MAS1403/ACE2013

Quantitative Methods for Business Management

Statistics for Marketing and Management

Semester 2 Revision, 2007–08

Semester 2 syllabus

1. **Constructing confidence intervals**

Semester 2 syllabus

1 Constructing confidence intervals

- when the population variance is *known*;

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.

Semester 2 syllabus

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**
   - when the population variance is *known*;

Semester 2 syllabus

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**
   - when the population variance is *known*;
   - when the population variance is *unknown*.

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
3. **Hypothesis tests for two means**

Semester 2 syllabus

1. **Constructing confidence intervals**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**
   - when the population variance is *known*;
   - when the population variance is *unknown*.
3. **Hypothesis tests for two means**
   - when both population variances are *known*;

Semester 2 syllabus

1. **Constructing confidence intervals**
    - when the population variance is *known*;
    - when the population variance is *unknown*.
2. **Hypothesis tests for one mean**
    - when the population variance is *known*;
    - when the population variance is *unknown*.
3. **Hypothesis tests for two means**
    - when both population variances are *known*;
    - when both population variances are *unknown*.

4 **Goodness–of–fit tests using the χ^2 distribution**

Semester 2 syllabus

4. **Goodness–of–fit tests using the χ^2 distribution**
5. **Tests of independence using the χ^2 distribution**

Semester 2 syllabus

4. **Goodness–of–fit tests using the χ^2 distribution**
5. **Tests of independence using the χ^2 distribution**
6. **Correlation and linear regression**

Semester 2 syllabus

4. **Goodness–of–fit tests using the χ^2 distribution**
5. **Tests of independence using the χ^2 distribution**
6. **Correlation and linear regression**
    - drawing/interpreting scatterplots;

4 **Goodness–of–fit tests using the χ^2 distribution**

5 **Tests of independence using the χ^2 distribution**

6 **Correlation and linear regression**

- drawing/interpreting scatterplots;
- calculating/interpreting the correlation coefficient;

4 **Goodness–of–fit tests using the χ^2 distribution**

5 **Tests of independence using the χ^2 distribution**

6 **Correlation and linear regression**

- drawing/interpreting scatterplots;
- calculating/interpreting the correlation coefficient;
- estimating the linear regression model $Y = \alpha + \beta X + \epsilon$;

Semester 2 syllabus

4 **Goodness–of–fit tests using the χ^2 distribution**

5 **Tests of independence using the χ^2 distribution**

6 **Correlation and linear regression**

- drawing/interpreting scatterplots;
- calculating/interpreting the correlation coefficient;
- estimating the linear regression model $Y = \alpha + \beta X + \epsilon$;
- using a linear regression model to make predictions.

8 **Time series and forecasting**

8 Time series and forecasting

- *describing* a time series;

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;
- forecasting.

Semester 2 syllabus

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;
- forecasting.

9 & 10 Linear programming

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;
- forecasting.

9 & 10 Linear programming

- *formulating* linear programming problems;

Semester 2 syllabus

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;
- forecasting.

9 & 10 Linear programming

- *formulating* linear programming problems;
- displaying linear programming problems *graphically*;

Semester 2 syllabus

8 Time series and forecasting

- *describing* a time series;
- isolating the *trend*;
- isolating the *seasonal effects*;
- forecasting.

9 & 10 Linear programming

- *formulating* linear programming problems;
- displaying linear programming problems *graphically*;
- *solving* linear programming problems.

Recall from Semester 1 two ways of summarising data:

- **graphical** summaries and
- **numerical** summaries.

Recall from Semester 1 two ways of summarising data:

- **graphical** summaries and
- **numerical** summaries.

When we summarise data numerically, we usually quote one measure of **location** (or average) and one measure of **spread**.

Recall from Semester 1 two ways of summarising data:

- **graphical** summaries and
- **numerical** summaries.

When we summarise data numerically, we usually quote one measure of **location** (or average) and one measure of **spread**.

The most popular measure of location is the **sample mean**, though if our data are skewed we often prefer to use the **sample median**.

The sample mean ($\bar{x}$) is a **point estimate** of the population mean (μ).

The sample mean ($\bar{x}$) is a **point estimate** of the population mean (μ).

In semester 2, we've looked at **interval estimates**, or **confidence intervals**, for the population mean.

1. Estimation

The sample mean ($\bar{x}$) is a **point estimate** of the population mean (μ).

In semester 2, we've looked at **interval estimates**, or **confidence intervals**, for the population mean.

