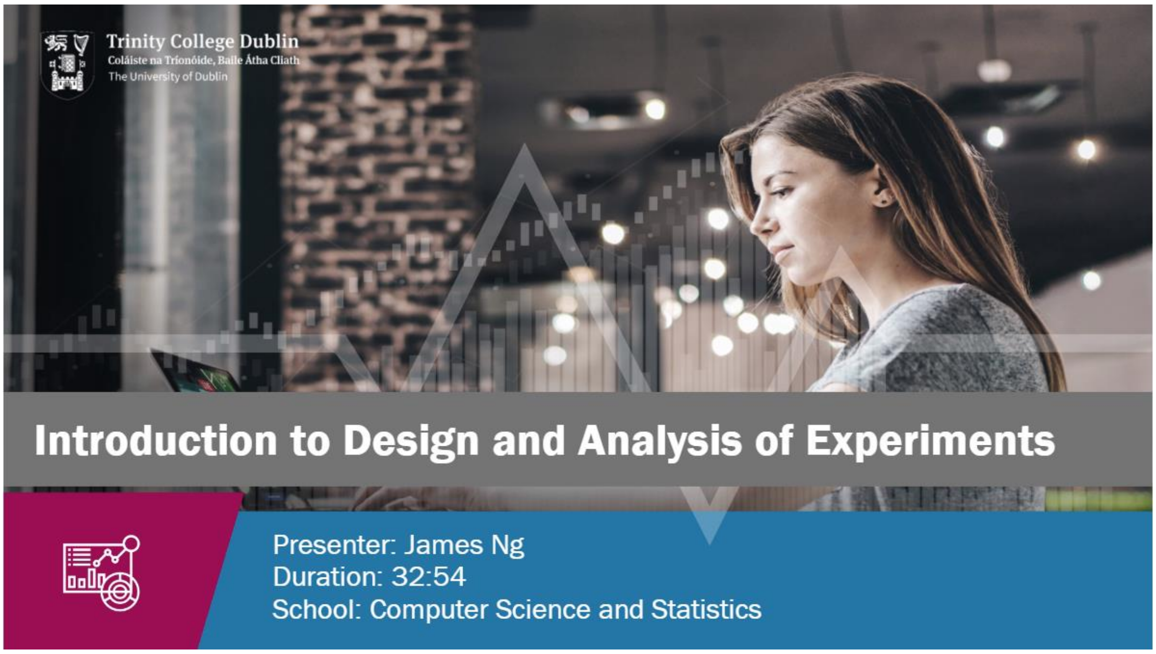
Introduction to Design and Analysis of Experiments

Contents

Slide 1: Introduction

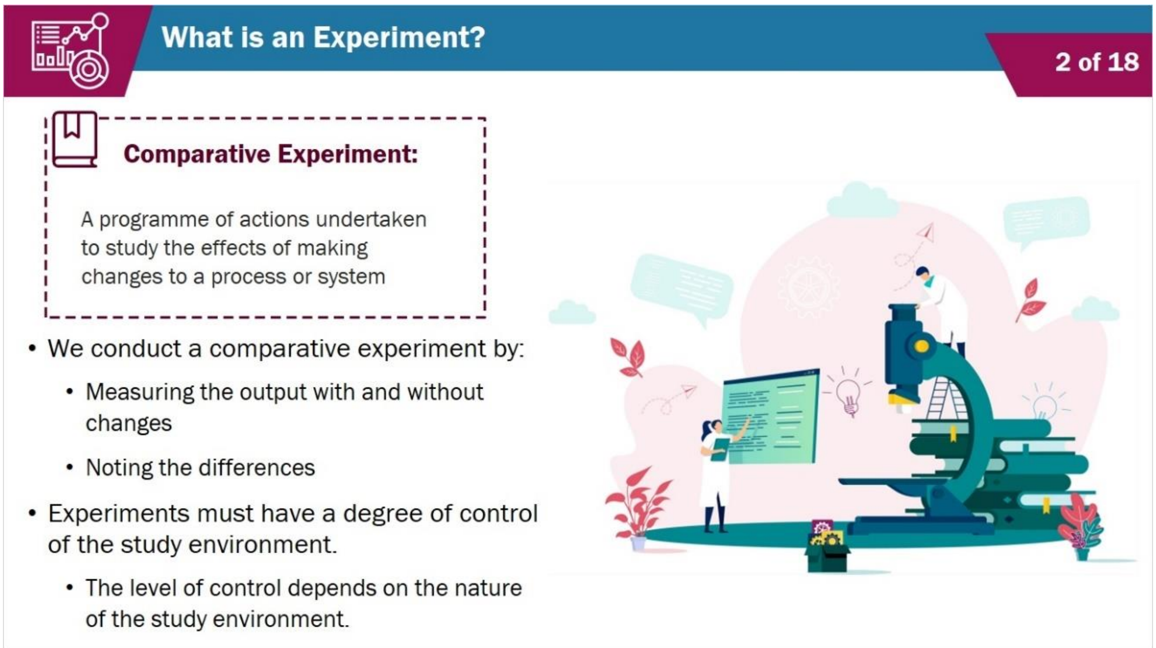


Hello and welcome. My name is James Ng and I will lead you through this presentation.

During this presentation I will introduce the basic concepts of design and analysis of experiments and present a few case studies.

We will examine the three main principles of design of experiments, namely, randomization, replication, and blocking. We will explore simple comparative experiments using a case study on manufacturing process improvement. We will compare multifactor experimental design with traditional design and illustrate the advantages of the former design. We will study the details of the multifactor experimental design and the procedure of analysis of variance through a case study on a filter membrane improvement project.

Slide 2: What Is an Experiment?

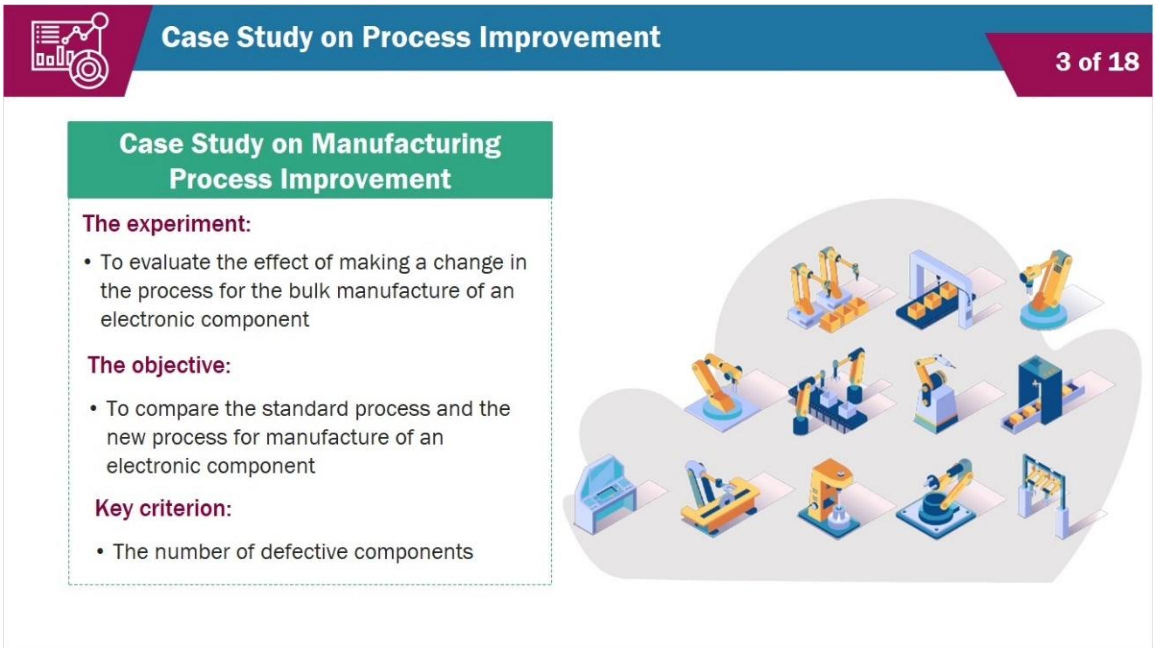


In this module, the focus is on comparative experiment. A comparative experiment is a programme of actions undertaken to study the effects of making changes to a process or system. Such an experiment is conducted by measuring the output of the process without and with the changes and noting the differences.

To find out what happens when you change something it is necessary to change it.

For comparative experiments to be successful, a key requirement is a suitable degree of control of the study environment. How much control can be exercised depends on the nature of the study environment.

Slide 3: Case Study on Process Improvement



To better understand comparative experiment, we are now going to look at a case study on manufacturing process improvement. An experiment is performed to evaluate the effect of making a change in a process for the bulk manufacture of an electronic component. The objective is to compare the standard (old) process and a new process for manufacture of electronic components based on the criterion number of defective components.

Slide 4: Experimental Design

Experimental Design

4 of 18

- Old and new processes were run on alternate days for eight weeks.
- The sequence was switched from week to week.
- 50 components were sampled per day and numbers of defective components counted.

	Week Number							
	1	2	3	4	5	6	7	8
Monday	Old	New	Old	New	Old	New	Old	New
Tuesday	New	Old	New	Old	New	Old	New	Old
Wednesday	Old	New	Old	New	Old	New	Old	New
Thursday	New	Old	New	Old	New	Old	New	Old
Friday	Old	New	Old	New	Old	New	Old	New
Saturday	New	Old	New	Old	New	Old	New	Old

The experiment was performed 6 days a week, from Monday to Saturday.

The old and new processes were run on alternate days for a period of eight weeks, switching the sequence (old followed by new or new followed by old) from week to week. Thus, during the first week, the old process was run on Monday, Wednesday and Friday with the new process being run on Tuesday, Thursday and Saturday while, during the second week, the new process was run on Monday, Wednesday and Friday with the old process being run on Tuesday, Thursday and Saturday, and continuing this pattern in subsequent pairs of weeks. On each day, a sample of 50 components was checked and the number of defective components counted.

Slide 5: Results

Results

5 of 18

Numbers of Defectives Per Daily Sample of 50 Components for 48 Days (8 Weeks)

Day	Defectives	Day	Defectives	Day	Defectives	Day	Defectives
1	0	13	1	25	0	37	2
2	0	14	0	26	0	38	0
3	6	15	3	27	0	39	0
4	3	16	1	28	2	40	0
5	3	17	0	29	0	41	0
6	3	18	2	30	0	42	0
7	4	19	0	31	1	43	1
8	1	20	1	32	1	44	0
9	0	21	2	33	0	45	2
10	2	22	0	34	0	46	0
11	0	23	1	35	0	47	0
12	0	24	3	36	2	48	0

- It appears that numbers of defectives are higher during the first four weeks.

The Numbers of defectives per daily sample of 50 for 48 days (8 weeks) are shown in the table. At first glance, it appears that the numbers of defectives are higher during the first 4 weeks compared to the last 4 weeks.

Slide 6: Comparison of Old and New Processes

Comparison of Old and New Processes

Comparison of Two Processes Over Eight Weeks. Number of Defectives in Samples of 50 Units

Day pair	Old Process	New Process	Difference (New - Old)	Day pair	Old Process	New Process	Difference (New - Old)
1	0	0	0	13	0	0	0
2	6	3	-3	14	0	2	+2
3	3	3	0	15	0	0	0
4	1	4	+3	16	1	1	0
5	2	0	-2	17	0	0	0
6	0	0	0	18	2	0	-2
7	1	0	-1	19	2	0	-2
8	3	1	-2	20	0	0	0
9	0	2	+2	21	0	0	0
10	1	0	-1	22	0	1	+1
11	0	2	+2	23	0	2	+2
12	3	1	-2	24	0	0	0

- The new process appears better but this may only be due to chance variation.

On this slide, the counts on successive *pairs* of days were recorded and tabulated as we see in this Table. On this evidence, it appears as if the new process is slightly better than the old. However, there is always the possibility that this apparent improvement is consistent with chance variation and that there is no *real* or long-term improvement.

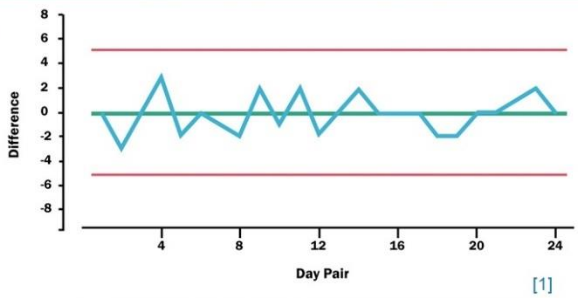
Slide 7: Differences in Numbers of Defectives, With Control Limits

Differences in Numbers of Defectives, With Control Limits

"Control" or "3-sigma" Limits

- We can expect some chance variation in processes, but we can put well-defined limits on that variation.
- If such limits are breached, there must be some assignable cause for the exceptional variation.

Differences in Numbers of Defective Components, With Control Limits



[1]

Upper/Lower 3-sigma
Difference between new/old processes

The Figure on this slide shows a line plot of the differences in the last column of the Table in the previous slide, with "control limits" representing a band of chance variation around a centre line at 0. These are "3-sigma" limits, based on Shewhart's idea that, in normal circumstances, we can expect some haphazard or chance variation in processes but that we can put more or less well defined limits on such variation and that, if such limits are breached, then we must conclude that there is some assignable cause for the exceptional variation.

Slide 8: Calculating the Control Limit

Calculating the Control Limit

To calculate the control limit:

- Compute the standard deviation of the differences (SD)
  - The control limit can be obtained as 0 +/- 3 SD(differences).

There is a link to these calculation details in the Extend section of your session homepage.

Comparison of Two Processes Over Eight Weeks.

Day pair	Old Process	New Process	Diff (New - Old)	Day pair	Old Process	New Process	Diff (New - Old)
1	0	0	0	13	0	0	0
2	6	3	-3	14	0	2	+2
3	3	3	0	15	0	0	0
4	1	4	+3	16	1	1	0
5	2	0	-2	17	0	0	0
6	0	0	0	18	2	0	-2
7	1	0	-1	19	2	0	-2
8	3	1	-2	20	0	0	0
9	0	2	+2	21	0	0	0
10	1	0	-1	22	0	1	+1
11	0	2	+2	23	0	2	+2
12	3	1	-2	24	0	0	0

- SD (Differences) = 1.57

Control limits:0 ± 3xSD = ±4.7

To calculate the control limit, we first compute the standard deviation of the differences SD (differences), and the control limit can be obtained as 0 +/- 3 SD (differences). You can find more details on this calculation in the appendix of Chapter 1, in the Extend section of your session homepage.

