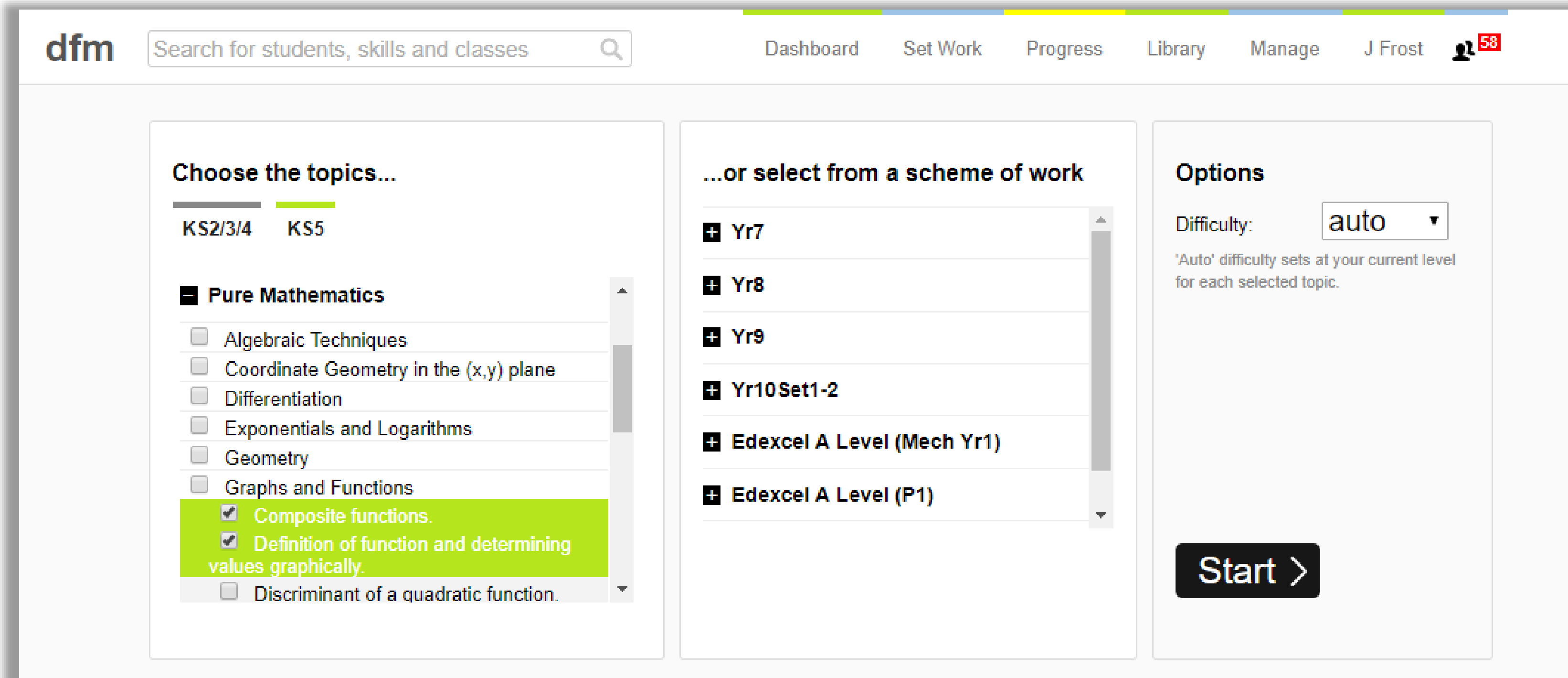
Stats1 Chapter 7 :: Hypothesis Testing

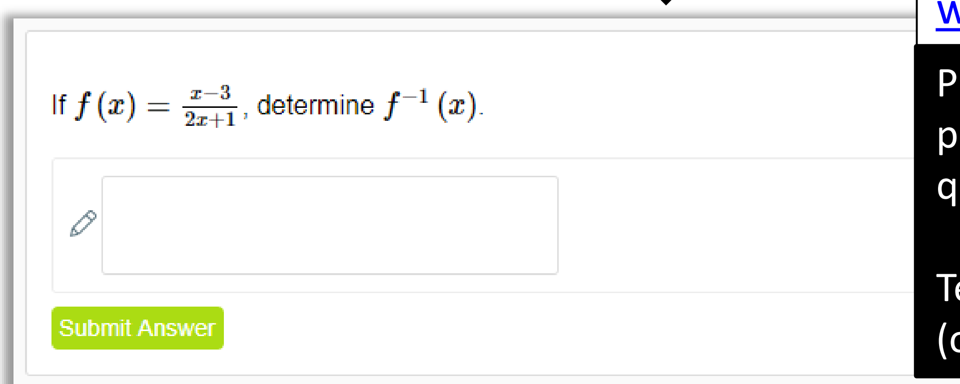
jfrost@tiffin.kingston.sch.uk

www.drfrostmaths.com
@DrFrostMaths

Last modified: 27th March 2019

Use of DrFrostMaths for practice





Register for **free** at:
www.drfrostmaths.com/homework

Practise questions by chapter, including past paper Edexcel questions and extension questions (e.g. MAT).

