WEEK 1

STATISTICS

1. Statistics and Describing Data, Measure of Tentencies

The goal of statistics (or any other science) is to understand the universe). Because of the time and financial limitations we do not have a chance to reach all the individuals of the universe. Therefore, we take a sample from the population (which we are interested in) and based on the sample information we try to make some inferences about the whole population. Understanding the population means that we try to get some information about the unknowns (we will call parameter or parameters). These unknowns (parameters) characterize the population. The parameters (we are going to study) will be mean, variance, percentiles, etc. We try to make inferences about the parameters by using statistical techniques.

In this class, we are going to study the basic idea of DATA and some measure of tendencies (mean mode, median variance, standard deviations etc.).

Definition 1 Statistics is a science of collecting, organizing and interpreting the numerical facts, which we call ***DATA***

That is statistics helps us to solve 3 problems in the nature

- collection of DATA
- organization (or analysis) of DATA
- interpretation of analysis

Figure 1.1

There are two different types of DATA

- Quantitative DATA (observations that are measured on a numerical scale)
- Qualitative DATA (if each measurements in a data set falls into one and only one of the set of categories, the data set is called qualitative or categorical data)

In order to collect data, first we need to perform an experiment. For example, tossing a coin twice, rolling a die, etc. However, when we perform an experiment, we may not observe numerical values.

Definition 2 All possible outcomes of an experiment is called sample space and denoted by either Ω or S.

Consider an experiment of tossing a coin twice. At each trial we observe either a head or tail. That is, a set of all possible outcomes is the sample space ($\Omega = \{HH, HT, TH, TT\}$). If we roll a die, the sample space is the set of die having number of dots appears. That is, $\Omega = \{w_1, w_2, w_3, w_4, w_5, w_6\}$ where w_i denotes a die i dots. As it is clear, the observations cannot be measured in numerical scales. As you can see, you cannot do any mathematical operations (or calculations) with these observations.

Figure 1.2

Consider an experiment of tossing a coin. In this experiment we can observe either a head or a tail. But we cannot do any mathematical operations (calculations or analysis) with heads and tails. However, if we transfer these events to a world that we know (world of mathematics) we get numerical observations. For example, if we observe a head, a function will match to zero and if we observe a tail a function will match one, we get numerical values like 0 and 1. Now, we can do analysis with these numerical values.

As we mentioned before, we have two types of data.

<u>a) Qualitative (or categorical) DATA</u>: measurements in the data set fall into one and only one of a set of categories. For example consider the test scores for BAS152 with categories are given below:

Scores	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Category	**FF**	**FD**	**DD**	**DC**	**CC**	**CB**	**BB**	**BA**	**AA**

If student get an average, say **75**, then he/she will get **BB** from the course. That is, his/her category is **BB**.

<u>b) Quantitative DATA:</u> Observations that are measured on a numerical scale. Monthly inflation rates, price of an item, test scores for an examination, weights of a person, etc.

Test scores: 75, 72, 68, 85, etc.

Weather temperatures: 10 ^{0}C, 15 ^{0}C, 25 ^{0}C, 22 ^{0}C, etc.

<u>Frequency:</u> the frequency for a category is the total number of measurements that fall in the category, the frequency for a particular category, say i, will be denoted by f_i .

<u>Relative Frequency:</u> f_i / n, where n is the total number of measurements in the sample and f_i is the number of measurements in the category i.

When you have a set of measurements $\{x_1, x_2, ..., x_n\}$ you can calculate some measure of tendencies (mean, median, mode, variance, standard deviation etc.) to get some information about the population.

Population	sample
mean $\rightarrow \mu$	sample mean $\rightarrow \bar{x}_n$
variance $\rightarrow \sigma^2$	sample variance $\rightarrow s_n^2$
median $\rightarrow M$	sample median $\rightarrow m$
etc.	etc.

<u>Sample mean:</u>

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + ... + x_n}{n}$$

<u>Sample variance:</u>

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,\bar{x}_n^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(x_1 + x_2 + ... + x_n)^2}{n}\right].$$

<u>The sample standard deviation</u> :

The positive squared root of the sample varaince, $s_n = +\sqrt{s_n^2}$.

