LECTURE 7-1: HYPOTHESIS TESTING
What is a hypothesis test
p values

Chris Robertson

Background and Examples

Hypothesis testing

Science News *from research organizations*

'Technoference': We're more tired and less productive because of our phones

Study finds problematic phone use getting worse in Australia

Date: March 25, 2019

Source: Queensland University of Technology

Summary: An Australian survey of 709 mobile phone users (aged 18 to 83) has found one in five women and one in eight men are losing sleep due to bad phone habits. The study identified other rising 'technoference' impacts, including physical aches and pains, and found 24% of women and 15% of men are now classified as "problematic users".

This is taken from an newspaper article reporting on problematic phone use and loss of sleep. I was interested in the topic as mobile phone use is widespread and there are reports of adverse effects.

https://www.sciencedaily.com/releases/2019/03/190325101428.htm

Hypothesis testing

One in five women and one in eight men are now losing sleep due to the time they spend on their mobile phones, according to new QUT-led research that has found a jump in 'technoference' over the past 13 years.

Researchers surveyed 709 mobile phone users across Australia aged 18 to 83 in 2018, using questions replicated from a similar survey back in 2005.

They then compared the findings and discovered significant increases in people blaming their phones for losing sleep, becoming less productive, taking more risks while driving and even getting more aches and pains.

Study leader Dr Oscar Oviedo-Trespalacios from QUT's Centre for Accident Research and Road Safety -- Queensland (CARRS-Q) said the survey results showed 24 per cent of women and 15 per cent of men could now be classified as "problematic mobile phone users."

Read the article carefully

then you come across the phrase

Significant Increase

This is a very specific statistical term and implies that a statistical hypothesis test has been carried out.

We are going to study the procedures and implications of statistical hypothesis testing in this section

https://www.sciencedaily.com/releases/2019/03/190325101428.htm

Hypothesis testing

- A statistical analysis is often used to test some matter of concern to do with a population.
- For example clinical trials have a hypothesis test on the main endpoint of the study
- Let's look at the current coronavirus vaccine trials
  - Individuals consent to take part in the study. They are randomly assigned to get a covid-19 vaccine or another vaccine.
  - After period of time the researchers count how many people in each vaccine group contracted Covid-19
  - If the covid-19 vaccine is effective it is expected that there will be fewer Covid-19 infections in the covid-19 vaccine group
  - The hypothesis to test – does the coronavirus vaccine have the same effect on preventing covid -19 disease as the control vaccine -

https://www.theguardian.com/business/2020/aug/25/astrazeneca-covid-19-antibody-drug-trial-volunteers-vaccine
https://www.astrazeneca.com/media-centre/articles/2020/university-of-oxford-potential-covid-19-vaccine-phase-iii-clinical-trial-initiated-in-the-us.html

Hypothesis testing

- In March 2020 there were no known treatments for coronavirus

- A number of clinical trials were started and the best of these is the Recovery Trial

https://www.recoverytrial.net/

A range of potential treatments have been suggested for COVID-19 but nobody knows if any of them will turn out to be more effective in helping people recover than the usual standard of hospital care which all patients will receive. The RECOVERY Trial is currently testing some of these suggested treatments:

- Low-dose Dexamethasone (now only recruiting children)
- Azithromycin (a commonly used antibiotic)
- Tocilizumab (an anti-inflammatory treatment given by injection)
- Convalescent plasma (collected from donors who have recovered from COVID-19 and contains antibodies against the SARS-CoV-2 virus)
- REGN-COV2 (a combination of monoclonal antibodies directed against coronavirus).

Data from the trial are regularly reviewed so that any effective treatment can be identified quickly and made available to all patients. Please see our **news page** for results that RECOVERY has already found. The RECOVERY Trial team will constantly review information on new drugs and include promising ones in the trial.

Hypothesis testing

- The Recovery trail is complex as it has a number of therapies to test but it has all the main characteristics of a clinical trial – patients are randomly allocated to a therapy to avoid bias.

- The simplest clinical trial compares two therapies – a new one compared to the standard current treatment.