Teachers: you can create student accounts (or students can register themselves).

Experimental

i.e. Dealing with collected data.

Chp1: Data Collection

Methods of sampling, types of data, and populations vs samples.

Chp2: Measures of Location/Spread

Statistics used to summarise data, including mean, standard deviation, quartiles, percentiles. Use of linear interpolation for estimating medians/quartiles.

Chp3: Representation of Data

Producing and interpreting visual representations of data, including box plots and histograms.

Chp4: Correlation

Measuring how related two variables are, and using linear regression to predict values.

Theoretical

Deal with probabilities and modelling to make inferences about what we ‘expect’ to see or make predictions, often using this to reason about/contrast with experimentally collected data.

Chp5: Probability

Venn Diagrams, mutually exclusive + independent events, tree diagrams.

Chp6: Statistical Distributions

Common distributions used to easily find probabilities under certain modelling conditions, e.g. binomial distribution.

Chp7: Hypothesis Testing

Determining how likely observed data would have happened ‘by chance’, and making subsequent deductions.

What is Hypothesis Testing?

To get a flavour of hypothesis testing, discuss how you would approach the following problem:

10% of the world's population are left-handed. On my holiday to Hawaii, I want to establish if the proportion of left-handed people in Hawaii is greater than the world average. I have a table of 20 people as my sample. I need to ensure any result I get is **statistically significant**.

1. On the table I count how many are left-handed. Suppose 5 people are.
2. If I were to assume that 10% of people were left-handed, I calculate (using Binomial distribution) the probability that 5 people (or more) would be left-handed by chance.
3. If this probability of at least 5 people being left-handed by chance is sufficiently low (say less than 5%), I conclude that that the proportion of Hawaiians who are left-handed is greater than the world proportion.

✎ A hypothesis is a statement made about the value of a **population parameter** that we wish to test by collecting evidence in the form of a sample.
(In this case, the population parameter is the proportion p of Hawaii's population who are left-handed, because this is the population under consideration. There are two possible hypotheses:)

✎ The **null hypothesis**, H_0 is the default position, i.e. that nothing has changed, unless proven otherwise.
(In this case, it's that Hawaii's proportion of left-handed people is consistent with the world proportion, i.e. 10%, and that seeing 5 left-handed people was a fluke.)

✎ The **alternative hypothesis**, H_1, is that there has been some change in the population parameter.
(i.e. Seeing 5 left-handed people wasn't a fluke – it was because the Hawaii's proportion of left-handed people is more than 10%)

✎ In a hypothesis test, the evidence from the sample is a **test statistic**.
(In this case, we've taken a sample by counting let-handed people, and found the test statistic of "number of lefties was 5")

✎ The **level of significance** α is the maximum probability where we would reject the null hypothesis. This is usually 5% or 1%.

What is Hypothesis Testing?

