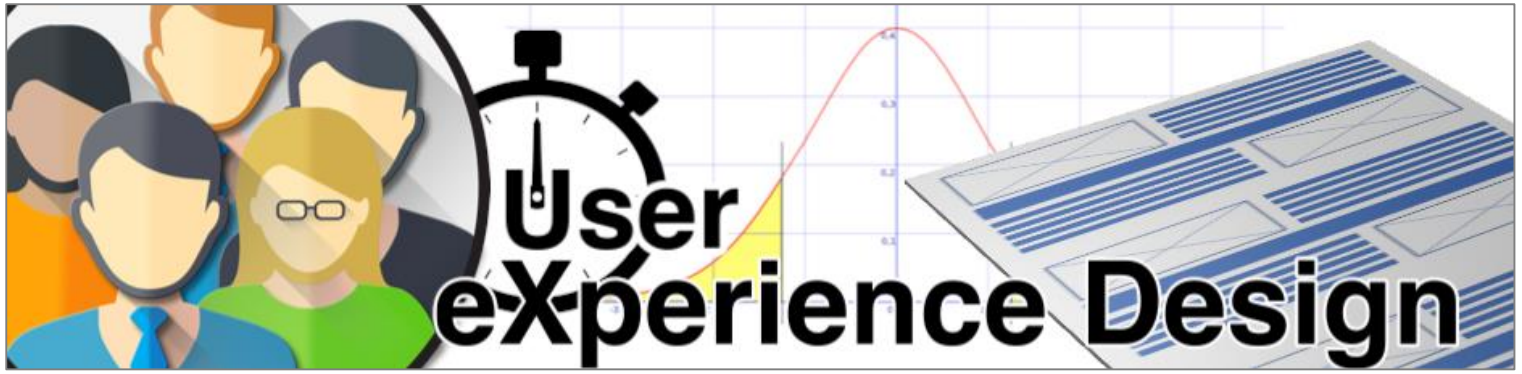
Faculty of Engineering and Technology
School of Computer Science and Mathematics

6110COMP
User eXperience Design

Andy Symons
a.symons@ljmu.ac.uk
Room 717, Byrom Street

Thomas Hughes-Roberts
t.hughesroberts@ljmu.ac.uk
Room 607b, Byrom Street



Lecture 2c – Normalised or Not?

Learning outcomes

- At the end of this session you should:
  - Understand what a normal distribution looks like.
  - Be able to determine if your dataset is normally distributed.

