DESCRIPTIVE

STATISTICS

POPULATION

Sample

OBSERVATIONS (DATA)

Quantitative
(Measured, Rational, some Ordinal)

Qualitative
(Descriptive, Nominal, some Ordinal)

PARAMETRIC
versus
NONPARAMETRIC METHODS

Frequency Distribution Tables

Histograms

Frequency Polygons

Ungrouped *versus* Grouped Data

- **variable, *X***

- ***n* observations,** $x_1, \cdots x_3 \cdots$

- **unique values,** $x(j)$, $j=1,k$, $(k \leq n)$

- **absolute frequency,** $f(j)$, *of* $x(j)$

- **relative frequency,** $f(j)/n$, $(0 \leq f(j) \leq 1)$

- **cumulative frequency**

- **cumulative relative frequency**

- **set of relative frequencies**
  ⇒ frequency (probability) function

- **set of cumulative relative frequencies**
  ⇒ distribution function

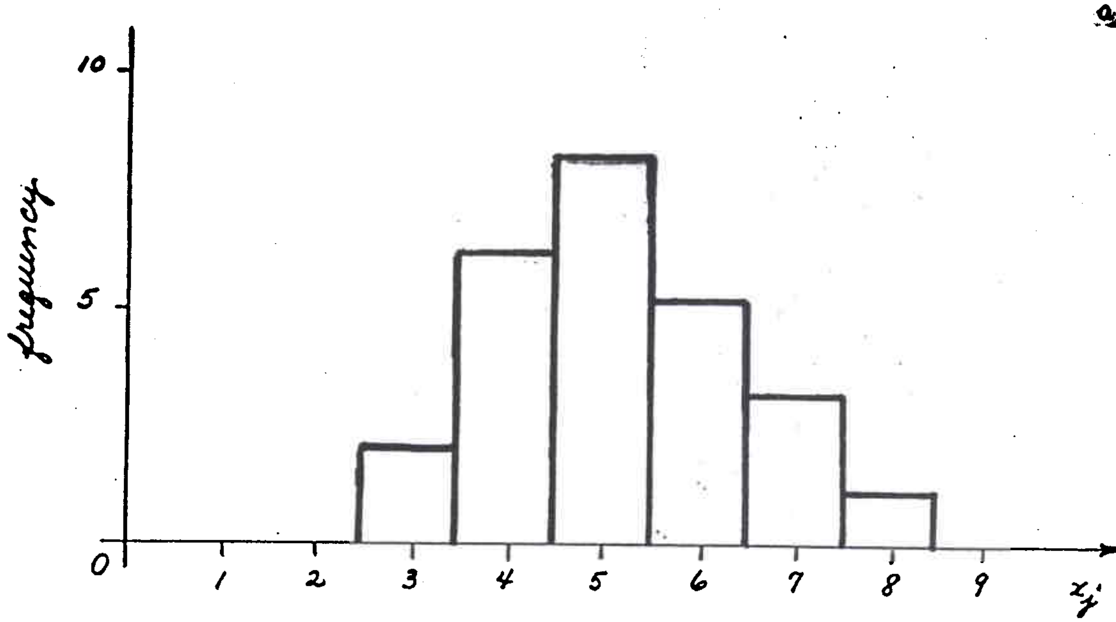
UNGROUPED DATA EXAMPLE

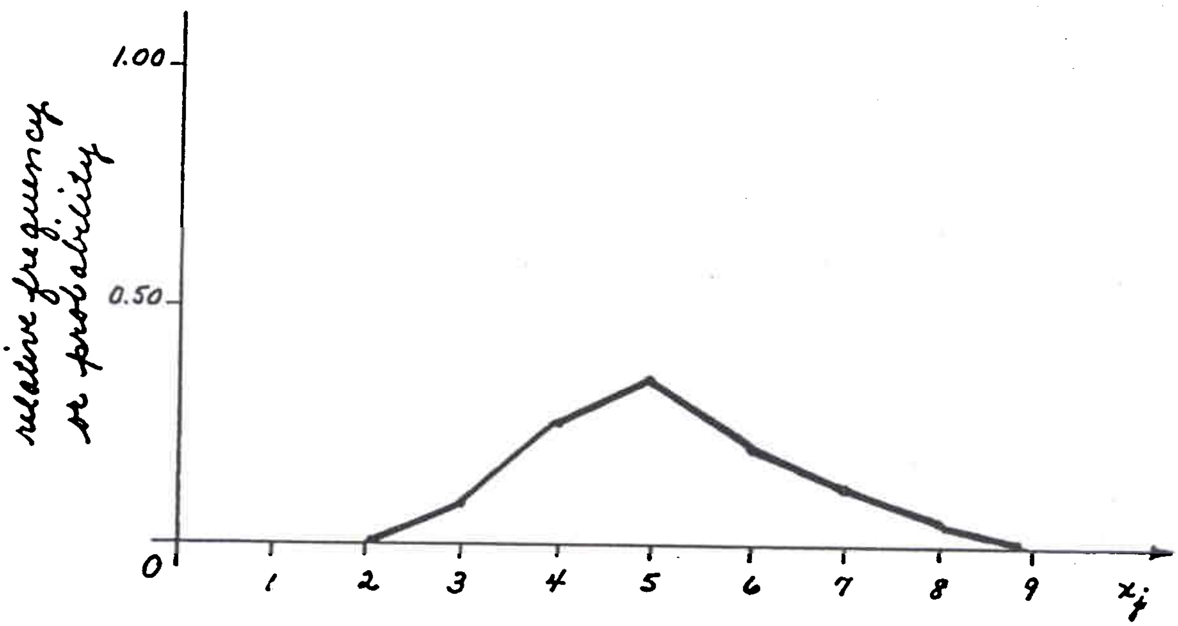
Assume a set of $n = 25$ observations x of the random variable X as follows:

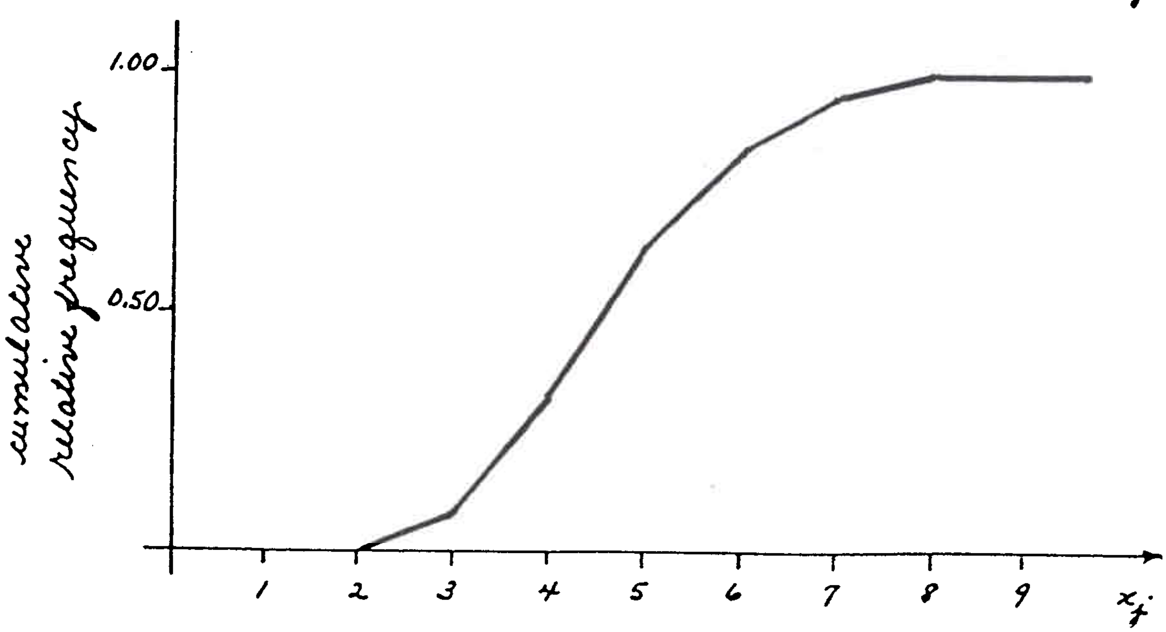
3	5	5	7	4
4	6	3	4	5
4	6	4	5	5
5	8	4	5	5
6	7	6	6	7

Frequency Table

variate $x(j)$	frequency $f(j)$	relative frequency (probability) $p(j) = f(j)/n$	cumulative relative frequency $F(j)$
3	2	0.08	0.08
4	6	0.24	0.32
5	8	0.32	0.64
6	5	0.20	0.84
7	3	0.12	0.96
8	1	0.04	1.00
		1.00	


10
5
0
frequency
1
2
3
4
5
6
7
8
9
x_j


1.00
0.50
0
relative frequency
or probability
1
2
3
4
5
6
7
8
9
x_j

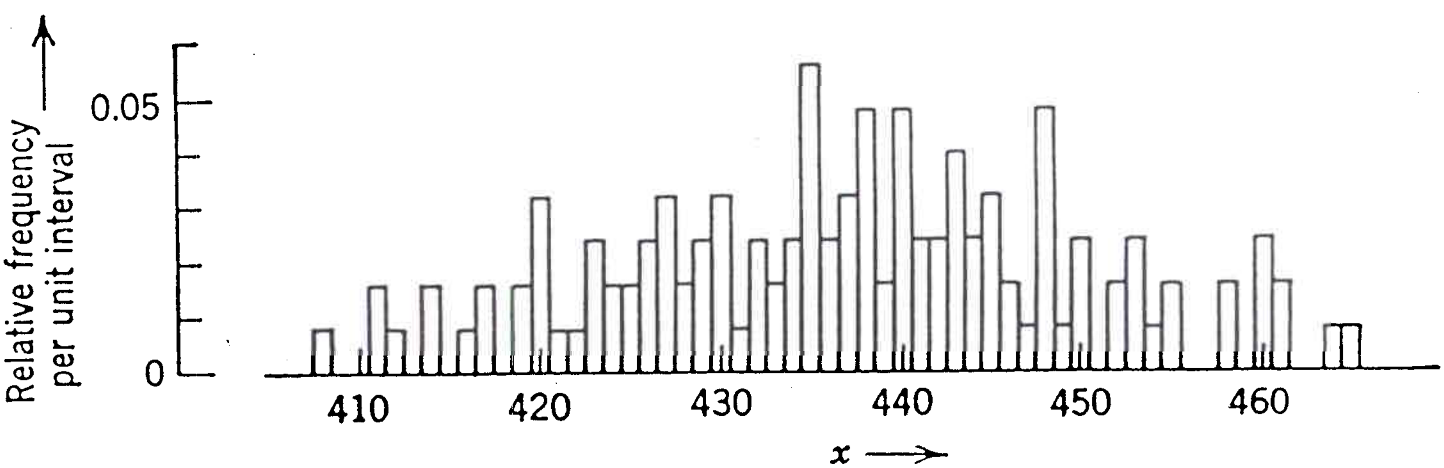

1.00
0.50
cumulative
relative frequency
1
2
3
4
5
6
7
8
9
x_j

UNGROUPED EXAMPLE
LARGE SAMPLE

Sample of 125 Values of the Splitting Tensile Strength of Concrete Cylinders

423	435	430	458	416	441	426	439	444	427	461	455	417
438	440	450	436	447	437	448	434	411	449	445	426	448
460	412	438	419	445	420	438	429	430	432	458	464	440
426	443	432	443	424	435	438	452	421	442	414	443	429
448	453	435	446	434	427	441	443	429	437	460	465	460
420	420	438	422	436	440	408	435	448	454	443	440	
452	424	444	427	450	448	419	440	453	417	423	428	
428	437	448	430	435	425	461	433	444	436	435	450	
432	442	411	440	455	425	427	442	423	446	453	445	
441	433	445	438	420	439	434	414	435	430	431	437	

RELATIVE FREQUENCY HISTOGRAM
Ungrouped Data



GROUPED DATA

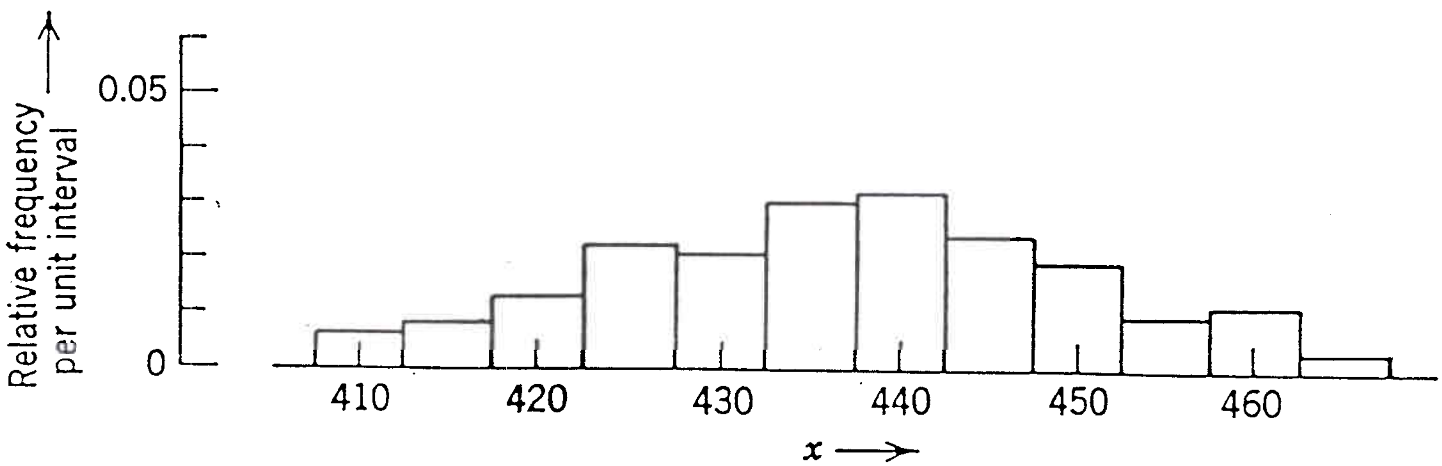
- Interval containing all data values
- Class Interval - subdivision of the data set Interval
- Bins Class Frequency - number of data values in a Class Interval
- Class Mark - midpoint of class interval

- Class Intervals should all have same length
- Class Marks should correspond to simple numbers
- Data values should not (if possible) coincide with Class Interval endpoints

Grouped Frequency Table of 125 Values of Splitting Tensile Strength of Concrete Cylinders

Class Interval	Class Mark	Class Frequency	Relative Class Frequency	Cumulative Relative Class Frequency
407.5 - 412.5	410	4	0.032	0.032
412.5 - 417.5	415	5	0.040	0.072
417.5 - 422.5	420	8	0.064	0.136
422.5 - 427.5	425	14	0.112	0.248
427.5 - 432.5	430	13	0.104	0.352
432.5 - 437.5	435	19	0.152	0.504
437.5 - 442.5	440	20	0.160	0.664
442.5 - 447.5	445	15	0.120	0.784
447.5 - 452.5	450	12	0.096	0.880
452.5 - 457.5	455	6	0.048	0.928
457.5 - 462.5	460	7	0.056	0.984
462.5 - 467.5	465	2	0.016	1.000

Grouped Frequency Histogram



DESCRIPTIVE MEASURES

Measures of Central Tendency

Measures of Variation or Dispersion

Measures of Central Tendency

Mode - that value which occurs most frequently

Median - that value "in the middle" (i.e., that value of X for which the cumulative relative frequency is 0.50)

Mean - the familiar arithmetic average, designated by

μ - for a population

$\bar{x}$ - for a sample

The expected value of the random variable, X

Computing the Sample Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

for a sample of $x_1, x_2, \cdots, x_n$

If the sample data are grouped,

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{K} x_j f_j$$

where K = the number of class intervals
x_j = the j^{th} class mark
and f_j = the j^{th} class frequency

The population mean, μ, is similarly determined - note that n may $\rightarrow \infty$, in which case integrals are required

Measures of Variation or Dispersion

Range - difference between largest and smallest values in the sample

Variance - the mean of the sum of squared deviations from the sample mean, denoted by s^2

- the mean of the sum of squared deviations from the population mean, denoted by σ^2

In terms of Expectation;

$$\mathcal{E}\{[x-\mu]^2\}$$

Population Variance
Computation; defined n

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

The square root of σ^2, σ,
is referred to as the
standard deviation

Computing the Sample Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

If the sample data are grouped,

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{K}(x_j - \bar{x})^2 f_j$$

$$= \frac{1}{n-1}\left[\sum_{j=1}^{K} x_j^2 f_j - \frac{1}{n}\left(\sum_{j=1}^{K} x_j f_j\right)^2\right]$$

where K = the number of Class Intervals

x_j = the j^{th} Class Mark

and f_j = the j^{th} Class Frequency

s = sample standard deviation