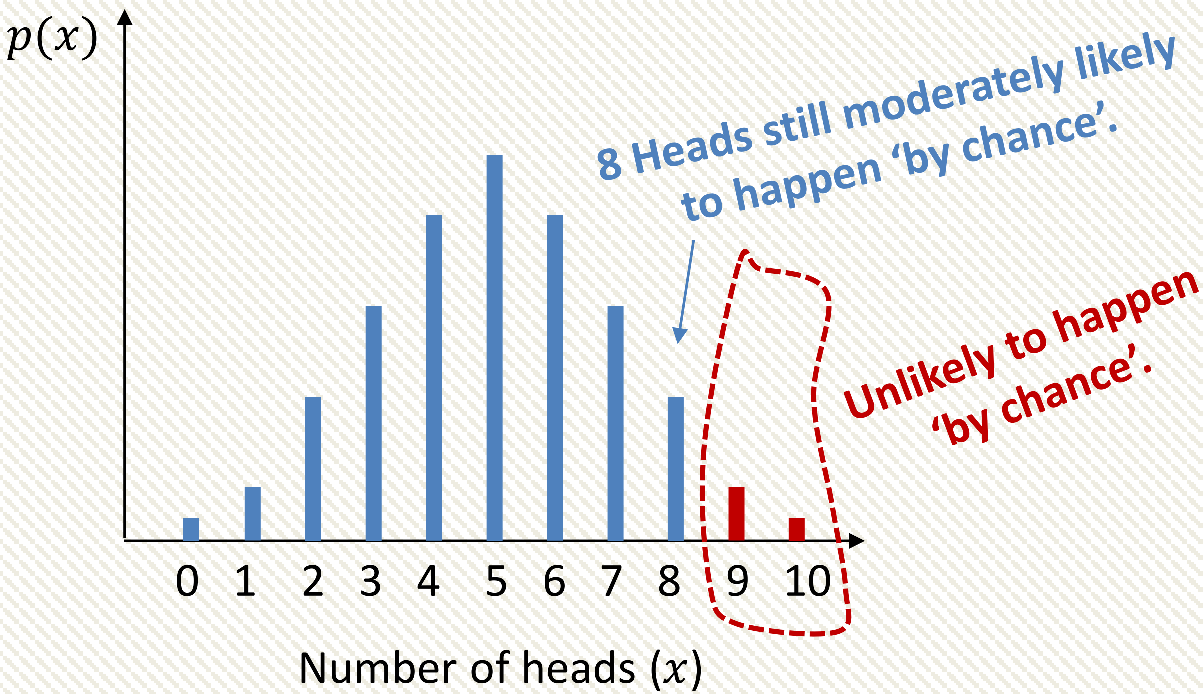
Hypothesis testing in a nutshell then is:

1. We have some hypothesis we wish to see if true (proportion of left-handed people in Hawaii is more than global average), so…
2. We collect some sample data (giving us our test statistic) and…
3. If that data is sufficiently unlikely to have emerged 'just by chance', then we conclude that our (alternative) hypothesis is correct.

What is Hypothesis Testing?

I throw a coin 10 times. For what numbers of heads might you conclude that the coin is biased towards heads? Why?

Our intuition is a large number of heads of low number of heads, far away from the 'expected' number of 5 heads out of 10. There is because the probability of this number of heads occurring 'by chance' (i.e. if the coin was in fact fair) is low.



What is Hypothesis Testing?

I throw a coin 10 times. For what numbers of heads might you conclude that the coin is biased towards heads? Why?

✎ A hypothesis is a statement made about the value of a **population parameter** that we wish to test by collecting evidence in the form of a sample.

✎ The **null hypothesis**, H_0 is the default position, i.e. that nothing has changed, unless proven otherwise.

✎ The **alternative hypothesis**, H_1, is that there has been some change in the population parameter.

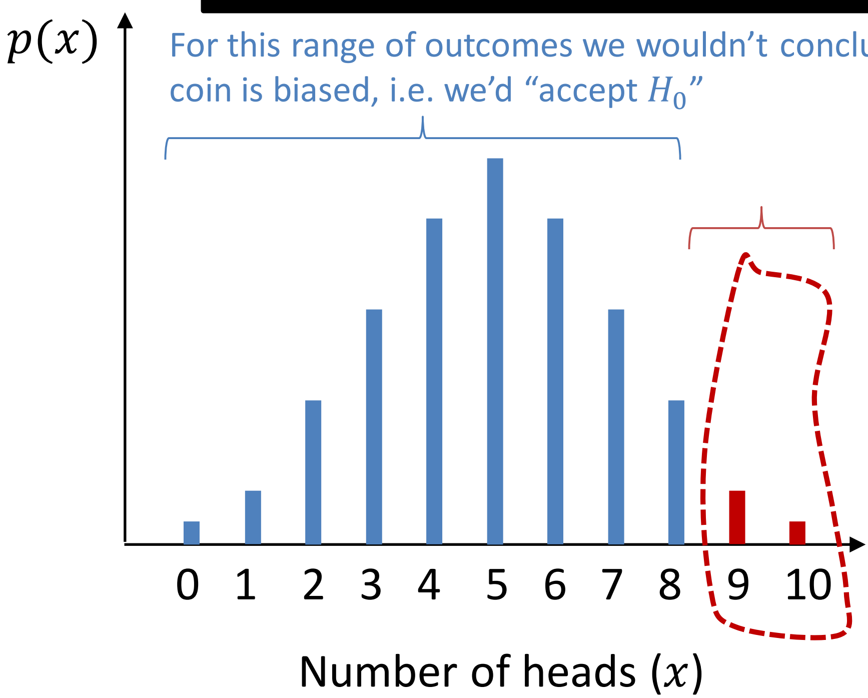
In this context…

?