Slide 9: Comparison of Two Processes Over Eight Weeks

Comparison of Two Processes Over Eight Weeks

To compute the old and new process over eight weeks:

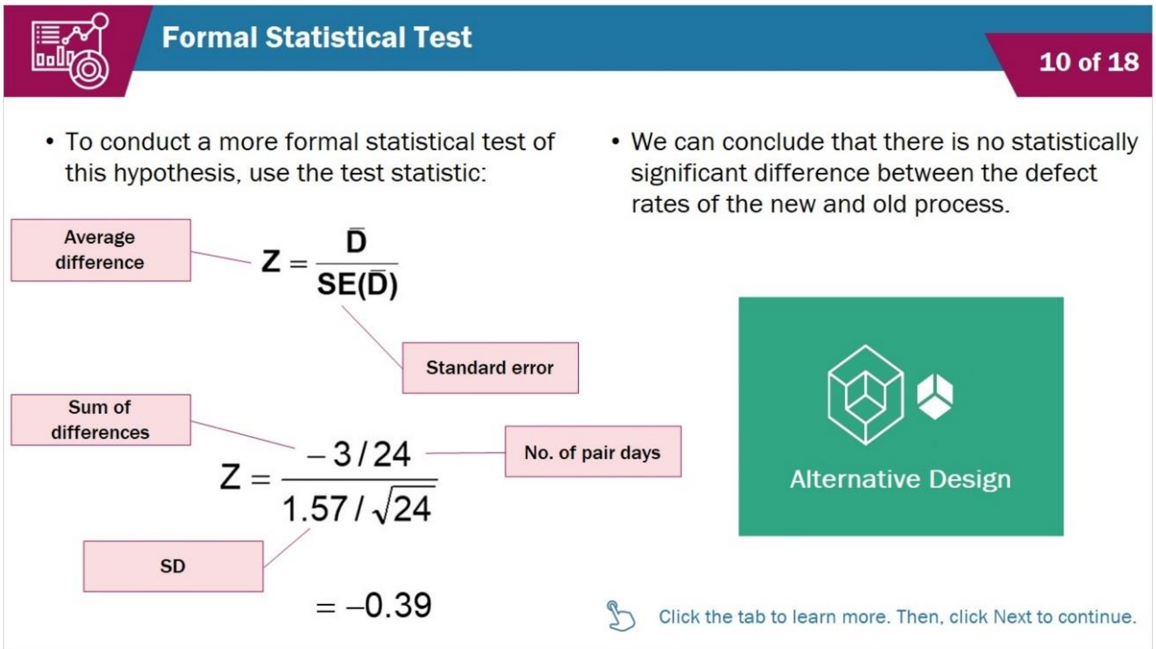
- Compute total and average number of defectives for the old and new process
- Calculate differences in totals and averages between the old and new processes

Number of Defectives Summary

	Old Process	New Process	Difference (New - Old)
Total	25	22	-3
8 week averages per cent	2.08	1.83	-0.25

We compute the total and average number of defectives for the old process and the new process over the eight-week period. The differences in the totals and the averages between the old and new processes are also calculated. For the old process, there are a total of 25 defectives, slightly more than the number of defectives for the new process.

Slide 10: Formal Statistical Test



We can conduct a more formal statistical test of this hypothesis.

A more formal test of this hypothesis supports this conclusion. The relevant test statistic is

$$Z = \frac{\bar{D}}{SE(\bar{D})}$$

where the numerator of the test statistic Z is the average difference, and the denominator is the standard error. The sum of differences equals to -3, as we previously computed. Therefore, the average difference is -3 divided by 24 where 24 is the number of day pairs. The standard deviation of differences is equal to 1.57, thus the standard error is 1.57 divided by square root of 24. We obtain a test statistic of Z = -0.39. Referred to a standard Normal frequency distribution, Z = −0.39 is not statistically significant.

The conclusion to be drawn from this experiment and its analysis is that there is no statistically significant difference between the defect rates of the new process and that of the old.

The protocol used to implement the experiment just described was quite complicated and resource intensive requiring, as it did, a process change every one or two days.

Click here to view an alternative design. Then click next to continue.

Tab 1: Alternative Design

Alternative Design

Alternative Design Proposed by Engineers

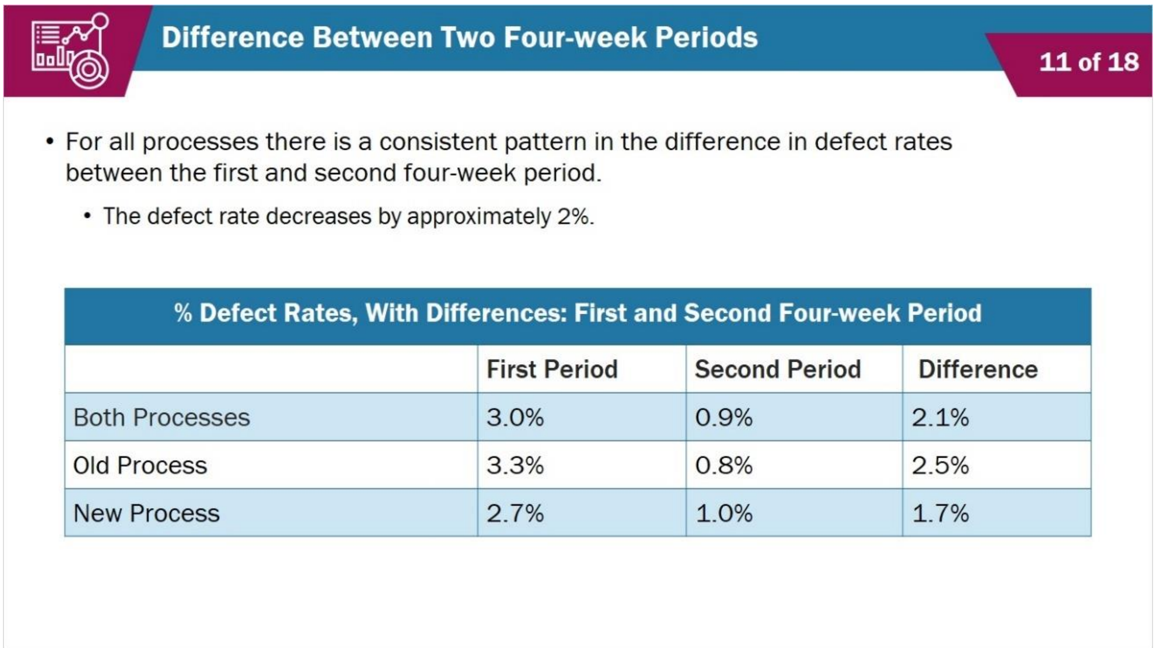
- An alternative approach is to:
  - Monitor the old process for four weeks
  - Introduce and monitor the new process for a further four weeks
  - Compare defect rates at the end of the eight-week period
- This is a flawed approach:
  - The process could be affected by another factor that changes with time

	Week Number							
	1	2	3	4	5	6	7	8
Monday	Old	Old	Old	Old	New	New	New	New
Tuesday	Old	Old	Old	Old	New	New	New	New
Wednesday	Old	Old	Old	Old	New	New	New	New
Thursday	Old	Old	Old	Old	New	New	New	New
Friday	Old	Old	Old	Old	New	New	New	New
Saturday	Old	Old	Old	Old	New	New	New	New

A much simpler approach would be to monitor the old process for a four-week period, then introduce the new process and monitor it for a further four weeks, and make a simple comparison of defect rates at the end of the eight-week period. In fact, this is what the engineers who were considering the process change proposed to do. However, this approach suffers from a serious flaw.

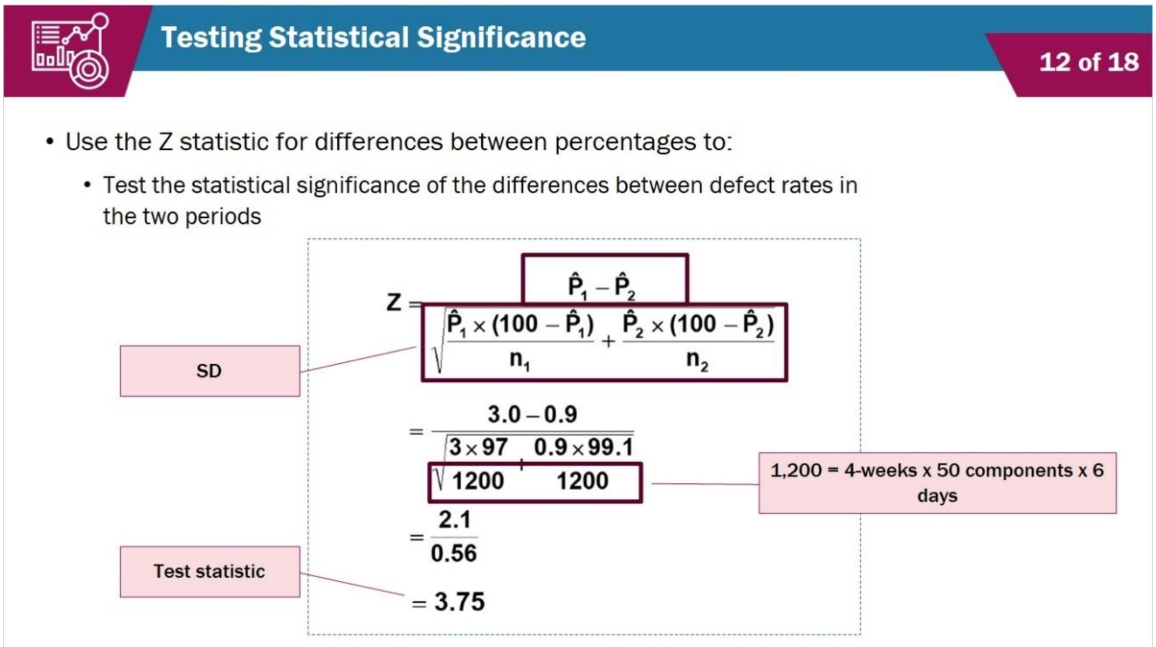
Conceivably, the process could be affected by some other factor that changes with time. For example, if the defect rate is sensitive to changes in ambient temperature, then the normal seasonal change in temperature, taken over a two month period, could influence the results of the experiment, so that any *perceived* difference between the old and new processes could well be due to seasonal change in temperature rather than process change.

Slide 11: Difference Between Two Four-week Periods



Going back to our original experiment we can see that The Table in this slide shows the difference in defect rates between the first four weeks and the second four weeks, for both processes and for each separately. The pattern is remarkably consistent, irrespective of process. We can see that the defect rate decreased by around 2% between the two periods.

Slide 12: Testing Statistical Significance

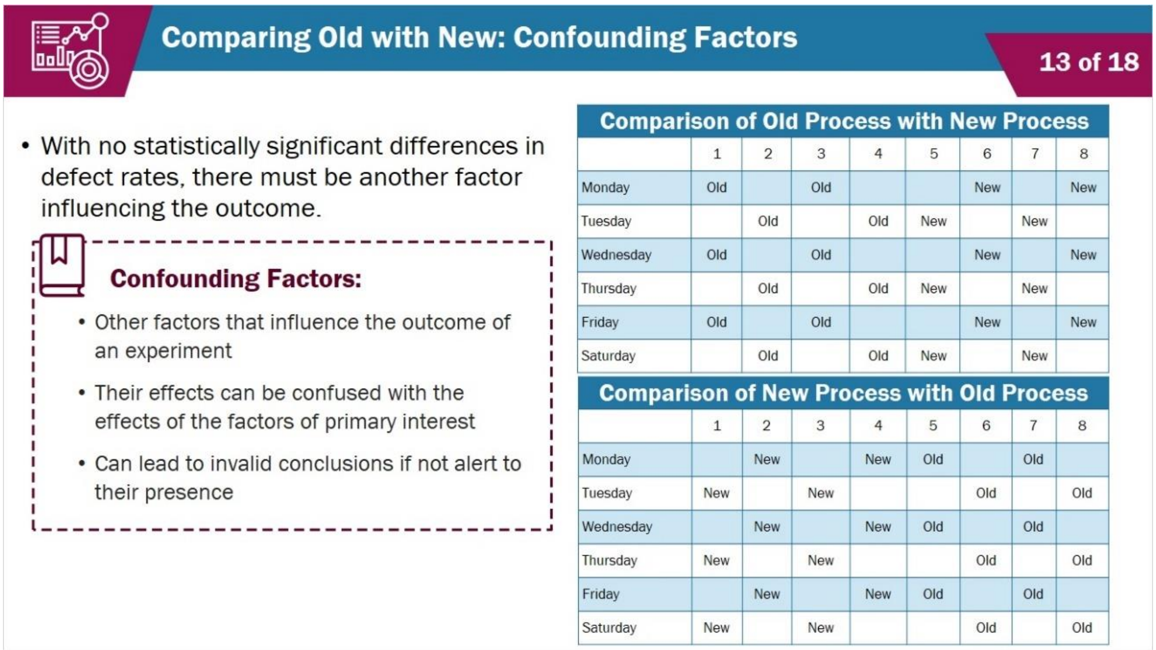


A formal test of the statistical significance of the observed differences between defect rates in the two periods may be based on the Z statistic for differences between percentages. The numerator of the test statistic is the difference of two percentages and the denominator is the square root of the sum of variances.

For both processes combined, the test statistic is 3.75, which is highly statistically significant. Note that the sample sizes are 1,200 because in each four-week period, 50 components were sampled per day for 6 days per week for 4 weeks

50 x 6 x 4 = 1,200.

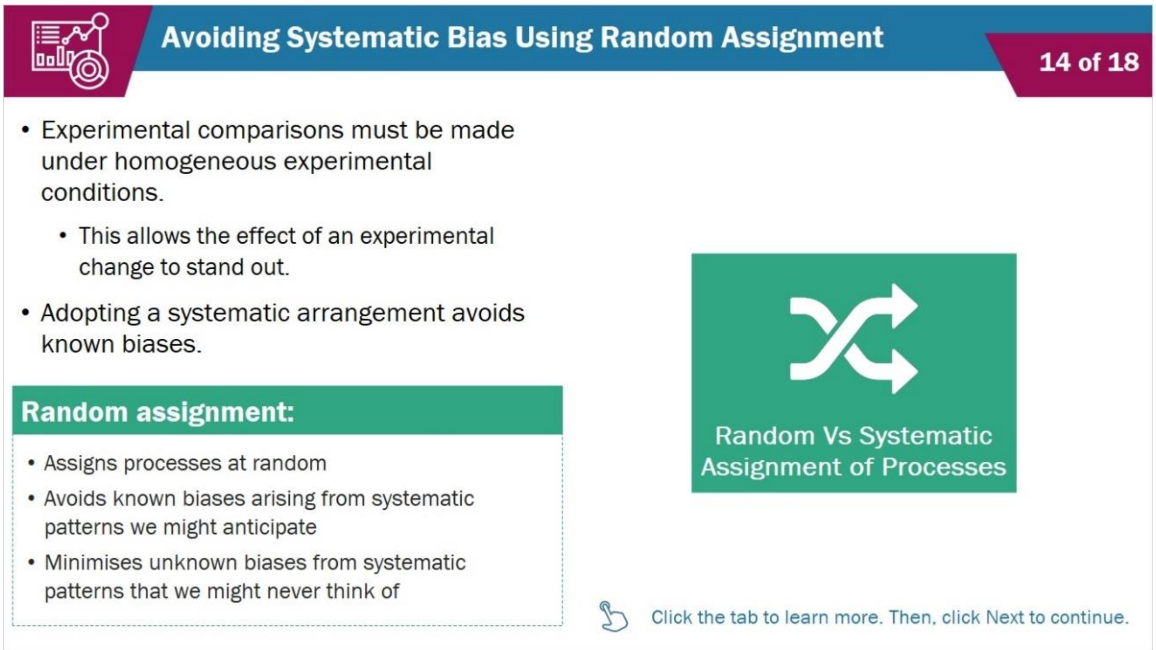
Slide 13: Comparing Old with New: Confounding Factors



Since the experiment actually carried out indicated that there was no statistically significant difference between the defect rates for the two processes, the "Old" and the "New", we can safely conclude that there must have been another factor influencing the outcome of the experiment.

