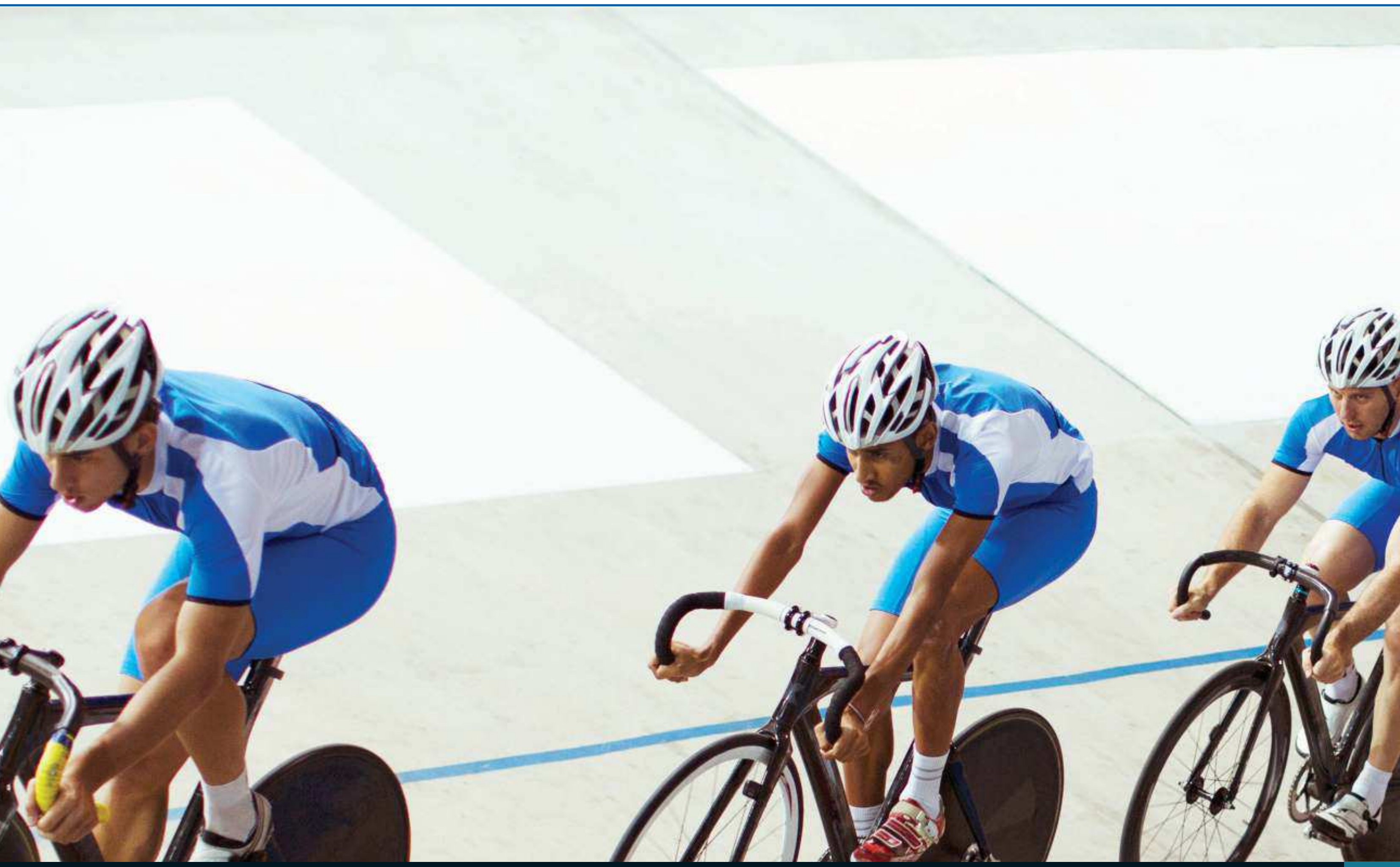
Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information



Chapter 1
Hypothesis testing

In this chapter you will learn how to:

- understand the nature of a hypothesis test; the difference between one-tailed and two-tailed tests, and the terms null hypothesis, alternative hypothesis, significance level, critical region (or rejection region), acceptance region and test statistic
- formulate hypotheses and carry out a hypothesis test in the context of a single observation from a population that has a binomial distribution, using:
  - direct evaluation of probabilities
  - a normal approximation to the binomial
- interpret outcomes of hypothesis testing in context
- understand the terms Type I error and Type II error in relation to hypothesis testing
- calculate the probabilities of making Type I and Type II errors in specific situations involving tests based on a normal distribution or direct evaluation of binomial probabilities.