?

?

For this range of outcomes we wouldn't conclude the coin is biased, i.e. we'd "accept H_0"



For this range of outcomes we'd conclude that this number of heads was too unlikely to happen by chance, and hence reject H_0 (i.e. that coin was fair) and accept H_1 (i.e. that coin was biased).

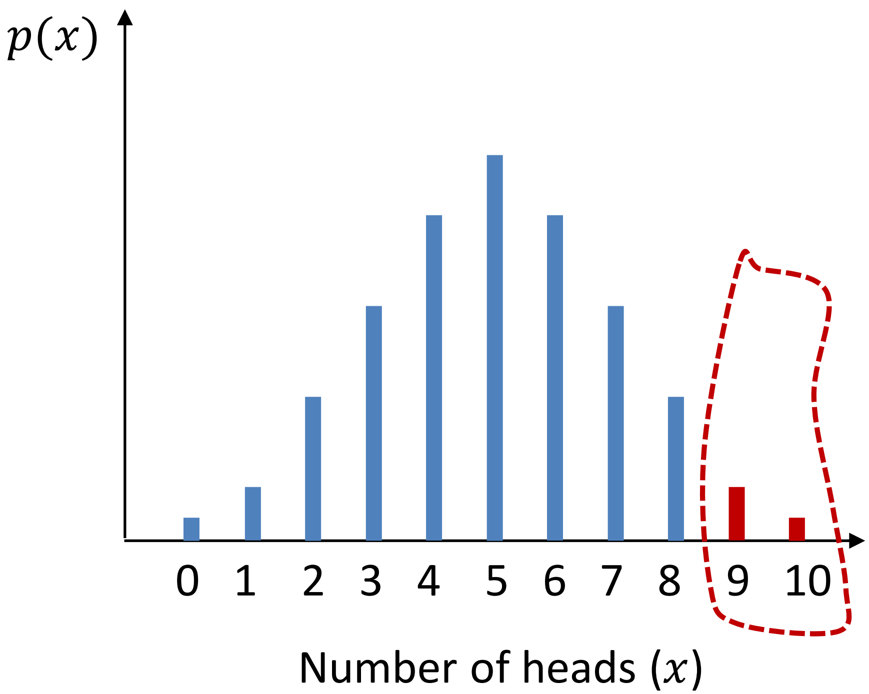
What is Hypothesis Testing?

I throw a coin 10 times. For what numbers of heads might you conclude that the coin is biased towards heads? Why?

In this context…

?

In a hypothesis test, the evidence from the sample is a **test statistic**.



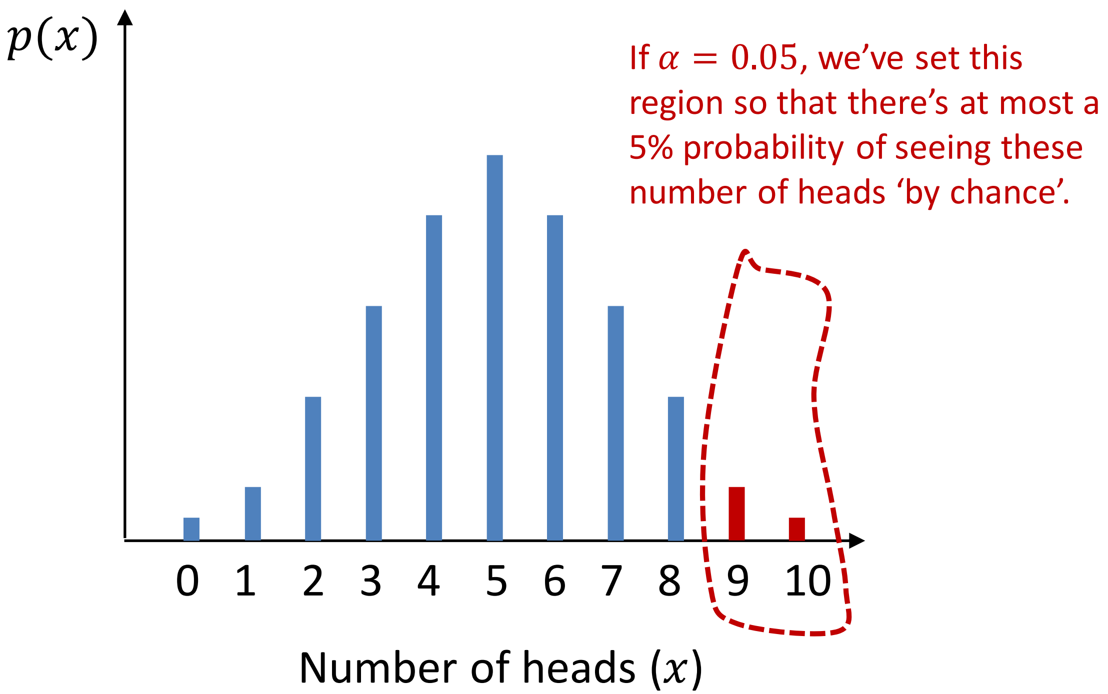
What is Hypothesis Testing?

