# Hypothesis Tests

*There are a lot of technical terms to learn over the next two pages. It might feel a bit hard-going, but don't despair — it's paving the way for a return to some old friends later on (and it'll come in really handy for the exam).*

## A **Hypothesis** is a **Statement** you want to **Test**

Hypothesis testing is about using **sample data** to **test statements** about **population parameters**. Unfortunately, it comes with a fleet of terms you need to know.

A parameter is a quantity that describes a characteristic of a population (e.g. mean or variance).

- **Null Hypothesis ($H_0$)** — a statement about the value of a population parameter. Your data may allow you to **reject** this hypothesis.
- **Alternative Hypothesis ($H_1$)** — a statement that describes the value of the population parameter if $H_0$ is rejected.
- **Hypothesis Test** — a statistical test that tests the claim that $H_0$ makes about the parameter against that made by $H_1$. It tests whether $H_0$ should be rejected or not, using evidence from sample data.
- **Test Statistic** — a statistic calculated from sample data which is used to decide whether or not to reject $H_0$.

1) For any hypothesis test, you need to **write two hypotheses — a null hypothesis** and an **alternative hypothesis**.
2) You often choose the **null hypothesis** to be something you actually **think is false**. This is because hypothesis tests can only show that **statements are false** — they **can't** prove that things are **true**. So you're aiming to find **evidence** for what you think is **true**, by **disproving** what you think is **false**.
3) $H_0$ needs to give a **specific value** to the parameter, since all your calculations will be based on this value. You **assume** this value holds **true** for the test, then see if your data allows you to **reject** it. **$H_1$** is then a statement that describes how you think the **value of the parameter differs** from the value given by $H_0$.
4) The **test statistic** you choose **depends on the parameter** you're interested in. It should be a **'summary'** of the sample data, and should have a sampling distribution that can be calculated using the parameter value specified by $H_0$.

**Example:** A 4-sided spinner has sides labelled A–D. Jemma thinks that the spinner is biased towards side A. She spins it 20 times and counts the number of times, $Y$, that she gets side A.
a) Write down a suitable null hypothesis to test Jemma's theory.
b) Write down a suitable alternative hypothesis.
c) Describe the test statistic Jemma should use.

a) If you assume the spinner is unbiased, each side has a probability of 0.25 of being spun. Let $p$ = the probability of spinning side A.
Then: **$H_0: p = 0.25$** ← By assuming the spinner is unbiased, the parameter, p, can be given the specific value 0.25. Jemma is then interested in disproving this hypothesis.

b) If the spinner is biased towards side A, then the probability will be greater than 0.25. So: **$H_1: p > 0.25$** ← This is what Jemma actually thinks.

c) The test statistic is $Y$, the number of times she gets side A. ← Assuming $H_0$ is true, the sampling distribution of Y is B(20, 0.25) — see p.162.

## Hypothesis Tests can be **One-Tailed** or **Two-Tailed**

The 'tailed' business is to do with the critical region used by the test — see the next page.

For $H_0: \theta = a$, where $\theta$ is a parameter and $a$ is a number:

1) The test is **one-tailed** if **$H_1$ is specific** about the value of $\theta$ compared to $a$, i.e. $H_1: \theta > a$, or $H_1: \theta < a$.
2) The test is **two-tailed** if **$H_1$ specifies only that $\theta$ doesn't equal $a$**, i.e. $H_1: \theta \neq a$.

Whether you use a one-tailed or a two-tailed test depends on how you define $H_1$. And that depends on **what you want to find out about the parameter** and any **suspicions** you might have about it.

E.g. in the example above, Jemma suspects that the probability of getting side A is **greater than 0.25**. This is what she wants to test, so it is sensible to define **$H_1: p > 0.25$**.
If she wants to test for **bias**, but is **unsure** if it's towards or against side A, she could define **$H_1: p \neq 0.25$**.

A very important thing to remember is that the results of a hypothesis test are either '**reject $H_0$**', or '**do not reject $H_0$**' — which means you haven't found enough evidence to **disprove** $H_0$, and **not** that you've proved it.

