

Introduction

- Topics covered
 - Experimental Design
 - Dependent and Independent Variables
 - The Null Hypothesis
 - Type I and Type II Errors
 - Statistical Significance
 - One-tailed and two-tailed tests
- Inferential statistics allow us to make predictions based upon the sampling of a large population

Experimental Design

- An experiment is a task that aims to predict the value of an *output variable* by introducing a change in the set of *input variables*
- It may also be necessary to identify some *control variables* that need to be kept constant to prevent external factors from influencing the result
- You are trying to show that
 - IF the control variables are kept constant
 - THEN changing the input variables
 - WILL CAUSE a change in the output variable

Experimental Credibility

- Experimental findings must be credible
- Credibility tests
 - **Reliability**
 - Will the experiment give the same result if run a second time? Or by a different experimenter?
 - **Validity**
 - Is this an appropriate way to explore this problem?
 - Do you have the right experimental conditions?
 - Are you collecting the right data?
 - **Generalisability**
 - Do the results apply globally?
 - Are there any conditions on where/when the results apply?

Independent and Dependent Variables

- When designing an experiment we need to define three things:
 - What we are going to measure
 - What conditions we're going to measure this under
 - What sort of a relationship do we expect there to be between the thing we're going to measure under the conditions we've defined
- This allows us to define three important experimental items
 - The *dependent variable* (the output variable)
 - The *independent variable(s)* (the input variable(s))
 - The experimental *hypothesis*

The Dependent Variable

- The dependent variable (or *DV*) is the output variable that will be measured
 - Ping time
 - Accuracy of an image classification model
 - Number of sales
 - Time required for a brute force attack to decrypt encrypted data
- When designing an experiment it is sometimes difficult to measure the thing directly
 - In this case we measure something that is likely to vary in the same manner as the thing we want to measure
 - This is known as a *proxy*
 - If you use a proxy you must explain why
 - Proxy serves as a stand in or substitute for a variable that is difficult or impossible to measure directly.

The Dependent Variable

- Proxy variable example

- 1.

- Variable of interest : Vulnerability of IoT devices to cyber attacks
 - Proxy variable : Anomaly detection score derived from network traffic patterns.

- 2.

- Variable of interest : Mental health status of Social media users
 - Proxy Variable : Frequency of positive and negative sentiment expressions in users post

The Independent Variable

- The independent variable (IV) is the input variable you will change in order to cause a change in the dependent variable
 - Hardware configuration (configuration 1/configuration 2)
 - System design (design 1/design 2)
 - Medical treatment (before/after, with/without)
 - ...
- Comparing two values of the IV is the simplest type of experiment, and is the easiest to analyse
- Experiments can have more than one IV, but if so then they must test all possible combinations of the different IVs
- Having more than two values of the IV or more than one IV means that the statistical analysis is more complicated

The Hypothesis

- The *hypothesis* is your theory of how the IV affects the DV
- The most basic hypothesis might be *changing the value of the IV changes the value of the DV*
- A more complex hypothesis might be *increasing the value of the IV causes an increase in the value of the DV*
- You would then design an experiment that would test whether the output results (DV) for condition 1 of the IV (Population A) differ from those for condition 2 of the IV (Population B)
 - This commonly requires the experimental to compare the mean value or the total values of the DVs under each of the two conditions
 - It is better to use the mean than the total, but this may require that tests are run multiple times

Example 1 – Using Simple Variables

Suppose we want to measure how many mistakes a user makes when using the old version and the new version of a piece of hardware

- What is the output variable (DV)?
- What is the input variable (IV)?
- What conditions of the IV should be included?
- What is the basic hypothesis?

Example 1 – Using Simple Variables

Suppose we want to measure how long a packet takes to reach a PC when using two different versions of a routing algorithm

- What is the output variable (DV)?
 - Packet transmission time
- What is the input variable (IV)?
 - Version of the routing algorithm
- What conditions of the IV should be included?
 - Version 1, Version 2
- What is the basic hypothesis?
 - Changing the version used leads to a change in the mean packet transmission time

Exercise

Suppose we want measure the effect of changing the website design on the number of items sold.

- What is the output variable (DV)?
- What is the input variable (IV)?
- What conditions of the IV should be included?
- What is the basic hypothesis?

Example 2 – Using Proxy Variables

Suppose we want to investigate the effectiveness of two types of firewall

- What is the output variable (DV)?
 - We cannot measure ‘effectiveness’ directly, so we need to use a proxy measure for it, e.g. number of bad packets passed through
- What is the input variable (IV)?
 - Firewall type
- What conditions of the IV should be included?
 - Firewall type 1, Firewall type 2
- What is the basic hypothesis?
 - The effectiveness (as measured using the number of bad packets passed through) will change depending on which firewall type is used

Exercise

Suppose we want to investigate the effectiveness of an antivirus programme.