I throw a coin 10 times. For what numbers of heads might you conclude that the coin is biased towards heads? Why?

In this context...

?

✎ The **level of significance α** is the maximum probability where we would reject the null hypothesis.
This is usually 5% or 1%.



What is Hypothesis Testing?

Hypothesis testing in a nutshell then is:

1. We have some hypothesis we wish to see if true (e.g. coin is biased towards heads), so…
2. We collect some sample data by throwing the coin (giving us our 'test statistic') and…
3. If that number of heads (or more) is sufficiently unlikely to have emerged 'just by chance', then we conclude that our (alternative) hypothesis is correct, i.e. the coin is biased.

Null Hypothesis and Alternative Hypothesis

[Textbook] John wants to see whether a coin is unbiased or whether **it is biased towards coming down heads**. He tosses the coin 8 times and counts the number of times X, it lands head uppermost.

We said that our two hypotheses are about the population parameter.

Suppose p is the probability of a coin landing heads.

Null hypothesis:	?
Alternative hypothesis:	?

Under the **null hypothesis** H_0, we **assume that the population parameter is correct**, in this case, that it is a normal coin and the probability of Heads is 0.5

Under the **alternative hypothesis** H_1, there has been an underlying change in the population parameter, in this case that the coin is actually biased towards Heads.

The latter is known as a '**one-tailed test**' because we're saying the coin is biased one way or the other (i.e. $p > 0.5$ or $p < 0.5$).
But we could also have had the hypothesis 'the coin is biased (either way)', i.e. $p \neq 0.5$. This is known as a **two-tailed test**.

Further Example

[Textbook] An election candidate believes she has the support of 40% of the residents in a particular town. A researcher wants to test, at the 5% significance level, whether the candidate is over-estimating her support. The researcher asks 20 people whether they support the candidate or not. 3 people say they do.

a) Write down a suitable test statistic.
b) Write down two suitable hypotheses.
c) Explain the condition under which the null hypothesis would be rejected.

a

?

For a hypothesis test involving the binomial distribution, the test statistic is always the **count of successes**.

b

?

The alternative hypothesis is that the candidate is **overestimating** her support, so we're interested where **less than 40%** support them (more than 40% would not undermine the candidate's claim).

c

?

This is the hard bit!
We always calculate the probability of seeing this outcome **<u>or more extreme</u>** (in this case, 'more extreme' meaning even fewer the 3 people, because this takes us even further from the expected number of people out of the 20 (i.e. 8) who would support them.
The "$p = 0.4$" bit is because, as discussed before, we calculate the probability of seeing the observed outcome of 3 people (or more extreme) if it occurred **purely by chance** (the null hypothesis), i.e. if the candidate **did** have 40% support.

Test Your Understanding

In the UK, 5% of students turn up late to school each day. Mr Hameed wishes to determine, to a 10% significance level if his school, Piffin School, has a problem with attendance. He stands at the front gate one day and finds that 6 of the 40 students who pass him are late.

a) Write down a suitable test statistic.
b) Write down two suitable hypotheses.
c) Explain the condition under which the null hypothesis would be rejected.

a ?

b ?

If there is indeed a <u>problem</u> with attendance (i.e. the school is not the norm) then we expect the proportion p of students late at Tiffin to be higher than 5%.

c ?

Just to reiterate again what's going on here:

a) We assume that the school is the norm (i.e. 5% of students are late), and calculate the probability that 6 or more students would be late under this assumption.
b) If the probability of this happening by chance is sufficiently unlikely (less than the 10% significance level), we conclude that it probably isn't just by chance, and the school's lateness rate is worse than 5%.