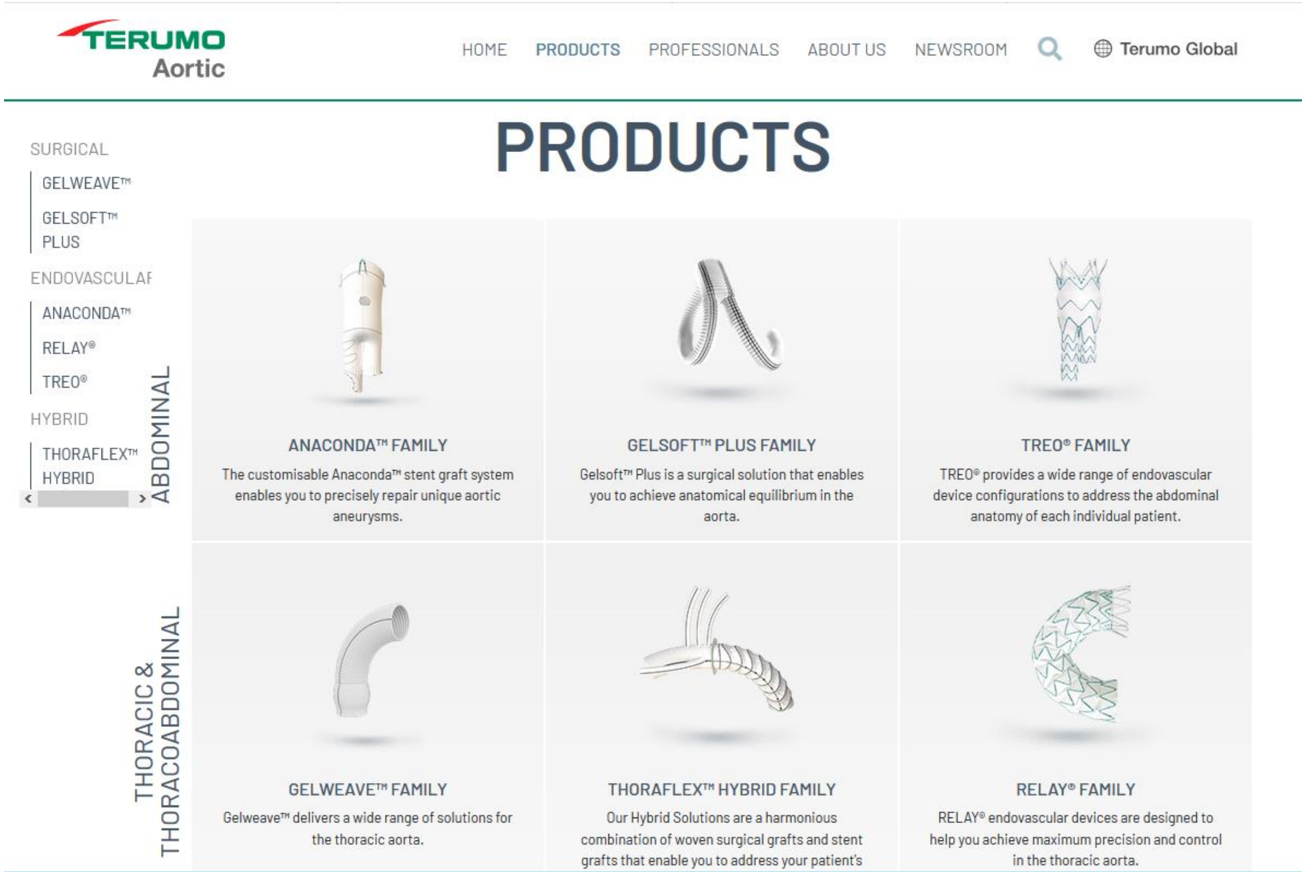
- Patients are assigned to therapy at random and this allocation is not subject to any subjective bias by clinicians

- The hypothesis to test - do the two therapies have the same effect

Hypothesis testing

- Quality control in manufacturing is an area where there is hypothesis testing
- Manufacturers have specifications for their goods – such as lengths, strengths

Hypothesis testing



Terumo Aortic is a Scottish company based in Inchinnan near Glasgow Airport. They are world leaders in the manufacturing of artificial grafts used in heart operations. Some of their products are stents to repair veins. They need to have exceptionally high quality standards

https://terumoaortic.com/

Hypothesis Testing Concepts

Hypothesis testing

- Is the strength of the fibres used to make replacement stents sufficiently high?
- To find out, we would select a random sample of products, put them through a stress test to find out the force needed to rip them
- Calculate the sample mean of the forces
- If the sample mean is below the specification then the evidence does not support the manufacturers claim.
- The hypotheses is that the mean breaking force of the product is equal to the specification level.