<u>Median:</u> The sample median is a number m that 50% of all measurements fall below m and 50% of all measurements fall above m.

Consider a set of measurements $\{x_1, x_2, ..., x_n\}$. In order to calculate the median, we need to order the measurements from smallest to largest $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ where

$x_{(1)} = \min\{x_1, x_2, ..., x_n\}$, $x_{(2)} =$ second smallest measurements of $\{x_1, x_2, ..., x_n\}$, ...

$x_{(n)} = \max\{x_1, x_2, ..., x_n\}$.

Then the median

$$m = \begin{cases} x_{((n+1)/2)} & , \quad n \text{ is odd} \\ 0.5[x_{(n/2)} + x_{((n/2)+1)}] & , \quad n \text{ is even.} \end{cases}$$

Using these ordered measurements, we can calculate the sample percentiles. For example, if 90% of all measurements are less than or equal to a number (say x_{90}) then the number x_{90} is called 90th percentile of the sample. Again, if 95% of all measurements are less than or equal to a number (say x_{95}) then the number x_{95} is called 95th percentile of the sample. In a similar way, we can calculate 99th, 1st, 5th, 10th percentiles of the sample. The 25th percentile of the sample is called the first quartile and denoted by Q_L (or Q_1) and 75th percentile is called the third quartile of the sample and denoted by Q_U (or Q_3). The range of the sample is the difference from the largest and the smallest measurement ($R = x_{(n)} - x_{(1)}$). The interquartile range is the difference between Q_3 and Q_1 and denoted by IQR ($IQR = Q_3 - Q_1$).

Mode: The most repeated measurements.

Example 1.1 Suppose we have $n = 10$ measurements in hand. The measurements and their ordered values are given below. The ordered values (from smallest to the largest) are indicated by brackets.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
3	2	4	6	1	6	7	4	5	2
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
1	2	2	3	4	4	5	6	6	7

Note that the sample mean

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{40}{10} = 4.$$

The sample variance

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2\right] = \frac{1}{9}\left[196 - 10(4)^2\right] = \frac{1}{9}[196 - 160] = \frac{36}{9} = 4$$

and thus, the standard deviation is $s_n = +\sqrt{s_n^2} = +\sqrt{4} = 2$. In order to calculate the median, we need to use ordered values. We have $n = 10$ measurements which is an even number. So the sample median is,

$$m = \frac{1}{2}[x_{(n/2)} + x_{((n/2)+1)}] = \frac{1}{2}\left[x_{(5)} + x_{(6)}\right] = \frac{1}{2}[4+4] = 4.$$

The sample mode is the most repeated measurements. In our sample the measurements 2,3 and 6 have been observed three times. Therefore, any one of these measurements can be considered as the sample mode.

Note that, the sample mean, variance and the median turned out to be the same. That is, $\bar{x}_n = s_n^2 = m = 4$. Does this say anything to us?

Example 1.2 Suppose we have $n = 30$ observations given below:

1.0	1.2	1.4	1.2	1.3	1.6	2.1	1.7	1.5	1.3
2.0	2.1	1.6	1.5	1.2	2.1	2.4	2.3	1.7	1.5
1.5	0.9	1.6	1.2	1.8	1.3	2.2	2.5	1.4	1.2

Note that, $x_{(n)} = 2.5$ and $x_{(1)} = 0.9$ and the range is $R = x_{(n)} - x_{(1)} = 2.5 - 0.9 = 1.6$. Moreover,

$$\sum_{i=1}^{n} x_i = 48.3 \text{ and } \sum_{i=1}^{n} x_i^2 = 83.07.$$

Therefore,

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{48.3}{30} = 16.1$$

and

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,\bar{x}_n^2\right] = \frac{1}{9}\left[83.07 - 30(1.61)^2\right] = \frac{1}{30}[83.07 - 77.763] = 0.183$$

and the sample standard deviation $s_n = +\sqrt{s_n^2} = +\sqrt{0.183} \cong 0.428$.