Remember, the sample mean will vary from sample to sample...

The sample mean ($\bar{x}$) is a **point estimate** of the population mean (μ).

In semester 2, we've looked at **interval estimates**, or **confidence intervals**, for the population mean.

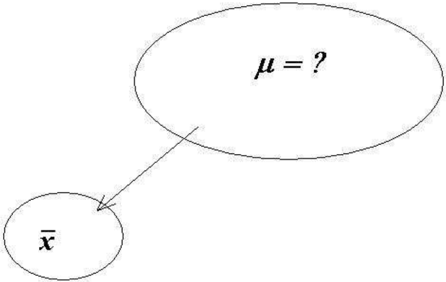
Remember, the sample mean will vary from sample to sample...

... we haven't got the resources to take many samples and see what happens to the sample mean, so we can use confidence intervals as a way of quantifying the variability of $\bar{x}$.

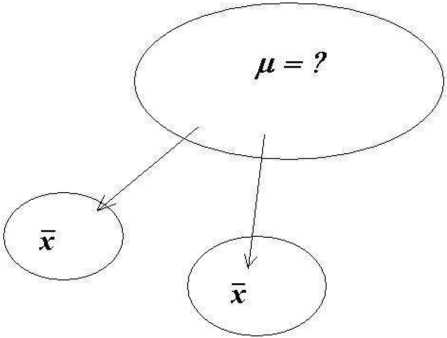
1. Estimation

$$\mu = ?$$

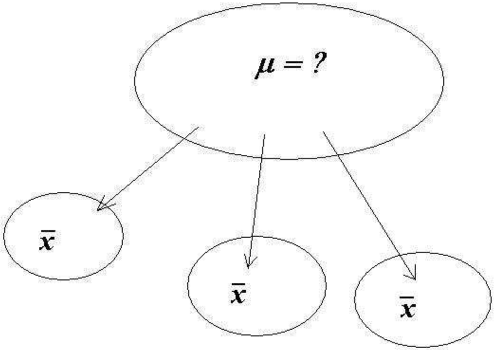
1. Estimation

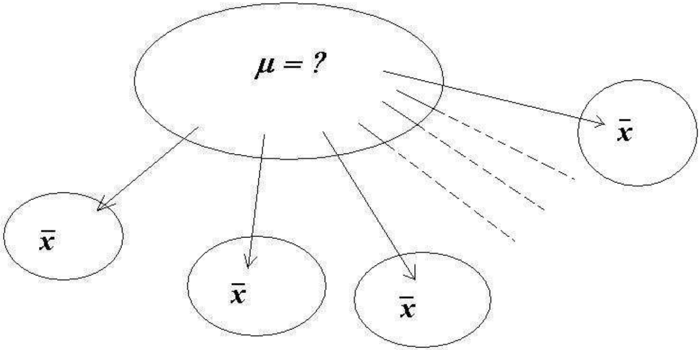


1. Estimation

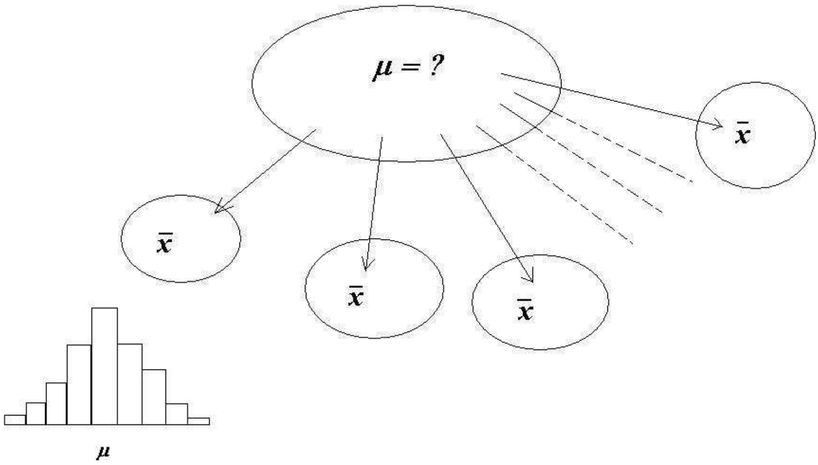


1. Estimation



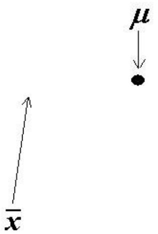

$\mu = ?$
$\bar{x}$
$\bar{x}$
$\bar{x}$
$\bar{x}$

1. Estimation

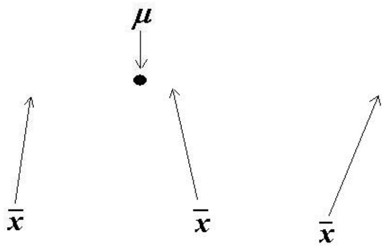


1. Estimation

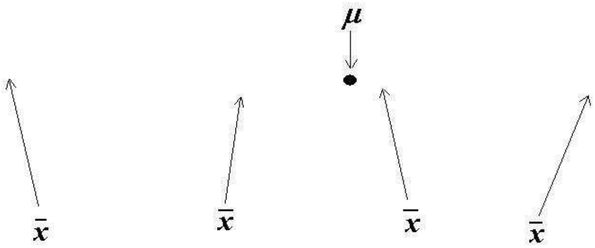
1. Estimation



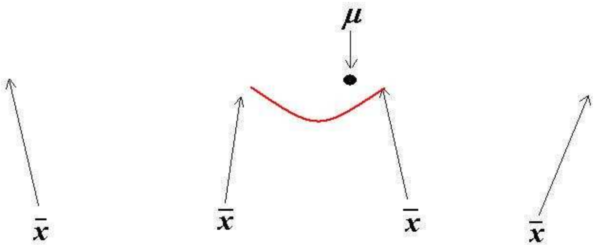
1. Estimation



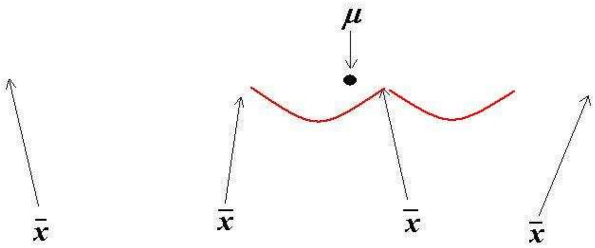
1. Estimation



1. Estimation



1. Estimation



1.1 Population variance σ^2 known

If the population variance σ^2 is **known** – i.e. the question says

- "the population variance is ..."
- "the process variability is known to be ..."
- "$\sigma^2 = \ldots$"

1.1 Population variance σ^2 known

If the population variance σ^2 is **known** – i.e. the question says

- "the population variance is . . ."
