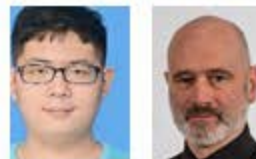


The implications of non-parametric distributions: what does it mean when your data's not normal?

Steven Ford, Gengyuan Miao, SIPBS, 161 Cathedral Street, G4 ORE



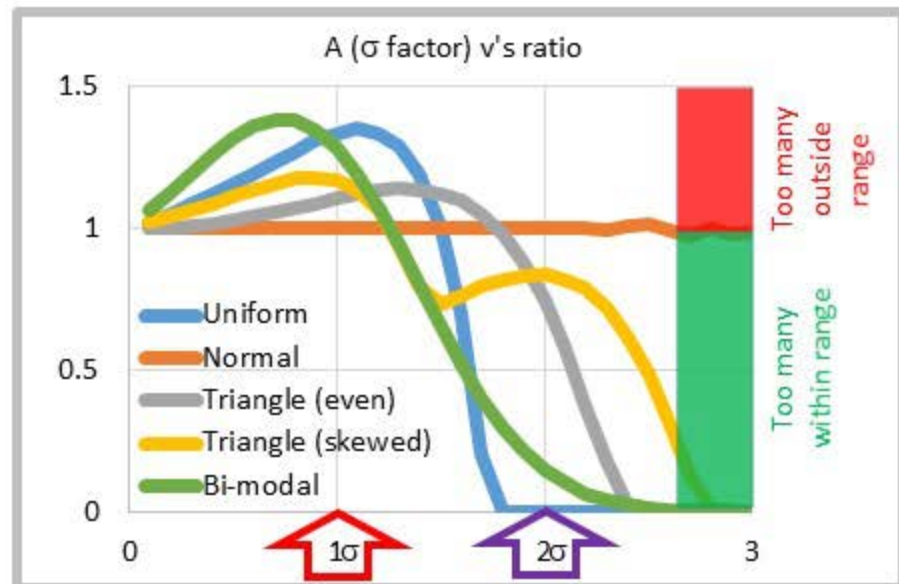
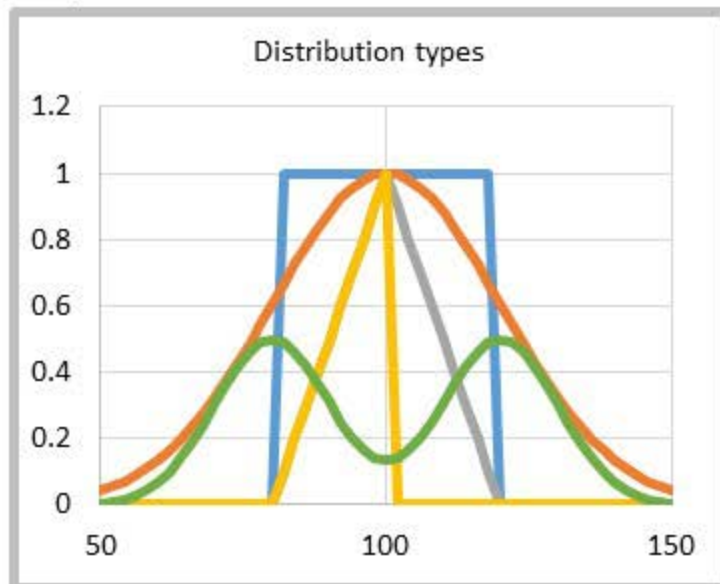
Introduction

Many scientists use mean (μ) and standard deviation (σ) to describe the centre point and 'range' of a measured set of data. However, this is only strictly true when dealing with normally distributed data. How well does μ and σ represent other (non-normal) distributions, and how many datapoints are 'outside' the range defined by μ and σ ? This is a key question for manufacturing (ie tablet production).

Methods

Several different types of distribution were modelled in Excel (see left figure). The number of datapoints outside the modelled range (the range is given by $\mu \pm A\sigma$, where $A = 0.1, 0.2, \dots, 3.0$) is compared to the expected number of datapoints outside the same range in a normal distribution, producing a ratio (right figure). 10^5 and 10^6 datapoints are required to reduce error at A values > 1.5 .

Results



These results suggest that considering $\mu \pm 2\sigma$ is a better guide to underestimate the number datapoints outside that range than $\mu \pm \sigma$.