Hypothesis testing

- A statistical hypothesis test is designed to determine whether there is evidence to support a hypothesis
- It starts by assuming a specific hypothesis and a decision is made to reject this hypothesis if the data observed is **sufficiently unlikely** to have come from the model representing the specific hypothesis
- The specific hypothesis is called the **null hypothesis (H_0)**
- The general hypothesis of interest is called the **alternative hypothesis (H_1)**

Examples of hypotheses

- A researcher wants to determine whether or not a new pain killing drug is more effective than the best treatment currently available

  H_0 : There is no difference between the two treatments

  H_1 : There is a difference between the two treatments

  H_0 is specific and states that the two drugs have exactly the same effect

  H_1 is general and encompasses any difference between the two drugs

  The generality of **H_1** encompases a wide variety of situations

  New drug more effective;

  New drug much more effective;

  New drug less effective

  New drug much less effective

Examples of hypotheses

- A researcher wants to determine whether or not there is a difference in the starting salaries of recent male and female graduates

H_0 : There is no difference between the starting salaries

H_1 : There is a difference between the starting salaries

H_0 is specific and states that men and women have the same average starting salary

H_1 is general and encompasses any difference in average salary between men and women

Men paid more on average;

Men paid a lot more on average;

Men paid less on average

Hypothesis testing

- Note: H_0 describes the situation of no effect while the H_1 describes the change or difference

The null hypothesis is always very specific and specifies that a population parameter takes a specific value

For example

$$H_0:\ \mu = 0$$
$$H_0:\ \mu = 24$$
$$H_0:\ p = 0.5$$

We will only look at hypothesis tests with one sample in this class but more complex tests are often used.

Hypothesis testing

- Note: H_0 describes the situation of no effect while the H_1 describes the change or difference

- H_1 can be one of the following
  - Is there a difference ($\neq$): two tailed test
  - One is greater than the other (>): upper (high) one tailed test
  - One is smaller than the other (<): lower (low) one tailed test

The alternative hypothesis is always non specific and specifies what you believe if the null hypothesis is not true.

For example $H_1\!:\ \mu \neq 0$

$$H_1\!:\ \mu > 24$$

$$H_1\!:\ p < 0.5$$

What you choose for the alternative hypothesis is important and you know what you choose from the context of the problem. If you have no information on direction then you choose not equal ($\neq$).

Hypothesis testing

- Note: H_0 describes the situation of no effect while the H_1 describes the change or difference

- H_1 can be one of the following
  - Is there a difference ($\neq$): two tailed test
  - One is greater than the other (>): upper (high) one tailed test
  - One is smaller than the other (<): lower (low) one tailed test

- It is impossible to 'prove' a hypothesis is true.

All that you can conclude is
(1) there is evidence to reject the null hypothesis, or
(2) there is no evidence to reject the null hypothesis

Hypothesis testing

- It is impossible to ‘prove’ a hypothesis is true.

- In order to make a decision we calculate a probability measure
  - the probability of the data, or a more extreme form of it, occurring by chance assuming H_0 is true
  - This probability is known as the p value of the test

- As a result, because we are dealing with probabilities, we shall **never** ‘prove’ the null hypothesis is true or that the alternative hypothesis is true

Hypothesis Testing Process

- Select a suitable outcome variable/measured response
- Use your research question to define an appropriate and testable null hypothesis involving this variable together with a suitable alternative hypothesis
- Collect the appropriate sample data and determine the relevant sample statistic
- Use a decision rule that will enable you to judge whether the sample evidence supports or does not support the null hypothesis
- Based on this evidence, reject or do not reject the null hypothesis

Going through the steps using an example

Example

The target breaking strength for a specific type of fibre is 24 Newtons. The company carries out this breaking strength experiment regularly and so knows that the standard deviation of the breaking strengths is 2 Newtons. A sample of 10 fibres from a new yarn yields a mean breaking strength of 23 Newtons. Is there evidence to suggest that the mean breaking strength of the fibres from the new yarn is less than the target? The breaking strengths are normally distributed.