- "the process variability is known to be . . ."
- "$\sigma^2 = \ldots$"

then we have the following formulae for a 90%, 95% and 99% confidence interval for μ:

If the population variance σ^2 is **known** – i.e. the question says

- "the population variance is . . ."
- "the process variability is known to be . . ."
- "$\sigma^2 = \ldots$"

then we have the following formulae for a 90%, 95% and 99% confidence interval for μ:

– **90% confidence interval for μ**

$$\bar{x} \quad \pm \quad \mathbf{1.645} \times \sqrt{\sigma^2/n}$$

– **95% confidence interval for μ**

$$\bar{x} \quad \pm \quad \mathbf{1.96} \times \sqrt{\sigma^2/n}$$

– **99% confidence interval for μ**

$$\bar{x} \quad \pm \quad \mathbf{2.576} \times \sqrt{\sigma^2/n}$$

1.2 Population variance σ^2 unknown

If the population variance is **unknown**, then a confidence interval for μ is given by

1.2 Population variance σ^2 unknown

If the population variance is **unknown**, then a confidence interval for μ is given by

$$\bar{x} \quad \pm \quad \mathbf{t} \times \sqrt{s^2/n}$$

where t is a value obtained from t–tables (Table 1.1 in the notes).

"Aphroditair" are an internet–based budget airline offering cheap flights to the Greek islands.

From a random sample of 14 customers with Aphroditair, the mean price of flights to Kefalonia in September was £136 with a standard deviation of £25.50.

Obtain a 95% confidence interval for the average price of flights to Kefalonia in September with Aphroditair.

