Lecture 2

HYPOTHESIS TESTS

Introduction

Recall that **confidence intervals** can be used to make inferences about the population mean μ:

- A confidence interval is "**better**" than a point estimate on it's own – we now have a *range* of values for μ;
- If the confidence intervals for two (independent) samples **overlap**, then we can be 'confident' that there is no real difference between the population means;
- If a confidence interval for μ **captures** a 'target value', then we can be 'confident' that the population mean might be equal to this value.

An alternative approach to statistical inference is through

hypothesis testing

Hypothesis tests

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.

We have two hypotheses:

- the **null hypothesis** (H_0), and
- the **alternative hypothesis** (H_1)

and we want to try to "prove"/"disprove" the null hypothesis.

- If we "prove" the null hypothesis to be correct, then the alternative is discarded.
- Conversely, if we "disprove" the null hypothesis, we have an alternative to go with!

An illustrative example

Suppose you are going on holiday to Sicily in March. Before you leave, a friend tells you that Sicily has an average of **10** hours sunshine a day during March.

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.

Thus, the null hypothesis would state that the average sunshine hours per day is 10 (as suggested by your friend):

$$H_0 \quad : \quad \mu = 10$$

An illustrative example

Your alternative hypothesis might state that the average sunshine hours per day *is less than* 10:

$$H_1 \quad : \quad \mu < 10$$

You might be tempted to go with the alternative hypothesis, since on your first three days you have observed less than 10 hours of sunshine.

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.

Other possible alternative hypotheses could have been:

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

Decision time...

We need to go with either the null or the alternative, and we can use statistics to help us decide!

Hypothesis tests can help us to choose between the null and the alternative.

Methodology for hypothesis testing

All hypothesis tests follow the same **basic framework**:

1. State the null hypothesis

Say what your null hypothesis *is*! If you're asked to test if the population mean is equal to 10, for example, then you'd write

$$H_0 \quad : \quad \mu = 10.$$

Or if you're asked to find out if two population means are equal, you'd write

$$H_0 \quad : \quad \mu_1 = \mu_2$$

2. State the alternative hypothesis

This is what else could happen! For example, you could use

$$H_1 \quad : \quad \mu \neq 10;$$

thus is known as a **two–tailed** or **general** alternative.

Or if your sample suggests that the mean is considerably lower, you might use

$$\begin{aligned} H_1 \quad &: \quad \mu < 10, \qquad \text{or perhaps} \\ H_1 \quad &: \quad \mu > 10 \end{aligned}$$

if you think it might be higher. These are known as **one–tailed** alternatives.

3. Calculate the test statistic

We use the data to calculate this. It usually has a similar nature to the population value in the null hypothesis.

4. Find the p–value

This is the probability of observing the data we have **if the null hypothesis is true**.

Obviously, if this probability is small, then the alternative hypothesis is more likely to be true.

You'll see how to obtain this value shortly, and we'll discuss what we mean by a "small" p–value.

5. Form a conclusion

Once we've used our test statistic to obtain our p–value, the following conventions can be used:

p–value	Interpretation
$p > 10\%$	no evidence against H_0
p lies between 5% and 10%	*slight* evidence against H_0
p lies between 1% and 5%	*moderate* evidence against H_0
$p < 1\%$	*strong* evidence against H_0

Write a sentence or two that's in the context of the question!

Testing one mean

From a **single** population we draw a **single sample**. We'd like to use the information in this sample to see how convincing a proposal for the population mean is.

Just like in the construction of confidence intervals, there are two situations which will determine the details of the test:

- The population variance (σ^2) is **known**;
- The population variance (σ^2) is **unknown**.

Case 1: Known population variance σ^2

If the population variance is *known*, we:

1. **State the null hypothesis**

$$H_0 \quad : \quad \mu = c.$$

2. **State the alternative hypothesis**
   We have three options:

$$\begin{aligned} H_1 &: \quad \mu \neq c; \\ H_1 &: \quad \mu > c, \qquad \text{or} \\ H_1 &: \quad \mu < c. \end{aligned}$$