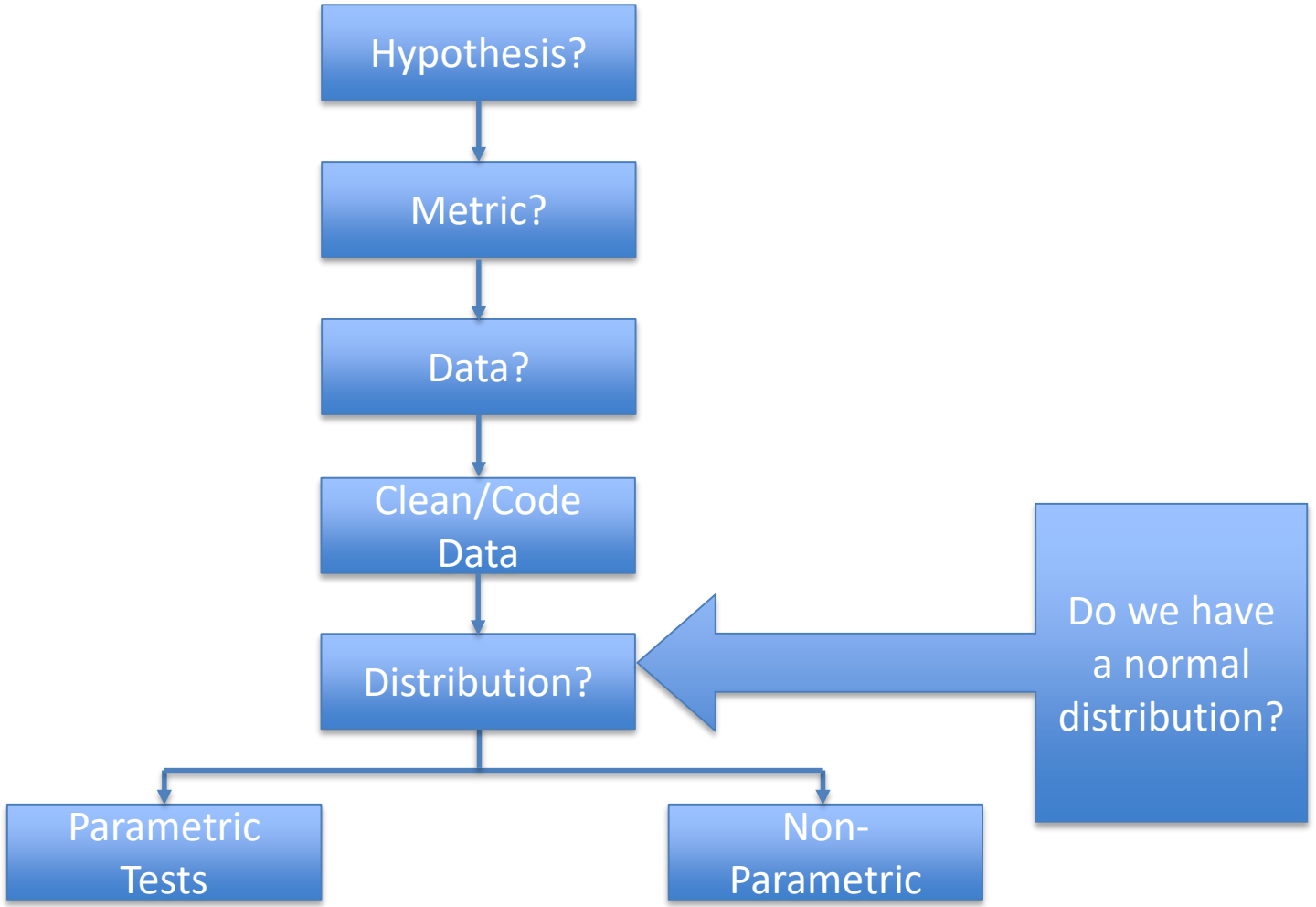
In this session…

– Data distribution

– Frequency graphs

– Kolmogorov-Smirnov tests for normality.

Process Overview



Normal Distribution – when?

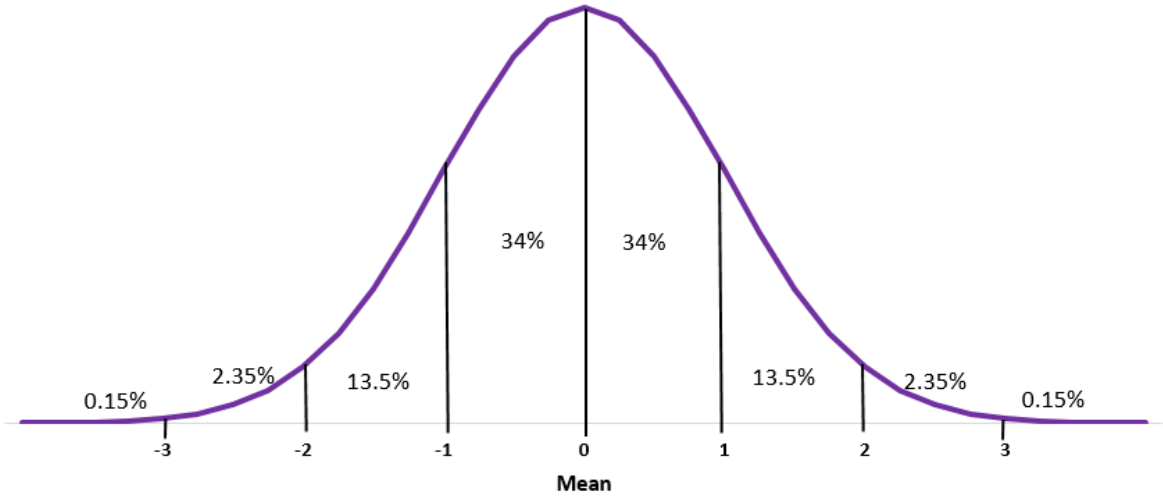
- In order to further identify what test we can run, we need to ascertain if our data is normally distributed – assuming it is a continuous variable.