We have

$$\begin{aligned} n &= 14 \\ \bar{x} &= £136 \\ s &= £25.50 \end{aligned}$$

1.3 Exam–style question: solution

We have

$$\begin{aligned} n &= 14 \\ \bar{x} &= \pounds 136 \\ s &= \pounds 25.50 \end{aligned}$$

This is **case 2** since we have the *sample* standard deviation and not the *population* value. So we use

We have

$$\begin{aligned} n &= 14 \\ \bar{x} &= £136 \\ s &= £25.50 \end{aligned}$$

This is **case 2** since we have the *sample* standard deviation and not the *population* value. So we use

$$\bar{x} \quad \pm \quad t \times \sqrt{\frac{s^2}{n}} \qquad \text{i.e.}$$

$$136 \quad \pm \quad t \times \sqrt{\frac{25.50^2}{14}}$$

1.3 Exam–style question: solution

We have

$$\begin{aligned} n &= 14 \\ \bar{x} &= £136 \\ s &= £25.50 \end{aligned}$$

This is **case 2** since we have the *sample* standard deviation and not the *population* value. So we use

$$\begin{aligned} \bar{x} &\pm t \times \sqrt{\frac{s^2}{n}} \quad \text{i.e.} \\ 136 &\pm t \times \sqrt{\frac{25.50^2}{14}} \\ 136 &\pm t \times \sqrt{\frac{650.25}{14}} \end{aligned}$$

From Table 1.1, we get $t = 2.160$ (using $\nu = 14 - 1 = 13$).

Thus, we have

From Table 1.1, we get $t = 2.160$ (using $\nu = 14 - 1 = 13$).

Thus, we have

$$136 \quad \pm \quad 2.160 \times \sqrt{\frac{650.25}{14}}$$

From Table 1.1, we get $t = 2.160$ (using $\nu = 14 - 1 = 13$).

Thus, we have

$$\begin{aligned} 136 &\pm 2.160 \times \sqrt{\frac{650.25}{14}} \\ 136 &\pm 14.7207 \end{aligned}$$

giving

1.3 Exam–style question: solution

From Table 1.1, we get $t = 2.160$ (using $\nu = 14 - 1 = 13$).

Thus, we have

$$\begin{aligned} 136 &\pm 2.160 \times \sqrt{\frac{650.25}{14}} \\ 136 &\pm 14.7207 \end{aligned}$$

giving

$$\mathbf{(\pounds 121.28,\ \pounds 150.72)}$$

In Chapters 2 and 3 of the notes we looked at an alternative approach to statistical inference for the population mean through **hypothesis tests**.

2. Hypothesis tests for the mean

In Chapters 2 and 3 of the notes we looked at an alternative approach to statistical inference for the population mean through **hypothesis tests**.

We discussed two scenarios:

2. Hypothesis tests for the mean

In Chapters 2 and 3 of the notes we looked at an alternative approach to statistical inference for the population mean through **hypothesis tests**.

We discussed two scenarios:

1. **Hypothesis tests for one mean**, where one sample mean is compared to a target/proposed/hypothesised value for the population mean.

2. Hypothesis tests for the mean

In Chapters 2 and 3 of the notes we looked at an alternative approach to statistical inference for the population mean through **hypothesis tests**.

We discussed two scenarios:

1. **Hypothesis tests for one mean**, where one sample mean is compared to a target/proposed/hypothesised value for the population mean.
2. **Hypothesis tests for two means**, where two sample means are directly compared to one another.

2. Hypothesis tests for the mean

In Chapters 2 and 3 of the notes we looked at an alternative approach to statistical inference for the population mean through **hypothesis tests**.

We discussed two scenarios:

1. **Hypothesis tests for one mean**, where one sample mean is compared to a target/proposed/hypothesised value for the population mean.
2. **Hypothesis tests for two means**, where two sample means are directly compared to one another.

Remember, the aim of such hypothesis tests is to use the information in our sample(s) to make conclusions about our population(s).

In this section, we focus on tests for **one** mean.

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

2. Hypothesis tests for the 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 for the population mean is, based on the information in our sample.

2. Hypothesis tests for the 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 for the population mean is, based on the information in our sample.

Recall that the details of the test will depend on whether we:

1. know the population variance/standard deviation (**case 1**)
2. Don't know the population variance/standard deviation but can estimate it with the *sample equivalent* (**case 2**)

Steps 1 and 2 *(hypotheses)*
The null hypothesis is

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

Steps 1 and 2 *(hypotheses)*
The null hypothesis is

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

Options for the alternative hypothesis include:

Steps 1 and 2 *(hypotheses)*
The null hypothesis is

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

Options for the alternative hypothesis include:

$$H_1 \quad : \quad \mu < c$$

Steps 1 and 2 *(hypotheses)*
The null hypothesis is

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

Options for the alternative hypothesis include:

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

Steps 1 and 2 *(hypotheses)*
The null hypothesis is

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

Options for the alternative hypothesis include:

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

Step 3 *(calculate the test statistic)*

Step 3 *(calculate the test statistic)*

Do we have **case 1** or **case 2**?

