Ungrouped Frequency Distributions

If there are only a few different values of a variable x it is useful to use frequency in order to present the data.

Example: Listed below are the number of times each customer returned a car for repairs in the first year after purchase. There is data for 50 cars.

3	1	2	4	1	2	3	2	3	5
1	2	4	3	3	2	4	1	1	3
4	3	0	2	1	1	2	4	3	4
5	1	0	1	3	0	1	2	0	2
2	0	3	1	2	3	0	3	5	1

Let's let the variable x be the number of returns. Then there are only six possible values for x.

We can present the data in an **ungrouped frequency distribution table.**

It is convenient to extend the distribution table

Computing s and s^2

In fact, it seems it would be useful to add another column.

Example: Given the following data set

1	2	4	7	3	1	3	2	1	7
2	1	1	7	4	4	3	3	7	2
7	7	7	2	1	3	4	2	2	1

Build a frequency distribution table and compute the mean and sample standard deviation.

•

Grouped Frequency Distribution Tables

Let's look at our lightbulb data again

Life of lightbulbs (hrs)

107	73	68	97	76	79	94	59	98	57
54	65	71	70	84	88	62	61	79	98
66	62	79	86	68	74	61	82	65	98

We want to put this data onto a frequency table but we will need to group this data into classes.

When dividing the data we must determine the appropriate number of classes.

To determine the number of classes for a data set of size n we will use either

$\sqrt{n}$ or $\log_2(2n)$

In our example:

Find the class width by

1. Dividing the range by the number of classes

2. Round up (even if the result is an integer)

Class Requirements

- Classes must be mutually exclusive: Each data measurement fits into only one class
- Classes must be exhaustive: Each measurement must be in a class.
- The first and last classes cannot be empty

Note that there may be more than one way to establish the class limits that satisfy the requirements.

We will use the <u>class boundaries</u>

$54 \le x < 65$

$65 \le x < 76$ Class width = 11

$76 \le x < 87$

$97 \le x < 98$

$76 \le x < 109$

And so we get

Class	Data
$54 \le x < 65$	
$65 \le x < 76$	
$76 \le x < 87$	
$97 \le x < 98$	
$76 \le x < 109$	

Now that we know how to group the data into classes we can build <u>grouped frequency tables.</u>

Class	Frequency (f)	Relative Frequency (rf)
$54 \le x < 65$		
$65 \le x < 76$		
$76 \le x < 87$		
$97 \le x < 98$		
$76 \le x < 109$		

It is often useful to have a single numerical value for each class that represents all of the data values that fall into that class. We will use the **class midpoint (or class mark)** which is the value that is exactly in the middle of each class.

Class	Class Mark
$54 \le x < 65$	
$65 \le x < 76$	
$76 \le x < 87$	
$97 \le x < 98$	
$76 \le x < 109$	

Notice that when we group data into classes we loose some information. Only when we have all the raw data do we know the exact values that were actually observed for each class.

Another way to express a frequency distribution is to use a <u>cumulative frequency distribution</u>.

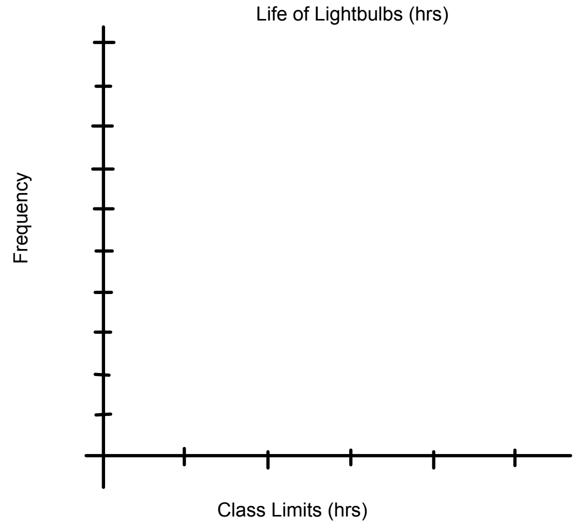
The **cumulative frequency** for any given class is the sum of the frequencies for that class and all of the classes of smaller values.