Doing a full one-tailed hypothesis test

We've done various bits of a hypothesis test, and haven't actually properly conducted one yet. Let's do an example!

John tosses a coin 8 times and it comes up heads 6 times. He claims the coin is **biased towards heads**. With a significance level of 5%, test his claim.

?

?

?

?

STEP 1: Define test statistic X (stating its distribution), and the parameter p.

STEP 2: Write null and alternative hypotheses.

STEP 3: Determine probability of observed test statistic (or 'more extreme'), assuming null hypothesis.
i.e. Determine probability we'd see this outcome just by chance.

STEP 4: Two-part conclusion:
1. Do we reject H_0 or not?
2. Put <u>in context of original problem</u>.

C.D.F. Binomial table: $p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
3	0.3633
4	0.6367
5	0.8555
6	0.9648
7	0.9961

NEW TO A LEVEL 2017: The probability of 'the observed value or more extreme' is known as the p-value.

Exercise 7A

Pearson Pure Mathematics Year 1/AS
Pages 100-101

Critical Regions and Values

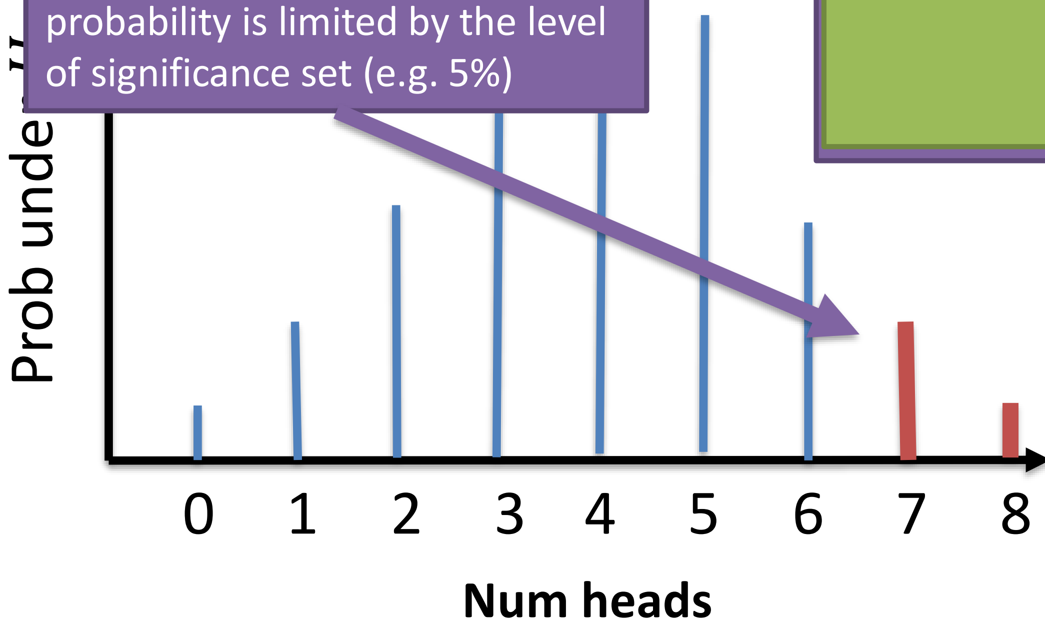
John wants to see whether a coin is unbiased or whether **it is biased towards coming down heads**. He tosses the coin 8 times and counts the number of times X, it lands head uppermost. **What values would lead to John's hypothesis being rejected?**

As before, we're interested how likely a given outcome is likely to happen 'just by chance' under the null hypothesis (i.e. when the coin is not biased).

However, there are values which collectively form a range of 'extreme values' where it would be unlikely that the coin would be unbiased. Their combined probability is limited by the level of significance set (e.g. 5%)

The probability of getting exactly 5 heads is only 22%, which is more likely to not happen than to happen. If we saw this number of heads, why would it not be sensible to think the coin is biased?

?



Critical Regions and Values

John wants to see whether a coin is unbiased or whether **it is biased towards coming down heads**. He tosses the coin 8 times and counts the number of times X, it lands head uppermost. **What values would lead to John's hypothesis being rejected**, if the significance level was 5%?

What's the probability that we would see **6 heads**, or an **even more extreme value**? Is this sufficiently unlikely to support John's claim that the coin is biased?

?

What's the probability that we would see **7 heads**, or an **even more extreme value**?

?