Step 3 *(calculate the test statistic)*

Do we have **case 1** or **case 2**?

If case 1, then the test statistic is

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

Step 3 *(calculate the test statistic)*

Do we have **case 1** or **case 2**?

If case 1, then the test statistic is

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

If case 2, the test statistic is

Step 3 *(calculate the test statistic)*

Do we have **case 1** or **case 2**?

If case 1, then the test statistic is

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

If case 2, the test statistic is

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

Step 4 *(find the p–value)*

Step 4 *(find the p–value)*

Draw a table which looks like this:

Step 4 *(find the p–value)*

Draw a table which looks like this:

p–value	10%	5%	1%
critical value			

Step 4 *(find the p–value)*

Draw a table which looks like this:

p–value	10%	5%	1%
critical value			

and obtain a range for your p–value. Remember, if we have case 2, $\nu =$

Step 4 *(find the p–value)*

Draw a table which looks like this:

p–value	10%	5%	1%
critical value			

and obtain a range for your p–value. Remember, if we have case 2, $\nu = n - 1$.

Step 5 *(form your conclusion)*

Step 5 *(form your conclusion)*

To get full marks, you need to

Step 5 *(form your conclusion)*

To get full marks, you need to

1. Use your p-value to state the **strength of evidence** against H_0 (see Table 2.1)

Step 5 *(form your conclusion)*

To get full marks, you need to

1. Use your p-value to state the **strength of evidence** against H_0 (see Table 2.1)
2. Say whether you're going to **keep or reject** H_0

Step 5 *(form your conclusion)*

To get full marks, you need to

1. Use your p-value to state the **strength of evidence** against H_0 (see Table 2.1)
2. Say whether you're going to **keep or reject** H_0
3. Write a sentence in the **context of the question**

3. Tests for two means

Here, we compare the means from two independent samples instead of comparing a single sample mean to a hypothesised value.

3. Tests for two means

Here, we compare the means from two independent samples instead of comparing a single sample mean to a hypothesised value.

Again, we have two situations to consider: the case when *both* population variances are known, and the case when *both* are unknown.

Steps 1 and 2 *(hypotheses)*
The null hypothesis is always

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

Steps 1 and 2 *(hypotheses)*
The null hypothesis is always

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

If the question wants to know if one population mean is bigger than the other, the alternatives might be

$$\begin{aligned} H_1 \quad &: \quad \mu_1 > \mu_2 \qquad \text{or} \\ H_1 \quad &: \quad \mu_1 < \mu_2. \end{aligned}$$

Steps 1 and 2 *(hypotheses)*
The null hypothesis is always

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

If the question wants to know if one population mean is bigger than the other, the alternatives might be

$$\begin{aligned} H_1 \quad &: \quad \mu_1 > \mu_2 \qquad \text{or} \\ H_1 \quad &: \quad \mu_1 < \mu_2. \end{aligned}$$

These are examples of **one–tailed alternatives**.

Steps 1 and 2 *(hypotheses)*
The null hypothesis is always

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

If the question wants to know if one population mean is bigger than the other, the alternatives might be

$$\begin{aligned} H_1 &\quad : \quad \mu_1 > \mu_2 \qquad \text{or} \\ H_1 &\quad : \quad \mu_1 < \mu_2. \end{aligned}$$

These are examples of **one–tailed alternatives**.

Otherwise, if we are just testing for a general difference between two populations, we have the **two–tailed alternative**

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

Step 3 *(calculate the test statistic)*

Step 3 *(calculate the test statistic)*

If we have **case 1** (i.e. both population variances known), the test statistic is

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Step 3 *(calculate the test statistic)*

If we have **case 1** (i.e. both population variances known), the test statistic is

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

If we have **case 2** (i.e. both population variances unknown), the test statistic is

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where s is the *pooled standard deviation*.

Step 4 *(find the p–value)*

Step 4 *(find the p–value)*
As always, draw a table up and locate your test statistic!
Remember, if we have case 2, $\nu =$

Step 4 *(find the p–value)*
As always, draw a table up and locate your test statistic!
Remember, if we have case 2, $\nu = n_1 + n_2 - 2$.

p–value	10%	5%	1%
critical value			

Step 4 *(find the p–value)*
As always, draw a table up and locate your test statistic!
Remember, if we have case 2, $\nu = n_1 + n_2 - 2$.

p–value	10%	5%	1%
critical value			

Step 5 *(form your conclusion)*
Interpretation of the p–value is exactly the same as always.

