Lecture 3

HYPOTHESIS TESTS

Announcement

CBA4 has already gone “live” in **practice mode**.

This will go “live” in **assessed mode** next Monday.

Deadline: Next Friday, 24th February – but really, the “ultimate deadline” is 23:59:59 on **Sunday 26th February**.

Introduction

We have seen that confidence intervals can be used to make inferences about population parameters.

Sometimes, you may be asked to assess whether or not a parameter takes a specific value. For example: whether the population mean $\mu = 5$.

One way of re-expressing this question is to ask whether the parameter value is plausible in light of the data.

A simple check to see whether the value is contained in a 95% confidence interval will provide an answer.

An alternative method, called a **hypothesis test**, is available. It is used extensively in reporting experimental results.

Introduction

A hypothesis test is a rule for establishing whether or not a set of data is consistent with a hypothesis about a parameter of interest.

The **null hypothesis** is a statement that a parameter has a certain value, and is usually written as H_0. For example:

- $H_0 : \mu = 2.7$
- $H_0 : \sigma^2 = 12.4$.

If the null hypothesis is not true, what alternatives are there?

Usually, the **alternative hypothesis** is written as H_1. Examples include:

- $H_1 : \mu \neq 2.7$
- $H_1 : \sigma^2 \neq 12.4$.

Introduction

Based on the information we have in our sample, we'd like to go with either the null hypothesis or the alternative hypothesis.

We might use our sample as evidence to suggest that, for example, the population mean could well be equal to 2.7; alternatively, the sample might give evidence to the contrary and suggest that the population mean is not equal to 2.7.

Quick example

Suppose you are going on holiday to Sicily in March.

A friend tells you that in March, Sicily has an average of **10 hours** sunshine a day.

On the first three days of your holiday there are **7**, **8** and **9** hours of sunshine respectively.

You consider that this is evidence that your friend is wrong.

Quick example

Thus,

- the null hypothesis would state that the average sunshine hours per day is 10
- your alternative hypothesis might state that the average sunshine hours per day *is less than* 10

In symbols,

- $H_0 : \mu = 10$ versus
- $H_1 : \mu < 10$

Quick example

You might be tempted to go with the alternative hypothesis.

However, this sample of three days could be a fluke result – you might have chosen the most miserable period in March for years for your holiday.

Your test results are not conclusive; they only give you evidence for or against a particular belief.

Methodology for hypothesis testing

All hypothesis tests follow the same basic methodology, although the actual calculations may vary depending on the data available.

1. **State the null hypothesis ($\mathbf{H_0}$)**
   We use a hypothesis test to throw light on whether or not this statement is true. For example, you might ask "**is the population mean equal to 10?**", or "**are the two population means equal?**"; such hypotheses are expressed in the following way:

$$
\begin{aligned}
H_0 &: \mu = 10, \qquad \text{and} \\
H_0 &: \mu_1 = \mu_2; \qquad \text{or maybe} \\
H_0 &: \mu = c,
\end{aligned}
$$

   Where c could be any constant.

2. **State the alternative hypothesis (H_1)**
   This is the conclusion to be reached if the null hypothesis is found to be false. For example, "**the population mean does not equal 10**", or even "**the population mean is less than 10**"; in symbols:

$$H_1 \quad : \quad \mu \neq 10 \qquad \text{or maybe}$$
$$H_1 \quad : \quad \mu < 10.$$

   To test for two different populations, we might say "**the two population means are different**". In symbols:

$$H_1 \quad : \quad \mu_1 \neq \mu_2.$$

Methodology for hypothesis testing

3. **Calculate the test statistic**
   The value calculated from the sample which is used to perform the test is called the **test statistic**.

   It usually has a similar nature to the population value mentioned in the null hypothesis.

Methodology for hypothesis testing

4. **Find the p–value of the test**

   The probability that such an extreme test statistic occurs, *assuming that* H_0 *is true*, is called the **p–value**.

   This can be found by comparing the test statistic to values from statistical tables.

5. **Reach a conclusion**

A small p–value suggests that our test statistic is unlikely to occur if H_0 is true, and so we reject H_0 in favour of the alternative H_1.

p–value	Interpretation
p is bigger than 10%	no evidence against the null hypothesis: stick with H_0
p lies between 5% and 10%	*slight* evidence against H_0, but not enough to reject it
p lies between 1% and 5%	moderate evidence against H_0: reject it, and go with H_1
p is smaller than 1%	strong evidence against H_0: reject it, and go with H_1

Testing one mean

Here, from a **single** population, we draw a **single** sample, and we estimate the population mean μ with the sample mean $\bar{x}$.

We'd then like to see how convincing a proposal (say c) for the population mean is, based on the information in our sample.

As with the construction of confidence intervals, the choice of test statistic in step 3 above depends on whether or not the population variance is known.

We will now demonstrate a test for one mean using two examples: one in which the population variance is **known** (case 1), and one in which it is **unknown** (case 2).

Case 1: Population variance known: Example

A chain of shops believes that the average size of transactions is £130, and the population variance is known to be £900.

The takings of one branch were analysed and it was found that the mean transaction size was £123 over the 100 transactions in one day.

Based on this sample, test the null hypothesis that the true mean is equal to £130.

Case 1: Population variance known: Example

Since σ^2 is known (we are given that $\sigma^2 = 900$), this corresponds to case 1: population variance known (think back to confidence intervals).