# Hypothesis Tests

## *If your Data is **Significant**, **Reject $H_0$***

1) You would **reject $H_0$** if the **observed value** of the test statistic is **unlikely** under the null hypothesis.
2) The **significance level** of a test ($\alpha$) determines **how unlikely** the value needs to be before $H_0$ is rejected. It also determines the **strength** of the **evidence** that the test has provided — the lower the value of $\alpha$, the stronger the evidence you have for saying $H_0$ is false. You'll usually be told what level to use — e.g. 1% ($\alpha = 0.01$), 5% ($\alpha = 0.05$), or 10% ($\alpha = 0.1$). For a **two-tailed** test, you want a level of $\frac{\alpha}{2}$ for **each tail**.
3) To decide whether your result is **significant**:
   - Define the **sampling distribution** of the **test statistic** under the **null hypothesis**.
   - Calculate the **probability** of getting a value that's **at least as extreme** as the **observed value** from this distribution — this is known as the ***p*-value**.
   - If the *p*-value is **less than or equal to $\alpha$** (or $\frac{\alpha}{2}$ for a two-tailed test), **reject $H_0$** in favour of $H_1$.

**Example:** Javed wants to test at the 5% level whether or not a coin is biased towards tails. He tosses the coin 10 times and records the number of tails, $X$. He gets 9 tails.
a) Define suitable hypotheses for $p$, the probability of getting tails.
b) State the condition under which Javed would reject $H_0$.

P(at least as extreme as 9) means 9 or more (this is the p-value).

Significance level

a) **$H_0$: $p = 0.5$** and **$H_1$: $p > 0.5$**. b) Under $H_0$, $X \sim B(10, 0.5)$. **If $P(X \geq 9) \leq 0.05$, Javed would reject $H_0$.**

## *The **Critical Region** is the **Set of Significant Values***

1) The **critical region** (CR) is the **set of all values of the test statistic** that would cause you to **reject** $H_0$. The first value that's **inside** the CR is called the **critical value**, so results **as extreme** (or **more**) as this are **significant**.
2) **One-tailed tests** have a **single** CR, containing the highest or lowest values. For **two-tailed tests**, the region is **split into two** — half at the lower end and half at the upper end. Each half has a probability of $\frac{\alpha}{2}$.
3) To **test whether your result is significant**, find the critical region and if it **contains the observed value**, reject $H_0$.

**Example (continued):** c) Find the critical region for the test, at the 5% level.

This is a **one-tailed** test with $H_1$: $p > 0.5$, so you're only interested in the **upper end** of the distribution.

Use the **binomial cdf** on your calculator (see p.163) to find the value of $x$ such that **$P(X \geq x) \leq 0.05$** (the significance level). As calculators usually give probabilities for $P(X \leq x)$ you need to use $P(X \geq x) = 1 - P(X < x)$.

$P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7) = 1 - 0.9453... = \mathbf{0.0546... > 0.05}$
$P(X \geq 9) = 1 - P(X < 9) = 1 - P(X \leq 8) = 1 - 0.9892... = \mathbf{0.0107... < 0.05}$

So the critical region is **$X \geq 9$** ← So values of 9 or 10 would cause you to reject $H_0$: $p = 0.5$.

The **acceptance region** is where you **don't** reject $H_0$ — i.e. $X \leq 8$.

The **actual significance level** of a test is the probability of **incorrectly rejecting $H_0$** — i.e. the probability of getting extreme data by chance when $H_0$ is true. This is often **different** from the level of significance originally asked for in the question. Here, the actual significance level is $P(X \geq 9) = 0.0107$, which is much lower than 0.05.

## *Practice Question*

Q1 In 2012, a survey found that 68% of residents in a town used the local library. In 2016, the proportion was found to be 53%. A hypothesis test was carried out, using $H_0$: $p = 0.68$ and $H_1$: $p < 0.68$, and the null hypothesis was rejected. Does this result support the claim that the percentage of local residents using the library fell by 15%?