Step 4 *(find the p–value)*
As always, draw a table up and locate your test statistic!
Remember, if we have case 2, $\nu = n_1 + n_2 - 2$.

p–value	10%	5%	1%
critical value			

Step 5 *(form your conclusion)*
Interpretation of the p–value is exactly the same as always.

Remember to write a sentence in the **context of the question**.

Exam–style question

"El Cheapo" are another internet–based flight company also offering cheap flights to Greece.

From a random sample of **16** customers with El Cheapo, the mean price of flights to Kefalonia in September was **£120** with a standard deviation of **£28.30**.

Test the null hypothesis that there's no difference between the price of flights to Kefalonia between Aphroditair and El Cheapo.

*[The pooled standard deviation for this problem is £**27.04**].*

From the two samples (Aphroditair and El Cheapo) we have:

	n	$\bar{x}$	s
Aphroditair	14	136	25.50
El Cheapo	16	120	28.30

From the two samples (Aphroditair and El Cheapo) we have:

	n	$\bar{x}$	s
Aphroditair	14	136	25.50
El Cheapo	16	120	28.30

We also know that the pooled standard deviation is 27.04.

Steps 1 and 2 *(hypotheses)*

We have

$$\begin{aligned} H_0 &: \mu_A = \mu_E \qquad \text{versus} \\ H_1 &: \mu_A \neq \mu_E \end{aligned}$$

Step 3 *(test statistic)*

This is case 2, since we have the sample standard deviations and not the population values.

Thus the test statistic is

Step 3 *(test statistic)*

This is case 2, since we have the sample standard deviations and not the population values.

Thus the test statistic is

$$t = \frac{|\bar{x}_A - \bar{x}_E|}{s \times \sqrt{\frac{1}{n_A} + \frac{1}{n_E}}}$$

Step 3 *(test statistic)*

This is case 2, since we have the sample standard deviations and not the population values.

Thus the test statistic is

$$\begin{aligned} t &= \frac{|\bar{x}_A - \bar{x}_E|}{s \times \sqrt{\frac{1}{n_A} + \frac{1}{n_E}}} \\ &= \frac{|136 - 120|}{27.04 \times \sqrt{\frac{1}{14} + \frac{1}{16}}} \end{aligned}$$

Step 3 *(test statistic)*

This is case 2, since we have the sample standard deviations and not the population values.

Thus the test statistic is

$$
\begin{aligned}
t &= \frac{|\bar{x}_A - \bar{x}_E|}{s \times \sqrt{\frac{1}{n_A} + \frac{1}{n_E}}} \\
&= \frac{|136 - 120|}{27.04 \times \sqrt{\frac{1}{14} + \frac{1}{16}}} \\
&= 1.6169
\end{aligned}
$$

Step 4 *(p–value)*

From table 2.3 in the notes (using $\nu = 14 + 16 - 2 = 28$) we get

Step 4 *(p–value)*

From table 2.3 in the notes (using $\nu = 14 + 16 - 2 = 28$) we get

p–value	10%	5%	1%
critical value	1.701	2.048	2.763

Step 4 *(p–value)*

From table 2.3 in the notes (using $\nu = 14 + 16 - 2 = 28$) we get

p–value	10%	5%	1%
critical value	1.701	2.048	2.763

Thus, p is bigger than 10%.

Step 5 *(conclusion)*

- There is **no** evidence against H_0

Step 5 *(conclusion)*

- There is **no** evidence against H_0
- Thus, we **retain** H_0

Step 5 *(conclusion)*

- There is **no** evidence against H_0
- Thus, we **retain** H_0
- There is insufficient evidence to suggest a real difference in the price of flights with the two companies

5. Tests of independence

It is usually fairly obvious if you are required to perform a χ^2 test of independence, since the data will be given in the form of a **contingency table**.

It is usually fairly obvious if you are required to perform a χ^2 test of independence, since the data will be given in the form of a **contingency table**.

This test is used to determine whether or not there is any **association** between two **categorical variables**.