CAMBRIDGE
Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information

PREREQUISITE KNOWLEDGE

Where it comes from	What you should be able to do	Check your skills
Probability & Statistics 1, Chapter 7	Calculate probabilities using the binomial distribution.	**1** $X \sim \text{B}(8,\ 0.1)$. Find: **a** $\text{P}(X = 2)$ **b** $\text{P}(X \leqslant 1)$ **c** $\text{P}(X > 3)$ **2** $X \sim \text{B}(15, 0.25)$. Find: **a** $\text{P}(X = 7)$ **b** $\text{P}(X < 2)$ **c** $\text{P}(2 \leqslant X < 4)$
Probability & Statistics 1, Chapter 8	Use normal distribution tables to calculate probabilities.	**3** Given that $X \sim \text{N}(22, 16)$, find: **a** $\text{P}(X < 24)$ **b** $\text{P}(X > 15)$ **c** $\text{P}(18 < X < 23)$ **4** Given that $X \sim \text{N}\left(30, 6^2\right)$, find: **a** $\text{P}(X < 23)$ **b** $\text{P}(X > 25)$
Probability & Statistics 1, Chapter 8	Approximate the binomial distribution to a normal distribution.	**5** Given that $X \sim \text{B}(80, 0.4)$, state a suitable approximating distribution and use it to find: **a** $\text{P}(X \leqslant 36)$ **b** $\text{P}(X > 30)$ **6** Given that $X \sim \text{B}(120, 0.55)$, state a suitable approximating distribution and use it to find: **a** $\text{P}(X > 70)$ **b** $\text{P}(X \leqslant 63)$

Why do we study hypothesis testing?

An opinion poll asks a sample of voters who they will vote for in an election with two candidates, candidate A and candidate B. If 53% of voters say they will vote for candidate A, can you actually be certain candidate A will win the election?

DNA evidence is presented in courtrooms. A judge and/or jury has to make a decision based on this evidence. How can the judge or jury be certain the evidence is true?

A pharmaceutical company testing a new drug treatment observes a positive effect in people using it. How can they test if the effect of the new drug treatment, compared with an older drug treatment, is significant?

These three situations are examples of difficult, real-life problems where statistics can be used to explain and interpret the data. Statistics, and more specifically hypothesis testing, provide a method that scientifically analyses data in order to reach a conclusion. A hypothesis test analyses the data to find out how likely it is that the results could happen by chance. In opinion polls, statisticians look at the sample chosen, whether it is representative, the sample size and how significant the poll results are. Voters looking at opinion poll results may make decisions based on what they read; however, the consequences for individuals are more pertinent in a court case with DNA evidence or when testing new drug treatments. Decisions based on this evidence can be life-changing for individuals.

In this chapter, you will learn how to calculate if the occurrence of events is statistically significant or whether they could have occurred by chance, using the statistical process called hypothesis testing.

DID YOU KNOW?

Hypothesis testing is widely used in research in the social sciences, psychology and sociology, in scientific studies and in the humanities. The conclusions reached rely not only on a clear grasp of this statistical procedure, but also on whether the experimental design is sound. To find out further information on experimental design you may like to look at the work of Ronald Fisher, an English statistician and biologist who, in the first half of the twentieth century, used mathematics to study genetics and natural selection.

1.1 Introduction to hypothesis testing

Suppose you have a set of dice numbered 1 to 6. Some of the dice are fair and some are biased. The biased dice have bias towards the number six. If you pick one of the dice and roll it 16 times, how many times would you need to roll a six to convince yourself that this die is biased? Write down your prediction of how many sixes imply the die is biased.

Carry out this experiment yourself and work through each step.

Roll your die 16 times and record how many times you get a six. How many did you get? Do you want to reconsider your prediction?

It is possible, although highly unlikely, to roll 16 sixes with 16 rolls of a fair die. If X is the number of sixes obtained with 16 rolls of the die then $X \sim \mathrm{B}\left(16, \frac{1}{6}\right)$ and $\mathrm{P}(X = 16) = \left(\frac{1}{6}\right)^{16}$.