**Exam Question**

Q1 One year ago, 43% of customers rated a restaurant as 'Excellent'. Since then, a new chef has been employed, and the manager believes that the approval rating will have gone up. He decides to carry out a hypothesis test to test his belief. Define the null and alternative hypotheses the manager should use. [1 mark]

## *I repeat, X has entered the critical region — we have a significant situation...*

*Don't mix up the p-value with the binomial probability p. I know, I know, it would have been nice if they'd used a different letter, but that's life. Remember to always divide $\alpha$ by 2 whenever you're doing a two-tailed test.*

# Hypothesis Tests and Binomial Distributions

*OK, it's time to pick your best 'hypothesis testing' foot and put it firmly forward. It's also a good time to reacquaint yourself with binomial distributions, which you met in Section 13. Have a look back there before you go any further.*

## Use a *Hypothesis Test* to *Find Out* about the *Population Parameter p*

The first step in exam questions is to work out **which distribution** to use to model the situation — you've a choice of binomial or normal. Words like '**proportion**', '**percentage**' or '**probability**' are clues that it's **binomial**.

Hypothesis tests for the binomial parameter $p$ all follow the **same general method** — this is what you do:

1) Define the **population parameter** in **context**
   — for a binomial distribution it's always $p$, a **probability** of success, or **proportion** of a population.
2) Write down the **null** hypothesis ($H_0$) — $\mathbf{H_0: p = a}$ for some constant $a$.
3) Write down the **alternative** hypothesis ($H_1$)
   — $H_1$ will either be $\mathbf{H_1: p < a}$ or $\mathbf{H_1: p > a}$ (one-tailed test) or $\mathbf{H_1: p \neq a}$ (two-tailed test).
4) State the **test statistic**, $X$ — always just the number of '**successes**' in the sample.
5) Write down the **sampling distribution** of the test statistic under $H_0$ — $X \sim B(n, p)$ where $n$ is the sample size.
6) State the **significance level, $\alpha$** — you'll usually be given this.
7) Test for **significance** or find the **critical region** (see previous page).
8) Write your **conclusion** — state whether or not you have **sufficient evidence** to **reject $H_0$**.

**Example:** In a past census of employees, 20% were in favour of a change to working hours. After making changes to staff contracts, the manager now believes that the proportion of staff wanting a change in their working hours has decreased. The manager carries out a random sample of 30 employees, and 2 are in favour of a change in hours.

Stating your hypotheses clearly, test the manager's claim at the 5% level of significance.

1) Let $p$ = **proportion of employees in favour of change to hours.**
2) Assume there's been **no change** in the proportion: $\mathbf{H_0: p = 0.2}$
3) The manager's interested in whether the proportion has **decreased**, so: $\mathbf{H_1: p < 0.2}$

*You assume there's been no change in the value of the parameter, so you can give it a value of 0.2. The alternative hypothesis states what the manager actually thinks.*

4) Let $X$ = the number of employees in the sample who are in favour of change.
5) Under $H_0$, $\mathbf{X \sim B(30, 0.2)}$. ← *The sampling distribution of the test statistic uses the value $p = 0.2$.*
6) The **significance level** is 5%, so $\boldsymbol{\alpha} = \mathbf{0.05}$.
7) Find the **$p$-value** — the probability of a value for your **test statistic at least as extreme** as the **observed value**. This is a **one-tailed test** and you're interested in the lower end of the distribution. So you want to find the probability of $X$ taking a value less than or equal to 2.
   Using the binomial cdf on your calculator:
   $P(X \leq 2) = $ **0.0441...**, and since **0.0441... < 0.05**, the result is **significant**.
8) Now write your **conclusion**: **There is evidence at the 5% level of significance to reject $H_0$ and to support the manager's claim that the proportion in favour of change has decreased.**

*Always say "there is evidence to reject $H_0$", or "there is insufficient evidence to reject $H_0$", never just "accept $H_0$" or "reject $H_1$".*

To find a **critical region**, your test would look the same except for step 7...

