Slide 1

Introduction to experimental design

Slide 2

Experimental design is used to provide the higher standard and confidence in decision-making.

In this presentation, we will have a look at an example through which we will learn the basics of experimental design.

Slide 3

A company with over one million customers considers adding a privacy alert to its website. The role of the alert is to make the customers aware of the fact that the company uses their data to analyze and to adapt their recommended new content.
The company’s Chief Marketing Officer stresses that this alert will make the customers trust the company and thus will promote their brand, which will make the customers more engaged on the website, while the head of the design team suggests it will just scare off new customers, and make them less engaged.
The head of the design team decides to test his assumption and tries to establish a causal relationship between the presence of the alert and a decrease in the engagement with the website.

Slide 4

How are we actually going to conduct this experiment such that it will establish a causal relationship?

The team decides to run an A/B test: during the next two weeks, half of the customers logging into the website will get a privacy alert, while the other half will not.
We call the half that gets the privacy alert “The Experiment group” because they are exposed to our experiment, or “Treatment group” when, for example, we test the effect of a clinical treatment, instead of a website’s effectiveness.
The group that gets to see the left side (without the alert) are the “Control group”, since their role is only to control the experiment.
Having a Control group is one of the most important components in a proper experiment design. If you don’t have anything to compare with, your conclusions will be basically flawed.

If the daily average number of clicks for the customers in the Experiment group turn out to be higher than the Control group, in a way that does not seem like it would just be a random chance, the company will conclude that the privacy alert is the cause and will decide to add it to the website.
If it turns out to be higher in the Control group, the company will decide to remove the privacy alert.

Slide 5

Let’s define the two variables that we will need to measure:
The Independent variable (sometimes called the Explanatory variable) is the variable that makes the effect, the one which splits the customers into the Experiment or Control group. The Independent variable in our case will be – whether a customer will be exposed to the privacy alert or not.
The Dependent variable (sometimes called the Response variable) is the one through which we measure the effect of the Independent variable. In our case it will be the daily average number of clicks within the website. This variable was chosen to be the most representative of users’ engagement with the website, which is what the company is interested to maximize.

Slide 6

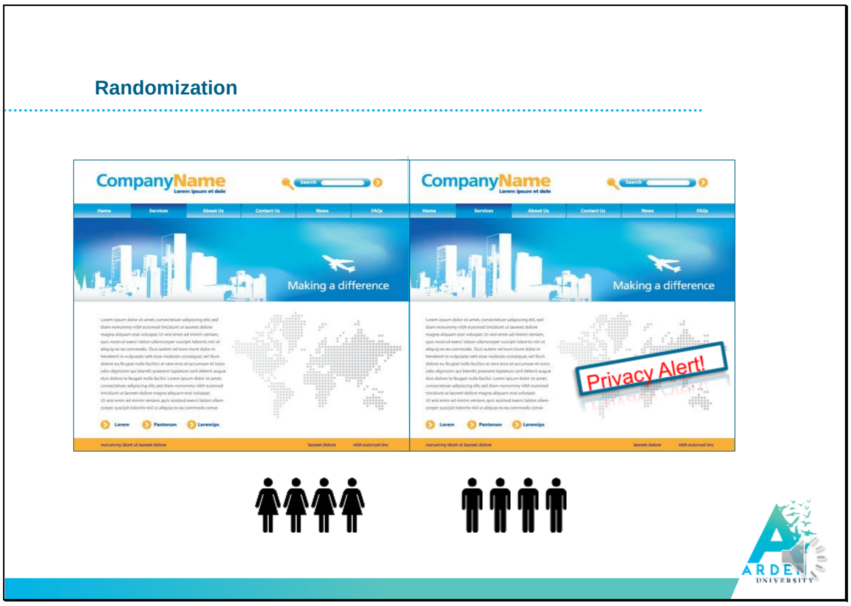
Confounding variables

Of course. we want to make sure that the privacy alert is the only possible cause for the difference between the groups. Otherwise, we wouldn't be able to say in confidence that the alert is indeed the cause.

Confounding occurs when there are other factors, besides the Independent variables, which might cause the difference in the levels of the Dependent variable. Thus, we must do whatever we can to avoid confounding variables.