The ordered values are given below:

0.9	1.0	1.2	1.2	1.2	1.2	1.2	1.3	1.3	1.3
1.4	1.4	1.4	1.5	1.5	1.5	1.6	1.6	1.6	1.7
1.7	1.8	2.0	2.1	2.1	2.1	2.2	2.3	2.4	2.5

We have $n = 30$ measurements which is an even number. So the sample median is,

$$m = \frac{1}{2}[x_{(n/2)} + x_{((n/2)+1)}] = \frac{1}{2}\left[x_{(15)} + x_{(16)}\right] = \frac{1}{2}[1.5+1.5] = 1.5.$$

That is, $\bar{x}_n = 1.61$ and $m = 1.5$ so that $m = 1.5 < 1.61 = \bar{x}_n$. That is, the sample median is smaller than the sample mean. We can write the followings:

a) If $m < \bar{x}_n$ then the data is skewed to the right,

b) If $m = \bar{x}_n$, then the data is symmetric,

c) If $m > \bar{x}_n$ then the data is skewed to the left.

$m < \bar{x}_n$	$m = \bar{x}_n$	$m > \bar{x}_n$

Figure 1.3.

What about the standard deviations? Let us calculate the following intervals:

$$I_1 = (\bar{x}_n - s_n, \bar{x}_n + s_n) \ , \ I_2 = (\bar{x}_n - 2s_n, \bar{x}_n + 2s_n) \text{ and } I_3 = (\bar{x}_n - 3s_n, \bar{x}_n + 3s_n)$$

Here, $\bar{x}_n$ is the sample mean and s_n is the standard deviation.

$$I_1 = \bar{x}_n \mp s_n = (\bar{x}_n - s_n, \bar{x}_n + s_n) = (1.61 - 0.428, 1.61 + 0.428) = (1.182, 2.038)$$

$$I_2 = \bar{x}_n \mp 2s_n = (\bar{x}_n - 2s_n, \bar{x}_n + 2s_n) = (1.61 - 2(0.428), 1.61 + 2(0.428)) = (0.754, 2.466)$$

$$I_3 = \bar{x}_n \mp 3s_n = (\bar{x}_n - 3s_n, \bar{x}_n + 3s_n) = (1.61 - 3(0.428), 1.61 + 3(0.428)) = (0.326, 2.894) .$$

Notice that there are 21 observations fall within the first interval (I_1), 29 observations fall within the second interval (I_2) and all observations (30) fall within the third interval. That is 70% of all observations fall within I_1, 95% of all observations fall within I_2 and 100% of all observations fall within the last interval.

	Interval	# of observations	Percentage
I_1	$(1.182, 2.038)$	21	%70
I_2	$(0.754, 2.466)$	29	96%
I_3	$(0.326, 2.894)$	30	100%

Note (Chebyshev's Theorem): If we have a symmetric data (approximately), then

a) Approximately 68% of all observations fall within I_1

b) Approximately 95% of all observations fall within I_2

c) Almost all (100%) of all observations fall within I_3

In our example, we have $\bar{x}_n = 1.61$ and $m = 1.5$. That is, $\bar{x}_n = 1.61 \cong 1.5 = m$. This means that the data is *nearly* symmetric.

<u>Graphical representation</u>:

	Interval	# of observations f_i	Relative Frequencies	Cumulative Relative Frequencies
I_1	0.85-1.05	2	2/30	2/30
I_2	1.05-1.25	5	5/30	7/30
I_3	1.25-1.45	6	6/30	13/30
I_4	1.45-1.65	6	6/30	19/30
I_5	1.65-1.85	3	3/30	22/30
I_6	1.85-2.05	1	1/30	23/30
I_7	2.05-2.25	4	4/30	27/30
I_8	2.25-2.45	2	2/30	29/30
I_9	2.45-2.65	1	1/30	30/30
		30	1.00	

Using these values, we can consruct a histogram in order to get some sensetive information about the shape of distrtibution.