We can calculate theoretical probabilities of rolling different numbers of sixes.

$$P(X=0)=\binom{16}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{16}=0.0541 \qquad P(X\leqslant 0)=0.0541=5.41\%$$

$$P(X=1)=\binom{16}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{15}=0.1731 \qquad P(X\leqslant 1)=0.2272=22.7\%$$

$$P(X=2)=\binom{16}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{14}=0.2596 \qquad P(X\leqslant 2)=0.4868=48.7\%$$

$$P(X=3)=\binom{16}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{13}=0.2423 \qquad P(X\leqslant 3)=0.7291=72.9\%$$

$$P(X=4)=\binom{16}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{12}=0.1575 \qquad P(X\leqslant 4)=0.8866=88.7\%$$

$$P(X=5)=\binom{16}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{11}=0.0756 \qquad P(X\leqslant 5)=0.9622=96.2\%$$

These probabilities tell us that in 16 rolls of a die, almost 89% of the time we would expect to roll at most 4 sixes, and for over 96% of the time we would expect to roll at most 5 sixes. This means that rolling 6 sixes or more will occur by chance less than 4% of the time.

How small must a probability be for you to accept a claim that something occurred by chance? That is, in 16 rolls of the die how many sixes do you need to be convinced the die is biased?

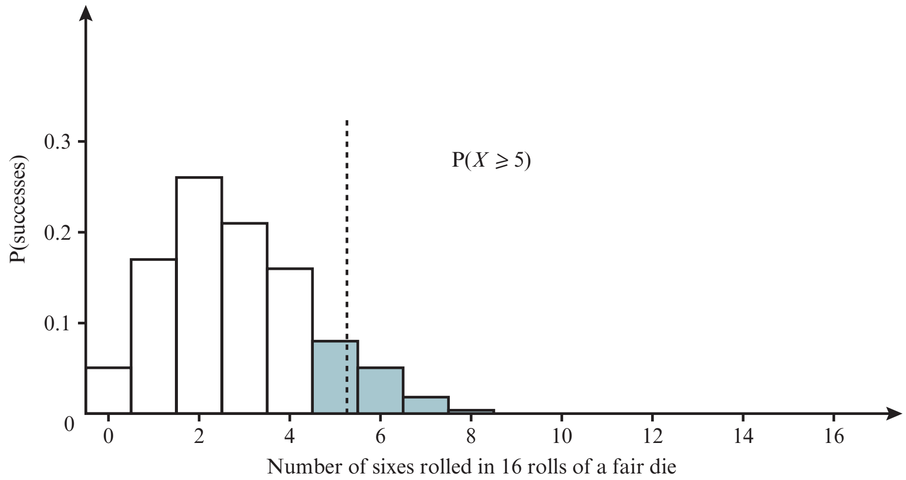
Do you want to reconsider your prediction?

Using the theoretical probability calculations, we can work out $P(X\geqslant 5)$ and $P(X\geqslant 6)$.

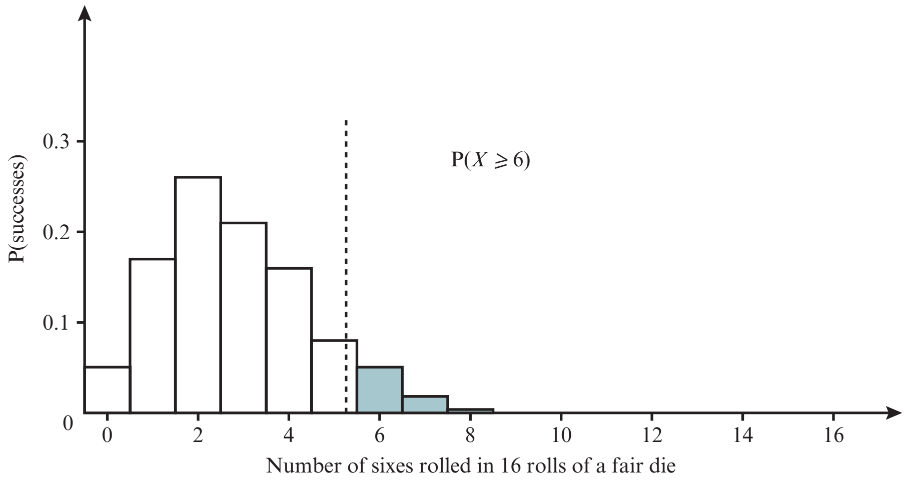
$$P(X\geqslant 5)=1-P(X\leqslant 4)=1-0.8866=0.1134=11.3\%$$

$$P(X\geqslant 6)=1-P(X\leqslant 5)=1-0.9622=0.0378=3.78\%$$

The shaded regions on these graphs show the regions for the probabilities. The first graph shows $P(X\geqslant 5)$ and the second shows $P(X\geqslant 6)$.