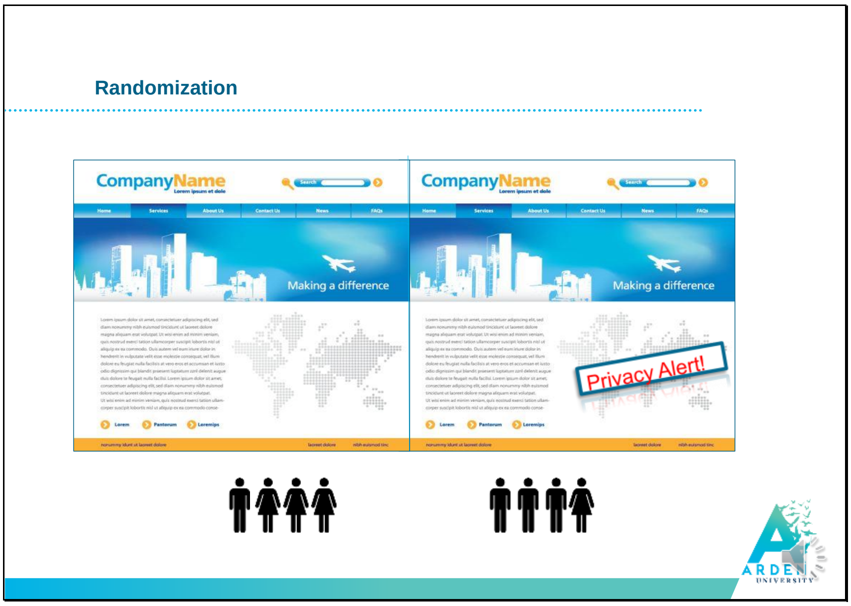
In our example, we must make sure both the groups operate at the same time, that the design of the website is identical (aside to the alert), that there are no external events that might affect the results (like a major news article about a privacy breach in the company), that the customers assigned to one group are not inherently more engaged than those in the second group, and that the same set of support and services are offered.

Slide 7



OK, so the groups now have the exact same conditions.
How would we split the customers between the groups? To avoid confounding factors, we don't want to split the customers by any factor like gender, login time of the day, age or anything similar.
If, for example, we split them by age and we find a difference in the number of clicks, it might be due to age, so age becomes a confounding factor.
To avoid that we use randomization: we randomly split the customers between the two groups.

Slide 8

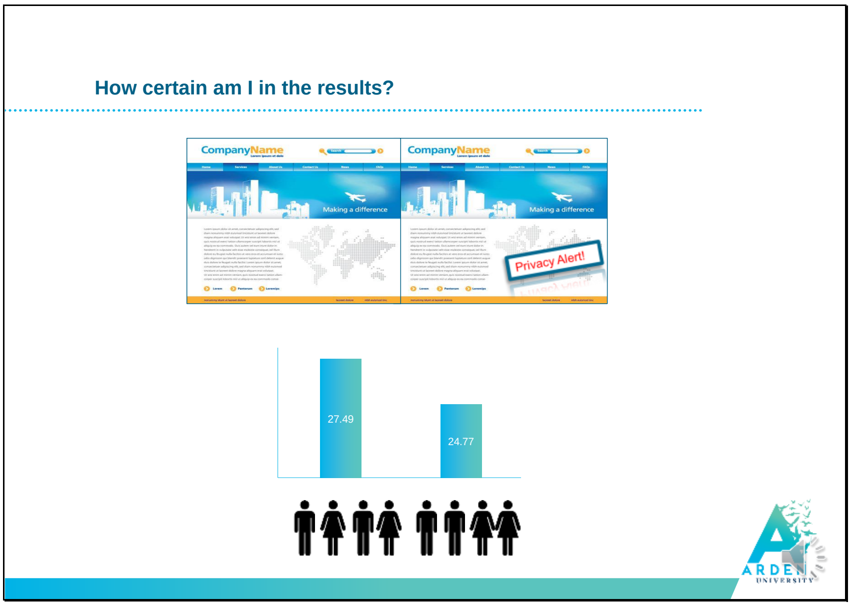


The customers are now randomly assigned to the two groups.
But what if, the random split resulted in for example, 75% women in the Control group and 75% men in the Experiment group? Then gender may become a confounding variable.
One of the solutions to this bias would be to pick randomly 50% of the women and 50% of the men to be in the Control group, and leave the rest 50% of each gender to be in the Experiment group.
That way – gender is less likely to introduce bias and serve as a confounding factor.