Such factors are referred to as "confounding" factors. Their effects can be confounded with or confused with the effects of the factors of primary interest and, if there is no awareness of their presence, invalid conclusions way be drawn concerning the effects of the factors of primary interest. In the case of the experiment under discussion, if the old process had been run in the first four-week period and the new in the second, the experimenters would have been inclined to conclude that the new process was better. This is because the defect rate was higher in the first four weeks compared to the next four weeks for both processes. Therefore, if we were to run the old process in the first four-week period and the new process in the second four-week period, the old process would likely have a higher defect rate due to this confounding factor.

Slide 14: Avoiding Systematic Bias Using Random Assignment



A fundamental principle is that comparisons be made in the most *homogeneous* conditions possible, so that any effect of an experimental change will *stand out* as clearly as possible. If we refer back to our experiment, Clearly, conditions were not homogeneous across the two four week periods and the simple design that assigns the old process to the first period and the new to the second would have failed because of this. Alternating on a daily basis as implemented to avoid this known bias. This is a form of systematic arrangement.

An alternative to the systematic alternating of weeks adopted in this case, with a view to avoiding time related biases, is *random* assignment of processes to days within pairs. Then, if there are trends or other systematic patterns that affect the system, the chances that such systematic patterns affect the results of the experiment are very small. The great advantage of random assignment is that, not only does it minimise the chances of biases arising from systematic patterns that we might anticipate, but also it minimises the chances of biases arising from systematic patterns that we might never think of.

Click the tab to learn more about random vs systematic assignment. Then click next to continue.

Tab 1: Random Vs Systematic Assignment of Processes

Random Vs Systematic Assignment of Processes (1/3)

Random Vs Systematic Assignment: "Other Factor"

- Suppose there is an additional "other factor", unknown to the experimenter.
  - This "other factor" has the settings:
    - Up
    - Down
  - These settings alternate every day, including Sunday.

Systematic Assignment of Processes

	Week Number							
	1	2	3	4	5	6	7	8
Monday	Old	New	Old	New	Old	New	Old	New
Tuesday	New	Old	New	Old	New	Old	New	Old
Wednesday	Old	New	Old	New	Old	New	Old	New
Thursday	New	Old	New	Old	New	Old	New	Old
Friday	Old	New	Old	New	Old	New	Old	New
Saturday	New	Old	New	Old	New	Old	New	Old

Let's have a more detailed comparison between the random and systematic assignment. The table shows the systematic assignment of processes as implemented in our original experiment. Suppose that there was another systematic pattern that was not known to the engineers in charge of this process. Suppose that there is an additional "other factor", unknown to the experimenters, that can be Up or Down, alternating every day, including Sunday.

Tab 1.1: Random Vs Systematic Assignment of Processes: Confounding Factors

Random Vs Systematic Assignment of Processes (2/3)

Random Vs Systematic Assignment: Confounding Factors

- The daily patterns of the two factors show:
  - Old and Up always coincide
  - New and Down always coincide
- They are confounded.
  - Any changes in process quality cannot be ascribed to the change in process.

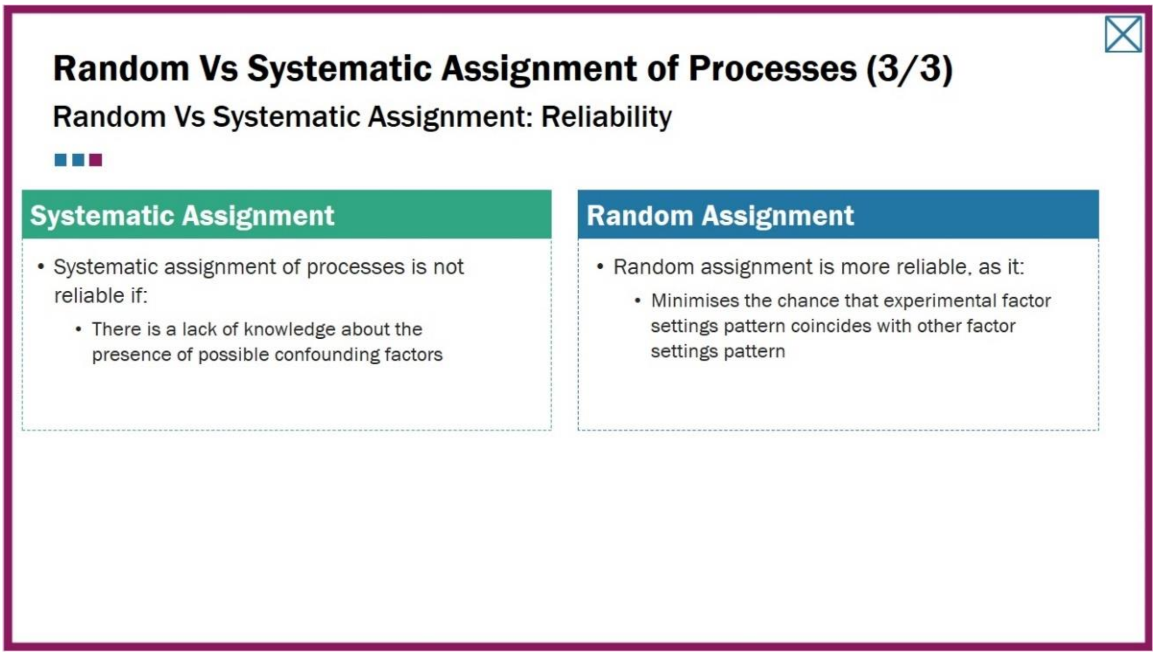
Confounding Factors

	Week 1		Week 2	
	Experimental Factor	Other Factor	Experimental Factor	Other Factor
Monday	Old	Up	New	Down
Tuesday	New	Down	Old	Up
Wednesday	Old	Up	New	Down
Thursday	New	Down	Old	Up
Friday	Old	Up	New	Down
Saturday	New	Down	Old	Up
Sunday		Up		

Assuming that the systematic assignment of Old and New processes to successive days outlined at the outset was in place, the combination of the known and unknown factors in the first two weeks would follow the pattern shown in the table.

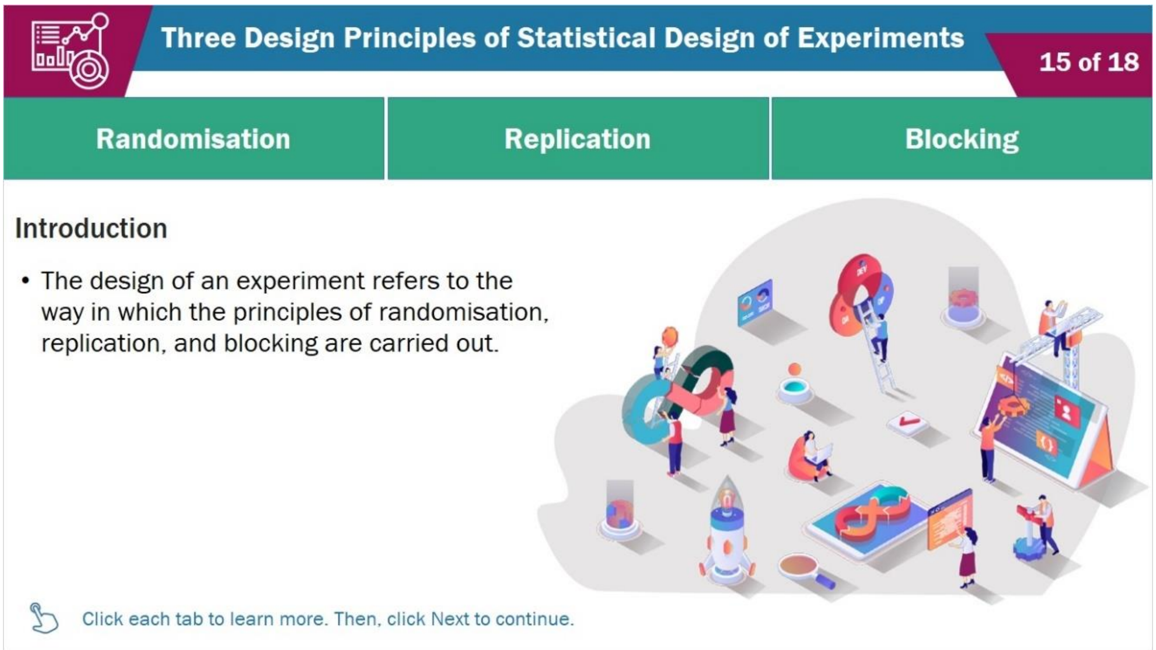
On inspection of the daily patterns of the two factors from Monday to Saturday, when the process is operational, it is seen that, whenever the Old process is run, the "other factor" is Up and whenever the New process is run, the "other factor" is Down. Thus, the two factors are irretrievably confounded and any observed change in process quality cannot be ascribed to the change in process solely.

Tab 1.2: Random Vs Systematic Assignment: Reliability



In the absence of knowledge or even suspicion of the presence of possible confounding factors, systematic assignment of processes to days is not reliable. On the other hand, random assignment is more likely to succeed by minimising the chances of variation patterns in an unknown factor coinciding with the variation patterns in the experimental factor.

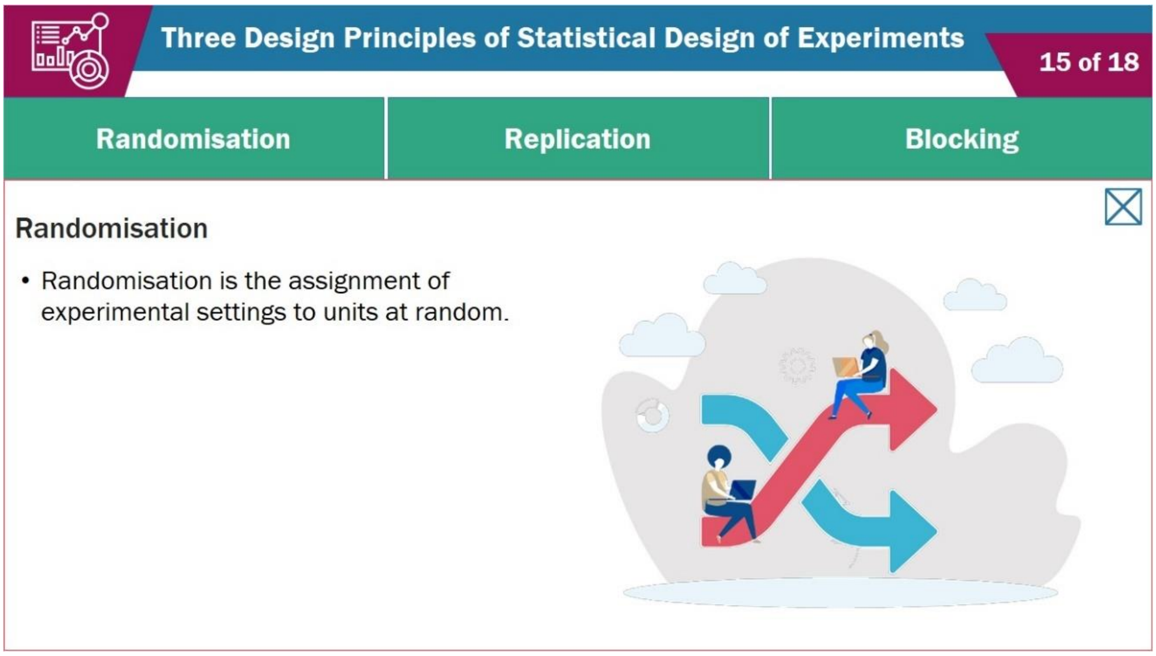
Slide 15: **Three Design Principles of Statistical Design of Experiments**



The three important design principles of statistical design of experiments are respectively randomization, replication, and blocking. When we talk about the design of an experiment, we are referring to the way in which these principles are carried out.

Click each tab to learn more. Then click next to continue.

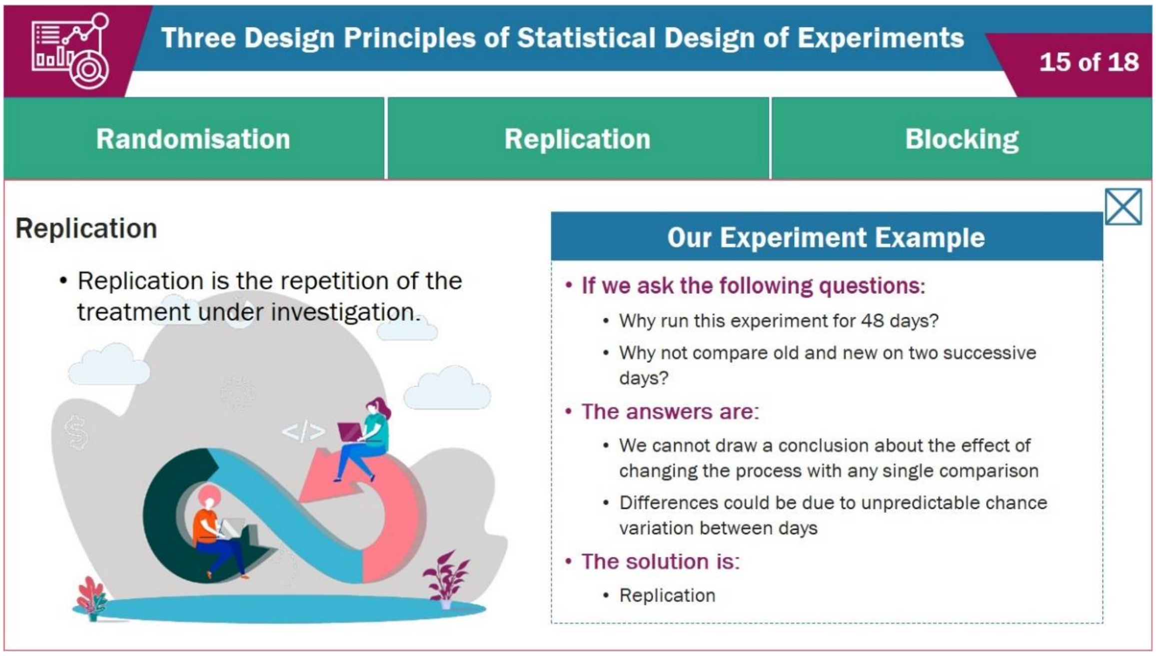
Tab 1: **Randomisation**



Randomisation is assigning experimental settings to units at random.

We have talked about randomization in the previous slide, which is the process of randomly assigning experimental units to one of the treatment groups.

