Chapter 7 - Statistics

Hypothesis Testing

Chapter Overview

1. Hypothesis Testing
2. Finding Critical Values
3. One-Tailed Tests
4. Two-Tailed Tests

Topics	What students need to learn:		
	Content		Guidance
5 **Statistical hypothesis testing** *continued*	5.2	**Conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context.**	
		Understand that a sample is being used to make an inference about the population. **and** **appreciate that the significance level is the probability of incorrectly rejecting the null hypothesis.**	**Hypotheses should be expressed in terms of the population parameter p** **A formal understanding of Type I errors is not expected.**
	5.3	Conduct a statistical hypothesis test for the mean of a Normal distribution with known, given or assumed variance and interpret the results in context.	Students should know that: If $X \sim \mathrm{N}(\mu, \sigma^2)$ then $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ and that a test for μ can be carried out using: $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathrm{N}(0, 1^2)$. No proofs required. Hypotheses should be stated in terms of the population mean μ. Knowledge of the Central Limit Theorem or other large sample approximations is not required.

1. Hypothesis Testing

What is Hypothesis Testing?

Vocabulary:

Hypothesis = statement about the value of a population parameter
Test Statistic = the result of the experiment, or the statistic that is calculated from the example
Null Hypothesis, H_0 = the hypothesis you assume is correct
Alternative Hypothesis, H_1 = tells you about the parameter if your assumption is shown to be wrong

Example

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) What is the hypothesis?

2) What is the population parameter?

In my sample of 20 people in Hawaii, I find that 5 are left-handed.
3) Suggest a null hypothesis (the hypothesis I assume is correct)

4) Suggest an alternative hypothesis (what happens to my parameter if my assumption is wrong)

5) What is the test statistic?

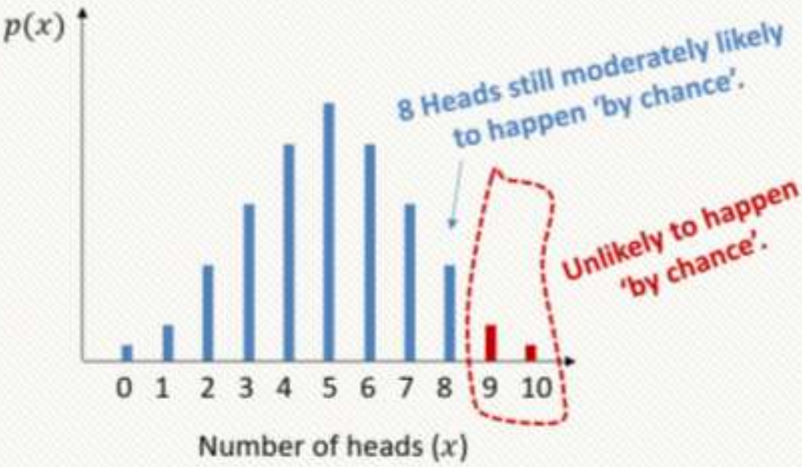
What is Hypothesis Testing?

Hypothesis testing in a nutshell then is:

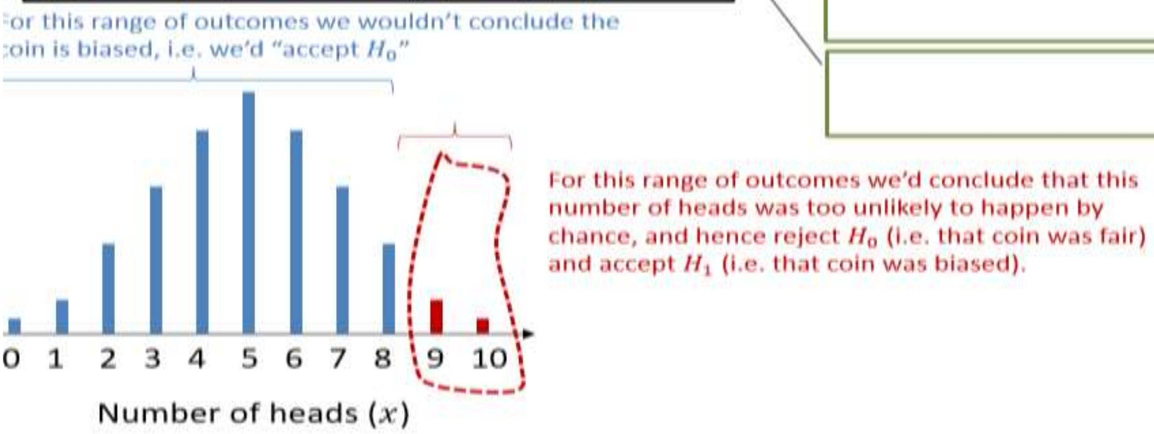
1. We have some hypothesis we wish to see if true (proportion of left-handed people in Hawaii is more than global average), so…
2. We collect some sample data (giving us our test statistic) and…
3. If that data is sufficiently unlikely to have emerged 'just by chance', then we conclude that our (alternative)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

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.



In this context…



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.

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.

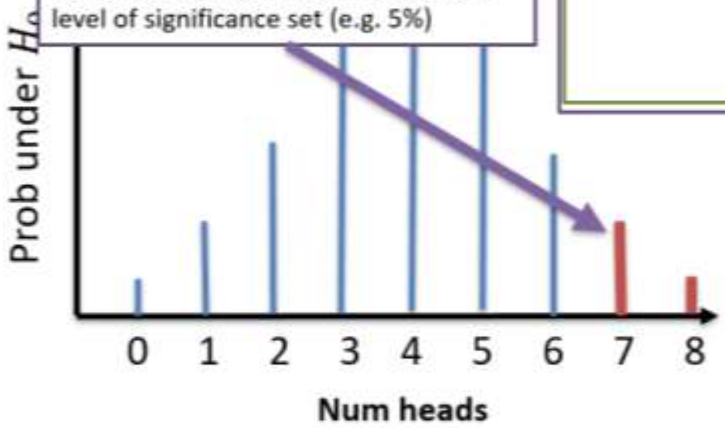
Critical Regions and Values

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

✎ 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

Quick fire 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 $P(X \geq 7)$?

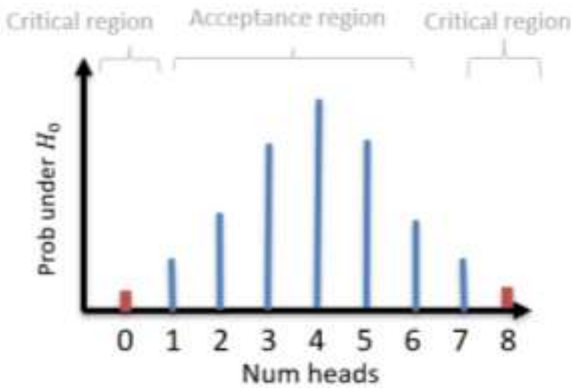
$\boldsymbol{P(X \geq 7)} =$

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%.

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.

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

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

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

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.

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.

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)**

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)**