This is a hypothesis test about a population mean

Example

The **target** breaking strength for a specific type of fibre is **24 Newtons**. The company carries out this breaking strength experiment regularly and so knows that the standard deviation of the breaking strengths is 2 Newtons. A sample of 10 fibres from a new yarn yields a mean breaking strength of 23 Newtons. Is there evidence to suggest that the mean breaking strength of the fibres from the new yarn is less than the target? The breaking strengths are normally distributed.

This bit of the problem is telling you that the null hypothesis is H_0: $\mu = 24$

This bit of the problem is telling you that the alternative hypothesis is

$$H_1: \ \mu < 24$$

Example

The **target** breaking strength for a specific type of fibre is **24 Newtons**. The company carries out this breaking strength experiment regularly and so knows that the standard deviation of the breaking strengths is 2 Newtons. A sample of 10 fibres from a new yarn yields a ean breaking strength of 23 Newtons. **Is there evidence to suggest that the mean breaking strength of the fibres from the new yarn is LESS than the target?** The breaking strengths are normally distributed.

This bit of the problem is telling you that the null hypothesis is H_0: $\mu = 24$

This bit of the problem is telling you that the alternative hypothesis is

$$H_1:\ \mu < 24$$

Example

The target breaking strength for a specific type of fibre is 24 Newtons. **The company carries out this breaking strength experiment regularly and so knows that the standard deviation of the breaking strengths is 2 Newtons.** A sample of 10 fibres from a new yarn yields a mean breaking strength of 23 Newtons. Is there evidence to suggest that the mean breaking strength of the fibres from the new yarn is less than the target? The breaking strengths are normally distributed.

This bit is giving information about the population as the company does this regularly and so knows σ

$$n = 10$$
$$\bar{x} = 23$$
$$\sigma = 2$$

Example

The target breaking strength for a specific type of fibre is 24 Newtons. **The company carries out this breaking strength experiment regularly and so knows that the standard deviation of the breaking strengths is 2 Newtons. A sample of 10 fibres from a new yarn yields a mean breaking strength of 23 Newtons.** Is there evidence to suggest that the mean breaking strength of the fibres from the new yarn is less than the target? The breaking strengths are normally distributed.

This bit is giving information about the population as the company does this regularly and so knows σ

Information about the sample size and the sample mean

$$n = 10$$
$$\bar{x} = 23$$
$$\sigma = 2$$

Example - Summary

$$H_0: \ \mu = 24$$
$$H_1: \ \mu < 24$$

From the question we have got the Null and Alternative hypotheses

$$n = 10$$
$$\bar{x} = 23$$
$$\sigma = 2$$

We have also got the data and the standard deviation

In order to make a decision we need to calculate a probability – the p-value

This is the probability of the data, or a more extreme form of it, occurring by chance assuming H_0 is true, known as the p value

So how do we do this?

We know how to calculate probabilities using a normal distribution
We know that the strengths are normally distributed
The sample mean is used so we will look at the results about the sampling distribution of the mean.

Example – Test Statistic

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

H_0: $\mu = 24 = \mu_0$
H_1: $\mu < 24$

From the question we have got the Null and Alternative hypotheses. μ_0 is just the value of the mean assuming H_0 is true

$n = 10$
$\bar{x} = 23$
$\sigma = 2$

Things we know from an earlier part of the class

$\mu_{\bar{X}} = \mu$

We know that the sample mean is normally distributed with a mean, $\mu_{\bar{X}} = \mu_0$, assuming H_0 is true.

We also know that the standard error of the sample distribution of the mean is $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

We also know how to convert any normal distribution into the standard normal distribution by subtracting off the mean and dividing by the standard deviation.