Tab 2: Replication

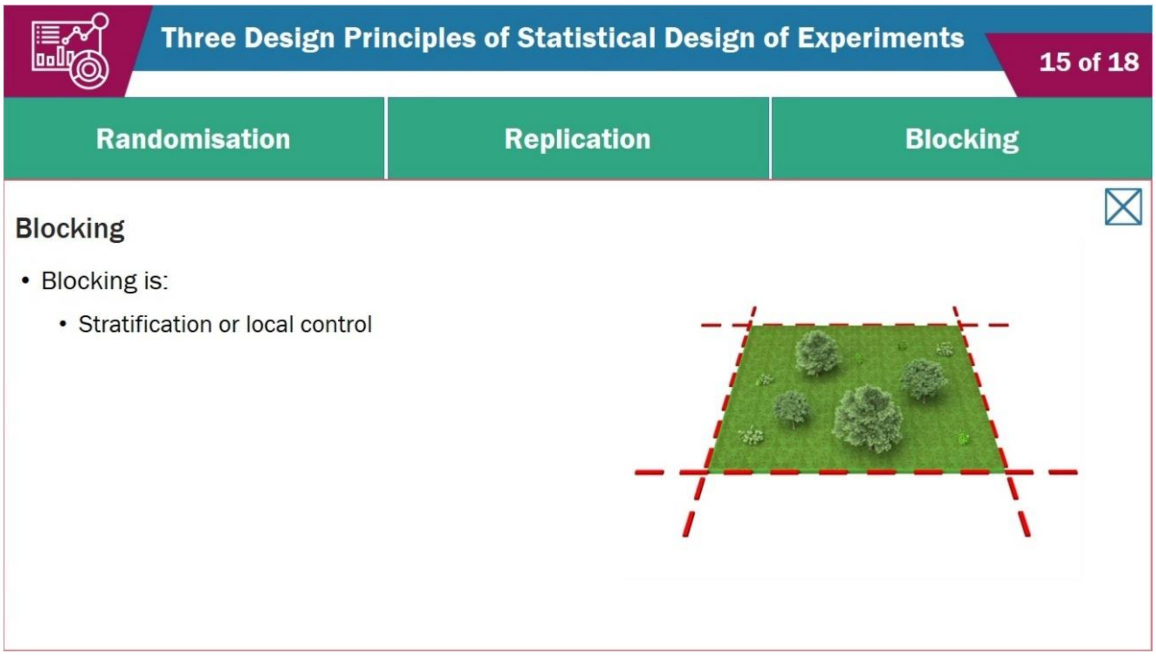


We may ask ourselves he question why this experiment needs to be run for 48 days, that is 24-day pairs resulting in 24 comparisons between Old and New? Why not just compare Old and New on two successive days, thereby making substantial savings.

The simple answer to this question is that a single comparison of Old and New on a pair of successive days will not allow any conclusion to be drawn about the effect of changing the process; any observed difference could just as well be due to the unpredictable chance variation between the days. The solution to this conundrum is *replication* whereby the comparison made on a pair of successive days is replicated on several other pairs of days, 24 in total in the case of this experiment.

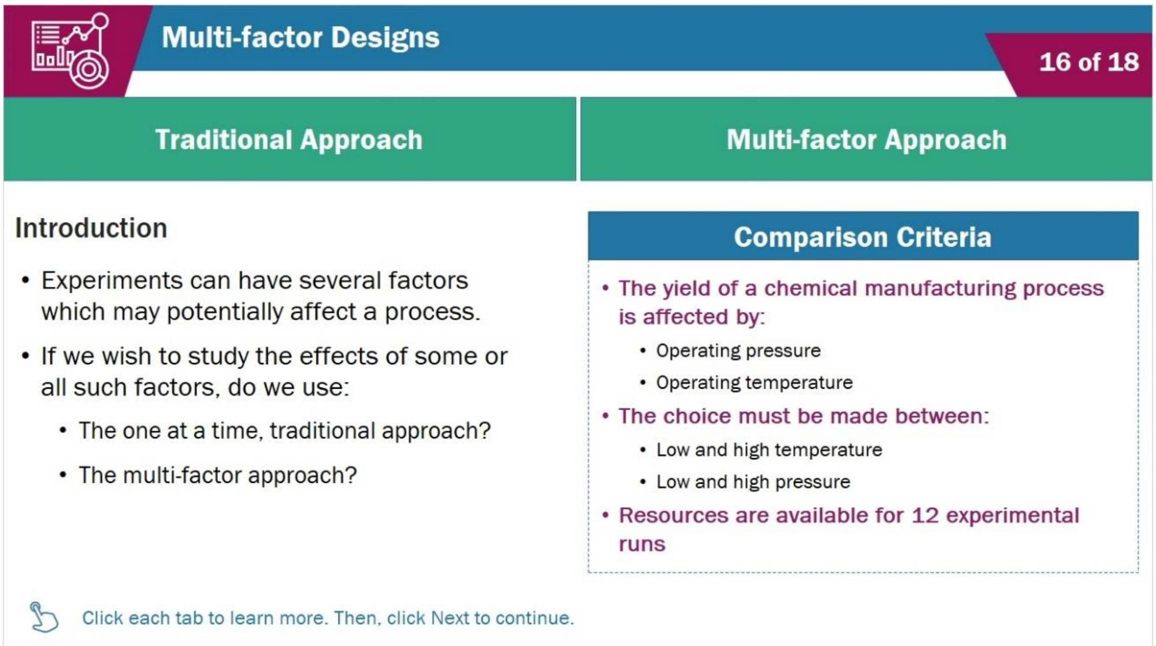
Replication is the repetition of the treatment under investigation. In this case study, the two treatments (new process and old process) were both replicated 24 times.

Tab 3: Blocking



We will study blocking in more details in other sessions. However, the pairing of Old and New processes on a pair of successive days is an example of what is more generally referred to as stratification or *local control*. It is a special case of *blocking*, a reference to contiguous plots of land used in agricultural experiments.

Slide 16: Multi-factor Designs



Up to this point, attention has been confined to just one experimental factor. In many cases, there are many factors that may potentially affect a process and we may wish to study the effects of some or all such factors.

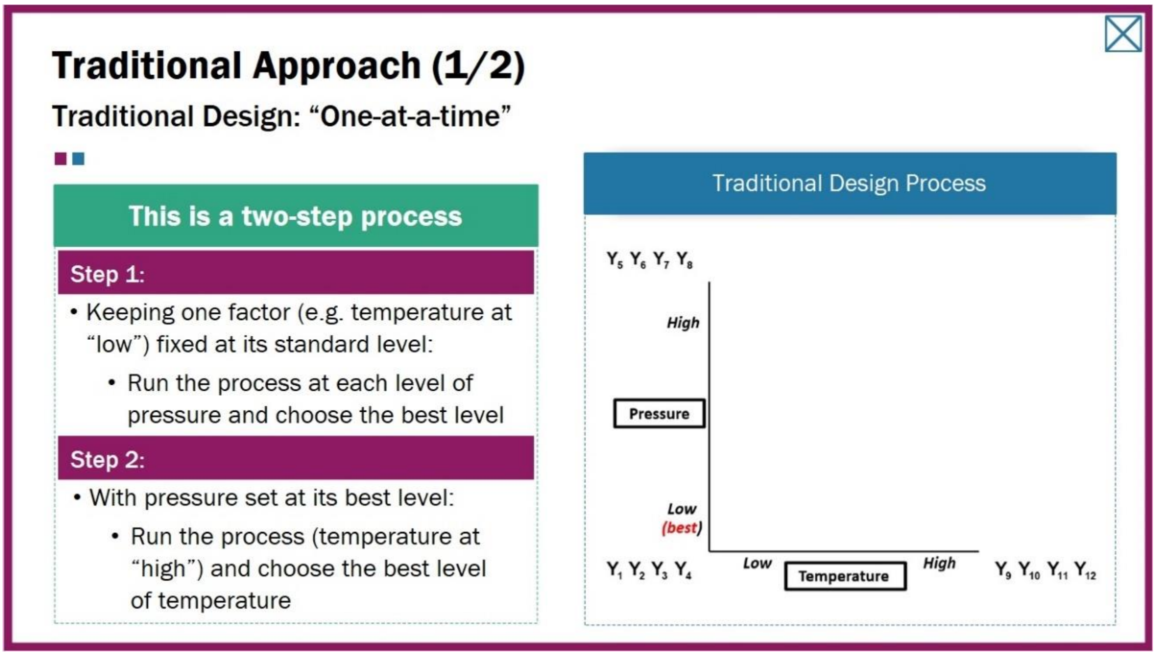
The first issue concerns whether factor effects should be studied one factor at a time or whether several factors should be studied simultaneously. The "change-one-factor-at-a-time" approach has been used traditionally in much scientific investigation, or the

the statistically recommended design approach which looks at all combinations of level of both factors in a single study?

Consider a process that may be affected by two factors, say a chemical manufacturing process where the yield of the process may be affected by operating pressure and operating temperature. A choice is to be made between two possible temperature levels, say "Low" and "High", as well as between low and high levels of pressure. We will also Suppose that available resources allow that the process may be run in experimental mode twelve times.

Click each tab to learn more. Then click next to continue.

Tab 1: Traditional Approach

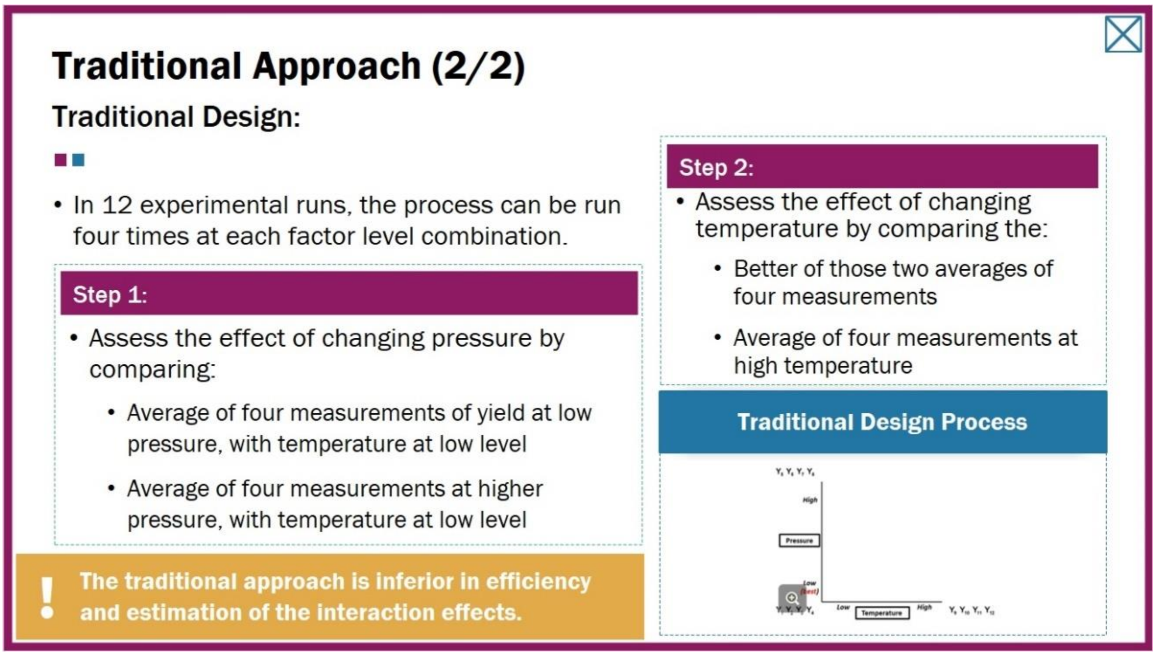


This traditional approach involves two steps:

keep one factor fixed at its standard level, say Temperature at "Low", run the process at each level of Pressure and choose the best level,

with Pressure set at its best level, run the process at the "High" level of Temperature and choose the best level of Temperature.

Tab 1.1: **Traditional Design**



With twelve experimental runs, it is natural to run the process four times at each factor level combination, that is

Low Temperature, Low Pressure,

Low Temperature, High Pressure,

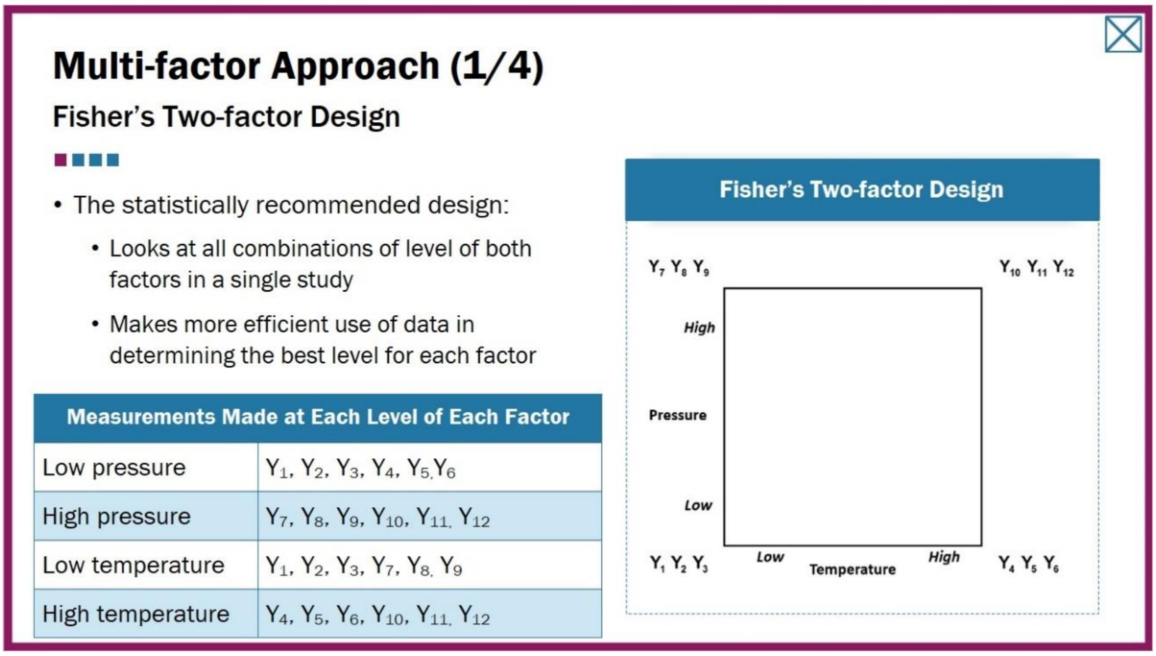
High Temperature, Best Pressure.

In the first step, the effect of changing Pressure is assessed by comparing the average of the four measurements of yield at Low Pressure with the average of the four measurements at High Pressure, with Temperature at its Low level in each case.

In the second step, the effect of changing Temperature is assessed by comparing the better of those two averages of four measurements with the average of the four measurements at High Temperature.

The traditional approach is inferior to the multi-factor approach, recommended by statisticians which we will look at next, in terms of efficiency and estimating the interaction effects.

Tab 2: **Multi-factor Approach: Fisher's Two-factor Design**



Here we will demonstrate that, compared to the traditional approach, the statistically recommended design looks at all combinations of level of both factors in a single study

Suppose available resources allow that the process may be run in experimental mode twelve times. It is easily demonstrated that the two factor (Fisher) approach makes more efficient use of the data.