- For example, task completion is not a continuous variable but categorical and therefore we do not run a test for normality – which automatically rules out certain statistical tests.

A Normal Distribution

- Continuous data that has a symmetric distribution – a bell shape.

- Get an idea from a graphical view of the data….

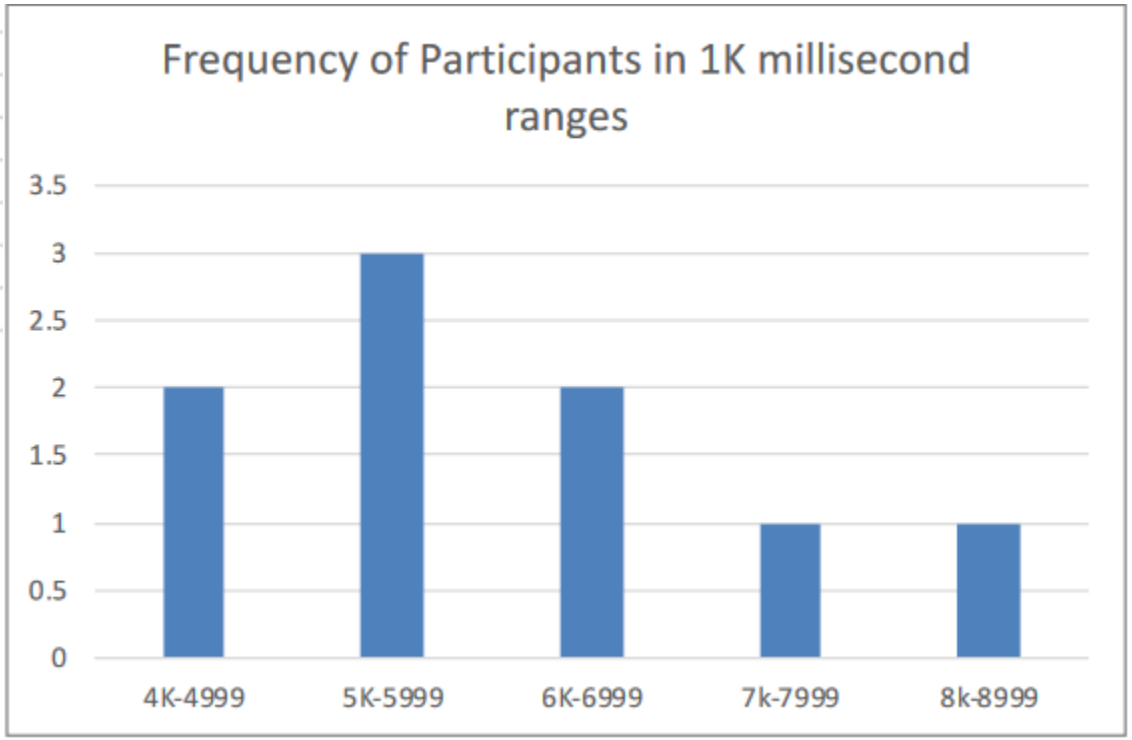


Frequency Distribution

- From some sample data (worksheet provided in this weeks session), we can plot a range of interaction times.

	A	B	C	D
			DATA	
			5432	
			8438	
			4827	
			4934	
			5905	
			6345	
			6396	
			7003	
			5354	
		MAX	8438	
		MEAN	6070.44	
		MIN	4827	

	A	B	C	D
15				
16		4K-4999	2	
17		5K-5999	3	
18		6K-6999	2	
19		7k-7999	1	
20		8k-8999	1	
21				
22				
23				



However…

- Just viewing the data in this format does not lend itself well to accuracy…

- We have formal tests we can run against the data to determine its distribution.

Kolmogorov-Smirnov Test

- A test that can determine if a single sample of data is normally distributed.