- What is the output variable (DV)?
- What is the input variable (IV)?
- What conditions of the IV should be included?
- What is the basic hypothesis?

The Null Hypothesis

- Normally researchers are interested in demonstrating the truth of a hypothesis of interest
 - This group is healthier
 - These shops sell more
 - This routing algorithm is faster
- This means that Population A must differ from Population B in some way
- Unfortunately it is very difficult to prove something true, so instead we try to prove that something is false
- We therefore state the exact opposite of what we want to demonstrate to be true and then prove it false
 - What remains (the thing we're interested in) must be true
- This 'opposite' hypothesis is known as the *null hypothesis* H_0

The Null Hypothesis

- Suppose we want to show that men's feet are bigger than women's feet
 - DV – foot size
 - IV – Gender of foot owner
 - Conditions – Male foot owners, Female foot owners
- We can't prove that without measuring the feet of every male on the planet to get μ_{Male}
- Instead we take a random sample of males and find out their shoe size to get $\bar{X}_{Male}$ (Population A)
- Similarly we need to measure a sample of female feet to get a value for $\bar{X}_{Female}$ (Population B)

The Null Hypothesis

- Suppose we want to show that men's feet are bigger than women's feet
- The thing we want to show is that men's feet are bigger than women's feet, i.e. $H_1: \mu_{Male} > \mu_{Female}$
- Our null hypothesis is the opposite of this: Men's feet are the same size as or smaller than women's feet, i.e. $H_0: \mu_{Male} \leq \mu_{Female}$
- We then use statistical techniques to reject (disprove) this hypothesis by showing that $\bar{X}_{Male} > \bar{X}_{Female}$

The Null Hypothesis

- We can now state our working hypotheses:
- H_0 : Men's feet are as small as or smaller than women's feet.
 - The mean foot size of a male is as the same as or less than the mean foot size of a female
- H_1 : Men's feet are larger than women's feet.
 - The mean foot size of a male is greater than the mean foot size of a female
- How can we reject H_0 ?
 - By showing that $\bar{X}_{Male} > \bar{X}_{Female}$

The Null Hypothesis

- Why is there a difference between male and female sample means?
- Two possible reasons:
 - There is a difference between the male and female mean shoe size, i.e. $\bar{X}_{Male} \neq \bar{X}_{Female}$
 - There is no difference between the male and female mean shoe size, i.e. $\bar{X}_{Male} = \bar{X}_{Female}$ and any variation is due to chance or to a sampling error
- The t-test and the ANOVA test will help to decide if the difference is due to chance or not

Exercise

- We want to show that the new version of a piece of hardware allows users to complete tasks more quickly than the previous version
- What are the DV, IV and IV conditions?
- Formulate a null hypothesis H_0 and an alternative hypothesis H_1 for this experiment
- What would you need to show to reject the null hypothesis?

Type I and Type II Errors

- The decision to reject (or not) the null hypothesis is subject to error because you have only studied a sample, not the entire population
 - You can never be completely sure whether H_0 is true or not
 - You can reject an H_0 that is actually true – a Type I error
 - You can accept an H_0 that is actually false – a Type II error

The Real Situation (unknown to the investigator)		
Investigator's decision	H_0 is True	H_0 is False
Reject H_0	Type I error	Correct decision

- | | | |
|--------------|------------------|---------------|
| Accept H_0 | Correct decision | Type II error |
|--------------|------------------|---------------|
- Inferential statistics help you decide whether to reject the null hypothesis and to estimate the probability of a Type I or Type II error, not tell you if you've made an error

Type I and Type II Errors

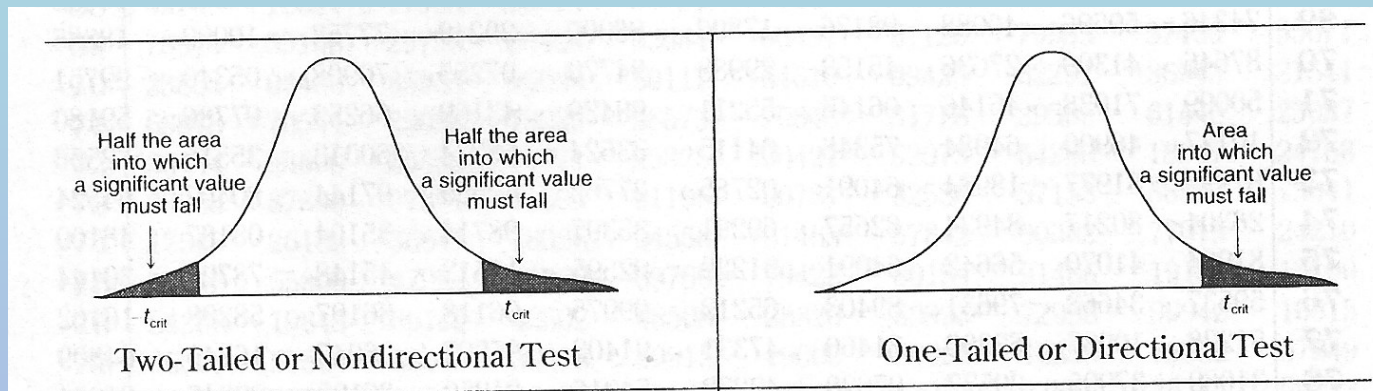
	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	 Type I Error (False positive)	 Correct Outcome! (True positive)
Fail to reject null hypothesis	 Correct Outcome! (True negative)	 Type II Error (False negative)

