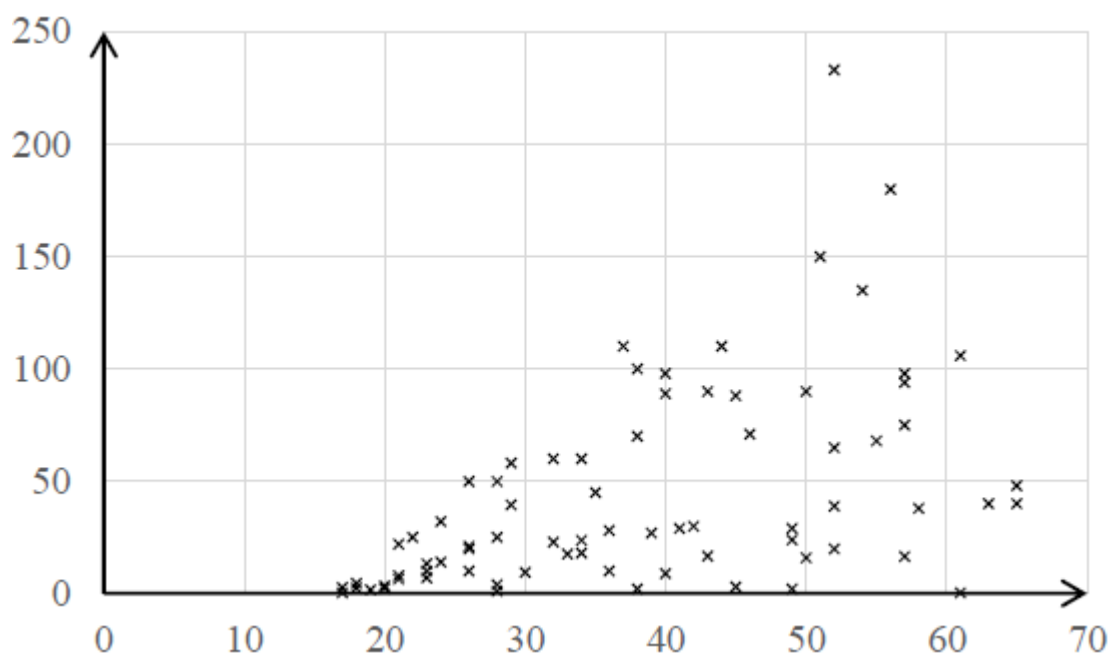
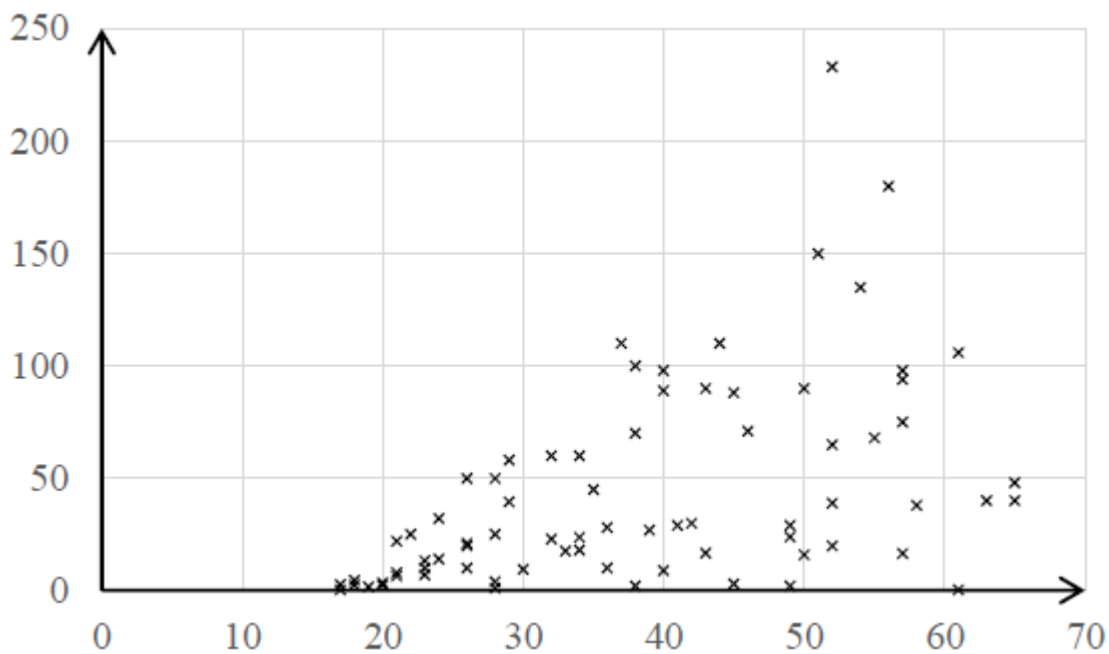


Correlation Q1 [Practice/E](8/6/21)

Data has been collected for two variables (as shown below), and possible association between the variables is to be investigated. What issues should be considered, to ensure that the test is valid?



Data has been collected for two variables (as shown below), and possible association between the variables is to be investigated. What issues should be considered, to ensure that the test is valid?



Solution

- Is this fresh data (rather than the data that prompted the investigation)?
- Are both variables random?
(Necessary for either rank correlation or PMCC).
- Should the outlier at (approx.) (52, 235) be excluded
(if either an error, or atypical)?
- The data points don't seem to form an ellipse (concentrated at its centre), and so a bivariate normal distribution is unlikely to be appropriate (so that a PMCC wouldn't be valid)
- Is there a linear association? If not, rank correlation would be appropriate, instead of PMCC.