C.D.F. Binomial table:
$p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
3	0.3633
4	0.6367
5	0.8555
6	0.9648
7	0.9961

Critical Regions and Values

John wants to see whether a coin is unbiased or whether **it is biased towards coming down heads**. He tosses the coin 8 times and counts the number of times X, it lands head uppermost. **What values would lead to John's hypothesis being rejected**, if the significance level was 5%?

✎ The **critical region** is the range of values of the test statistic that would lead to you rejecting H_0

?

✎ The value(s) on the boundary of the critical region are called **critical value(s)**.

Critical value:

?

C.D.F. Binomial table:
$p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
3	0.3633
4	0.6367
5	0.8555
6	0.9648
7	0.9961
8	1

Quickfire Critical Regions

Determine the critical region when we throw a coin where we're trying to establish if there's the specified bias, given the specified number of throws, when the level of significance is 5%.

Coin thrown 5 times. Trying to establish if biased towards heads.

$p = 0.5, n = 5$

x	$P(X \leq x)$
0	0.0312
1	0.1875
2	0.5000
3	0.8125
4	0.9688

Critical region:

?

Coin thrown 10 times. Trying to establish if biased towards heads.

$p = 0.5, n = 10$

x	$P(X \leq x)$
0	0.0010
1	0.0107
2	0.0547
…	…
7	0.9453
8	0.9893
9	0.9990

Critical region:

?

Coin thrown 10 times. Trying to establish if biased towards tails.

$p = 0.5, n = 10$

x	$P(X \leq x)$
0	0.0010
1	0.0107
2	0.0547
…	…
7	0.9453
8	0.9893
9	0.9990

Critical region:

?

Fro Reminder: At the positive tail, use the value AFTER the first that exceeds 95% (100 - 5).

At the negative tail, we just use the first value that goes under the significance level.

Actual Significance Level

John wants to see whether a coin is unbiased or whether **it is biased towards coming down heads**. He tosses the coin 8 times and counts the number of times X, it lands head uppermost. **What values would lead to John's hypothesis being rejected**, if the significance level was 5%?

We saw earlier that the critical region was $X \geq 7$, i.e. the region in which John would reject the null hypothesis (and conclude the coin was biased).

We ensured that $P(X \geq 7)$ was less than the significance level of 5%.
But what actually is $P(X \geq 7)$?

$\boldsymbol{P(X \geq 7) =}$?

This is known as the actual significance level, i.e. the probability that we're in the critical region. We expected this to be less than, but close to, 5%.

The **actual significance level** is the actual probability of being in the critical region.

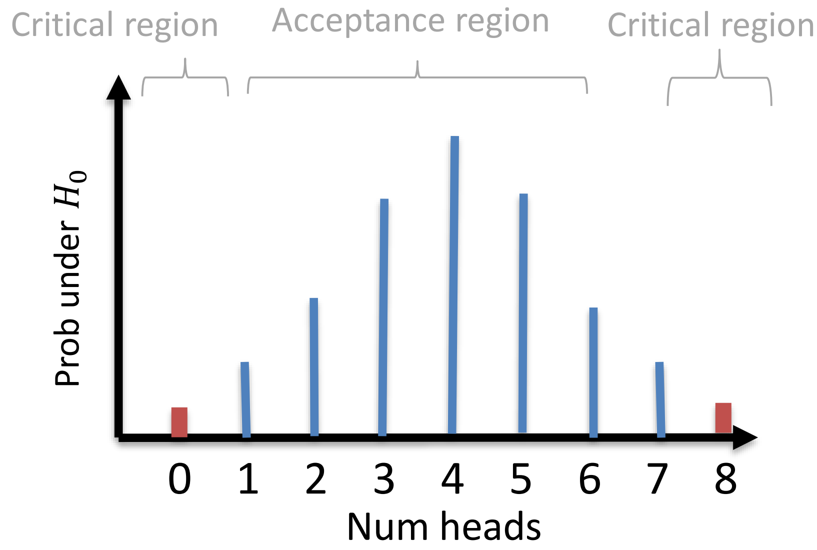
C.D.F. Binomial table:
$p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
3	0.3633
4	0.6367
5	0.8555
6	0.9648
7	0.9961
8	1

Two-tailed test

Suppose I threw a coin 8 times and was now interested in how may heads would suggest it was a **biased coin** (i.e. either way!). How do we work out the critical values now, with 5% significance?

?



Critical region at positive tail:

?

Critical region at negative tail:

?