In this design, there are six measurements made at each level of each factor:

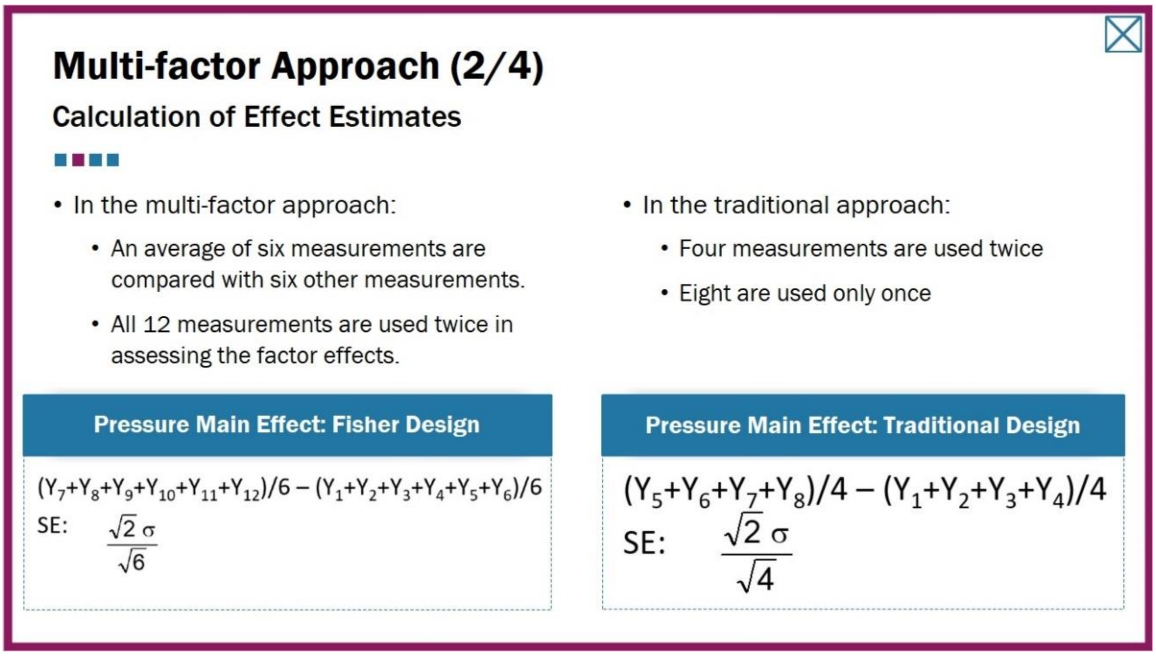
Y_1, Y_2, Y_3, Y_4, Y_5 and Y_6 at Low Pressure, compared to

Y_7, Y_8, Y_9, Y_{10}, Y_{11} and Y_{12} at High Pressure;

Y_1, Y_2, Y_3, Y_7, Y_8 and Y_9 at Low Temperature, compared to

Y_4, Y_5, Y_6, Y_{10}, Y_{11} and Y_{12} at High Temperature.

Tab 2.1: **Calculation of Effect Estimates**



It is seen that the effect of changing the levels of each factor is assessed by comparing an average of six measurements with an average of six other measurements. This represents a considerable improvement on the comparison of four with four employed in the traditional approach

To achieve the same quality of comparison with the traditional approach, eighteen measurements, divided into three subsets of six, would be required. Looking at it in another way, with the two-factor design, all twelve measurements are used twice in assessing the factor effects whereas, with the one-at-a-time approach, four measurements are used twice while eight are used only once. Thus, the two-factor approach makes much more efficient use of the twelve measurements available.

Tab 2.2: Multi-factor Approach: Revealing Interaction

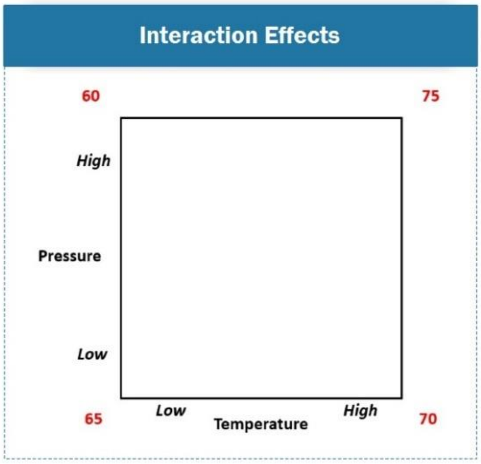
Multi-factor Approach (3/4)

Multi-factor Designs: Revealing Interaction

- Statistical design detects interaction effects among multiple factors.
- Factors interact when the effect of changing one factor depends on the level of the other.
  - This is an interaction between the factors.

Pressure Effect	
Low T	60 – 65 = -5
High T	75 – 70 = +5
Diff.	5 – (-5) = 10

Temperature Effect	
Low P	70 – 65 = 5
High P	75 – 60 = 15
Diff.	15 – 5 = 10



Another advantage of the statistical design over traditional design is its ability to detect interaction effects among the multiple factors.

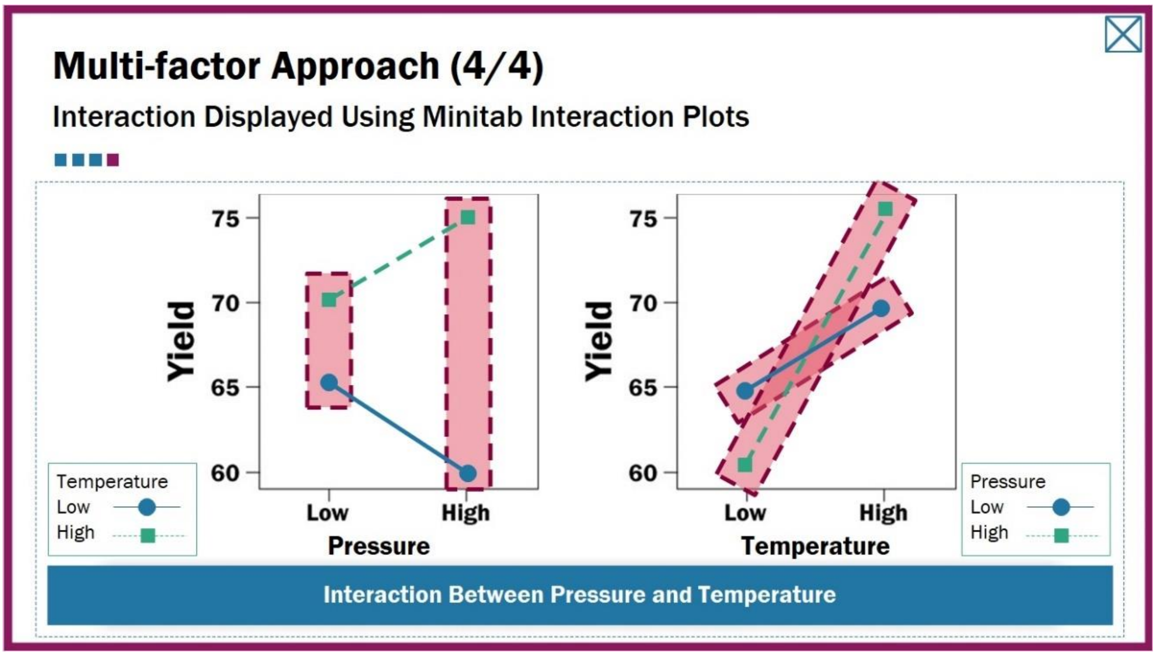
When the effect of changing one factor depends on the level of another factor, we say that these factors interact.

The interaction effect between Temperature and Pressure can be illustrated in this Figure.

Note that process yield increases by 5, from 65 to 70, when Temperature changes from Low to High at the Low level of Pressure. However, at the High level of Pressure, the effect of changing from Low to High Temperature is 15, that is, from 60 to 75. Correspondingly, at the Low level of Temperature, yield decreases by 5, from 65 to 60, when Pressure is changed from Low to High, whereas, at the High level of Temperature, yield increases by 5, from 70 to 75.

In short, the effect of changing the level of one factor depends on the level of the other factor. In statistical terminology, this is referred to as an *interaction* between the factors.

Tab 2.3: Interaction Displayed Using Minitab Interaction Plots



We can visualize the interaction between pressure and temperature with the help of Minitab. At low temperature, increasing the pressure level reduces the yield whereas at high temperature, increasing the pressure level results in an increase in the yield. If there were no interaction, the two lines should be approximately parallel. Therefore, to detect potential interaction effect, we should check whether or not the lines are parallel in the interaction plots.

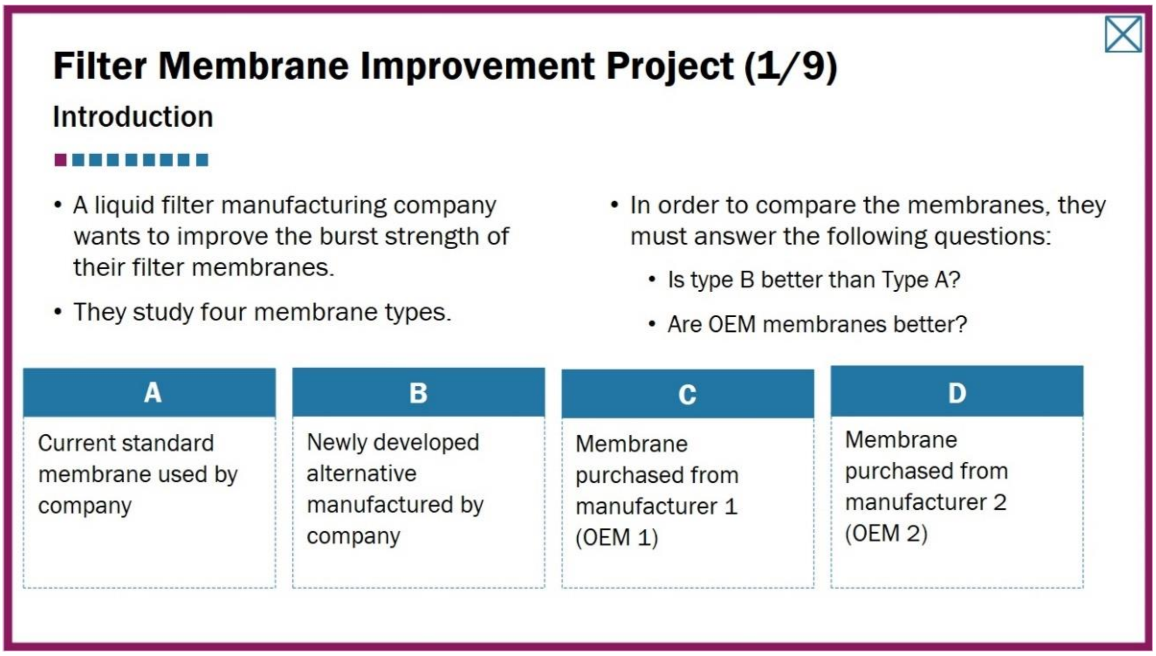
Alternatively, we can visualize the effect of temperature at low- and high-pressure level. We can see that while the effect of temperature is positive at both pressure levels, it is much larger at high pressure level. Again, the two lines are non-parallel which indicates potential interaction effect.

Slide 17: A Multi-level Experimental Factor

So far, we have only considered experimental factors with two levels. We are now going to look at a case study where the experimental factor has more than two levels. The project we are going to explore intended to improve the burst strength of filter membranes used to filter liquids in the pharmaceuticals and food industries.

Click the tab to learn more. Then click next to continue.

Tab 1: Filter Membrane Improvement Project



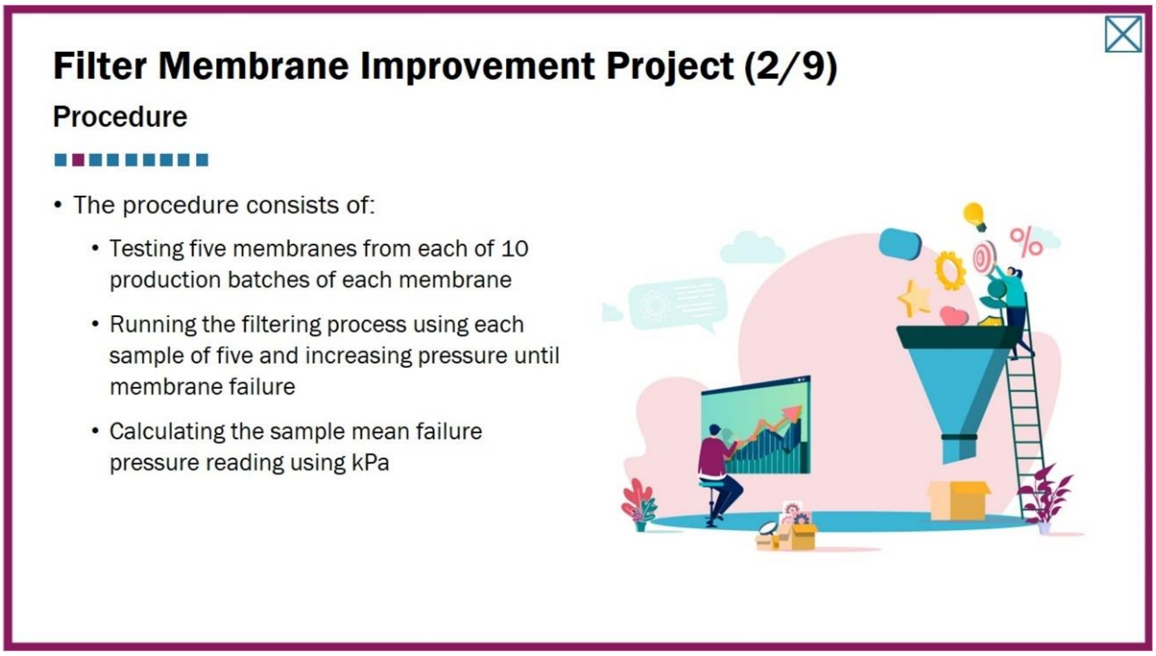
A company that manufactures liquid filters is concerned with improving the burst strength of the membranes which constitute the critical part of the filter. They have conducted a study of four types of filter membrane, labelled A, B, C and D. Membrane A is the standard type currently used by the company. Membrane B is an alternative produced by the company using a new material they have developed. Membranes C and D were purchased from other manufacturers.

In order to compare the membranes, there are two key questions that we would like to answer:

is Type B better than Type A?