The steps involved in a χ^2 test for independence are:

The steps involved in a χ^2 test for independence are:

Steps 1 and 2 *(hypotheses)*

No matter what the two categorical variables are, if you are testing to see whether they are independent or not the hypotheses are

$$\begin{aligned} H_0 &: \quad \text{There is no association between the two categorical variables} \\ H_1 &: \quad \text{There \textit{is} an association between the two categorical variables.} \end{aligned}$$

Step 3 *(calculate the test statistic)*

The test statistic (as for the χ^2 goodness–of–fit test) is

$$X^2 = \sum \frac{(O - E)^2}{E},$$

where O and E represent **observed** and **expected** frequencies (respectively).

We can get directly to our expected frequencies by using the formula

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}$$

for each cell in the contingency table.

Once we have the expected frequencies, the test statistic can be calculated very easily by drawing up the following table:

Once we have the expected frequencies, the test statistic can be calculated very easily by drawing up the following table:

Observed (O)	Expected (E)	$\frac{(O-E)^2}{E}$
$\vdots$	$\vdots$	$\vdots$

and then adding up the values in the final column.

Step 4 *(find the p–value)*

Step 4 *(find the p–value)*

The 10%, 5% and 1% critical values are found in Table 4.1.

Step 4 *(find the p–value)*

The 10%, 5% and 1% critical values are found in Table 4.1. This time, the degrees of freedom is given by

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1).$$

Step 4 *(find the p–value)*

The 10%, 5% and 1% critical values are found in Table 4.1. This time, the degrees of freedom is given by

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1).$$

Step 5 *(form your conclusion)*

Step 4 *(find the p–value)*

The 10%, 5% and 1% critical values are found in Table 4.1. This time, the degrees of freedom is given by

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1).$$

Step 5 *(form your conclusion)*

This is getting boring now – exactly as before!

The following table includes data on the number of days sick leave taken by managerial and non–managerial employees of a particular organisation.

Is there an association between type of employee and number of days sick leave?

	Non–managerial	Managerial	**Total**
0–10 days	22	24	46
11–20 days	28	16	44
21 or more days	50	10	60
Total	100	50	150

Steps 1 and 2 *(hypotheses)*

H_0 : There is no association between employee type and sick leave

H_1 : There *is* an association between employee type and sick leave

Step 3 *(test statistic)*

Step 3 *(test statistic)*

For "cell 1" (0–10 days and non–managerial) we have

$$E_1 = \frac{46 \times 100}{150} = \mathbf{30.67}$$

Step 3 *(test statistic)*

For "cell 1" (0–10 days and non–managerial) we have

$$E_1 = \frac{46 \times 100}{150} = \mathbf{30.67}$$

Proceeding in the same way for the other cells, we get:

Step 3 *(test statistic)*

For "cell 1" (0–10 days and non–managerial) we have

$$E_1 = \frac{46 \times 100}{150} = \mathbf{30.67}$$

Proceeding in the same way for the other cells, we get:

	Non–managerial	Managerial
0–10 days	**30.67**	15.33
11–20 days	29.33	14.67
21+	40	20

Comparing **observed** and **expected** we have

Comparing **observed** and **expected** we have

O	E	$\frac{(O-E)^2}{E}$
22	30.67	2.451
24	15.33	4.903
28	29.33	0.060
16	14.67	0.121
50	40	2.5
10	20	5
		15.035

Comparing **observed** and **expected** we have

O	**E**	$\frac{(O-E)^2}{E}$
22	30.67	2.451
24	15.33	4.903
28	29.33	0.060
16	14.67	0.121
50	40	2.5
10	20	5
		15.035

Thus, we have $X^2 = \sum \frac{(O-E)^2}{E} = 15.035$.

5.2 Exam–style question: solution

Step 4 *(p–value)*

We use

$$\begin{aligned} \nu &= (\text{no. of rows} - 1) \times (\text{no. of columns} - 1) \\ &= (3-1) \times (2-1) \\ &= 2 \end{aligned}$$

5.2 Exam–style question: solution

Step 4 *(p–value)*

We use

$$\begin{aligned}\nu &= (\text{no. of rows} - 1) \times (\text{no. of columns} - 1) \\ &= (3-1)\times(2-1) \\ &= 2\end{aligned}$$

to obtain the following values from Table 4.1:

Step 4 *(p–value)*

We use

$$\begin{aligned}\nu &= (\text{no. of rows} - 1) \times (\text{no. of columns} - 1) \\ &= (3-1) \times (2-1) \\ &= 2\end{aligned}$$

to obtain the following values from Table 4.1:

p–value	10%	5%	1%
critical value	4.61	5.99	9.21

Step 4 *(p–value)*

We use

$$\begin{aligned} \nu &= (\text{no. of rows} - 1) \times (\text{no. of columns} - 1) \\ &= (3-1) \times (2-1) \\ &= 2 \end{aligned}$$

to obtain the following values from Table 4.1:

p–value	10%	5%	1%
critical value	4.61	5.99	9.21

Thus, our p–value is smaller than 1%.