Class	Frequency (f)	Cumulative Frequency (cf)
$54 \le x < 65$		
$65 \le x < 76$		
$76 \le x < 87$		
$97 \le x < 98$		
$76 \le x < 109$		

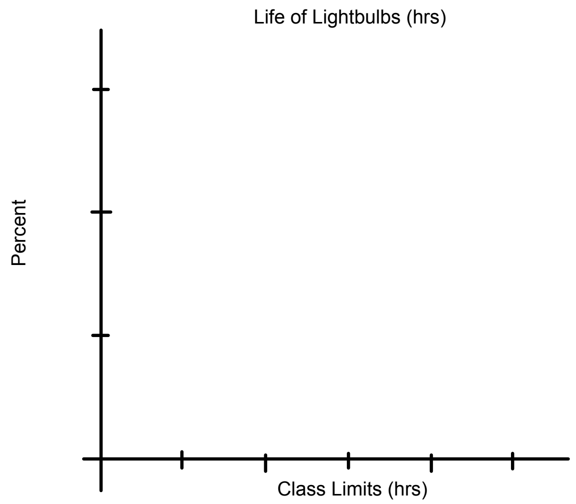
<u>Histograms</u>

- Bar chart
- Horizontal axis has classes
- Vertical axis has frequency or relative frequency

Frequency Histogram:



Relative Frequency Histogram:



Shapes of Histograms

- Symmetrical: Both sides of the distribution are identical
- Normal: A symmetrical distribution that is mounded up at the mean and becomes sparse at the extremes
- Uniform (Rectangular): Every Value appears with equal frequency.
- Skewed: One tail is stretched out longer than the other. The direction of the skewness is on the side of the longer tail.
- J-shaped: The is no tail on the side of the class with the highest frequency
- Bimodal: The two most populous classes are separated by one or more classes. (This situation often implies that two populations are being sampled).

Ogives

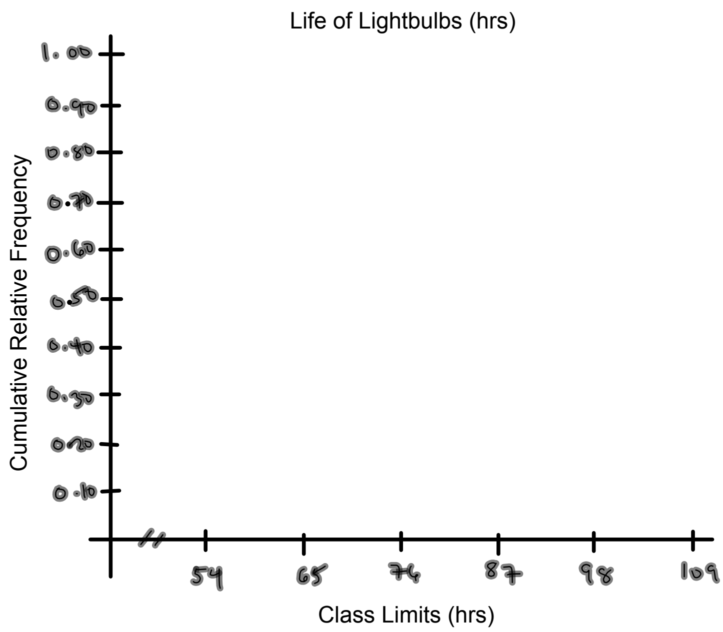
Ogives are a graphical representation of a cumulative frequency distribution.

First we calculate the **cumulative relative frequency.**

Recall:

Class	Frequency (f)	Cumulative Frequency (cf)	Cumulative Relative Frequency (crf)
$54 \le x < 65$	7	7	
$65 \le x < 76$	9	16	
$76 \le x < 87$	7	23	
$97 \le x < 98$	3	26	
$76 \le x < 109$	4	30	

We construct an Ogive in the following way



Attachments

Flash.pdf