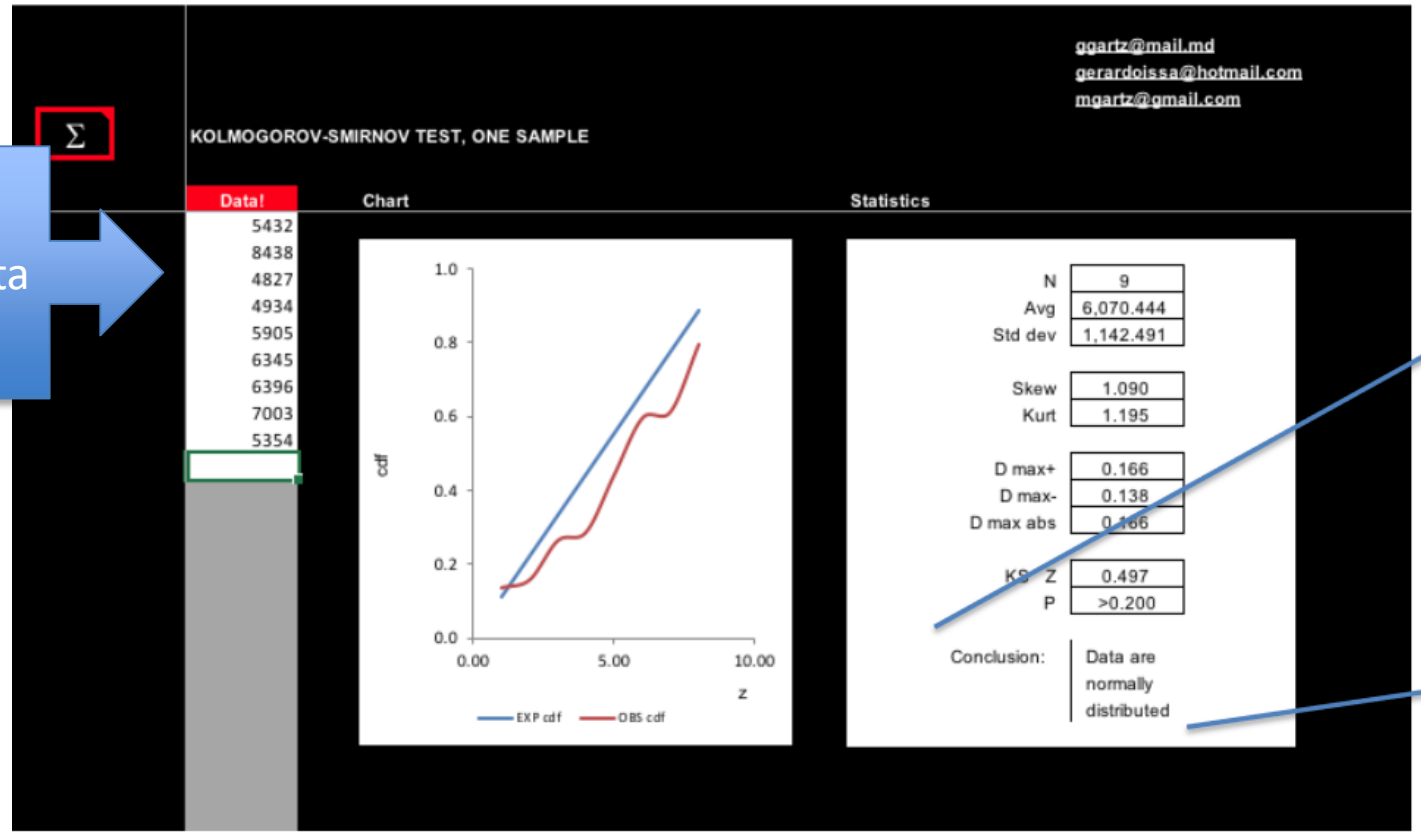
- Using a Hypothesis/null hypothesis testing:

H1: The data is not normally distributed.

H0: The data is normally distributed.

Using Excel

- We have provided a spreadsheet sourced online for you to run a KS test on your data.