C.D.F. Binomial table:
$p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
…	…
6	0.9648
7	0.9961
8	1

Test Your Understanding

A random variable X has binomial distribution $B(40, p)$. A single observation is used to test $H_0: p = 0.25$ against $H_1: p \neq 0.25$.

The $\neq$ indicates bias either way, i.e. two-tailed.

a) Using the 2% level of significance, find the critical region of this test. The probability in each tail should be as close as possible to 0.01.

b) Write down the actual significance level of the test.

This means you find the closest to 0.01 (even if slightly above) rather than the closest under 0.01

a

?

b

?

C.D.F. Binomial table: $p = 0.25, n = 40$

x	$P(X \leq x)$
2	0.0010
3	0.0047
4	0.0160
5	0.0433
16	0.9884
17	0.9953
18	0.9983
19	0.9994

Warning: Textbook has several typos in this example.

Exercise 7B

Pearson Pure Mathematics Year 1/AS
Pages 103-105

Alternative method using critical regions

We can also find the critical region and see if the test statistic lies within it.

John tosses a coin 8 times and it comes up heads 6 times. He claims the coin is **biased towards heads**. With a significance level of 5%, test his claim.

X is number of heads.
p is probability of heads.
$X \sim B(8, p)$

STEP 1: Define test statistic X (stating its distribution), and the parameter p.

$H_0: p = 0.5$
$H_1: p > 0.5$

STEP 2: Write null and alternative hypotheses.

?

?

STEP 3 (Alternative): Determine critical region.

STEP 4: Two-part conclusion:
1. Do we reject H_0 or not?
2. Put in context of original problem.

C.D.F. Binomial table: $p = 0.5, n = 8$

x	$P(X \leq x)$
0	0.0039
1	0.0352
2	0.1445
3	0.3633
4	0.6367
5	0.8555
6	0.9648
7	0.9961

More on p-values

(Note that this is not covered in the Pearson textbook, but **is** in the specification)

Sheila wants to know if a coin is biased towards heads and throws it a large number of times, counting the number of heads. The p-value is less than 0.03. Conduct a hypothesis test at the 5% significance level.

?

Froflections: Ordinarily we'd calculate the probability of seeing the observed number of heads 'or more extreme'. But this has already been done for us (i.e. the p-value), so we just need to compare this against the threshold.

Further Example

[Textbook] The standard treatment for a particular disease has a $\frac{2}{5}$ probability of success. A certain doctor has undertake research in this area and has produced a new drug which has been successful with 11 out of 20 patients. The doctor claims the new drug represents an improvement on the standard treatment.
Test, at the 5% significance level, the claim made by the doctor.

?

STEP 1: Define test statistic X (stating its distribution), and the parameter p.

?

STEP 2: Write null and alternative hypotheses.

?

STEP 3: Determine probability of observed test statistic (or 'more extreme'), assuming null hypothesis.

?

STEP 4: Two-part conclusion:
1. Do we reject H_0 or not?
2. Put in context of original problem.

Test Your Understanding

Edexcel S2 Jan 2011 Q2

A student takes a multiple choice test. The test is made up of 10 questions each with 5 possible answers. The student gets 4 questions correct. Her teacher claims she was guessing the answers. Using a one tailed test, at the 5% level of significance, test whether or not there is evidence to reject the teacher's claim.

State your hypotheses clearly. **(6)**

?

Exercise 7C

Pearson Pure Mathematics Year 1/AS
Pages 106-107

Two-Tailed Tests

We have already seen that if we're interest in bias 'either way', we have two tails, and therefore have to split the critical region by **halving the significance level at each end**.

Over a long period of time it has been found that in Enrico's restaurant the ratio of non-veg to veg meals is 2 to 1. In Manuel's restaurant in a random sample of 10 people ordering meals, 1 ordered a vegetarian meal. Using a 5% level of significance, test whether or not the proportion of people eating veg meals in Manuel's restaurant is different to that in Enrico's restaurant.

?

Test Your Understanding

Edexcel S2 Jan 2006 Q7a

A teacher thinks that 20% of the pupils in a school read the Deano comic regularly.

He chooses 20 pupils at random and finds 9 of them read the Deano.

(*a*) (i) Test, at the 5% level of significance, whether or not there is evidence that the percentage of pupils that read the Deano is different from 20%. State your hypotheses clearly.

(ii) State all the possible numbers of pupils that read the Deano from a sample of size 20 that will make the test in part (*a*)(i) significant at the 5% level. **(9)**

(a)(i)

?

(ii)

?

Exercise 7D

Pearson Pure Mathematics Year 1/AS
Pages 108-109