7) Find the **critical region** for a test at this level of significance. This is a **one-tailed test** and you're interested in the lower end of the distribution. The critical region is the biggest possible set of 'low' values of $X$ with a total probability of $\leq 0.05$.
   Using the binomial cdf on your calculator:
   Try $X \leq 2$: $P(X \leq 2) = $ **0.0441... < 0.05**. Now try $X \leq 3$: $P(X \leq 3) = $ **0.1227... > 0.05**.
   So **CR is $X \leq 2$**. These results fall in the CR, so the result is **significant**.

Marjorie and Edwin were ready to enter the critical region.

# Hypothesis Tests and Binomial Distributions

## You might be asked to find a *Critical Region* or *Actual Significance Level*

**Example:** Records show that the proportion of trees in a wood that suffer from a particular leaf disease is 15%. Chloe thinks that recent weather conditions might have affected this proportion. She examines a random sample of 20 of the trees.

a) Using a 10% level of significance, find the critical region for a two-tailed test of Chloe's theory. The probability of rejection in each tail should be less than 0.05.

b) Find the actual significance level of a test based on your critical region from part a).

Chloe finds that 8 of the sampled trees have the leaf disease.

c) Comment on this finding in relation to your answer to part a) and Chloe's theory.

a) Let $p$ = proportion of trees with the leaf disease.

**$H_0: p = 0.15$ $H_1: p \neq 0.15$**

Let $X$ = number of sampled trees with the disease. Under $H_0$, **$X \sim B(20, 0.15)$**.

**$\alpha = 0.1$**, and since the test is **two-tailed**, the probability of $X$ falling in each tail should be 0.05, at most.

This is a two-tailed test, so you're interested in both ends of the sampling distribution.
The lower tail is the biggest possible set of 'low' values of $X$ with a total probability of $\leq 0.05$.
The upper tail is the biggest possible set of 'high' values of $X$ with a total probability of $\leq 0.05$.

Using a calculator: **Lower tail**:
$P(X \leq 0) = 0.0387... < 0.05$
$P(X \leq 1) = 0.1755... > 0.05$

**Upper tail**:
$P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9326... = 0.0673... > 0.05$
$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9780... = 0.0219... < 0.05$

So **CR is $X = 0$ or $X \geq 7$**.

b) The **actual significance level** is: $P(X = 0) + P(X \geq 7) = 0.0387... + 0.0219... =$ **0.0607** or **6.07%** (3 s.f.)

c) The observed value of **8** is in the critical region. **So there is evidence at the 10% level of significance to reject $H_0$ and to support Chloe's theory that there has been a change in the proportion of affected trees.**

## Practice Questions

Q1 Carry out the following tests of the binomial parameter $p$.
Let $X$ represent the number of successes in a random sample of size 20:

a) Test $H_0: p = 0.2$ against $H_1: p \neq 0.2$, at the 5% significance level, using $x = 1$.

b) Test $H_0: p = 0.4$ against $H_1: p > 0.4$, at the 1% significance level, using $x = 15$.

Q2 Find the critical region for the following test where $X \sim B(10, p)$:
Test $H_0: p = 0.3$ against $H_1: p < 0.3$, at the 5% significance level.

**Exam Question**

Q1 Over a long period of time, the chef at an Italian restaurant has found that there is a probability of 0.2 that a customer ordering a dessert on a weekday evening will order tiramisu. He thinks that the proportion of customers ordering desserts on Saturday evenings who order tiramisu is greater than 0.2.

a) State the name of the probability distribution that would be used in a hypothesis test for the value of $p$, the proportion of Saturday evening dessert eaters ordering tiramisu. [1 mark]

A random sample of 20 customers who ordered a dessert on a Saturday evening was taken.
7 of these customers ordered tiramisu.

b) (i) Stating your hypotheses clearly, test the chef's theory at the 5% level of significance. [6 marks]

(ii) Find the minimum number of tiramisu orders needed for the result to be significant. [1 mark]

## *My hypothesis is — this is very likely to come up in the exam...*

*Remember, to make sure that you're not using the binomial pdf by accident (if your calculator has one). Of course, you can use the pdf if you need to find the probability of a single value, as long as you're doing it for the right reasons.*