3. **Calculate the test statistic**
   For this test, the test statistic is

$$z \quad = \quad \frac{|\bar{x} - \mu|}{\sqrt{\sigma^2/n}}.$$

4. Find the p–value

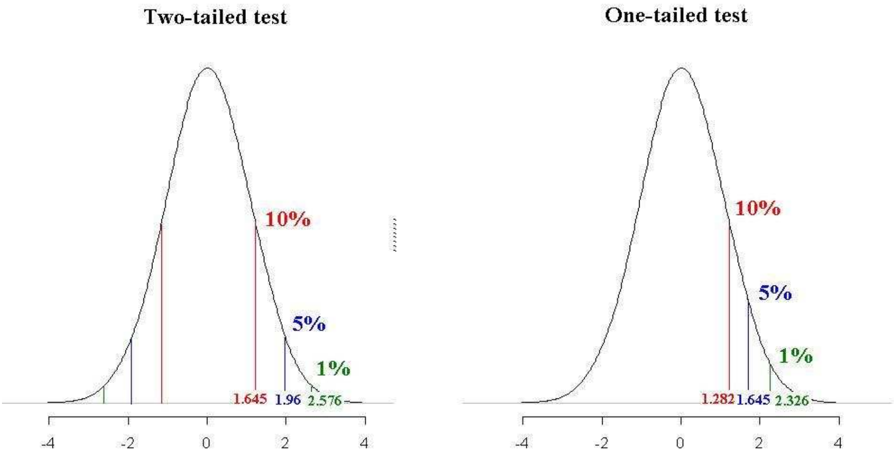
We use statistical tables for this. Just like for confidence intervals, since σ^2 is known, we can use standard normal tables (table 2.2 in notes).

For a two–tailed test, we have the following **critical values**:

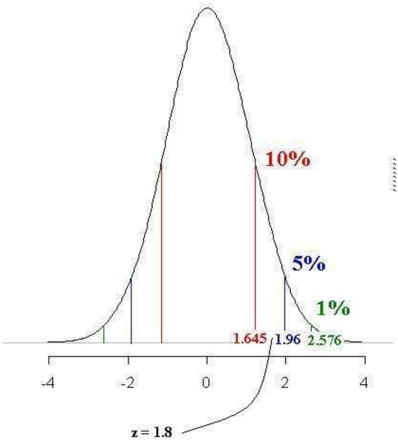
Significance level	**10%**	**5%**	**1%**
Critical value	1.645	1.96	2.576

Similarly, for a one–tailed test we have:

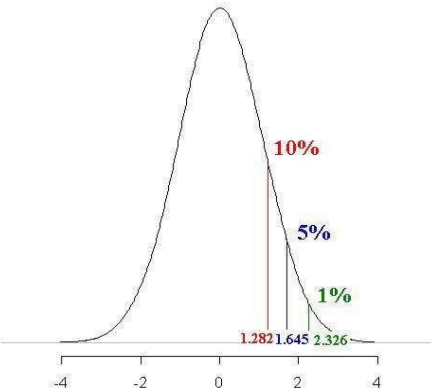
Significance level	**10%**	**5%**	**1%**
Critical value	1.282	1.645	2.326

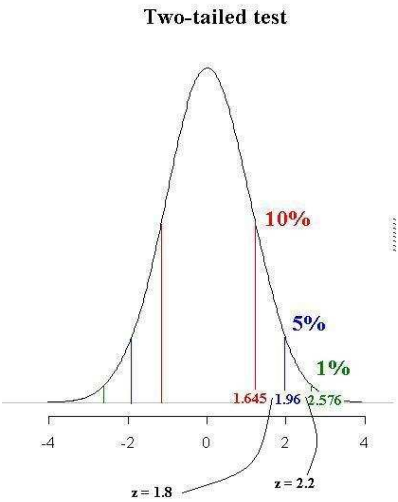

Two-tailed test
10%
5%
1%
1.645
1.96
2.576
-4
-2
0
2
4
One-tailed test
10%
5%
1%
1.282
1.645
2.326
-4
-2
0
2
4

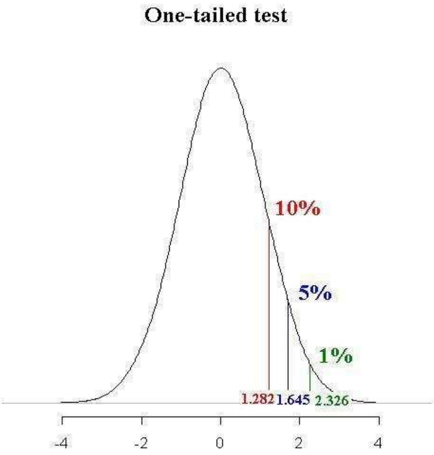
Two-tailed test



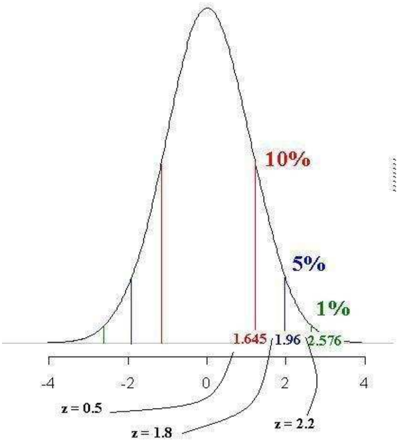
One-tailed test



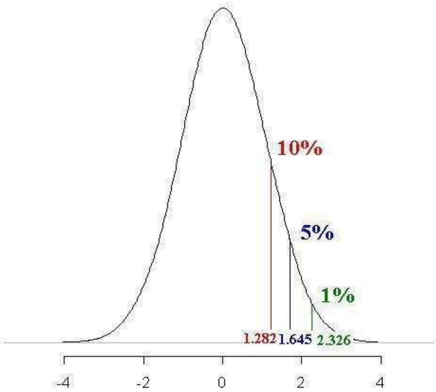

Two-tailed test
10%
5%
1%
1.645
1.96
2.576
-4
-2
0
2
4
z = 1.8
z = 2.2


One-tailed test
10%
5%
1%
1.282
1.645
2.326
-4
-2
0
2
4

Two-tailed test



One-tailed test



5. **Form a conclusion**

   Use table 2.1 in the notes to go with the null hypothesis or the alternative hypothesis!