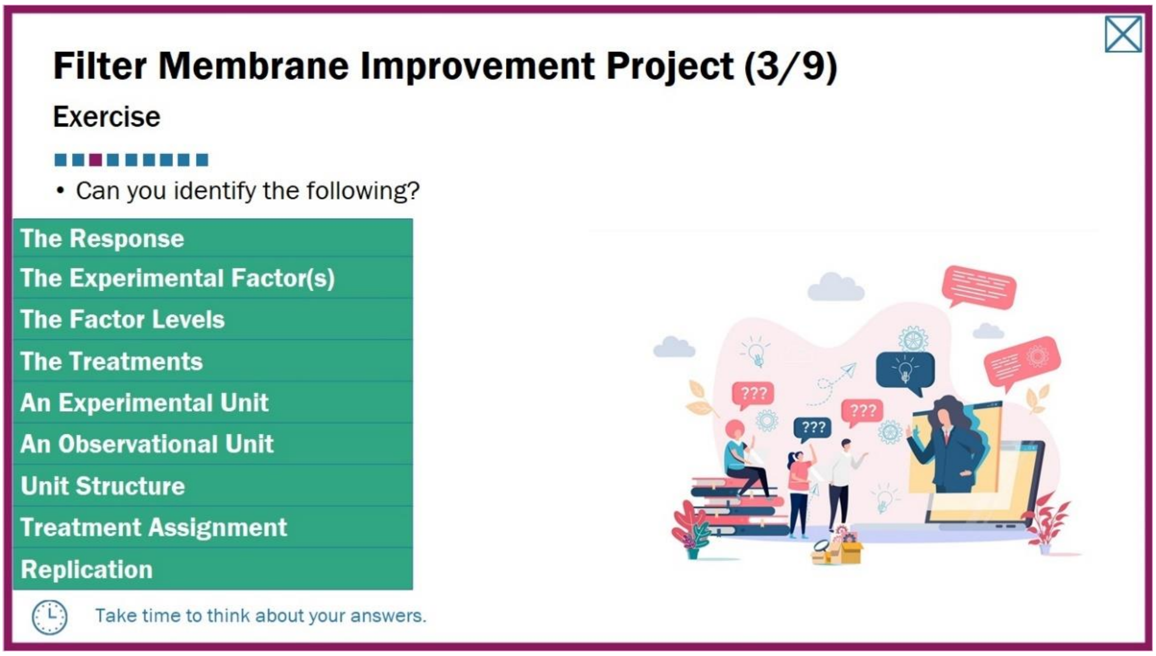
are OEM membranes better?

Tab 1.1: **Procedure**



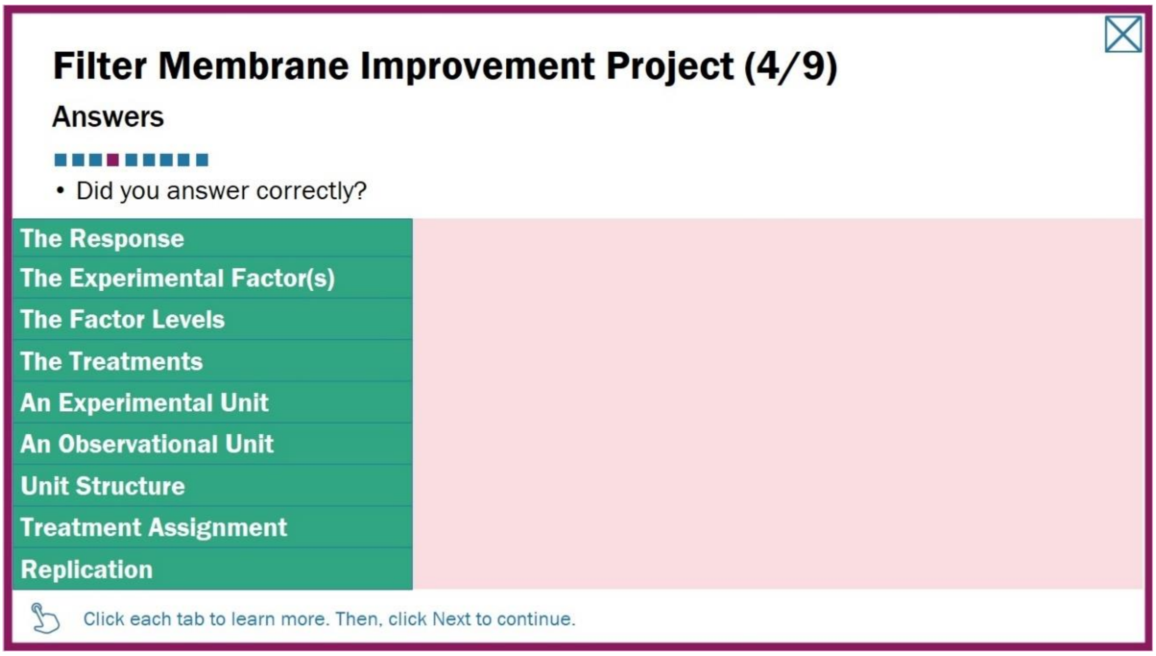
Following a review of historical data, it was decided to test five membranes sampled from each of ten batches of each membrane type. The standard measure of burst strength involved increasing the pressure of liquid through each of the five filters from a batch until the filter failed and averaging the 5 fail pressure readings. The measurement unit was kilopascal (thousands of Pascals, kPa).

Tab 1.2: Exercise



I would like you now to identify the following: the response, the experimental factor(s), the factor levels, the treatments, an experimental unit, an observational unit, unit structure, treatment assignment, and replication for this case study. Take time to think about your answers. Then, click next to continue.

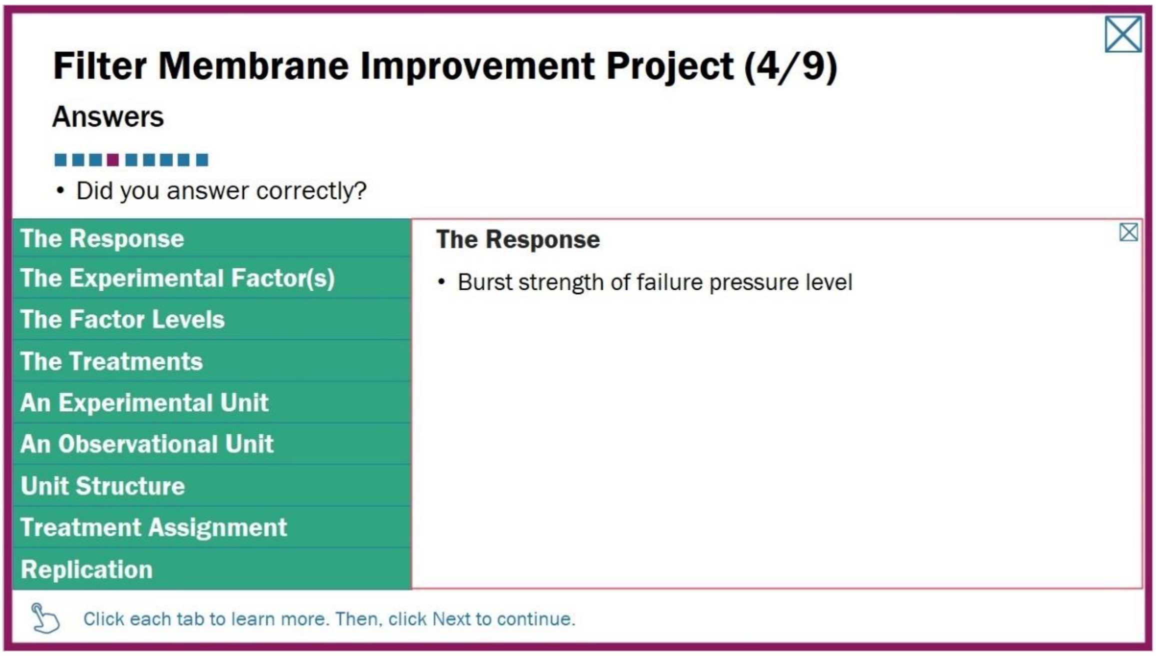
Tab 1.3: Case Study Answers



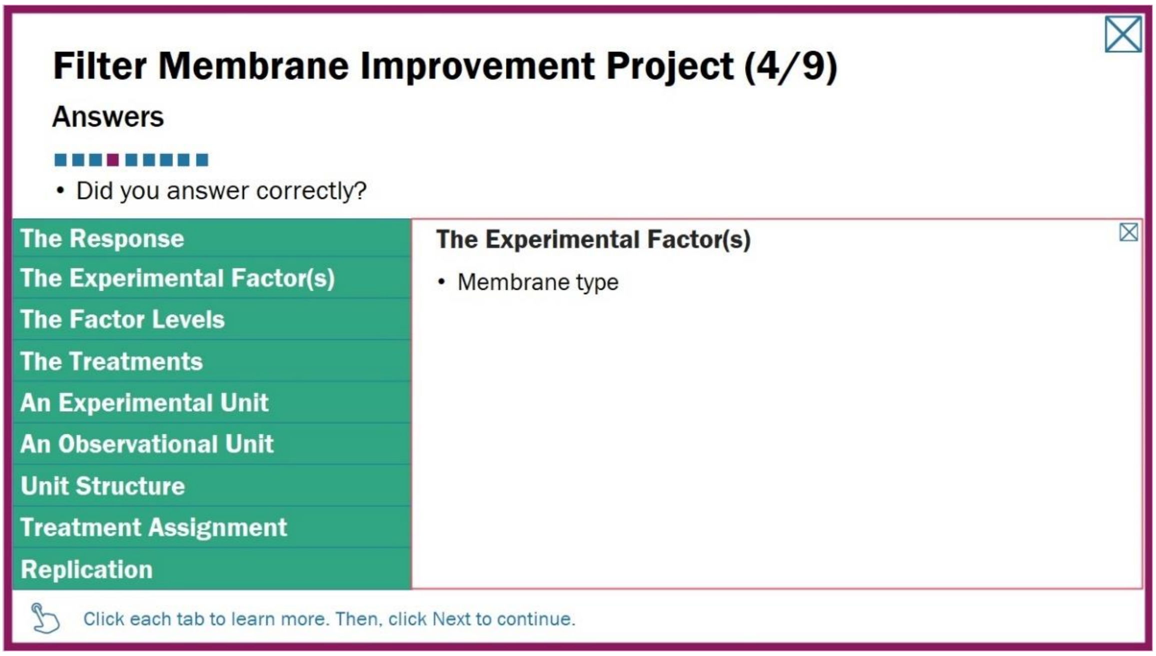
Now that you have considered the previous questions, it's time to look at the answers.

Click each tab to check if you answered correctly.

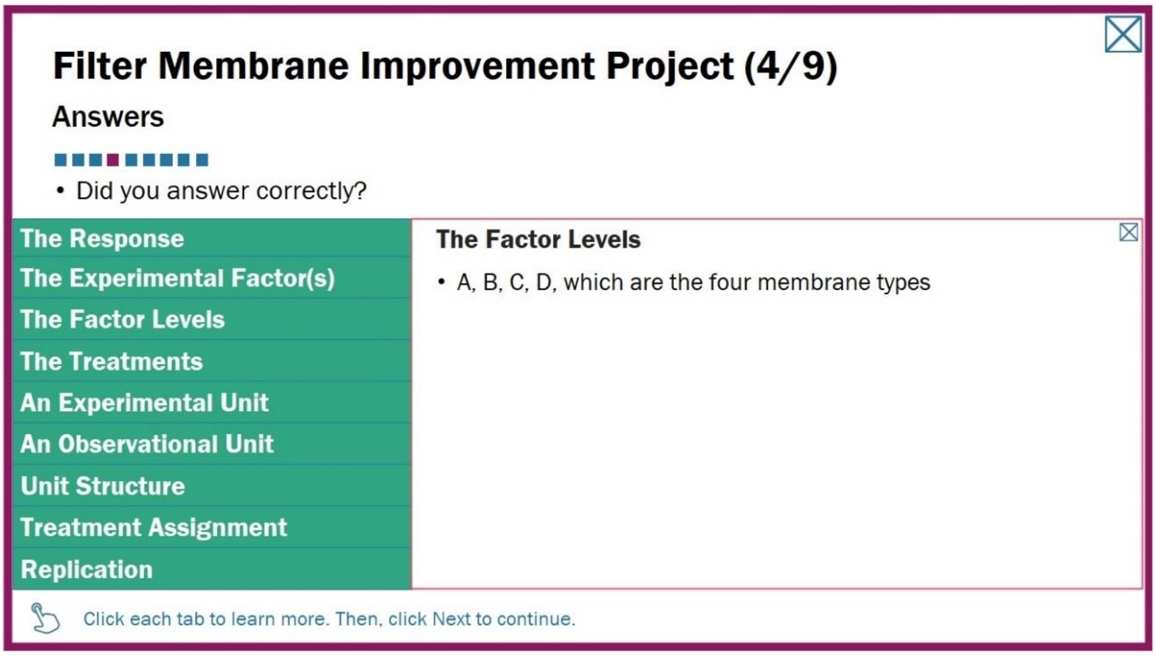
The response is burst strength or failure pressure level.



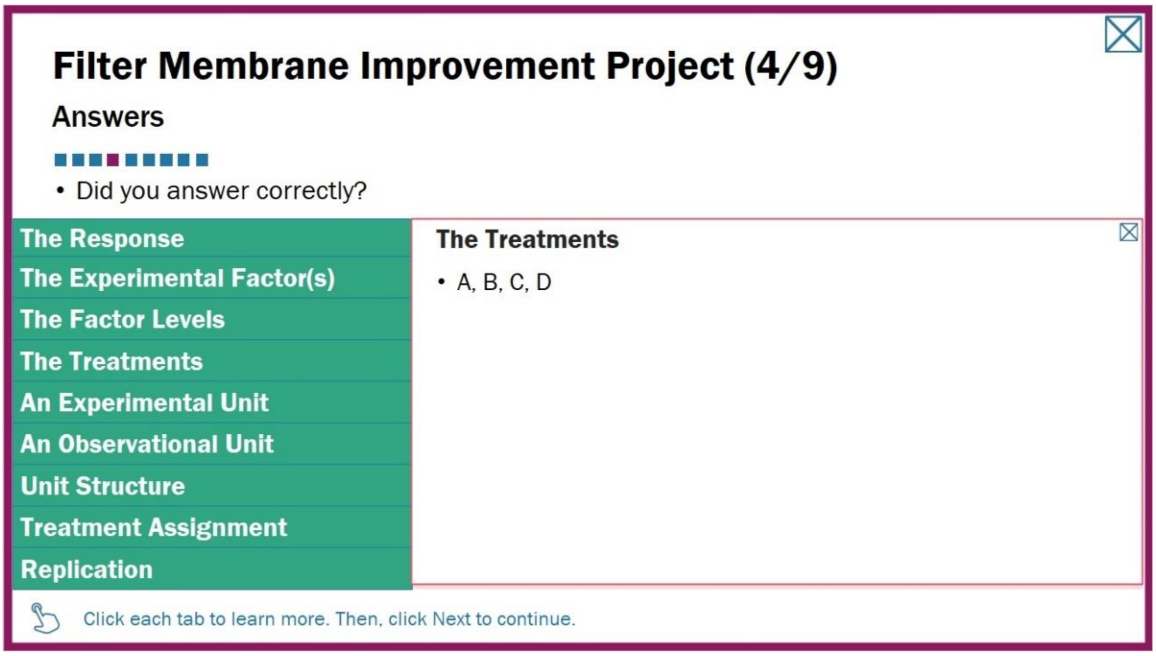
The experimental factor is membrane type.