CAMBRIDGE
Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information



In both graphs, the region to the right of the dotted line represents 5% of the total probability.

The second of the previous two graphs shows that $P(X \geqslant 6)$ is completely to the right of the dotted line, which we would expect as $P(X \geqslant 6)$ is less than 5%. However, the first graph shows the dotted line dividing $P(X \geqslant 5)$ into two parts, which is as expected since $P(X \geqslant 5)$ is greater than 5%. If we choose 5% as the critical percentage at which the number of sixes rolled is significant (that is, the **critical value** at which the number of sixes rolled is unlikely to occur by chance), then rolling 6 sixes is the critical value we use to decide if the die is biased.

The percentage value of 5% is known as the **significance level**. The significance level determines where we draw a dotted line on the graph of the probabilities. This region at one end of the graph is the **critical region**. The other region of the graph is the **acceptance region**.

You can choose the percentage significance level, although in practice 5% is used most often.

KEY POINT 1.1

The significance level is the probability of rejecting a claim. A claim cannot be proven by scientific testing and analysis, but if the probability of it occurring by chance is very small it is said that there is sufficient evidence to reject the claim.

The range of values at which you reject the claim is the critical region or rejection region.

The value at which you change from accepting to rejecting the claim is the critical value.

DID YOU KNOW?

This example has links with the philosopher Karl Popper's falsification theory. Popper believed that the claim most likely to be true is the one we should prefer. Consider this example; while there is no way to prove that the Sun will rise on any particular day, we can state a claim that every day the Sun will rise; if the Sun does not rise on some particular day, the claim will be falsified and will have to be replaced by a different claim. This has similarities with the example of trying to decide if the dice is biased. Until we have evidence to suggest otherwise, we accept that the original claim is correct.

Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information

Sample data, such as the data you obtain in the dice experiment, may only partly support or refute a supposition or claim. A **hypothesis** is a claim believed or suspected to be true. A **hypothesis test** is the investigation that aims to find out if the claim could happen by chance or if the probability of it happening by chance is statistically significant. It is very unlikely in any hypothesis test to be able to be absolutely certain that a claim is true; what you can say is that the test shows the probability of the claim happening by chance is so small that you are statistically confident in your decision to accept or reject the claim.

KEY POINT 1.2

In a hypothesis test, the claim is called the **null hypothesis**, abbreviated to H_0.

If you are not going to accept the null hypothesis, then you must have an **alternative hypothesis** to accept; the alternative hypothesis abbreviation is H_1.

Both the null and alternative hypotheses are expressed in terms of a parameter, such as a probability or a mean value.

The following example models a hypothesis test.

WORKED EXAMPLE 1.1

Experience has shown that drivers on a Formula 1 racetrack simulator crash 40% of the time. Zander decides to run a simulator training programme to reduce the number of crashes. To evaluate the effectiveness of the training programme, Zander allows 20 drivers, one by one, to use the simulator. Zander counts how many of the drivers crash.

Four drivers crash. Test the effectiveness of Zander's simulation training programme at the 5% level of significance.

Answer

Let X be 'the number of drivers that crash'.
Then $X \sim B(20, 0.4)$.

H_0: $p = 0.4$
H_1: $p < 0.4$

Define a **random variable** and its **parameters**. Define the null and alternative hypotheses.

The null hypothesis is the current value of p; the alternative hypothesis is training is effective, so p is smaller than 0.4.

The null hypothesis, H_0, is the assumption that there is no difference between the usual outcome and what you are testing. The usual outcome is crashing 40% of the time; that is, $p = 0.4$. The alternative hypothesis, H_1, is $p < 0.4$ since Zander's training programme aims to reduce the probability of crashes.

Four drivers crash; you calculate the test statistic, the region $P(X \leqslant 4)$:

$$P(X \leqslant 4) = \binom{20}{0}0.4^0\,0.6^{20} + \binom{20}{1}0.4^1\,0.6^{19} + \binom{20}{2}0.4^2\,0.6^{18} + \binom{20}{3}0.4^3\,0.6^{17} + \binom{20}{4}0.4^4\,0.6^{16}$$
$$= 0.050952 = 5.10\%$$

The **test statistic** is the calculated probability using sample data in a hypothesis test.

You compare your calculated test statistic with what is the expected outcome from the null hypothesis.

Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information

$P(X \leqslant 4) > 5\%$, so 4 is not a critical value.
Therefore, accept H_0. There is insufficient evidence at the 5% significance level to suggest that the proportion of drivers crashing their simulators has decreased.

Compare the calculated probability with the percentage significance level. Decide to accept or reject the null hypothesis and comment in context.

In this case, the calculated probability must be less than the percentage significance level to say there is sufficient evidence to reject the null hypothesis.

Your conclusion should be in context of the original question.

Consider this alternative scenario: suppose only three of the 20 drivers crashed after attending Zander's simulator training programme. At the 5% level of significance, what would you conclude?

Answer

$$P(X \leqslant 3) = \binom{20}{0}0.4^0\,0.6^{20} + \binom{20}{1}0.4^1\,0.6^{19} + \binom{20}{2}0.4^2\,0.6^{18} + \binom{20}{3}0.4^3\,0.6^{17} = 0.01596$$

$P(X \leqslant 3) = 1.6\% < 5\%$

This is less than the percentage significance level.

At the 5% level of significance there is sufficient evidence to reject H_0 and accept H_1.

The evidence suggests that the proportion of drivers that crash has decreased and the simulator training programme is effective.

Consider this alternative scenario: suppose four drivers crashed and you were testing the effectiveness of Zander's simulator training programme at the 10% level of significance.

Answer

$P(X \leqslant 4) = 5.10\% < 10\%$

As this is less than the percentage significance level, you reject H_0 and accept H_1. At the 10% significance level you can conclude that there is sufficient evidence to say the simulator training programme is effective.

EXPLORE 1.1

Use the simulator training programme example to write out the biased dice experiment described earlier in this chapter, as a hypothesis test. Define the random variable and its parameters for the biased dice experiment.

Define the null and alternative hypotheses. Explain why the default belief is that $p = \frac{1}{6}$.

What are you claiming might be true about p?

What do you think would be a sensible size for your critical region?

Calculate the test statistic for your chosen critical region.

What do you conclude?

Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information

KEY POINT 1.3

When you investigate a claim using a hypothesis test, you are not interested in the probability of an exact number occurring, but the probability of a range of values up to and including that number occurring. In effect that is the probability of a region. The test statistic is the calculated region using sample data in a hypothesis test.

The range of values for which you reject the null hypothesis is the critical region or **rejection region**.

The value at which you change from accepting the null hypothesis to rejecting it is the critical value.

WORKED EXAMPLE 1.2

Studies suggest that 10% of the world's population is left-handed. Bailin suspects that being left-handed is less common amongst basketball players and plans to test this by asking a random sample of 50 basketball players if they are left-handed.

a Find the rejection region at the 5% significance level and state the critical value.

b If the critical value is 2, what would be the least integer percentage significance level for you to conclude that Bailin's suspicion is correct?

Answer

Let X be the random variable 'number of basketball players' and p be the probability of being left-handed; then $X \sim B(50, 0.1)$.

> Define the distribution and its parameters; this is a binomial distribution as there is a fixed number of trials and only two outcomes.
>
> State H_0, H_1 and the significance level of the test.

a H_0: $p = 0.1$

H_1: $p < 0.1$

Significance level: 5%

The rejection region is $X \leqslant k$ such that

$P(X \leqslant k) \leqslant 5\%$ and $P(X \leqslant k+1) > 5\%$.

k	0	1	2
$P(X = k)$	0.0052	0.0286	0.0779

> It can be helpful to show the probabilities in a table.
>
> To calculate the probabilities, use
>
> $P(X = r) = \binom{50}{r} 0.1^r 0.9^{50-r}$.

$P(X \leqslant 1) = 0.0338 = 3.38\% < 5\%$

$P(X \leqslant 2) = 0.1117 = 11.17\% > 5\%$

The rejection region is $X \leqslant 1$; the critical value is 1.

> Ensure that the question is answered fully.

b $P(X \leqslant 2) = 11.17\% < 12\%$, so with a critical value of 2 the significance level needs to be 12% to reject the null hypothesis.

> For Bailin to reject the null hypothesis the probability of the region must be less than the significance level.

