

Lecture 20

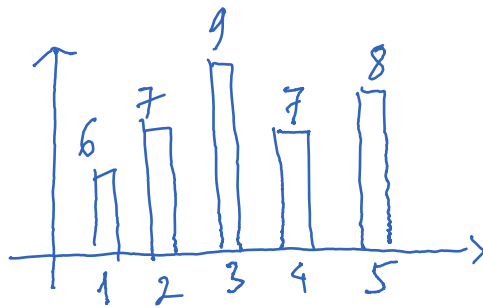
Sunday, October 27, 2024 1:51 AM

Finish the worksheet from last time: the histogram and distribution problem

Visual explanation of mean, median, and mode on the histogram:

Suppose the data set is not binned (so, every data value is counted separately from each other). Then mean is the average area of the bars. Median is the location on the horizontal line where the total area on the left is equal to the total area on the right. Mode is the location on the horizontal line where bar is tallest. This interpretation works mostly the same for binned data in which the bins are of equal widths.

The histogram encodes everything about the statistics of the data. For example:



The sorted data is 1, 1, ..., 1, 2, 2, ..., 2, 3, 3, ..., 3, 4, 4, ..., 4, 5, 5, ..., 5

$$\text{Mean} = \frac{1 \times 6 + 2 \times 7 + 3 \times 9 + 4 \times 7 + 5 \times 8}{6 + 7 + 9 + 7 + 8} = \frac{116}{37} \approx 3.14$$

Median = the 19th data value = 3

Mode = 3

Measures of variation:

Min value, max value, range, five-number summary (min value, low quartile, middle quartile, high quartile, max value), boxplot.

Example: the data set 3, 5, 2, 4, 6, 2, 13, 5, 8, 32, 6, 3