Slide 9

But after randomly splitting the experiment subjects, what if the customers know in advance that they are being tested for precisely this effect? If a customer knows in advance that she is in the ‘group that is not exposed to the alert’, it might affect her response, even without her being aware of this effect on her. Even if customers just find out they are in ‘competition’, they might tend to engage more (or less).
In order to prevent this, we want to do it in a way that neither group knows which version they are exposed to and that there is another design.
We call that a blind experiment.
In addition, what if the sales, marketing or support teams will treat differently (even by accident) the customers in the two groups? That will also turn into a confounding variable.
To avoid that type of thing happening, we are conducting a double-blind experiment: neither the customers, nor the company’s personnel knows which groups were the customers assigned to.

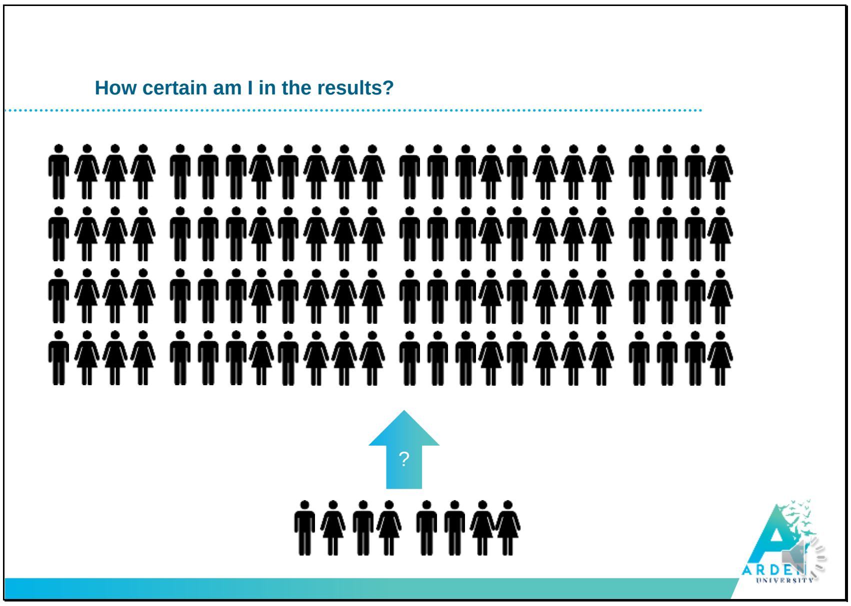
Slide 10



At the end of the two weeks, the company collected descriptive statistics of 100 customers. Two of the customers' data were outliers, showing an abnormal level of activity, so were taken out of the analysis.

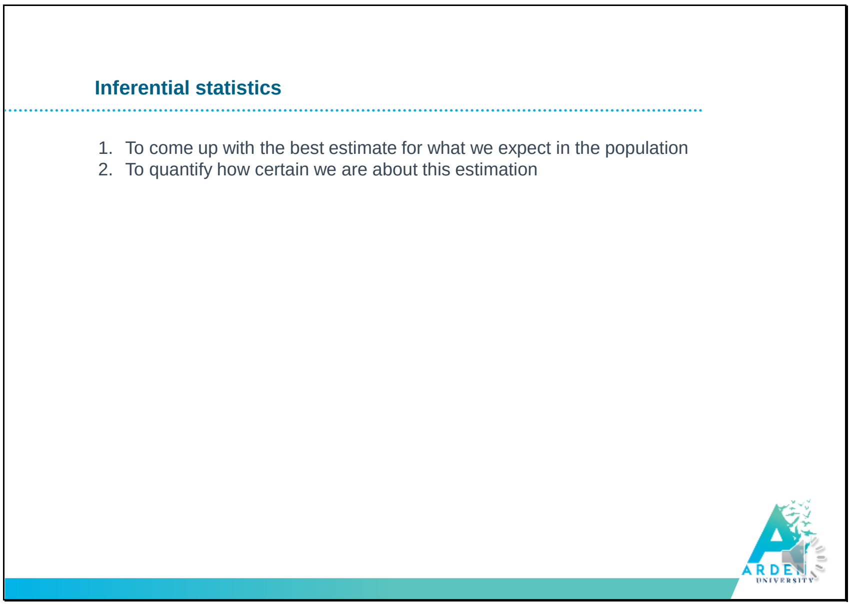
For the remaining 98 customers, the company finds out that the average number of clicks in the Experiment group is lower than the average in the Control group.

