

# *Basics of Statistics*

## **1.1 Definition of statistics**

Science of collection, presentation, analysis, and reasonable interpretation of data.

Or

Procedures for organising, summarizing, and interpreting information

The purpose of this course is to introduce you to the subject of statistics as a science of data. There is data abound in this information age; how to extract useful knowledge and gain a sound understanding in complex data sets has been more of a challenge. In this course, we will focus on the fundamentals of statistics, which may be broadly described as the techniques to collect, clarify, summarize, organize, analyze, and interpret numerical information.

## **1.2 How Statistics Helps Predict the Weather**

No need for fortune tellers to predict the weather, we've got the power of statistics.

In fact, meteorologists rely on one thing statisticians know a lot about—applying rigorous scientific techniques to extract information and knowledge from data.

Data first became crucial to predicting the weather in the mid-1800s, with the invention of the telegraph. The telegraph allowed forecasters to receive and record observations from much greater distances. It was during this time that the collection of weather-related data began to grow and change the way weather was predicted.

Today, weather-related data and the scientific discipline of statistics are much more sophisticated. The data, for example, are now collected through satellites, barometers, radars, weather balloons and many other instruments that are located on land, in water and the atmosphere. The data these instruments are collecting includes atmospheric pressure, temperature, speed, rain, humidity, and many other components of current weather conditions.

The Weather Company's prediction process begins with the aggregation of data from over 195,000 personal weather stations across the globe. Then they combine

the data with forecasting and statistical models to provide the hyperlocal weather forecasts that we turn to on our computers, tablets and phones each day.

Meteorologists also rely on these statistical techniques to provide weather predictions that extend from three to seven days out, one-month and even three-month average forecasts that are distributed by the Climate Prediction Center – NOAA. Inherent in the predictions is the likelihood of an event—e.g., 80 percent chance of rain, two to four inches of snow, etc.—which is another hallmark of the use of good statistical science.

Though some still feel weather predictions are no more than educated guesses, in reality, they are the result of extensive statistical analysis. So, on your next snow day be sure to thank statistics.

### 1.3 Type of statistics

Statistics could be divided into two broad areas:

1. descriptive statistics
2. inferential statistics.

Both are relevant to the atmospheric sciences.

#### A. *Descriptive Statistics:*

Tools for summarising, organising, simplifying data

- ❖ *Tables & Graphs*
- ❖ *Measures of Central Tendency*
- ❖ *Measures of Variability*

*Examples:*

Average rainfall in Manchester last year

Number of car thefts in last year

Your test results

Percentage of males in our class

## B. Inferential Statistics:

Data from *sample* used to draw inferences about *population*

*Generalising* beyond actual observations

*Generalise* from a *sample* to a *population*

### 1.4 Statistical terms

#### ❖ Population

- complete set of individuals, objects or measurements
- Sample
- a sub-set of a population

#### ❖ Variable

- a characteristic which may take on different values

#### ❖ Data

- numbers or measurements collected

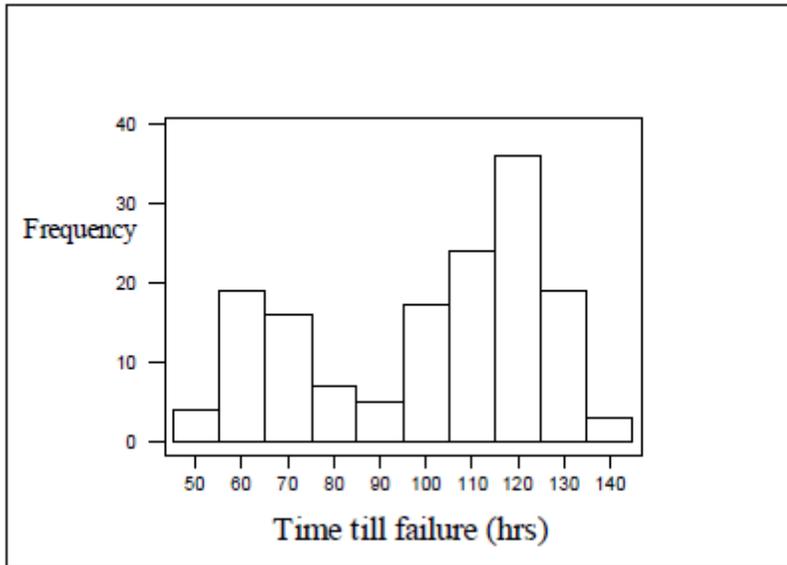
#### ❖ There are various types of variate .