   Don't forget to write a sentence or two in the context of the question!

Example 2.4.1 (page 18)

A chain of shops believes that the average size of transactions is £130, and the 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.

Since σ^2 is **known** (we are given that $\sigma^2 = 900$), we use the approach just discussed.

Example 2.4.1 (page 18)

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 (**two–tailed**) alternative, i.e.

$$H_1 \quad : \quad \mu \neq £130.$$

Example 2.4.1 (page 18)

Step 3 *(calculating the test statistic)*
Since σ^2 is known, the test statistic is

$$\begin{aligned} z &= \frac{|\bar{x} - \mu|}{\sqrt{\sigma^2/n}}, \qquad \text{i.e.} \\ z &= \frac{|123 - 130|}{\sqrt{900/100}} \\ &= \frac{7}{\sqrt{9}} \\ &= 2.33. \end{aligned}$$

Example 2.4.1 (page 18)

Step 4 *(finding the p–value)*
Since σ^2 is known, we use normal distribution tables (table 2.2) to obtain a range for our p–value. 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%.

Example 2.4.1 (page 18)

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 the alternative hypothesis H_1; it appears that the population mean transaction size *is not equal to* £130.

Example 2.4.1 (page 18)

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 in step 2, i.e we could have tested

$$\begin{aligned} H_0 &: \mu = £130 \qquad \text{against} \\ H_1 &: \mu < £130. \end{aligned}$$

Example 2.4.1 (page 18)

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

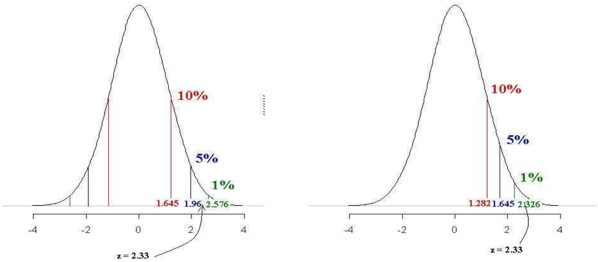
Significance level	10%	5%	1%
Critical value	1.282	1.645	2.326

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 in this more specific test there is *strong* evidence against H_0.

Example 2.4.1 (page 18)

Again, this can be seen more clearly with a diagram:



Notice that this one–tailed test is **more specific** than the two–tailed test previously carried out.

If you're not sure whether you should perform a one–tailed test or a two–tailed test, it's usually safer to test against the more general two–tailed alternative.

Unknown population variance σ^2

If the population variance is *unknown*, things are more awkward!
Steps 1 and 2 are the same, i.e.

1. **State the null hypothesis**
2. **State the alternative hypothesis**

However, since the variance is now unknown, we can no longer use the standard normal distribution. Instead, we have to use the t distribution (last week), which accounts for our unknown variance.

3. **Calculate the test statistic**
   The test statistic is now

$$t = \frac{|\bar{x} - \mu|}{\sqrt{s^2/n}}$$

   This is exactly the same as before, but

   - is now called t instead of z;
   - now has s^2 instead of σ^2.

4. Find the p–value

This is the same as when we calculated confidence intervals for the “unknown variance” situation: we need to use t–tables, with degrees of freedom $\nu = n - 1$.

For example, if our sample size was 12, then $\nu = 12 - 1 = 11$, and so, for a two–tailed test, we'd get:

Significance level	**10%**	**5%**	**1%**
Critical value	1.796	2.201	3.106

As before, we locate our test statistic in the above table to find a range for our p–value.

5. **Form a conclusion**

   Exactly the same as before – use table 2.1 to help you!

Example 2.4.2 (page 20)

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.

Example 2.4.2 (page 20)

Steps 1 and 2 *(hypotheses)*

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

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

Example 2.4.2 (page 20)

Step 3 *(calculating the test statistic)*

The population variance σ^2 is now unknown – the question does **not** say

- "the population variance is . . . "
- "the population standard deviation is . . . "
- "$\sigma = \ldots$"
- the process variance is . . . "

However, the *sample* standard deviation is given, based on a sample of size 16, and so we can use the t distribution.

Example 2.4.2 (page 20)

Thus, the test statistic is given by

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

Example 2.4.2 (page 20)

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%.

Example 2.4.2 (page 20)

Step 5 *(conclusion)*
Using table 2.1 to interpret our p–value, we see that there is only *slight* evidence against the null hypothesis and certainly not enough grounds to reject it, so we retain H_0.

It appears that, on average, the batteries *do* last for 20000 hours.