Slide 11



But how well can the difference between the two averages spotted by those 98 sampled customers be generalized to the whole population of one million customers?
Descriptive statistics can only describe the difference shown in the sample of the 98 customers.
Since the company cannot afford to continue this experiment and measure all their customers' responses, we must find a way to generalize from the difference in the sample to the difference in the population.
Specifically, in order to come to scientifically proven results, that we can be confident in reporting and acting upon, it is not only enough to make sure that we have no confounding variables, we must also find out and report:
What is the probability that this result (of difference between the groups) could have happened due to random chance?
In other words, what is the chance that if we would have continued the experiment for the whole population, we would have achieved the same results?

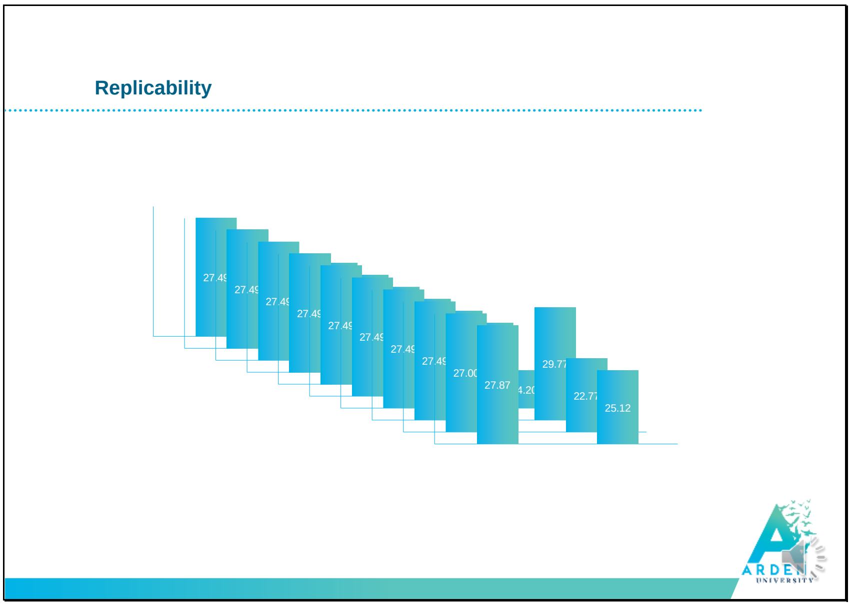
Slide 12



That is the role of inferential statistics:
first, to come up with the best estimate for what we expect in the population;
and second, to quantify how certain or uncertain we are about this estimation.

This uncertainty is due to the fact that the result is merely an estimation, drawn from a sample that might not be representative.
We will learn how to do this later in this lesson.

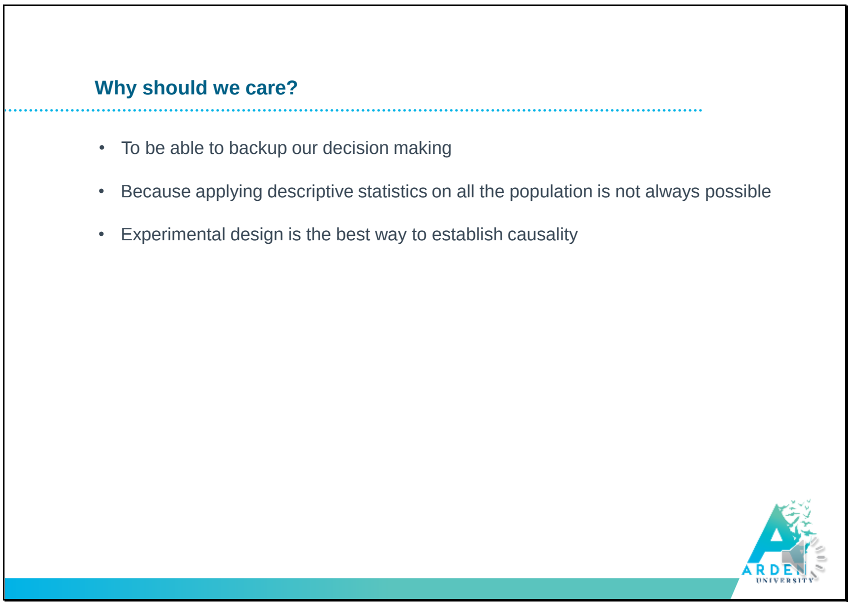
Slide 13



Experimental Error is the random variation present in all experimental results.
If we repeat this experiment dozens or hundreds of times, we will get different results to some extent.
It does not mean that our experiment was wrong.

