Section 1: Introducing hypothesis testing

Notes and Examples

These notes contain subsections on

Setting up a hypothesis test

In statistical work you often wish to find out if something that occurs is within the normal range of expectations or is an unusual occurrence.

Look at these three examples.

Example 1
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 4 consecutive times. Andy complains that the coin must be biased. Is his complaint justified?

Example 2
A student takes a multiple-choice test. There are 10 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 5 questions right.
Is the student's method of missing lessons and guessing a good strategy?
The student claims to be an inspired guesser. Is this true?

Example 3
The probability that a student at a certain school passes an exam is 0.8.
A new teacher is appointed and takes the next group. Out of a group of 10 students only 6 students now pass.
Should the Head of Maths be concerned?

These three examples will help you understand how hypothesis testing works.

The ideal hypothesis test involves in this order:

- Establish the null and alternative hypotheses
- Decide on the significance level
- Collect suitable data using a random sampling procedure that ensures the items are independent.
- Conduct a test, doing the necessary calculations.
- Interpret the results in terms of the original claim.

There are lots of situations where we cannot carry out a test as rigorously as this.

In all the examples 1, 2 and 3 the data has already been collected before the hypotheses are set up.

However, it is important to read what you are investigating, rather than looking at the data to set up the hypothesis test.

Example 1
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 4 consecutive times. Andy complains that the coin must be biased. Is his complaint justified?

Solution
Let p be the probability of getting Heads.
$H_0: p = \frac{1}{2}$ (equal chances of getting Heads or Tails)
$H_1: p < \frac{1}{2}$ (Heads is shown for less than half of the times)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

4 throws of coin, $n = 4$
Coin shows Heads 0 times.

Collect data

Let X be the number of Heads shown.
If H_0 is true then we are using $p = \frac{1}{2}$ and $X \sim B(4, 0.5)$

Conduct the test

$$P(X = 0) = {}^4C_0 \times 0.5^0 \times 0.5^4$$
$$= 0.0625$$

Interpret the results

Since $0.0625 > 0.05$ we accept H_0.
There is not sufficient evidence at 5% level to show that the coin is biased.

(Think of 0.0625 as a large probability compared to 5%, therefore the event is quite likely if p is $\frac{1}{2}$)

Example 2
A student takes a multiple-choice test. There are 10 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 5 questions right.
Is the student's method of missing lessons and guessing a good strategy?
The student claims to be an inspired guesser. Is this true?

Solution

Let p be the probability the student guesses correctly

Establish the null and alternative hypotheses

H_0: $p = \frac{1}{4}$ (equal chances of choosing each response)

H_1: $p > \frac{1}{4}$ (guesses correct in more than quarter of the questions)

Significance level set as 5%

Decide on the significance level

10 questions attempted, $n = 10$

Guesses correct 5 times.

Collect data

Let X be the number of correct guesses.

If H_0 is true then we are using $p = \frac{1}{4}$ and $X \sim B(10, \frac{1}{4})$

Conduct the test

$P(X \geq 5) = 1 - P(X \leq 4)$

$= 1 - 0.9219 = 0.0781$

We are investigating the probability of guessing correctly 5 times out of 10.
But guessing 6, 7, 8, 9 or 10 times would be an even more surprising result.
So we calculate the tail probability $P(X \geq 5)$.
This is the *p*-value for the test

Interpret the results

Since $0.0781 > 0.05$ we accept H_0.

There is not sufficient evidence at the 5% level that the student is performing any better than through random guessing; he/she would be advised to attend lessons.

Think of 0.0781 as a large probability compared to 5%, therefore the event is quite likely if p is $\frac{1}{4}$

IMPORTANT

It is important that in examples like this one you find the probability of obtaining **the trial result or a more extreme result**, in this case $P(X \geq 5)$ rather than $P(X = 5)$. This is a key concept in this chapter: when conducting a hypothesis test look for a region of probabilities.

Example 3

The probability that a student at a certain school passes an exam is 0.8.

A new teacher is appointed and takes the next group. Out of a group of 10 students only 6 students now pass.

Should the Head of Maths be concerned?