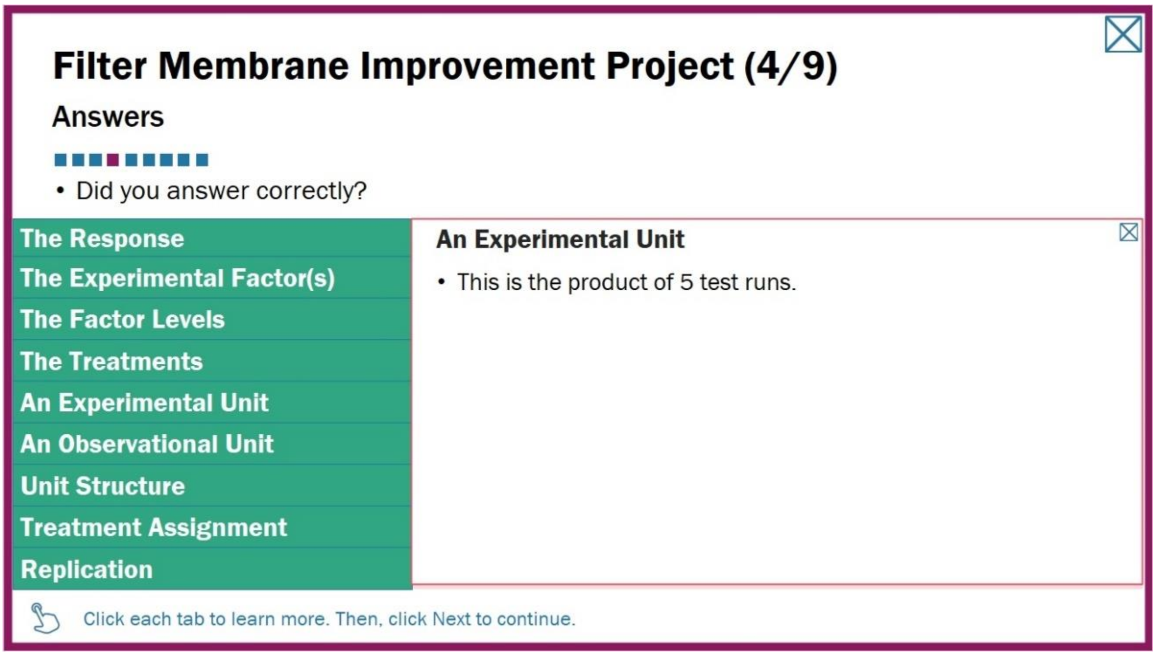
The factor levels are membrane A, B, C, D which are the four membrane types.



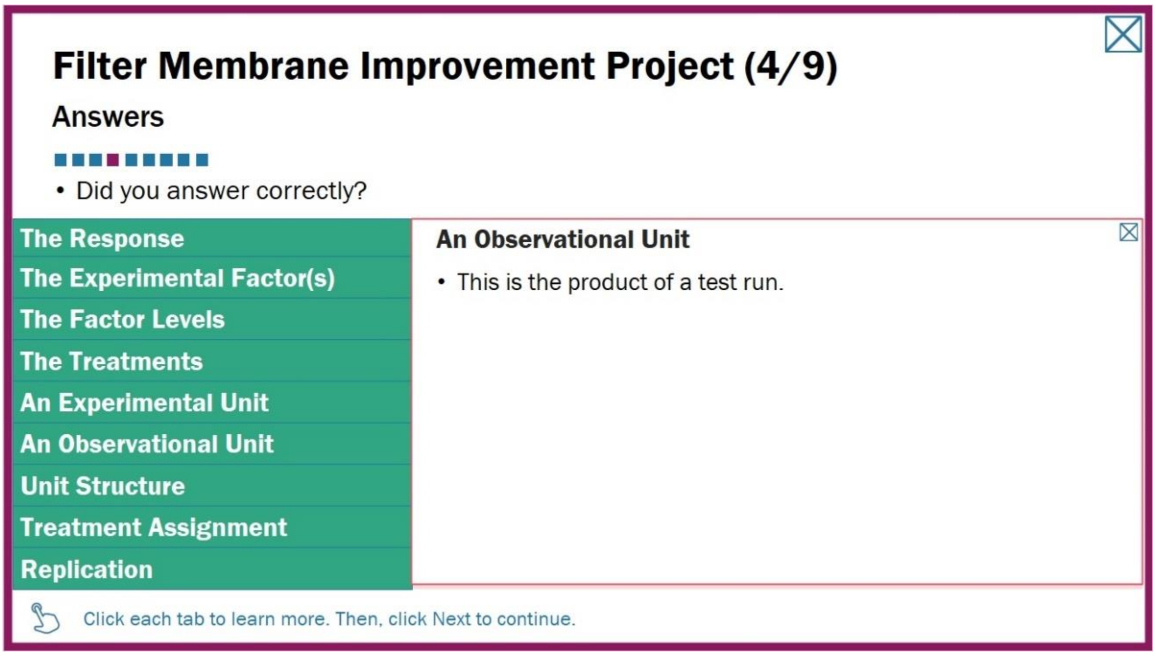
The treatments are also membrane A, B, C, D. When there is a single factor, treatments and factor levels coincide.



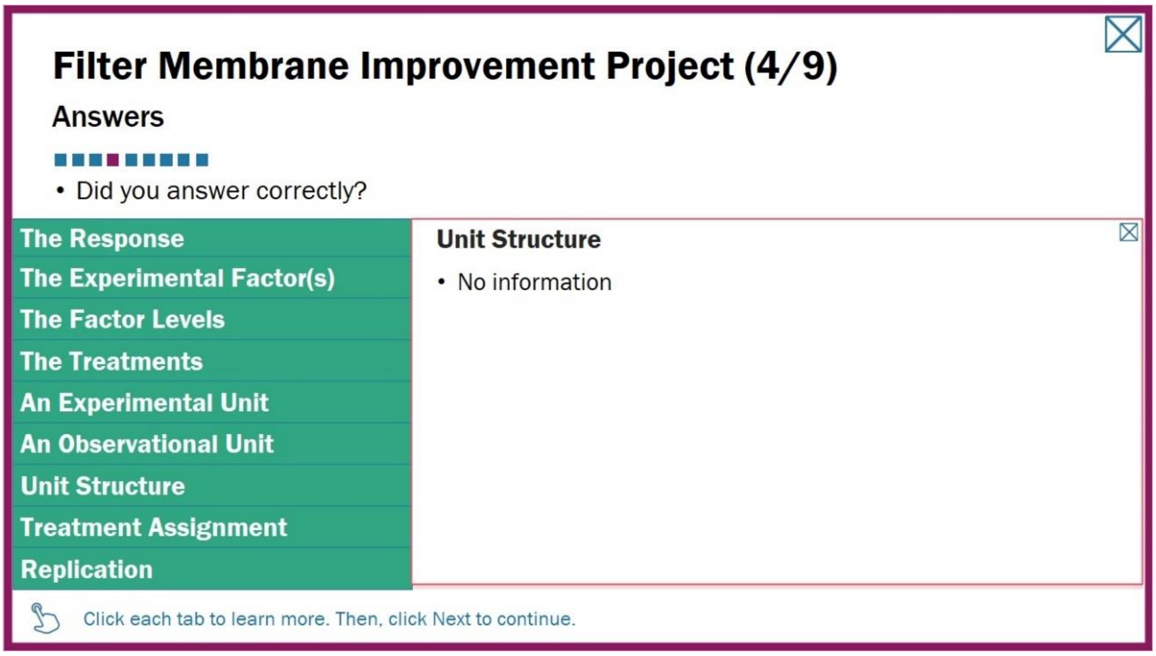
An experimental unit is the product of 5 test runs. This is the smallest unit to which a treatment is applied.



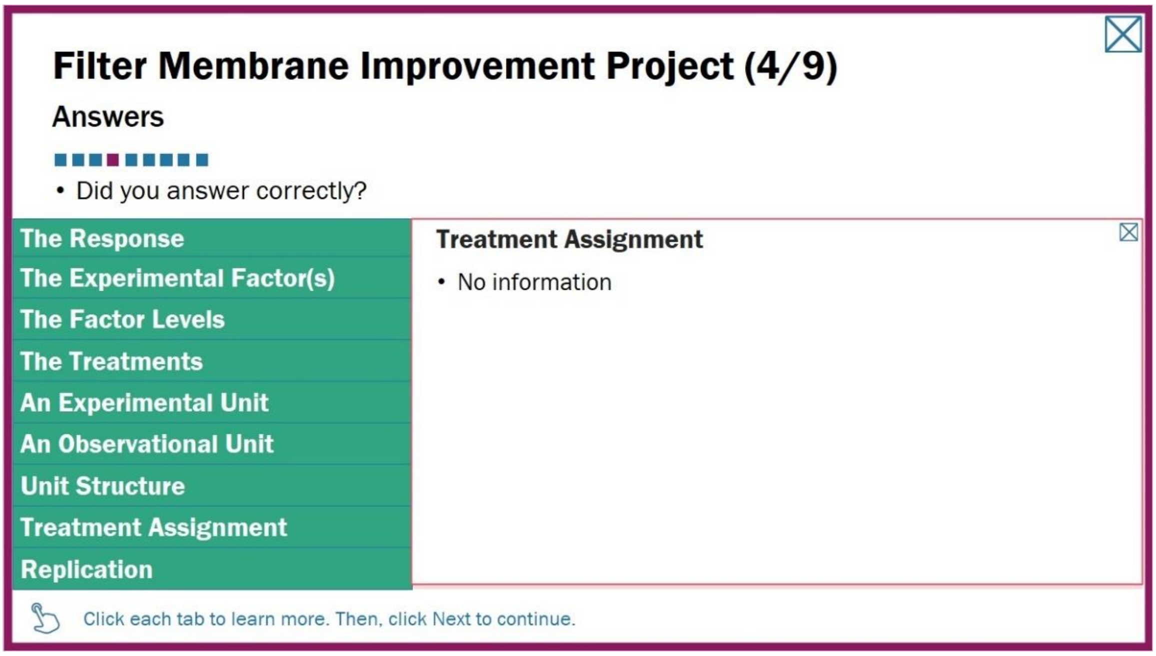
An observational unit is the product of a test run – the smallest unit on which a response is measured.



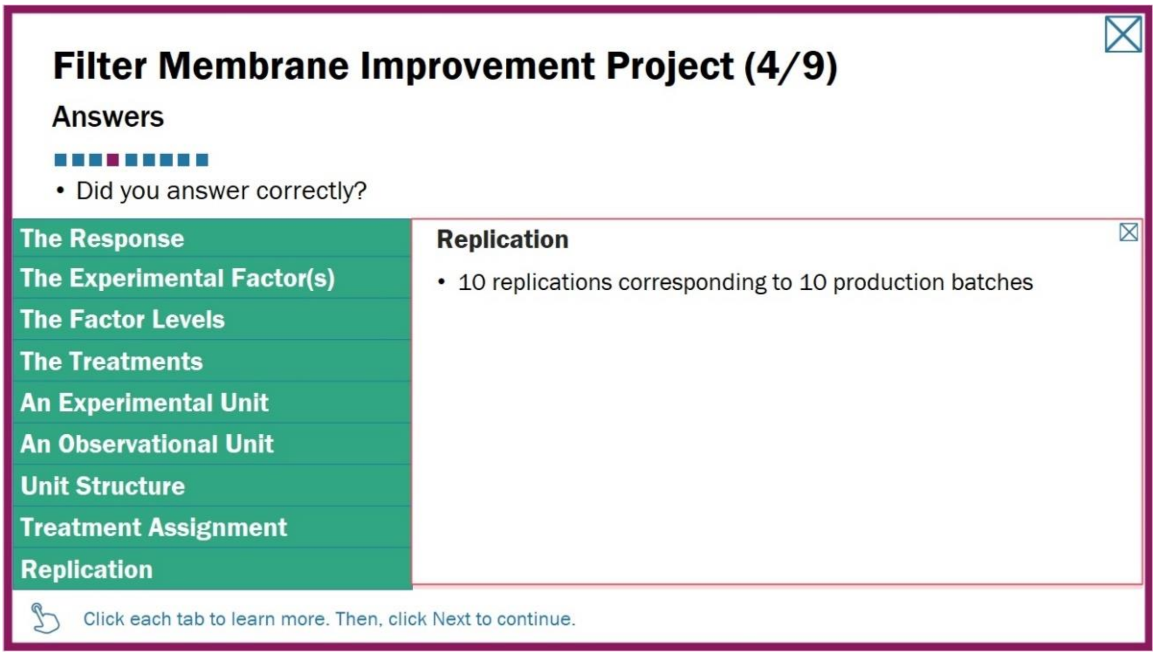
There is no information given on the unit structure.



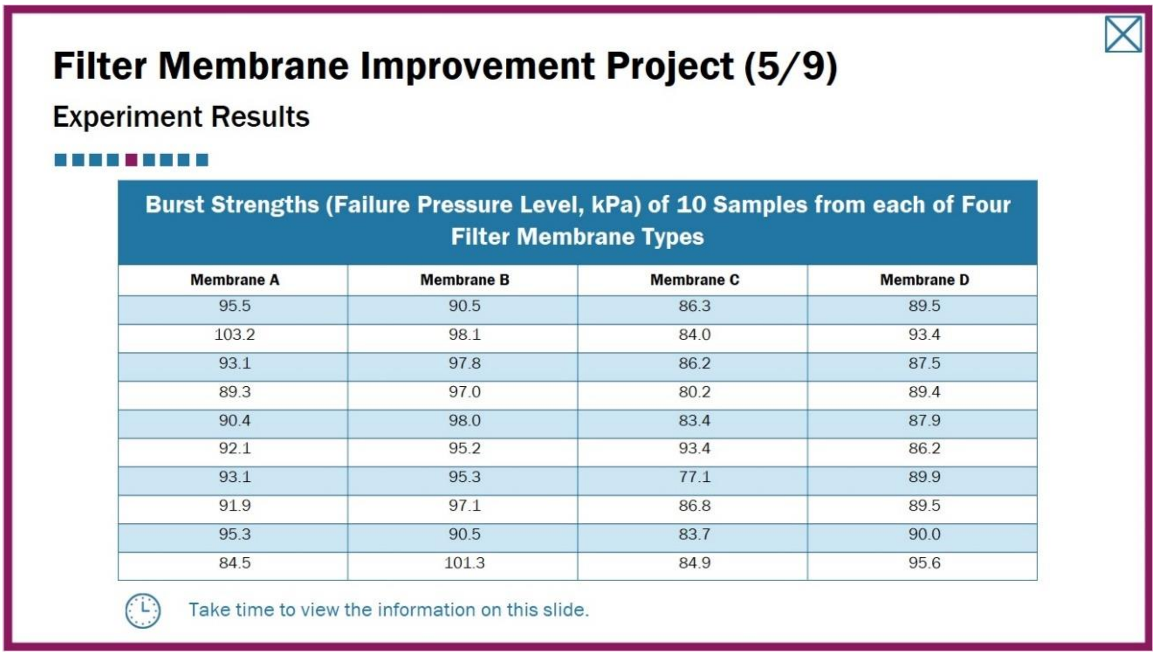
There is no information given on treatment assignment.



There are 10 replications corresponding to the 10 production batches.