Ideally, we would be able to repeat the experiment to make sure we ended up with the right conclusions.
In any case, a good scientific practice is to document the experiment in detail, to report any suspicion of bias in order for us (or others) to at least be able to replicate the experiment.

Slide 14

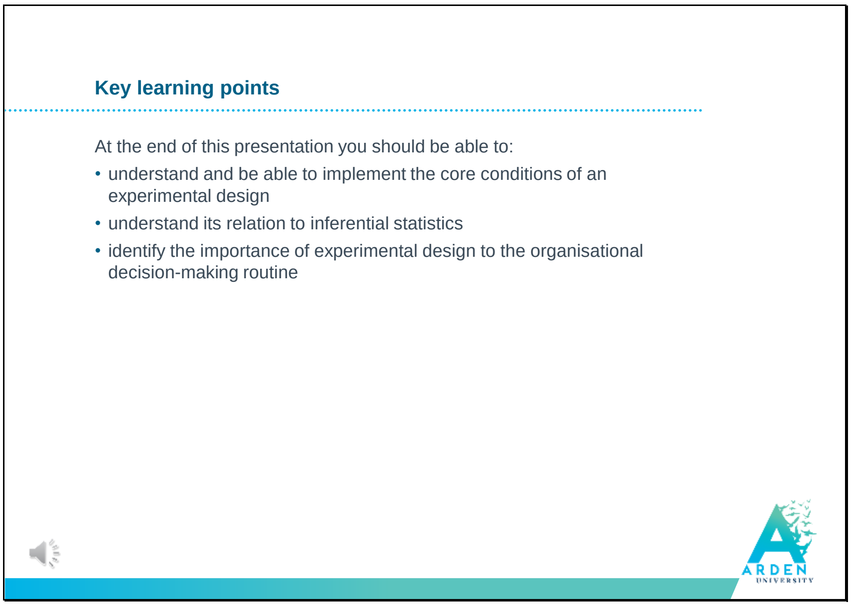


Why is all of this important?
First and foremost because we want to be able to provide the higher standard and confidence in decision-making. Performing an experiment requires us to pay close attention to all aspects of the study design, and to make sure that everything (from the data cleaning to statistical analysis) is tested and can be backed up. We need to make sure that we can report any issues that might make our results subject to reconsideration.

Second, as we discussed already, applying descriptive statistics on all the population is not always possible. It is either too expensive, or just not possible due to logistical or ethical reasons.

Third, experimental design is the best way to establish causality. If we find a correlation between two variables, that would not tell us anything about whether one of them caused the other. Experimental design can make sure we are truly establishing causality.

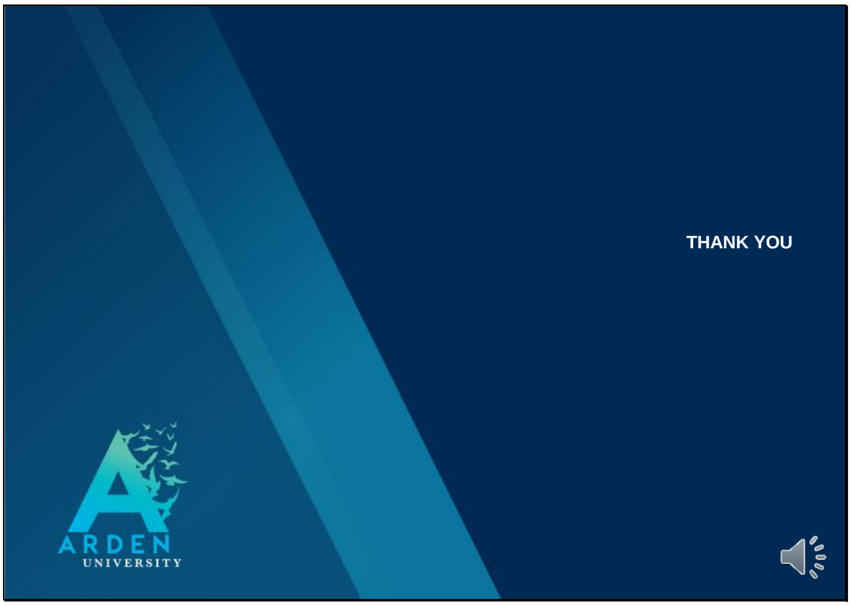
Slide 15



At the end of this presentation you should be able to:

- understand and be able to implement the core conditions of an experimental design
- understand its relation to inferential statistics
- identify the importance of experimental design to the organisational decision-making routine

Slide 16



Thank you.