Statistical Significance and Type I Errors

- Statistical decisions are normally made so as to minimise the likelihood of Type I errors
 - Which might mean that lots of Type II errors occur
- Statistical results are often reported as “statistically significant at the 0.05 level”
 - This means that there is a 5% chance of a Type I error
 - 5 times in 100 that error will still occur
- If you increase the level to 1% you increase the chance of making a Type II error
 - You are refusing to reject a hypothesis when the evidence that it should be rejected is pretty solid

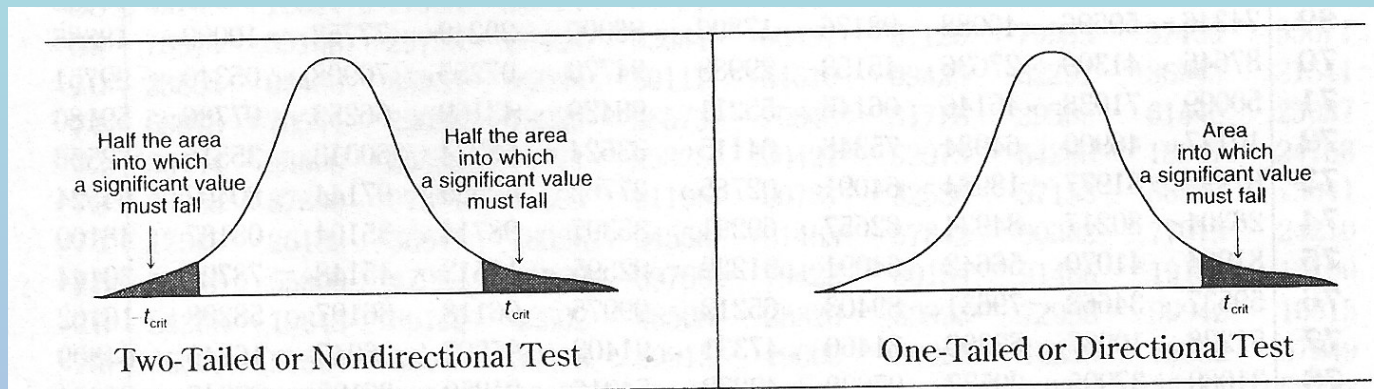
Directional v. Non-directional Tests

- Sometimes we are interested in showing that one sample's mean is specifically higher or lower than the other, rather than it just being not equal to the other
- To do this we use a directional or one-tailed test
- In a non-directional test the critical value represents 0.05 as 0.025 at each end of the distribution (a *two-tailed distribution*)
- In a directional test the critical value represents 0.025 at one end of the distribution (a *one-tailed distribution*)
- You must decide whether you want a one-tailed or two-tailed distribution prior to data collection



Directional v. Non-directional Tests

- If you want to show that there is a difference between two means but you don't care which of the two conditions gives the higher value, use the two-tailed test
- If you specifically want to show that one thing is higher than the other, use the one-tailed test



Exercise

- For each of the following situations, should you use a one-tailed or two-tailed test?
 - A company wants to see if there is any difference between the ease of use of the old version and the new version of their hardware
 - A network engineer wants to know what difference changing a setting on the firewall has on the security of their system
 - A company wants to know if a new “2024 model” sticker on their wares will lead to more sales

Summary

- Topics covered
 - Experimental Design
 - Dependent and Independent Variables
 - The Null Hypothesis
 - Type I and Type II Errors
 - Statistical Significance
 - One-tailed and two-tailed tests
- Understand
 - The t Test for Independent Samples
 - The Critical Value of t : t_{crit}
 - The t Test for Dependent (Matched) Samples
 - Directional vs. Non-directional Tests
 - ANOVA tests for >2 variables

- **Homework:** Read chapter 11 of *Statistics for the Terrified*.