Figure 1.4 Histogram

Now, consider 2 samples having the same number of measurements. For example, sample 1 contains 5 measurements ($x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5$) and the other sample contains the same number of measurements (say, $y_1 = 2, y_2 = 3, y_3 = 3, y_4 = 3, y_5 = 4$). Notice that $\bar{x}_n = \bar{y}_n = 3$. That is, both samples have the same mean. However, $s_{n,x}^2 = 2.5$ and $s_{n,y}^2 = 0.5$. That is, $s_{n,y}^2 < s_{n,x}^2$. If the variance smaller, then the data is more concentrated around the mean.

Sample 1	Sample 2
$\bar{x}_n = 3$, $s^2_{n,x} = 2.5$	$\bar{y}_n = 3$, $s^2_{n,y} = 0.5$
Figure 1.5	

z-scores:

As we have mentioned before, we use sample values (observed by experiments) to get some information about the population.

	Population: mean $\rightarrow \mu$ and variance $\rightarrow \sigma^2$
	Sample : mean $\rightarrow \bar{x}_n$ and variance $\rightarrow s_n^2$
Figure 1.6	

The sample z-score for a measurement x : $z = (x - \bar{x}_n) / s_n$

and

population z-score for a measurement x is : $z = (x - \mu) / \sigma$.

A set of data given above: $y_1 = 2, y_2 = 3, y_3 = 3, y_4 = 3, y_5 = 4$ and note that $\bar{y}_n = 3$, $s^2_{n,y} = 0.5$.

The sample z-scores for these observations are given below:

$$z_1 = \frac{(2-3)}{\sqrt{0.5}} \cong -1.414 \ , \ z_2 = \frac{(3-3)}{\sqrt{0.5}} = 0 \ , \ z_3 = \frac{(3-3)}{\sqrt{0.5}} = 0 \ ,$$

$$z_4 = \frac{(3-3)}{\sqrt{0.5}} = 0 \ , \ z_5 = \frac{(4-3)}{\sqrt{0.5}} \cong 1.414$$

Interpretation of z-scores:

a) Approximaley, 68% of all measurements will have z-scores between -1 and $+1$

a) Approximaley, 95% of all measurements will have z-scores between -2 and $+2$

c) All (or almost all) measurements will have z-scores between -3 and $+3$

Percentiles: Let $x_1, x_2, ..., x_n$ be a set of measurements arranged in increasing (or decreasing) order. The pth percentile is a number x that p% of all observations fall below the number x and (100-p)% of all observations fall above x. For example, the 90th percentile is a number x that when you order the data from smallest to the largest, 90% of all measurements are less than or equal to this number x. Note that, the percentile of a sample does not have to be in the set of measurements. Usually, we use, 1%, 5%, 10%, 90%, 95% and 99% values.

	90% of all measurements are less than or equal to the number x
Figure 1.7	

The lower quartile (Q_L) is the 25th percentile and the upper quartile (Q_U) is the 75th percentile. The interquartile range is the difference between Q_U and Q_L, $IQR = Q_U - Q_L$. And the middle quartile is the median.

Figure 1.8

Example (revisited) : Consider the previous example given above. Some of the percentiles and quartiles are calculated as follows:

Quartiles Percentiles

100% Max		100%	99%	2.5
2.5			95%	2.4
75%	Q_U	2.0	90%	2.25
50%	M	1.5	10%	1.2
25%	Q_L	1.3	5%	1.0
0% Min	0%	0.9	1%	0.9

The range : $R = x_{(n)} - x_{(1)} = 2.5 - 0.9 = 1.6$

The interquartile range : $IQR = Q_U - Q_L = 2.0 - 1.3 = 0.7$

The Mode: most repeated observation : 1.2

Moments

N	**30**	Sum Weights	30
Mean	1.61	Sum Observations	48.3
Std Deviation	0.42778499	Variance	0.183
Skewness	0.52374231	Kurtosis	-0.6414316
Uncorrected SS	83.07	Corrected SS	5.307
Coeff Variation	26.5704964	Std Error Mean	0.0781025

****

Basic Statistical Measures

Location		Variability	
Mean	**1.610000**	Std Deviation	**0.42778**
Median	**1.500000**	Variance	**0.18300**
Mode	**1.200000**	Range	**1.60000**
		Interquartile Range	**0.70000**

****

Quantile		Estimate
100%	**Max**	**2.50**
99%		2.50
95%		2.40
90%		2.25
75%	**Q3**	**2.00**
50%	**Median**	**1.50**
25%	**Q1**	**1.30**
10%		1.20
5%		1.00
1%		0.90
0%	**Min**	**0.90**

In the above example, we calculated some percentiles (99^{th}, 95^{th},90^{th}, 10^{th},5^{th} and 1^{st}). We can also calculate any percentile values and produce an histogram to get a tentative distributional property. In the following example we are going to calculate some other percentiles values and the histogram.

Example: A producer of an electronic divese want to put a warranty on on. In order to put a reasonable warranty period the producer conducted an experiment that he/she randomly selects 100 device and measure their life time (in months). The life-time of randomly selected device are given in the following table.

55.2	59.5	52.9	57.2	56.0	53.1	63.9	64.2	57.2	64.1	62.2	55.9	63.2	59.2	66.9
65.8	65.9	62.7	59.5	61.6	51.9	57.7	58.5	52.2	55.5	63.0	59.1	63.5	60.1	60.1
54.7	58.7	49.2	65.0	63.5	60.3	66.3	56.4	62.2	62.8	63.6	63.2	67.4	64.0	55.3
59.6	60.8	62.2	67.6	52.7	56.5	58.3	67.3	58.8	71.6	67.9	58.5	62.0	51.5	59.3
51.1	63.1	62.1	57.9	57.8	60.5	60.1	67.6	57.5	62.5	50.0	55.7	57.4	61.5	59.7
59.5	64.2	58.7	58.4	55.2	64.0	65.2	66.5	57.6	67.4	56.3	70.4	67.9	58.3	59.4
53.6	57.5	57.4	57.5	54.4	60.4	62.9	62.9	55.9	59.0					

In order to calculate these percentile values we need to order the data from smallest to the largest. The ordered values are in the following table.

49.2	50.0	51.1	51.5	51.9	52.2	52.7	52.9	53.1	53.6	54.4	54.7	55.2	55.2	55.3
55.5	55.7	55.9	55.9	56.0	56.3	56.4	56.5	57.2	57.2	57.4	57.4	57.5	57.5	57.5
57.6	57.7	57.8	57.9	58.3	58.3	58.4	58.5	58.5	58.7	58.7	58.8	59.0	59.1	59.2
59.3	59.4	59.5	59.5	59.5	59.6	59.7	60.1	60.1	60.1	60.3	60.4	60.5	60.8	61.5
61.6	62.0	62.1	62.2	62.2	62.2	62.5	62.7	62.8	62.9	62.9	63.0	63.1	63.2	63.2
63.5	63.5	63.6	63.9	64.0	64.0	64.1	64.2	64.2	65.0	65.2	65.8	65.9	66.3	66.5
66.9	67.3	67.4	67.4	67.6	67.6	67.9	67.9	70.4	71.6					

There are 100 observations in the sample. Therefore 1% of all observations are less than or equal to 49.2 and thus the 1st percentile is 49.2. Similarly, 2% of all observations are less than or equal to 50.55 (average of second and third observations) and thus the 2nd percentile is 50.55. If we want to calculate 15th percentile, we want to get a number (say a_{15}) such that 15% of all of all observations will be less than or equal to a_{15}. This number can be found as an average of 15th and 16th observations ($a_{15} = (x_{(15)} + x_{(16)})/2 = 55.4$). This means that 15% of all observations are less than or equal to 55.4. Similarly, some other percentiles are calculated below.