Establish the null and alternative hypotheses

Solution
Let p be the probability a student passes the exam
H_0: $p = 0.8$ (probability a student passes the exam is 0.8)
H_1: $p < 0.8$ (probability a student passes the exam has been lowered)

Significance level set as 5%

Decide on the significance level

10 students take the exam, $n = 10$
6 students pass.

Collect data

Let X be the number of students passing.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(10, 0.8)$
$P(X \leq 6) = 0.1209$

Conduct the test

We are investigating the probability of getting 6 passes out of 10.
But getting 0, 1, 2, 3, 4 or 5 would be an even more surprising result.
So we calculate the tail probability $P(X \leq 6)$

Interpret the results

The p-value for the test is 0.1209.
Since $0.1209 > 0.05$ we accept H_0.
There is insufficient evidence at the 5% level that the students are under-performing with the new teacher, based on this sample of students.

Think of 0.1209 as a large probability compared to 5%, therefore the event is quite likely if p is 0.8

Examples using larger samples

In examples 2 and 3 we calculated tail probabilities, i.e. a region.
Note: in example 1 we were calculating $P(X = 0)$ which of course is the same as $P(X \leq 0)$.

So although we may be surprised with the number of times the cricket captain guessed wrong, the number of questions the student got right and the number of students who passed the exam in examples 1, 2 and 3 respectively, the results turned out to be not significant: in all three cases the null hypothesis was accepted.

However, in all these cases the sample sizes were relatively small. What happens if we investigate results from larger samples?

Example 4
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 10 consecutive times. Andy complains that the coin must be biased against Heads. Is his complaint justified?

Example 5
A student takes a multiple-choice test. There are 20 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 10 questions correct.
The student claims to be an inspired guesser. Is this true?

Example 6
The probability that a student at a certain school passes an exam is 0.8.
A new teacher is appointed and takes the next group. Out of a group of 20 students only 12 students now pass.
Should the Head of Maths be concerned?

Note the proportions of success in examples 5 and 6 have stayed the same as in 2 and 3, i.e. 50% and 60% respectively. Do we come to the same conclusions?

Example 4
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 10 consecutive times. Andy complains that the coin must be biased against Heads. Is his complaint justified?

Solution
Let p be the probability of getting Heads
$H_0: p = \frac{1}{2}$ (equal chances of getting Heads or Tails)
$H_1: p < \frac{1}{2}$ (Heads shows in less than half of the times)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

10 throws of coin, $n = 10$
Heads shown 0 times.

Collect data

Let X be the number of times Heads is shown.
If H_0 is true then we are using $p = \frac{1}{2}$ and $X \sim B(10, 0.5)$
$P(X = 0) = {}^{10}C_0 \times 0.5^0 \times 0.5^{10}$
$= 0.0010$ (4 s.f.)

Conduct the test

The p-value is 0.0010
Since $0.0010 < 0.05$ we reject H_0.
The evidence suggests that the coin is biased against Heads.

Interpret the results

Think of 0.0010 as a very small probability compared with 5%, therefore it is very unlikely that p is $\frac{1}{2}$

Example 5

A student takes a multiple-choice test. There are 20 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 10 questions correct.

The student claims to be an inspired guesser. Is this true?

Solution

Establish the null and alternative hypotheses

Let p be the probability the student guesses correctly.

H_0: $p = \frac{1}{4}$ (equal chances of choosing each response)

H_1: $p > \frac{1}{4}$ (guesses correct in more than quarter of the questions)

Significance level set as 5%

Decide on the significance level

20 questions attempted, $n = 20$

Guesses correct 10 times.

Collect data

Let X be the number of correct guesses.

If H_0 is true then we are using $p = \frac{1}{4}$ and $X \sim B(20, \frac{1}{4})$

Conduct the test

$P(X \geq 10) = 1 - P(X \leq 9)$

$= 1 - 0.9861 = 0.0139$

We are investigating the probability of guessing correctly 10 times out of 20. But guessing 11, 12, 13, ….20 times would be an even more surprising result.

So we calculate the tail probability $P(X \geq 10)$.

Interpret the results

Since $0.0139 < 0.05$ we reject H_0.

There is evidence to suggest that the student is an inspired guesser!