We can find probabilities from the standard normal distribution from tables

Example – Test Statistic

H_0: $\mu = 24 = \mu_0$

H_1: $\mu < 24$

$n = 10$
$\bar{x} = 23$
$\sigma = 2$

Standardising the distribution of the sample mean

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

If the null hypothesis is true then we replace μ by μ_0 to give

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

This is the test statistic for the null hypothesis as its distribution is known exactly if H_0 is true – standard normal

We can calculate z as all the components are known

If H_0 is not true then the alternative says that H_1: $\mu < \mu_0 = 24$ and so values of $\bar{X}$ that are further away from $\mu_0 = 24$ are more extreme

Example – Test Statistic and p value

$H_0\colon\ \mu = 24 = \mu_0$

$H_1\colon\ \mu < 24$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{23 - 24}{2/\sqrt{10}} = \frac{-1}{0.632456} = -1.5811$$

$n = 10$

$\bar{x} = 23$

$\sigma = 2$

If H_0 is not true then the alternative says that $H_1\colon\ \mu < \mu_0 = 24$ and so values of $\bar{X}$ that are further away from $\mu_0 = 24$ are more extreme

Values of the sample mean less that 23 are more extreme from 24 in the direction of the alternative so count as more evidence against the null hypothesis

The p value is thus $P(Z < -1.5811) = 0.0569$

This is the probability of getting a value of z which is more extreme than the z value calculated from the observed data.

It represents how likely the data, or a more extreme form of the data, are, assuming that the null hypothesis is true

Example – p value and decision

H_0: $\mu = 24 = \mu_0$

H_1: $\mu < 24$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{23 - 24}{2/\sqrt{10}} = \frac{-1}{0.632456} = -1.5811$$

p-value = $P(Z < -1.5811) = 0.0569$

$n = 10$
$\bar{x} = 23$
$\sigma = 2$

The decision about whether or not to reject H_0 is made on the basis of how small this p-value is

This is achieved by specifying a significance level for the test - α

If $\text{p} - \text{value} \leq \alpha$ conclude that there is evidence to reject H_0
If $\text{p} - \text{value} > \alpha$ conclude that there is no evidence to reject H_0

Example – p value and significance level

H_0: $\mu = 24 = \mu_0$

H_1: $\mu < 24$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{23 - 24}{2/\sqrt{10}} = \frac{-1}{0.632456} = -1.5811$$

$n = 10$
$\bar{x} = 23$
$\sigma = 2$

p-value = $P(Z < -1.5811) = 0.0569$

This is achieved by specifying a significance level for the test - α

If $\text{p} - \text{value} \leq \alpha$ conclude that there is evidence to reject H_0
If $\text{p} - \text{value} > \alpha$ conclude that there is no evidence to reject H_0

The usual significance level to choose for any test is $\alpha = 0.05$, this corresponds to a 5% significance level

Here p-value = 0.0569 > 0.05 so we conclude that there is no evidence to reject the hull hypothesis that the average breaking strength is 24 Newtons.

Summary of Key points

P-value

- To decide whether or not the sample evidence does or does not support the null hypothesis we use a **p-value**

- The p-value can be interpreted as follows
  - The probability of getting the results you have observed, **or more extreme results**, assuming the null hypothesis is true
  - P-values are an area under a distribution curve of the **test statistic** being considered
    - Test statistic is a single number calculated from the sample data which has a known distribution assuming that the null hypothesis is true

The significance level α

- The significance level α is the threshold on the probability of our event/difference occurring by chance, that we are willing to consider as part of the experiment.

- Set by the researcher before experiments are conducted and data collected, usually set to 5% (i.e. 0.05).
- Depending on the experiment, sometimes the significance level is 1% or 10%.
  - 10% is often used when testing a new cancer therapy in a small study for a potential effect
  - 1% (or smaller) is often used when investigating hundreds of genetic markers

P-value vs. Significance level

- If $p \leq \alpha$ then
  - The probability of our event happening by chance is low in comparison to what we were willing to consider as a probability of the event happening by random chance – so must be significant!
  - There is evidence to reject the null hypothesis

- If $p > \alpha$ then
  - The probability of our event happening by chance is high in comparison to what we were willing to consider as a probability of the event happening by random chance – not significant!
  - There is no evidence to reject the null hypothesis