$a_5 = (x_{(5)} + x_{(6)})/2 = (51.9 + 52.2)/2 = 52.05$, $a_{10} = (x_{(10)} + x_{(11)})/2 = (53.6 + 54.4)/2 = 54.0$

$a_{20} = (x_{(20)} + x_{(21)})/2 = (56.0 + 56.3)/2 = 56.15$, $a_{25} = (x_{(25)} + x_{(26)})/2 = (57.2 + 57.4)/2 = 57.3$

$a_{30} = (x_{(30)} + x_{(31)})/2 = (57.5 + 57.6)/2 = 57.55$, $a_{40} = (x_{(40)} + x_{(41)})/2 = (58.7 + 58.7)/2 = 58.7$

$a_{50} = (x_{(50)} + x_{(51)})/2 = (59.5 + 59.6)/2 = 59.55$, $a_{60} = (x_{(60)} + x_{(61)})/2 = (61.5 + 61.6)/2 = 61.55$

$a_{70} = (x_{(70)} + x_{(71)})/2 = (62.9 + 62.9)/2 = 62.9$, $a_{75} = (x_{(75)} + x_{(76)})/2 = (63.2 + 63.5)/2 = 63.35$

$a_{80} = (x_{(80)} + x_{(81)})/2 = (64.0 + 64.0)/2 = 64.0$, $a_{90} = (x_{(90)} + x_{(91)})/2 = (66.5 + 66.9)/2 = 66.7$

$a_{95} = (x_{(95)} + x_{(96)})/2 = (67.6 + 67.6)/2 = 67.6$, $a_{99} = (x_{(99)} + x_{(100)})/2 = (70.4 + 71.6)/2 = 71.0$.

The sample mean and variance are

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{6010}{100} = 60.1\,, \quad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{2169.78}{99} = 21.9169697 \cong 21.92$$

and the median of the sample is $m = 59.55$. Note that $m = 59.55 < 60.1 = \bar{x}_n$. That is, the mean is larger than the median (actually theye are very close to each other) and thus the data is skewed to the right.

Moments

Mean	60.1	Sum Observations	6010
Std Deviation	4.68155633	Variance	21.9169697
Skewness	0.01543639	Kurtosis	-0.3346606
Uncorrected SS	363370.78	Corrected SS	2169.78

****

Basic Statistical Measures

Location		Variability	
Mean	60.10000	Std Deviation	4.68156
Median	59.55000	Variance	21.91697
Mode	57.50000	Range	22.40000
		Interquartile Range	6.05000

NOTE: The mode displayed is the smallest of 4 modes with a count of 3.

****

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.991836	Pr < W	0.8093
Kolmogorov-Smirnov	D	0.054045	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.045653	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.266135	Pr > A-Sq	>0.2500

****

Quantiles (Definition 5)

Quantile	Estimate
100% Max	71.60
99%	71.00
95%	67.60
90%	66.70
75% Q3	63.35
50% Median	59.55
25% Q1	57.30
10%	54.00
5%	52.05
1%	49.60
0% Min	49.20

****

```
                 Normal Probability Plot
71.5+                                  +*
    |                               *++
    |                               ++
    |                             +++
    |                           *+* *
    |                          **+
    |                        **+
    |                       **+
    |                      ***
    |                    ***+
    |                   **+
60.5+                   +**
    |                 ***
    |                ***
    |              ****
    |             **+
    |            ***
    |          +*+
    |         +**
    |        +***
    |      ***
    |    *+
49.5+* ++
    +----+----+----+----+----+----+----+----+----+----+
       -2       -1        0       +1       +2
```

<u>Contruction of the Histogram</u>:

Notice that the range of the sample is 22.4. Consider the following classes and count the number of observations fall with in these classes.

	Interval	# of observations f_i	Relative Frequencies	Cumulative Relative Frequencies
I_1	48.15-50.35	2	2/100	2/100
I_2	50.35-52.55	4	4/100	6/100
I_3	52.55-54.75	6	6/100	12/100
I_4	54.75-56.95	11	11/100	23/100
I_5	56.95-59.15	21	21/100	44/100
I_6	59.15-61.35	12	12/100	56/100
I_7	61.35-63.55	21	21/100	77/100
I_8	63.55-65.75	9	9/100	86/100
I_9	65.75-67.95	12	12/100	98/100
I_{10}	67.95-70.15	1	1/100	99/100

I_{11}	70.15-72.35	1	1/100	100/100
		100	1.00	

Histogram by hand	Histogram by computer