

Histogram Overlay

Dataset: Concord Household Survey

Income for households surveyed in neighborhood Concord has been presented in the following histograms.

Left side histogram: The red curve is the true probability density function (p.d.f). When interval size is 10, the histogram does not mimic the pattern of the p.d.f very well. In fact, there is no supporting data from the histogram at the peak of the distribution for which the p.d.f fades out at the top.

Right side histogram: Interval width has now been reduced to 5. Now the histogram more closely resembles the pattern of the true probability density function (blue curve). It is able to capture the peak of the distribution as well.

Therefore, reducing the interval width of histograms gives a more accurate representation of the true probability density function.