We now proceed with the five steps outlined in the previous section.

Steps 1 and 2 *(hypotheses)*

Here, we state our null and alternative hypotheses. The null hypothesis is given in the question – i.e.

$$H_0 \quad : \quad \mu = £130.$$

We could test against a general alternative, i.e.

$$H_1 \quad : \quad \mu \neq £130.$$

Case 1: Population variance known: Example

Step 3 *(calculating the test statistic)*

When σ^2 is known, we use following test statistic

$$z = \frac{|\bar{x} - \mu|}{\sqrt{\sigma^2/n}}, \qquad \text{i.e.}$$

$$\begin{aligned} z &= \frac{|123 - 130|}{\sqrt{900/100}} \\ &= \frac{7}{\sqrt{9}} \\ &= 2.33. \end{aligned}$$

Case 1: Population variance known: Example

Step 4 *(finding the p–value)*

Recall the **Central Limit Theorem** from Chapter 1; this tells us that the quantity

$$\frac{\bar{x} - \mu}{\sqrt{\sigma^2 / n}}$$

follows a standard Normal distribution.

Thus, the value we obtain from our test statistic formula above will be from the positive half of the standard Normal distribution.

Case 1: Population variance known: Example

We can therefore compare our test statistic to critical values from the standard Normal distribution to find our p–value, or at least a range for our p–value.

Remember, this is the probability of observing our data, or anything more extreme than this, if the null hypothesis is true; thus, the smaller the p–value, the more evidence there is **against** H_0.

Case 1: Population variance known: Example

Our alternative hypothesis is **two–tailed** (i.e. $\neq$ rather than $<$ or $>$), and so our values are:

Significance level	10%	5%	1%
Critical value	1.645	1.96	2.576

Our test statistic $z = 2.33$ lies between the critical values of 1.96 and 2.576, and so our p–value lies **between 1% and 5%**.

We can see this more clearly on a diagram.

Case 1: Population variance known: Example

Step 5 *(conclusion)*

Using table 2.1 to interpret our p–value, we see that:

- There is **moderate** evidence against H_0
- Thus, we should **reject** H_0 in favour of H_1
- It appears that the population mean transaction size **is not equal to** £130

Case 1: Population variance known: Example

Alternatively, since our sample mean $\bar{x} = £123$ is smaller than the proposed value of £130, we could have set up a **one–tailed** alternative hypothesis:

$$H_0 : \mu = £130 \quad \text{against}$$
$$H_1 : \mu < £130.$$

This is now a one–tailed test and the critical values from table 2.2 are

Significance level	10%	5%	1%
Critical value	1.282	1.645	2.326

Case 1: Population variance known: Example

The test statistic is (as before) 2.33, which now lies "to the right" of the last critical value in the table (2.326).

Thus, our p–value is now smaller than 1%, and so, using table 2.1, we see that

- there is **strong** evidence against H_0
- We would **reject** H_0 in favour of H_1
- There is evidence to suggest that the population mean **is less than** £130

Case 2: Unknown population variance: Example

The batteries for a fire alarm system are required to last for 20000 hours before they need replacing.

16 batteries were tested; they were found to have an average life of 19500 hours and a standard deviation of 1200 hours.

Perform a hypothesis test to see if the batteries do, on average, last for 20000 hours.

Case 2: Unknown population variance: Example

Steps 1 and 2 *(hypotheses)*

Using a **one–tailed test**, our null and alternative hypotheses are:

$$\begin{aligned} H_0 &: \mu = 20000 \qquad \text{versus} \\ H_1 &: \mu < 20000. \end{aligned}$$

We use a one–tailed test because we are interested in whether the batteries are effective or not; there is no problem if they last longer than 20000 hours.

Case 2: Unknown population variance: Example

Step 3 *(calculating the test statistic)*

Unlike the previous example, the population variance σ^2 is unknown.

However, the **sample** standard deviation is given, based on a sample of size 16, and so we need to use a slightly different test statistic.

In fact, we do what we did last week when we were constructing confidence intervals – i.e. we replace σ^2 with s^2 and then use tables of values from Student's t distribution instead of the standard Normal distribution.

Case 2: Unknown population variance: Example

Thus, the test statistic is given by

$$\begin{aligned} t &= \frac{|\bar{x} - \mu|}{\sqrt{s^2/n}} \\ &= \frac{|19500 - 20000|}{\sqrt{1200^2/16}} \\ &= \frac{500}{\sqrt{1440000/16}} \\ &= 1.667. \end{aligned}$$

Case 2: Unknown population variance: Example

Step 4 *(finding the p–value)*

Since σ^2 is unknown, we use t–distribution tables (table 2.3) to obtain a range for our p–value.

The degrees of freedom, $\nu = n - 1 = 16 - 1 = 15$, and under a one–tailed test this gives the following critical values:

Significance level	10%	5%	1%
Critical value	1.341	1.753	2.602

Our test statistic of $t = 1.667$ lies between the critical values of 1.341 and 1.753, and so the corresponding p–value lies between 5% and 10%.

Case 2: Unknown population variance: Example

Step 5 *(conclusion)*

Using table 2.1 to interpret our p–value, we see that there is

- only **slight** evidence against the null hypothesis
- this is not enough grounds to reject it, so we **retain** H_0
- There is insufficient evidence to suggest there is a problem with these batteries.