One and two tailed tests

$H_0\text{:}\;\; \mu = \mu_0$ Null Hypothesis

Three possible versions of the alternative hypothesis

$H_1\text{:}\;\; \mu < \mu_0$

$H_1\text{:}\;\; \mu > \mu_0$ One Tailed

$H_1\text{:}\;\; \mu \neq \mu_0$ Two Tailed

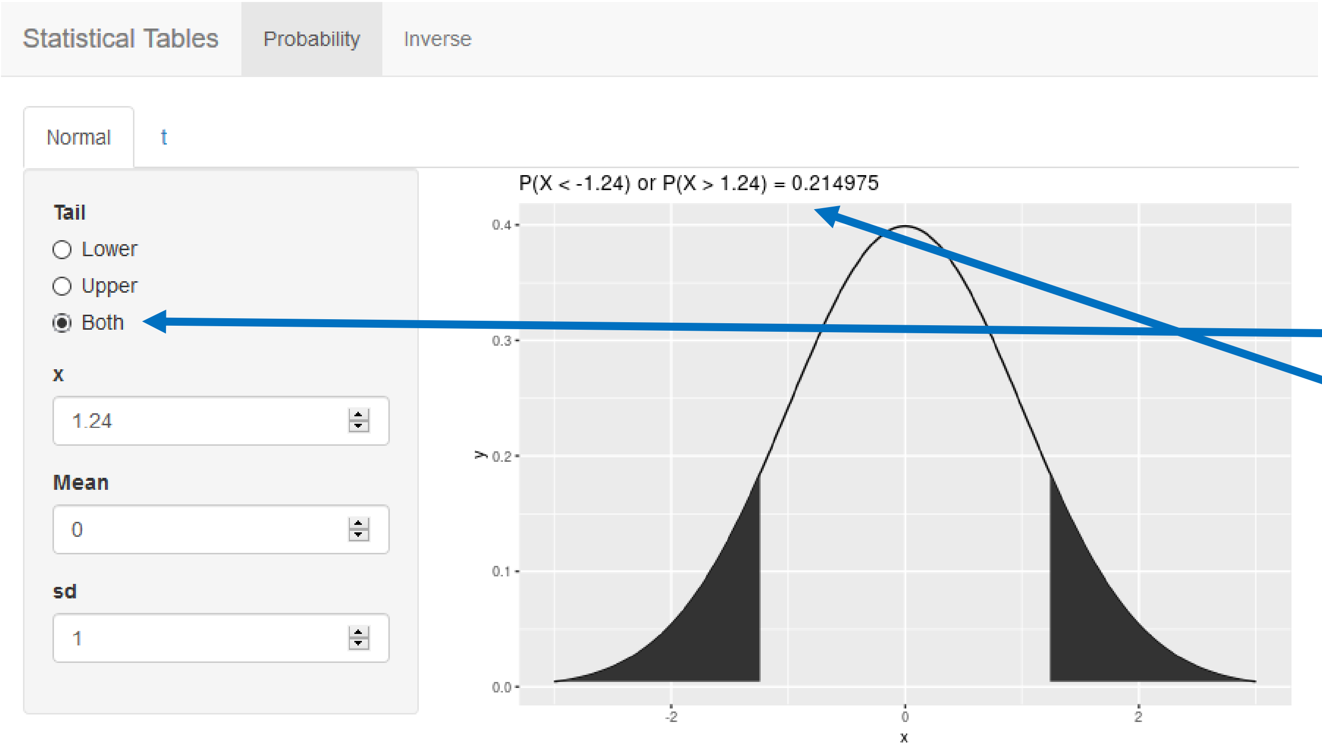
One Tailed alternative
use the upper or lower tail of the normal distribution
Upper for $H_1\text{:}\;\; \mu > \mu_0$ and lower for $H_1\text{:}\;\; \mu < \mu_0$

Two Tailed alternative
Use both the upper and lower tails of the distribution to calculate the p value

Two tailed tests

H_0: $\mu = \mu_0$ Null Hypothesis

H_1: $\mu \neq \mu_0$ Suppose the calculated value of the z statistic is 1.24



In the tables make sure that you check the Both box for a two tailed test.

The p value is 0.21

Key Points

Hypothesis tests are some of the most (over) used statistical techniques

Very important to specify the null hypothesis and alterative and the significance level

Choose $\alpha = 0.05$, unless specifically told not to

5% significance level corresponds to a 1 in 20 chance of wrongly rejecting the null hypothesis when it is true

Calculate the test statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

where μ_0 is the specified value for the mean under the null hypothesis

Calculate the p value paying close attention to the alternative hypothesis for the appropriate tails of the distribution to use

If $\mathrm{p-value} \leq \alpha$ conclude that there is evidence to reject H_0

If $\mathrm{p-value} > \alpha$ conclude that there is no evidence to reject H_0