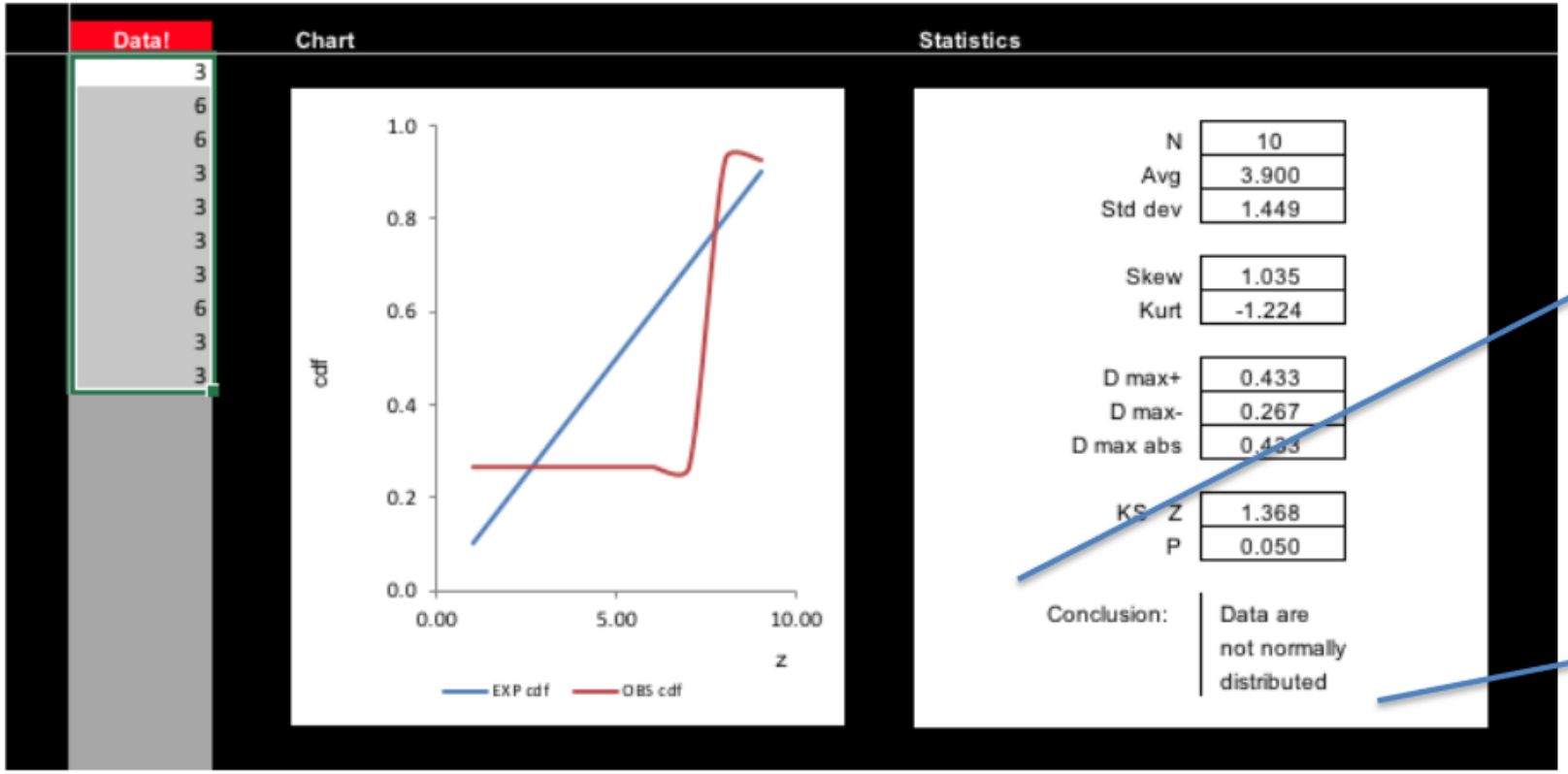
Our data

Check out the 'Conclusion'

Conclusion: | Data are normally distributed

KS Tests continued

- Significant Value:



Conclusion: | Data are not normally distributed

Note…

- This test tends to be used on larger sample sizes.
- However, for the purposes of this module it provides a good starting point for evaluating normality.
- Shapiro Wilk is a similar test that could be utilised.

Be aware…

- We will cover the statistical tests the module will deliver next week.
- For now, note that all tests have a set of assumptions that we must meet in order to utilise that particular test.
- These assumptions include (but are not limited to) the data type (nominal, ordinal etc.) and the distribution of the data.
- E.g. https://statistics.laerd.com/statistical-guides/independent-t-test-statistical-guide.php

Summary

- In this lecture, we have learned:
  - To identify normalised data.

- From the lectures this week, we should know:
  - What kind of data we have.
  - How to “treat” the data.
  - What our data distribution is.