



statstutor community project

encouraging academics to share statistics support resources

All stcp resources are released under a Creative Commons licence

stcp-marshallowen-1b

Mass Customisation Scenario Video Transcript

00:00 to 00:30 Title screen is displayed followed by student's email

00:30 Discussion begins:

Tutor: "If you can first of all give me an outline of what your research is about and what it is you would like some help from me today then that will be great, if you can do that first?"

Student: "My research is about Mass Customisation and Computative Design. Mass customisation is about customising a product in order to fulfil customer needs or desires."

Tutor: "Customising a product did you say?"

Student: "Yes."

Tutor: "So this is your data?" (*Laptop is shown on screen displaying data*)

Student: "Yes this is the data?"

Tutor: "Okay."

Student: "These are the mass customisation ..." (student pauses)

Tutor: "Are they the features? So the column headings you've got are the features yep." (*Excel spreadsheet showing data is displayed on screen*)



www.statstutor.ac.uk

© Ellen Marshall/Alun Owen
University of Sheffield/University of Worcester

Reviewer: Jean Russell
University of Sheffield

Student: “Yes and they are more than these they are about 16 of them.”

Tutor: “Okay.”

Student: “I just brought, for example, seven of them.”

Tutor: “So this is just a sample of what you’ve got to show me today? Yep. Okay.”

Student: (*Excel spreadsheet showing data is again displayed on screen*) “Each participant will score them based on the five scale.”

Tutor: “Yes so a five-point likert scale?”

Student: “It’s like this.” (Student shows the tutor the questions his participants answered)

Tutor: “Oh I see so you said ‘Not Very Important’ at the left and then the ‘Very Important’ to the right, but you didn’t sort of indicate what the others meant?” (*Actual questions now displayed on screen*).

Student: “Yes I didn’t.”

Tutor: “You just gave them a sort of heading for the left-most option and a heading for the right-most option, but you didn’t give any headings for the three centre options?”

Student: “No.”

Tutor: “Okay, and it was ‘Not Very Important’ and ‘Very Important’, they were the two options at the two extremes?” (Student agrees)

Tutor: “And when you scored these 1 to 5, does a 1 mean they responded was ‘Not Very Important’ and a 5 is ‘Very Important’ ?”

Student: “Yes.”

Tutor: “Okay. Okay yep if you want to carry on?”



Student: “And then based on the mean value that I calculated.”
(*Student points to the mean values displayed in his Excel spreadsheet*)

Tutor: “Okay.”

Student: “For each feature then I can sort the features based on the mean value.” (*Excel spreadsheet showing data is again displayed on screen*)

Tutor: “Okay.”

Student: “So the higher mean value the most important the feature. So here, importance rating based on the mean value, the first one for example is Guidance here.” (*Excel spreadsheet now has ‘Guidance highlighted and its mean value of 4.50 to show this is the highest rated*)

Tutor: “Is that this one that was 4.5?”

Student: “Yes 4.5”

Tutor: “Yes I see so you’ve rated, yep I see what you have done. Okay. And your supervisor was saying that they would like you to do more than just look at mean values they want you to do something else?”

Student: “Yes.”

Tutor: “Did they mention Analysis of Variance is that what you said in your email?”

Student: “Yes.”

03:20 Discussion Point is displayed

03:33 Helpful Tip is displayed

03:44 Discussion continues:

Tutor: “You could use Analysis of Variance to compare the features, but.., And there might be good reasons why you would



want to use Analysis of Variance, for example if your supervisor or your external examiner was expecting you to use Analysis of Variance, or your peer group uses Analysis of Variance in, you know sort of, published publications with this type of data”.

04:10 Helpful Tip is displayed

04:20 Discussion continues:

Tutor: “Now because you’ve got ordinal data, and it’s a rather controversial area, I think even amongst the statistics community there are some people who would view Analysis of Variance as being an acceptable way of analysing this type of data. I would take the view that personally, I would argue that it is not an appropriate way, because its ordinal data. The problem is that the Analysis of Variance assumes that the data is normally distributed. Okay, do you know what I mean by data being normally distributed? If you were to take the first feature...” (*Tutor does not allow time for the student to answer!*)

04:58 Warning X is displayed!

05:02 Helpful Tip is displayed

05:13 Question screen is displayed asking “How would you explain what is meant by data being normally distributed?”

05:17 Discussion continues:

Excel spreadsheet showing data is again displayed on screen

Tutor: “If I scroll up. This first feature ‘Visual Feedback’, if you took those values, the two 5s, the 3 and the 4 and then the 5, 2 and so on, take those ten values, if you drew a histogram of them, your histogram, I’ll draw it on a piece of paper you can take this away with you. You’d, what you’d be looking for is a histogram that looked something like that.” (*Tutor sketches a histogram on paper*).