Step 5 *(conclusion)*

Since $p < 1\%$,

Step 5 *(conclusion)*

Since $p < 1\%$,

- We have **strong** evidence against H_0

Step 5 *(conclusion)*

Since $p < 1\%$,

- We have **strong** evidence against H_0
- We therefore **reject** H_0 in favour of H_1

Step 5 *(conclusion)*

Since $p < 1\%$,

- We have **strong** evidence against H_0
- We therefore **reject** H_0 in favour of H_1
- There is evidence to suggest an association between employee type and the number of days sick leave taken

We now review the linear programming material we covered in weeks 9 and 10 (Chapters 9 and 10 in the notes).

Since it's probably still quite fresh in your head, the main points will be briefly highlighted before we look at an exam–style question.

9.1 Quick review

In an exam question on linear programming, you'll probably be asked to do three things:

1. **Formulate** a real–life scenario as a linear programming problem.
2. **Draw** a suitable diagram to enable the problem to be solved graphically.
3. **Solve** the problem (using the graph and/or by solving simultaneous equations).

9.1 Quick review

To **formulate** a linear programming problem, remember to:

- State clearly the **decision variables**;
- identify your **constraints**;
- identify the **objective function**.

Remember, it might help if you draw up a table which summarises all the information given in the question.

To represent the problem **graphically**, all you have to do is plot the inequalities you identified as "constraints".

Remember to shade out all the unwanted bits and clearly label the **feasible region**.

You should also plot the **objective line**, and draw an arrow which shows the direction of this line.

9.1 Quick review

To **solve** the problem, you just have to identify the point in your feasible region which optimises your objective.

This is usually at the intersection of two lines on your plot.

Once you have done this, the solutions can just be "read off"!

You should also know how to solve these problems **algebraically**.

9.2 Exam–style question

A chocolate manufacturer produces two types of chocolate bar, Asteroids and Blackholes.

Production of an Asteroid bar uses 10g of cocoa and 1 minute of machine time, whereas a Blackhole bar requires 5g of cocoa and 4 minutes of machine time.

Altogether, 2000g of cocoa and 480 minutes of machine time are available each day.

The manufacturer must make at least 50 Asteroid and 50 Blackholes each day to keep up with demand.

The manufacturer makes 10p profit from each Asteroid bar and 20p profit from each Blackhole bar.

(a) Formulate the chocolate manufacturers situation as a linear programming problem.

(b) Draw a suitable diagram to enable the problem to be solved graphically, indicating the feasible region and the direction of the objective line.

(c) Use your diagram to find the company's minimum and maximum profit, £P.

(d) Now solve this problem algebraically to verify your solution to part (c).

Draw up a table!

9.2 Exam–style question: solution (a)

Draw up a table!

	Cocoa	Machine time	Profit
Asteroids	10	1	10
Blackholes	5	4	20
Limits	2000	480	

9.2 Exam–style question: solution (a)

Draw up a table!

	Cocoa	Machine time	Profit
Asteroids	10	1	10
Blackholes	5	4	20
Limits	2000	480	

Step 1: Decision variables

x : No. of asteroids to make

y : No. of blackholes to make

Step 2: Constraints

$$10x + 5y \leq 2000$$

Step 2: Constraints

$$\begin{aligned} 10x + 5y &\leq 2000 \\ x + 4y &\leq 480 \end{aligned}$$

Step 2: Constraints

$$\begin{aligned} 10x + 5y &\leq 2000 \\ x + 4y &\leq 480 \\ x &\geq 50 \end{aligned}$$

Step 2: Constraints

$$\begin{aligned} 10x + 5y &\leq 2000 \\ x + 4y &\leq 480 \\ x &\geq 50 \\ y &\geq 50 \end{aligned}$$

Step 2: Constraints

$$\begin{aligned} 10x + 5y &\leq 2000 \\ x + 4y &\leq 480 \\ x &\geq 50 \\ y &\geq 50 \end{aligned}$$

Step 3: Objective function

$$P = 10x + 20y$$