Qualitative or nominal; described by a word or phrase (e.g. blood group, colour, male or female ).

Quantitative; described by a number (e.g. time till cure, number of calls arriving at a telephone exchange in 5 seconds).

### 1.5 Displaying data

It is nearly always useful to use graphical methods to illustrate your data. We shall describe in this section just a few of the methods available..



### What is a Frequency Distribution Table?

*Frequency* tells you **how often something happened**. The frequency of an observation tells you the number of times the observation occurs in the data.

### 1.6 Construction of a Frequency Table

Let us consider 30 test scores of students in Statistics.

The results are as follows:

98	97	92	90	87
82	85	99	93	91
90	89	81	76	88
87	81	83	88	89
90	92	90	89	77
80	90	85	86	95

In constructing a frequency table, we must follow certain steps.

1. Compute for the Range:

$$\text{Range} = \text{highest} - \text{lowest}$$

$$\text{Range} = 99 - 76$$

$$\text{Range} = 23$$

2. Compute for k (desired number of class interval)

$$k = 1 + 3.3 \log n$$

Where:  $n$  = number of observations

$$n = 30$$

$$k = 1 + 3.3 \log 30$$

$$k = 5.87 \approx 6$$

3. Compute for C (class size)

$$C = \text{Range} \div k$$

$$C = 23 \div 6$$

$$C = 3.83 \approx 4$$

4. Class Intervals

$$76 - 79$$

$$80 - 83$$

$$84 - 87$$

$$88 - 91$$

$$92 - 95$$

$$96 - 99$$

The next column is the class boundary. To

construct the class boundary, subtract 0.5 from the lower interval and add 0.5 to the higher interval.

Class Intervals	Class Boundaries
76 – 79	75.5 – 79.5
80 – 83	79.5 – 83.5
84 – 87	83.5 – 87.5
88 – 91	87.5 – 91.5
92 – 95	91.5 – 95.5
96 – 99	95.5 – 99.5

6. next is the Class Mark represented by  $X_i$ . It is computed by getting the average of the Class Interval.

Class Intervals	$X_i$
76 – 79	77.5
80 – 83	81.5
84 – 87	85.5
88 – 91	89.5
92 – 95	93.5
96 – 99	97.5

7. Tally all the observations according to their respective intervals.

Class Intervals	Tally
76 – 79	II
80 – 83	IIII
84 – 87	IIII
88 – 91	IIII – IIII – 1
92 – 95	IIII
96 – 99	II

8. Count the tally per interval and *write its numerical equivalence in the next column.*

Tally	Frequency
LI	2
LIII	5
LIII	5
IIII – IIII – I	11
IIII	4
III	3

9. Construct the less than and greater than cumulative frequency. The less than cumulative frequency is constructed by copying the first frequency and adding it to the next frequency. The result is to be added to the next frequency until you reach the last cumulative frequency which is equal to the total number of observations. The greater than cumulative frequency follows the same procedure, its just that it starts from the last frequency and accumulates upward and the first greater than cumulative frequency is equal to the total number of observations.

class Intervals	Freq	F<	F>
76 – 79	2	2	30
80 – 83	5	7	28
84 – 87	5	12	23
88 – 91	11	23	18
92 – 95	4	27	7
96 – 99	3	30	3
n = 30			

10. Construct the relative frequency and relative frequency percentage.

Relative frequency ( rf ) is computed by dividing each individual frequency by the total frequency. While the rf% is computed by multiplying rf by 100.

Frequency	rf	rf %
2	0.0667	6.67
5	0.1667	16.67
5	0.1667	16.67
11	0.3667	36.67
4	0.1333	13.33
3	0.1000	10.00
n = 30	1.0001 $\approx$ 1.0	100.01 $\approx$ 100

### Summary of the Frequency Table

<i>C.L</i>	<i>C.B.</i>	<i>Xi</i>	<i>Tally</i>	<i>Freq</i>	<i>F&lt;</i>	<i>F&gt;</i>	<i>f</i>	<i>f%</i>
76 -79	75.5-79.5	77.5	I	2	2	30	0.0667	6.67
80 -83	79.5-83.5	81.5	00	5	7	28	0.1667	16.67
84-87	83.5-87.5	85.5	00	5	12	23	0.1667	16.67
88-91	87.5-91.5	89.5	00 -00 -1	11	23	18	0.3667	36.67
92 -95	91.5-95.5	93.5	llll	4	27	7	0.1333	13.33
96-99	95.5 -99.5	97.5	lll	3	30	3	0.1000	10.00
				n=30			1.0001	100.01