Math 146 2.1 - Frequency Distributions

When working with large sets of data, it's difficult to see important characteristics of the data; what is the distribution or shape of the data, are there unusually large or unusually small values (outliers), is the data symmetric or stretched to one side (skewed), what is the range of the data, etc. Frequency distributions and histograms can help us understand the nature of the distribution of a data set.

Consider the following 50 subject's systolic blood pressure measurements in mm Hg:

100 112 134 126 114 134 118 138 114 124 138 110 116 134 100 92 124 100 106 158
132 116 122 132 138 128 134 112 122 134 106 104 158 94 144 128 120 112 100 116
162 110 122 130 130 112 154 92 108 128

A *frequency distribution* groups the data into several categories or classes, typically between 5 and 20. The frequency for a particular class is the number of data values that fall into that class.

Step 1: Determine the number of classes, and the width of each class. If n is the number of desired classes, then: class width = $\frac{(\text{max value})-(\text{min value})}{n}$

We have: max value = 162 and min value = 92. Suppose we want around 8 classes:

class width = $\frac{162-92}{8}$ = 8.75 Let's round this up to 10 for convenience.

Step 2: Find the *lower limit* of each class. We could start at 92, but it makes more sense to start at 90. The next classes lower limit is 90 + 10 = 100, the next lower limit is 110, then 120, 130, etc.

Step 3: The *upper class* limit is one less than the next classes lower limit: 99, then 109, 119, etc. This gives:

Class	Frequency
90 – 99	
100 – 109	
110 – 119	
120 – 129	
130 – 139	
140 – 149	
150 – 159	
160 – 169	

Step 4: Use tally marks (𝍸) to count the data values in each class, and then replace the tally with the numerical value:

Systolic (mmHg)	Frequency
90 – 99	3
100 – 109	8
110 – 119	12
120 – 129	10
130 – 139	12
140 – 149	1
150 – 159	3
160 – 169	1

Other numerical features of a frequency distribution are **class midpoints** and **class boundaries**.

- The *class midpoints* are the middle values between a classes upper limit and lower: $\frac{\text{lower limit} + \text{upper limit}}{2}$.
  The class midpoints for this frequency distribution are 94.5, 104.5, 114.5,… 164.5.
- The *class boundaries* are the values halfway between an upper limit and the next classes lower limit. In this case, 99.5, 109.5, 119.5. Also, the lower class boundary of the first class would be 89.5, while the upper class boundary for the last class is 169.5.

Other Frequency Distributions

A **cumulative frequency distribution** has each frequency the sum of the frequencies of that class and the proceeding classes.

A **relative frequency distribution** converts each frequence into a percentage of the total. For this frequency distribution, since there where 50 data values, divide each frequency by 50 and convert to a percentage.

Systolic (mmHg)	Cumulative Frequency
90 – 99	3
100 – 109	11
110 – 119	23
120 – 129	33
130 – 139	45
140 – 149	46
150 – 159	49
160 – 169	50

Systolic (mmHg)	Relative Frequency
90 – 99	6%
100 – 109	16%
110 – 119	24%
120 – 129	20%
130 – 139	24%
140 – 149	2%
150 – 159	6%
160 – 169	2%

Note: we say a distribution is *normally distributed* if the frequencies start small, increase to a maximum in the middle classes, and then decrease to small in the upper classes. Does the distribution of systolic measurements appear normal?

Example 1 Create a *frequency distribution* and a *relative frequency distribution* for the following Math 146 Exam 1 scores. Use a class width of 5.

42 42 38 42 46 34 40 23 41 38
39 46 50 49 45 46 36 52 34 43
44 26 41 45 37