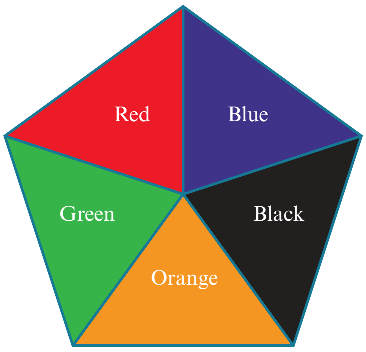
KEY POINT 1.4

To carry out a hypothesis test:

- Define the random variable and its parameters.
- Define the null and alternative hypotheses.
- Determine the critical region.
- Calculate the test statistic.
- Compare the test statistic with the critical region.
- Write your conclusion in context.

WORKED EXAMPLE 1.3

A regular pentagonal spinner is spun ten times. It lands on the red five times.
Test at the 4% level of significance if the spinner is biased towards red.



Answer

Let random variable X be 'number of spins landing on red'. Then $X \sim \text{B}(10, 0.2)$.

Define the random variable and its parameters.

H_0: $p = 0.2$.

H_1: $p > 0.2$

Define the null and alternative hypotheses.

4% significance level

You are looking for a region with probability less than 0.04.

$\text{P}(X \geqslant 5) = 1 - \text{P}(X \leqslant 4)$

Calculate the test statistic.

$$= 1 - \left(\binom{10}{0} 0.8^{10} + \binom{10}{1} 0.2\ 0.8^{9} + \binom{10}{2} 0.2^{2}\ 0.8^{8} + \binom{10}{3} 0.2^{3}\ 0.8^{7} + \binom{10}{4} 0.2^{4}\ 0.8^{6} \right)$$

$$= 0.0328 = 3.28\%$$

$3.28\% < 4\%$

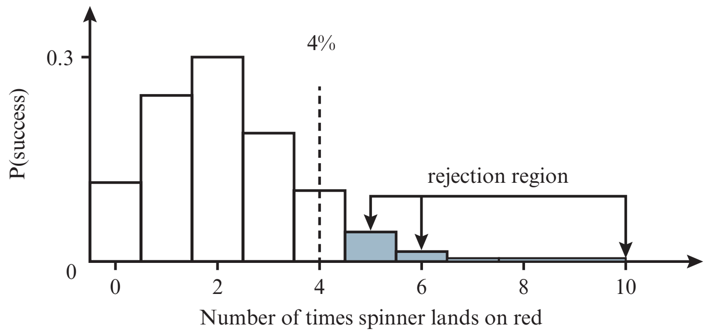
Compare with the significance level.

Therefore, reject the null hypothesis. There is evidence to show the spinner is biased towards red.

Interpret the result in the context of the problem.

CAMBRIDGE
Cambridge University Press
978-1-108-63305-5 — Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook with Cambridge Online Mathematics (2 Years)
Jayne Kranat , Edited by Julian Gilbey
Excerpt
More Information

The graph shows the distribution of the spinner landing on red when it is spun ten times.



The probability 0.328 or 3.28% is completely in the 4% tail of the graph; that is, the critical, or rejection, region; the null hypothesis is rejected and the alternative hypothesis is accepted.

Note that if the spinner had landed on red four times in ten spins, then $P(X \geqslant 4) = 0.121 = 12.1\% > 4\%$ and the null hypothesis would be accepted; there would not be enough evidence to say the spinner is biased.

KEY POINT 1.5

The lower the percentage significance level, the smaller the rejection region and the more confident you can be of the result.

EXPLORE 1.2

Why would the significance level be especially important when exploring the data from trials of new medicines?

Discuss the levels of significance you think would be appropriate to use with the following hypotheses.

- Treatment with drug X, which has no known side effects, may not make a surgical option less likely.
- Compound Y, which is expensive, has no effect in preventing colour fading when added to clothes dye.
- Ingredient Z, which is inexpensive, has no effect on the appearance of a food product.

For a large sample, you can approximate a **binomial distribution** to a **normal distribution** and then carry out a hypothesis test.

REWIND

In Chapter 8 of the Probability & Statistics 1 Coursebook, we learnt that we need to use a continuity correction when approximating a binomial distribution by a normal distribution.

WORKED EXAMPLE 1.4

Arra is elected club president with the support of 52% of the club members. One year later, the club members claim that Arra does not have as much support any more. In a survey of 200 club members, 91 said they would vote for Arra. Using a suitable approximating distribution, test this claim at the 5% significance level.

