Statistics I

Zheng Wei
wei@math.umass.edu

Department of Mathematics and Statistics
University of Massachusetts Amherst
(Example-1) Count/discrete data: the daily number of computer stoppages over 30 days:

1 3 1 1 0 1 0 1 1 0 2 2 0 0 0
1 2 1 2 0 0 1 6 4 3 3 1 2 4 0

How one can get frequency \( f \), relative frequency \( r \) and its histogram for discrete data?

1 Determine the sample size, \( n \).

2 Determine the frequency \( f \) by counting the number of distinct value.

3 Calculate relative frequency by dividing the frequency \( f \) with \( n \).

4 Draw the relative frequency histogram: center a rectangle with base of length \( k \) (usually 1) at each observed integer value and make the height equal to relative frequency.
We can summarize their frequencies and relative frequencies in a frequency table:

Also we can construct a relative frequency histogram based on frequency table:
(Example-2) Continuous data (or discrete data with many possible values):

2.1, 2.4, 2.2, 2.3, 2.7, 2.5, 2.4, 2.6, 2.6, 2.9

How one can get frequency (f), relative frequency (r) and its histogram for continuous data?

1. Determine the sample size, n and find the minimum and maximum of the data, \( x_{\text{min}} \) and \( x_{\text{max}} \).

2. Group the data into k class of equal length that cover the range of data without overlapping, i.e., \([a_1, b_1), \ldots, [a_j, b_j), \ldots, [a_k, b_k)\) where \( j = 1, \ldots, k \),

- \( a_1 < x_{\text{min}} < b_1 < a_2 < \ldots < a_j < b_j < \ldots < b_{k-1} < a_k < x_{\text{max}} < b_k \)
- \([a_j, b_j)\) : class intervals
- \( a_j \) and \( b_j \) : class endpoints or class boundaries
- \( b_j - a_j \) : length of class interval

3. Determine the frequency (f) by counting the number of data in a particular class.

4. Calculate relative frequency by dividing the frequency (f) with n.

5. Draw the relative frequency histogram: draw a rectangle for each class with the class interval as the base and the height equal to relative frequency.
We can summarize their frequencies and relative frequencies in a frequency table: Here, we choose the length of class interval \( (b_j - a_j) \) as 0.2, commencing with 2.05 \( (=a_1) \) (so \( k \) is 5).

Also we can construct a relative frequency histogram based on frequency table: