How two random samples from a population of numbers can vary

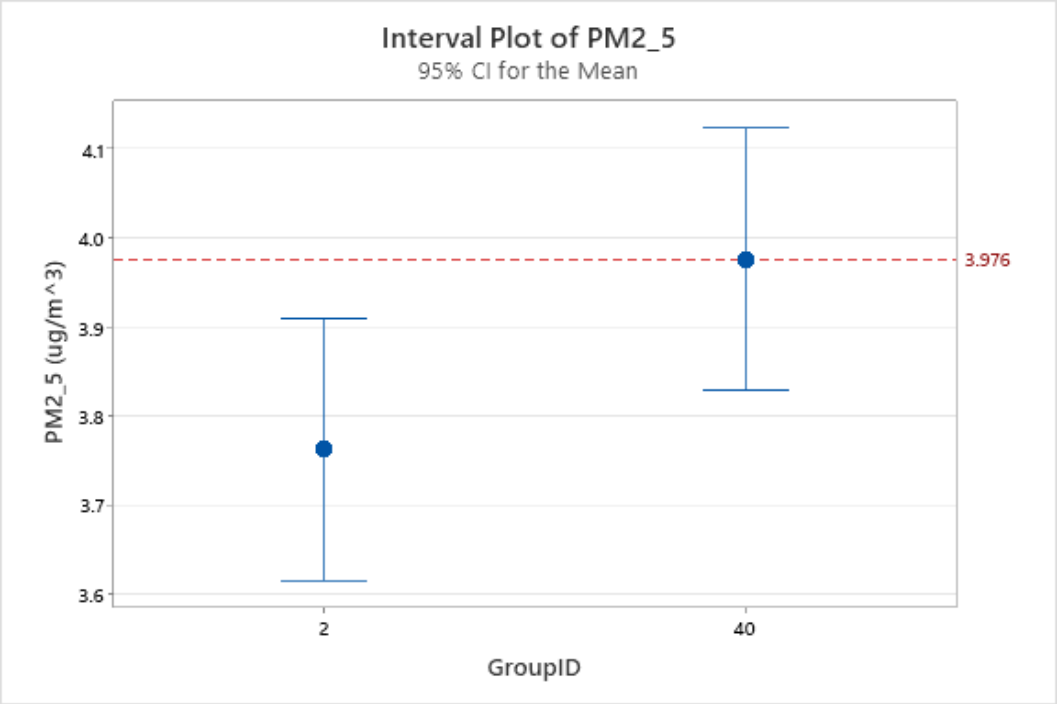

Figure: Sample means (blue dots) and 95% z-intervals of the population mean (crossbars) for each group, and the population mean (red dashed line). Confidence intervals were calculated using pooled variance.

- Open University students were recruited to collect data on atmospheric particulate matter. In total, 107 840 air quality readings were taken and randomly allocated to 40 equally-sized groups, with a GroupID from 1 to 40.
- Of these readings, the PM2_5 variable is of particular interest, being the concentration (in μgm^{-3}) of particles smaller than 2.5 micrometers in diameter in a particular sample.
- The figure above shows the sample means for GroupIDs 2 and 40. Despite being samples from the same population, GroupID=2 has a lower mean (3.762 μgm^{-3}) than GroupID=40 (3.976μgm^{-3}).
- GroupIDs 1 and 40 have similar standard deviations of 3.711 μgm^{-3} and 4.077 μgm^{-3}, respectively, and so pooled variance was used to calculate the z-intervals shown in the Figure.
- Despite being drawn from the same population, only one of the samples' confidence intervals contains the population mean of 3.976 μgm^{-3}. If this was repeated for all samples, we would expect 95% of the intervals to include the population mean.

Source: Open University M248 module website
Produced for TMA04, 15th February 2025