the mean and standard deviation of the distribution of a sample proportion?
- Describe the steps in a significance test for a proportion.
- Describe the steps in a significance test for the difference between proportions.
- When should you use a one-tail test and when a two-tail test?
- Would you be more confident in a significant result from a 5% significance test or a 1% significance test? Explain why.