Tab 1.4: **Experiment Results**



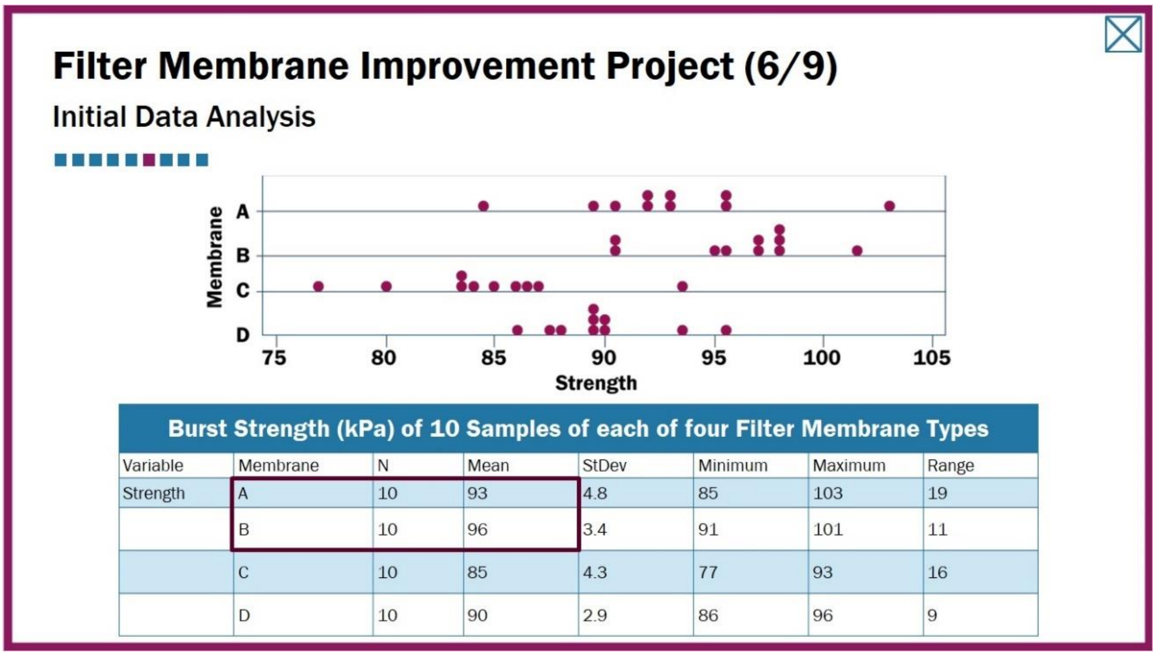
Burst Strengths (Failure Pressure Level, kPa) of 10 Samples from each of Four Filter Membrane Types

Membrane A	Membrane B	Membrane C	Membrane D
95.5	90.5	86.3	89.5
103.2	98.1	84.0	93.4
93.1	97.8	86.2	87.5
89.3	97.0	80.2	89.4
90.4	98.0	83.4	87.9
92.1	95.2	93.4	86.2
93.1	95.3	77.1	89.9
91.9	97.1	86.8	89.5
95.3	90.5	83.7	90.0
84.5	101.3	84.9	95.6

The results of this experiment are shown in the table where we have 10 measurements from each of the 4 membrane types.

Take time to view the information on this slide. Then click next to continue.

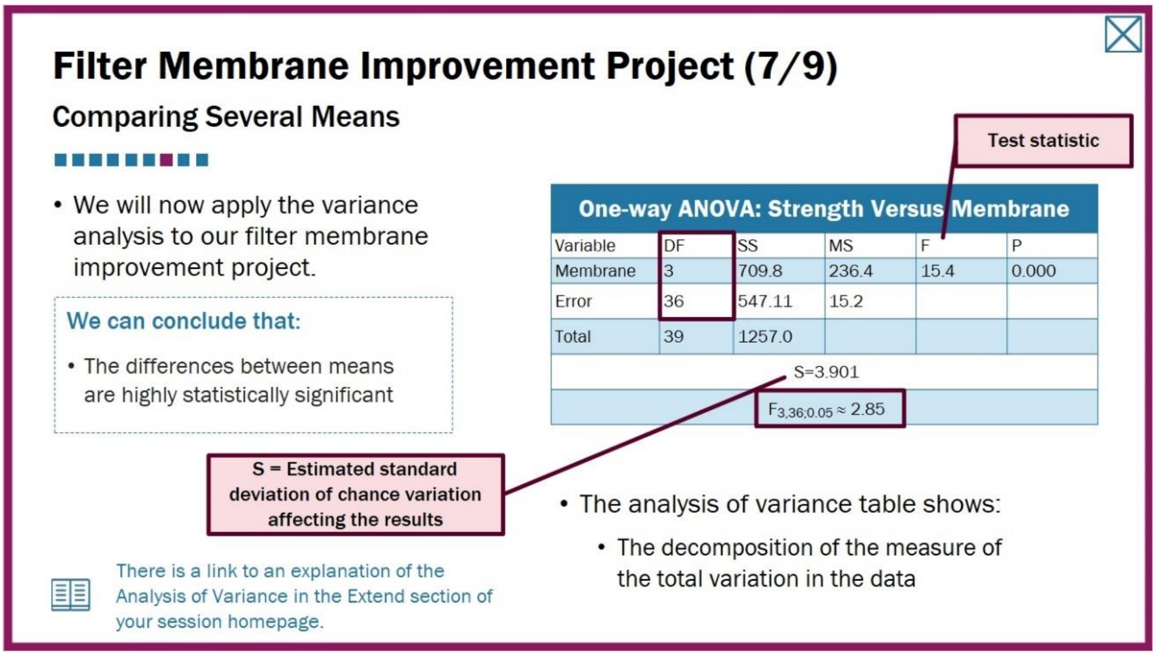
Tab 1.5: **Initial Data Analysis**



We can visualize the burst strength of the 10 samples of each of four membrane types. Summary statistics are also given in the slide.

Membrane types A and B, produced by the company, appear to be marginally better than Type D and considerably better than Type C, produced by other manufacturers. Type B, the new type produced by the company, appears marginally better than Type A, the existing type produced by the company. To assess whether these apparent effects are real or are merely artefacts produced by chance, a more formal analysis is needed.

Tab 1.6: **Comparing Several Means**



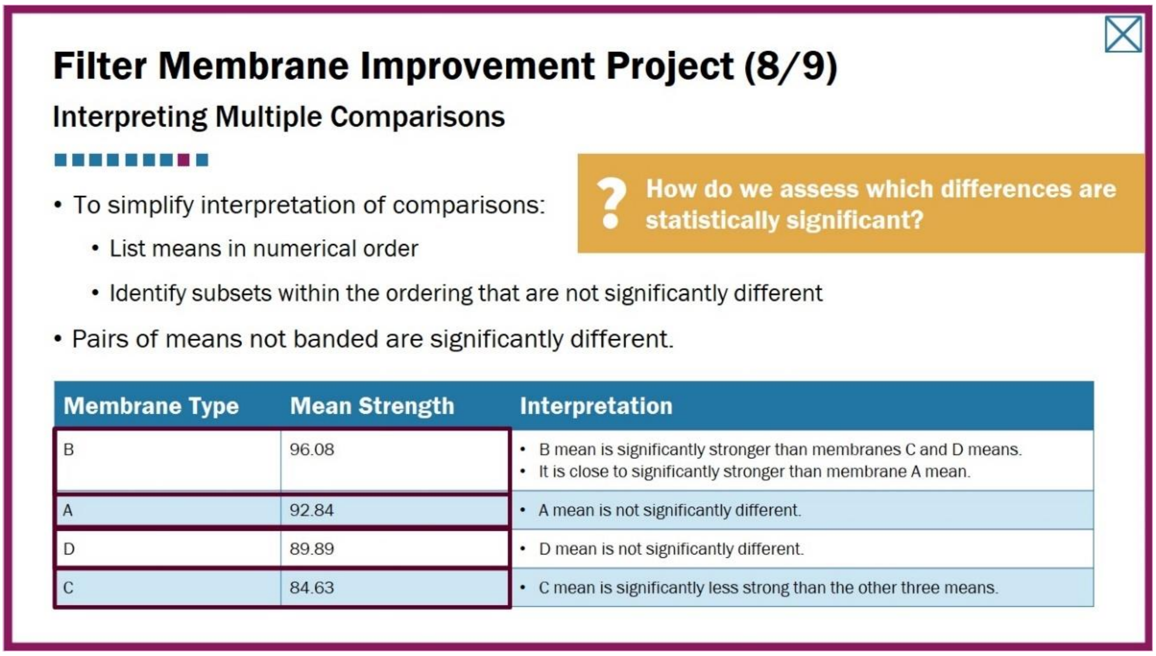
The application of the Analysis of Variance to the problem of comparing several means is discussed extensively in chapter 5 of [the Base Module: Analysis of Variance: an] introduction to one-way ANOVA. Here, it is reviewed in the context of the filter membrane improvement project example.

The One-Way Analysis of Variance procedure of Minitab produces the following (slightly edited) Analysis of Variance table. The two key numbers in this table are the value of F (equivalently, p) and the value of s. s is the estimated standard deviation of chance variation affecting the results. F is the test statistic for testing the statistical significance of differences between membrane mean burst strengths. The value of F, 15.4, should be compared to the critical value of the F distribution with 3 and 36 degrees of freedom. Reference to the relevant table of critical values indicates that, in this case, the critical value corresponding to the customary 5% statistical significance level is between 2.8 and 2.9. Clearly, the calculated value for F exceeds this and so the conclusion is that the differences between membrane mean burst strengths are statistically significant. Equivalently, this conclusion could have been drawn from the fact that the p-value, recorded in the table as 0.000, is less than 0.05.

The Analysis of Variance table shows the decomposition of a measure of the total variation in the data into two components, one representing the chance variation between results from different experimental units that would arise if the same membrane type was used throughout, and the other representing variation (in addition to chance variation) between the effects on the results of the four membrane types.

There is a link to a more detailed explanation of the Analysis of Variance in the Extend section of your session homepage.

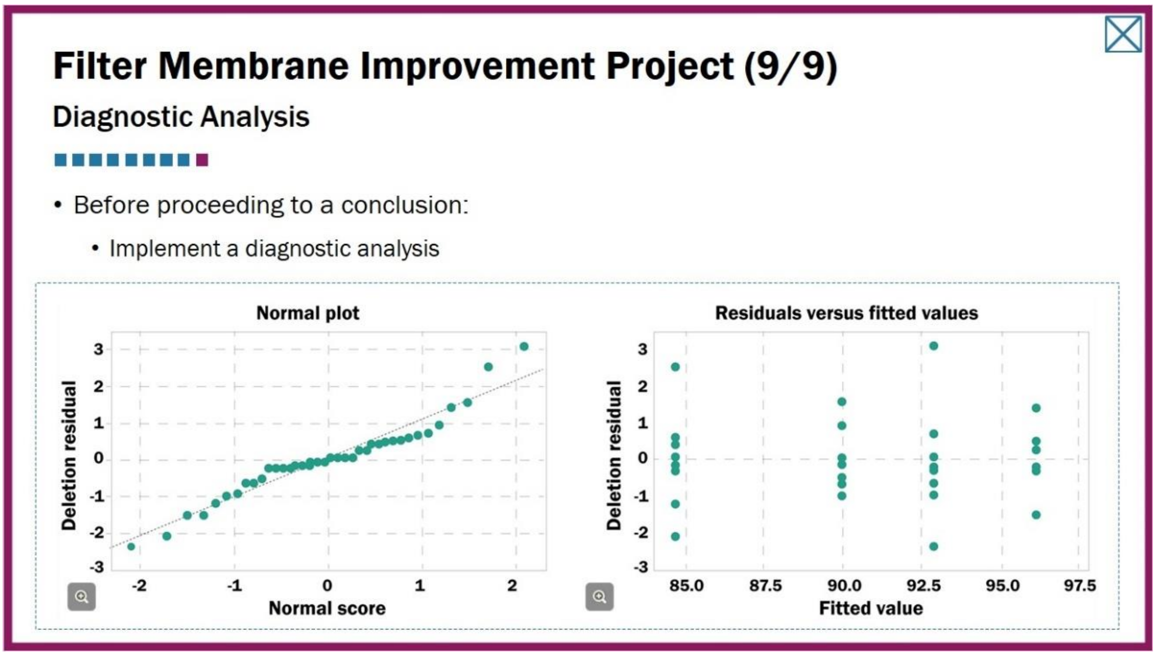
Tab 1.7: Interpreting Multiple Comparisons



Having established that differences between membrane mean burst strengths are statistically significant, the next step is to attempt to assess which of those differences are statistically significant. A simplifying first step is to list the means in numerical order. Then, from the computer output, identify subsets within the ordering that are not significantly different and band them with a line, as shown in the Table. Pairs of means that are not banded are significantly different.

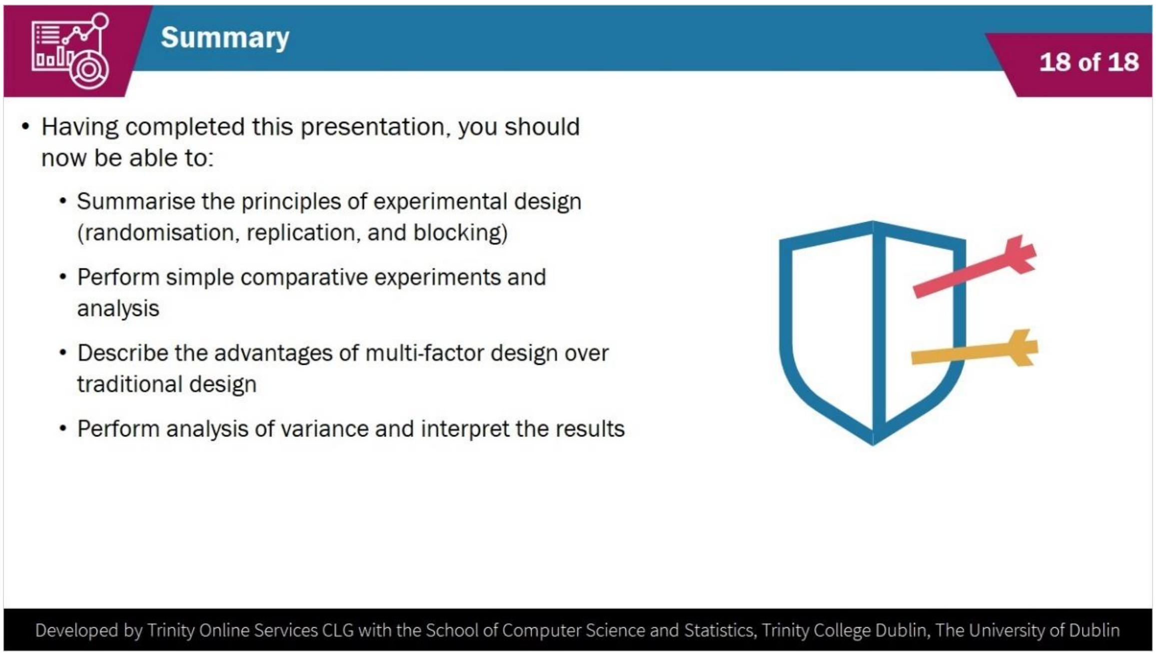
This analysis confirms our initial suggestion that the differences in mean strength between Type C and the other types are statistically significant and also indicates that Type B is statistically significantly different from Type D but does not allow us to distinguish between Types A and B or Types A and D.

Tab 1.8: Diagnostic Analysis



Before proceeding to a conclusion, a diagnostic analysis should be implemented. The standard diagnostic plots, Normal plot and Residuals versus Fitted Values plot are shown in the Figures. Neither of these suggests any significant deviations from standard assumptions.

Slide 18: ## Summary



Having completed this presentation, you should now be able to:

Summarise the principles of experimental design (randomisation, replication and blocking),

Perform simple comparative experiments and analysis,

Describe the advantages of multifactor design over traditional design, and

Perform analysis of variance and interpret the results.