Think of 0.0139 as a very small probability compared to 5%, therefore it is very unlikely that p is $\frac{1}{4}$

Example 6

The probability that a student at a certain school passes an exam is 0.8.

A new teacher is appointed and takes the next group. Out of a group of 20 students only 12 students now pass.

Should the Head of Maths be concerned?

Solution

Let p be the probability a student passes the exam

H_0: $p = 0.8$ (probability a student passes the exam is 0.8)

H_1: $p < 0.8$ (probability a student passes the exam has been lowered)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

20 students take S1, $n = 20$

12 students pass.

Collect data

Let X be the number of students passing.

If H_0 is true then we are using $p = 0.8$ and $X \sim B(20, 0.8)$

Conduct the test

$P(X \leq 12) = 0.0321$

We are investigating the probability of getting 12 passes out of 10.But getting 0, 1, 2, 3,….,11 would be an even more surprising result.
So we calculate the tail probability $P(X \leq 12)$

Interpret the results

Since $0.0321 < 0.05$ we reject H_0.

There is evidence to suggest that the students are under performing with the new teacher, based on this sample of students.

Think of 0.0321 as a very small probability, therefore it is very unlikely that p is 0.8

However, there could be several other factors: did the students also perform badly in other subjects, was it a very difficult exam paper etc?

Notice that the outcomes in examples 4, 5 and 6 are now all significant.

As sample sizes get larger, the conclusion may change even if the ratio of success stays the same. The chance of guessing 10 or more questions correct out of 20 is much smaller than the chance of guessing 5 or more questions correct out of 10.

IMPORTANT

Notice from the examples above that the conclusion should always be given in terms of the problem. First state whether H_0 is to be accepted or rejected, then make a statement beginning “there is evidence to suggest that …” or “there is not sufficient evidence to suggest that …”. You should **NOT** write “this proves that ….” or “so the claim is right”. You are not proving anything, only considering evidence.

Significance levels

When conducting hypothesis tests there are two types of error that occur.

In a situation where H_0 is correct, we may reject H_0.
The probability of making this error is equal to the significance level.
The other error is in a situation where we should reject H_0 but we accept it.

If we increase the significance level we will increase the chance of rejecting H_0 when we should be accepting it.

Further example

Example 7
Using recent data provided by the low-cost airline Brianair, the probability of a flight arriving on time is estimated to be 0.9.

On three different occasions I am taking a flight with Brianair.

(i) What is the probability that I arrive on time on all 3 flights?
(ii) What is the probability that I arrive on time on exactly 2 occasions?
(iii) After some rescheduling, Brianair state that their performance has improved. In a recent survey 19 out of 20 flights arrived on time. Using a significance level of 5%, is Brianair's conclusion correct?

Solution
Let X be the number of times a flight is on time.
$n = 3$, $p = 0.9$, $q = 0.1$
$X \sim \mathrm{B}(3, 0.9)$.

(i)
$$\begin{aligned} \mathrm{P}(X=3) &= {}^3C_3 \times 0.9^3 \times 0.1^0 \\ &= 0.9^3 \\ &= 0.729 \end{aligned}$$

(ii)
$$\begin{aligned} \mathrm{P}(X=2) &= {}^3C_2 \times 0.9^2 \times 0.1^1 \\ &= 3 \times 0.9^2 \times 0.1 \\ &= 0.243 \end{aligned}$$

(iii) Let p be the probability a flight is on time
H_0: $p = 0.9$ (probability a flight is on time is 0.9)
H_1: $p > 0.9$ (probability a flight is on time has increased)

Significance level set as 5%

20 flights, $n = 20$
19 on time

Let X be the number of times a flight is on time.
If H_0 is true then we are using $p = 0.9$ and $X \sim B(20, 0.9)$.

$$\begin{aligned} P(X \geq 19) &= 1 - P(X \leq 18) \\ &= 1 - 0.6083 \\ &= 0.3917 \end{aligned}$$

Since $0.3917 > 0.05$ we accept H_0.

There is not sufficient evidence to suggest that Brianair's performance has improved.
They need to take a larger sample size to investigate further.

Note that 95% landed on time, but the test shows that this evidence is not sufficient to suggest that there has been an improvement.