Tutor: “What you want to see is this kind of bell-shaped curve.” (*Tutor’s diagram is shown on screen*)

Tutor: “So you’d want to see your histogram having that type of shape. Now there are two problems; One is it’s very difficult to get



that kind of normal shape, normal distribution shape with data that can only take one of five discrete values, and the second problem you've got is that you've only got 10 participants. For those reasons, I would argue that you should use an equivalent technique to Analysis of Variance that doesn't make those assumptions. Okay?"

Student: "Okay."

Tutor: "And that technique, it's what's called a non-parametric technique, and because you've got each person to rate each feature, you've got what's called repeated measures, you've got a repeated measurement from each person."

06:35 Discussion Point is displayed

06:45 Discussion continues:

Tutor: "Because you've got repeated measures you want to do a repeated measures analysis and a non-parametric technique you can use to do that is something called a Friedman test. F.R.I, I'll let you take this piece of paper away." (*Tutor writes down the name of the test*)

07:01 Discussion Point is displayed

07:13 Next screen explains that the tutor decided to show the student how to do the analysis in SPSS and asks how you could do this without doing the full analysis for them.

07:21 Helpful Tip is displayed

07:31 Practical Exercise is displayed

07:36 Discussion continues:

Tutor: "So, would like to do the menu clicks and I'll show you how to do it? It's 'Analyze' ", pause, (*Screen displays the SPSS screen the student is working on*), " 'Nonparametric tests' and if you go to the Legacy Dialogs, we want 'k-Related Samples', and it's 'K-Related Samples' because the features are all related to each other, they're not independent, you've got more than two of them that's why it's k. Okay. And can you see where it says test



variables? You need to drag each of the features across, so we've called our features Variable 1, Variable 2, Variable 3. And you, with one of them highlighted you can click on the blue button like that, that moves one of them across. Or if you want to you can drag them like you would normally drag. You've got two ways of doing it."

Student: "Okay."

Tutor: "Just write this down for you." (*Tutor writes down the menu clicks the student has just done*)

08:27 Helpful Tip is displayed

08:38 Discussion continues:

Tutor: "And can you see that it says Friedman, and that's ticked by default? So it's doing a Friedman test for you. Okay so leave that alone." (*SPSS screen is displayed with the Friedman option ticked and student agrees*)

Tutor: "And if you click OK, it will do the test for you, and it will tell you if there are any significant differences between them. What we are looking for at the moment is this value here which is called the sig value." (*Screen shows the sig value of .358 highlighted*) "Can you see that your sig value is 0.358?"

Student: "Yes."

Tutor: "Have you done any hypothesis testing before or significance testing? And have you come across things called p-values or, ever done anything like that?"

Student: "No."

Tutor: "Okay."

09:15 Discussion Point is displayed

09:28 Discussion continues:

Tutor: "Well if I explain how we use that sig value. What you're doing is you're choosing between two possible true situations, and



these are your hypotheses. So your hypothesis, if you like your research hypothesis that you are investigating, is that there is a difference in the, I'll use the word 'average', although we're not really testing the average but I'll explain in a second, the difference in the average rating for the three features. So your, sort of research hypothesis is that there is a difference in the rating, the sort of average rating if you like between the three features." (*Tutor writes this hypothesis down on paper for the student*)

Student:

Tutor: "The other situation is that there's no difference between them. We want to decide whether, what we're seeing are just random chance differences so that really there is no inherent difference out there amongst the, population of people that would use this mass customisation toolkit. Do we think that H_0 , what we call the Null Hypothesis, is that true? Or do we think that the data provides evidence to support our research hypothesis which I've labelled H_A ? I've used A because people often call it the Alternative hypothesis, but you can think about it as your kind of research hypothesis, what it is you want to find evidence to prove. And that's the way that we usually set this up, that there's a null hypothesis H_0 , which is basically that there's no relationship or no difference, or something like that, and in your case we are saying no difference. And then you've got your research hypothesis that says there is a difference. And you then say does the data provide evidence to support my research hypothesis? If the sig value, if that's less than 0.05 then that is providing evidence to support your research hypothesis. Now can you see that, H_0 the, sorry, the sig value here in your case is 0.358? So that isn't less than 0.05 is it? Does that make sense?"

Tutor: "So you don't have evidence that those three that we've looked at are significantly different. So that does not provide evidence to support your research hypothesis." (*Tutor writes this conclusion down for the student*)

Tutor: "Obviously what you need to do is to do that for larger numbers of features".

12:12 Closing remark is displayed.

