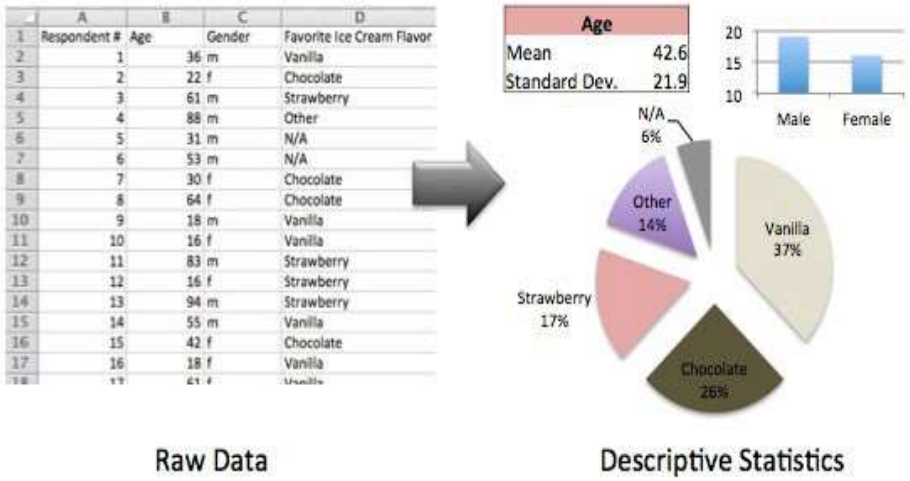
Statistics: Descriptive and Inferential

- Statistics plays a main role in the field of research. It helps us in the collection, analysis and presentation of data.
- The field of statistics is composed of two broad categories- Descriptive and inferential statistics.
- Both of them together gives us a powerful tool for description and prediction.
- Population is the group that is targeted to collect the data from. Our data is the information collected from the population.
- Sample population which is selected randomly for the study. The sample should be selected such that it represents all the characteristics of the population.

Descriptive Statistics



	A	B	C	D
1	Respondent #	Age	Gender	Favorite Ice Cream Flavor
2	1	36	m	Vanilla
3	2	22	f	Chocolate
4	3	61	m	Strawberry
5	4	88	m	Other
6	5	31	m	N/A
7	6	53	m	N/A
8	7	30	f	Chocolate
9	8	64	f	Chocolate
10	9	18	m	Vanilla
11	10	16	f	Vanilla
12	11	83	m	Strawberry
13	12	16	f	Strawberry
14	13	94	m	Strawberry
15	14	55	m	Vanilla
16	15	42	f	Chocolate
17	16	18	f	Vanilla

Raw Data

Descriptive Statistics

- To summarize and graph the data for a group that you choose.
- You simply take a group that you're interested in, record data about the group members, and then use summary statistics and graphs to present the group properties.
- There is no uncertainty because you are describing only the people or items that you actually measure.
- You're not trying to infer properties about a larger population.
- Data can be summarized and represented in an accurate way using charts, tables and graphs.

Common tools of descriptive statistics

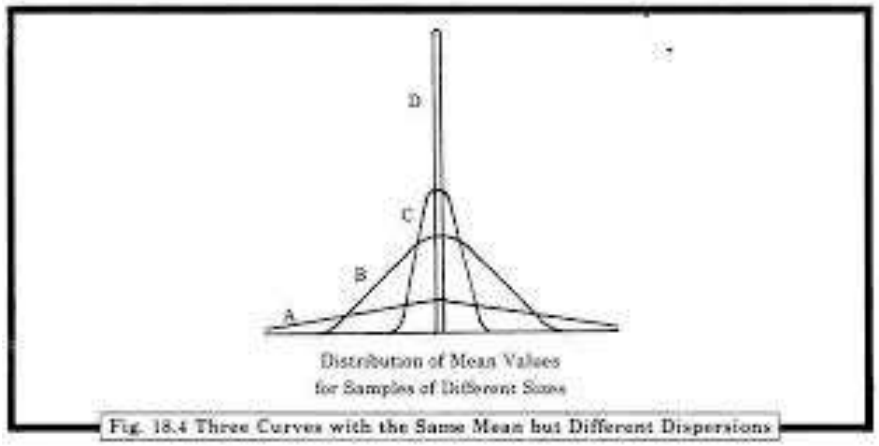
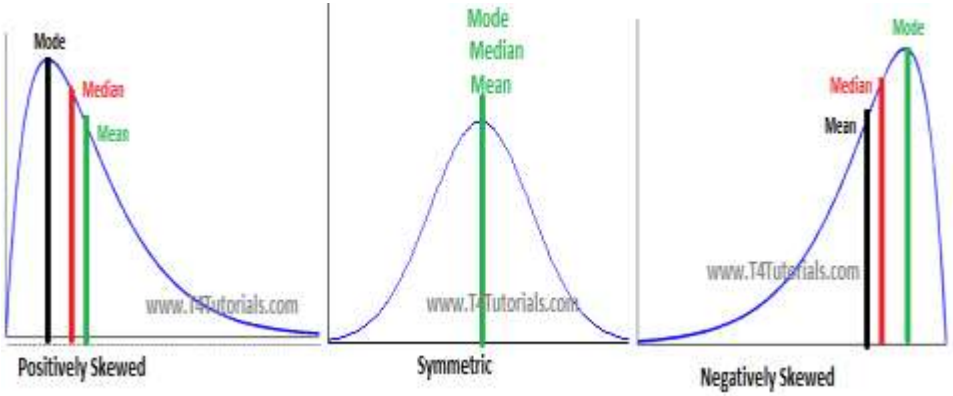


Fig. 18.4 Three Curves with the Same Mean but Different Dispersions

Descriptive statistics provides us the tools to define our data in a most understandable and appropriate way:

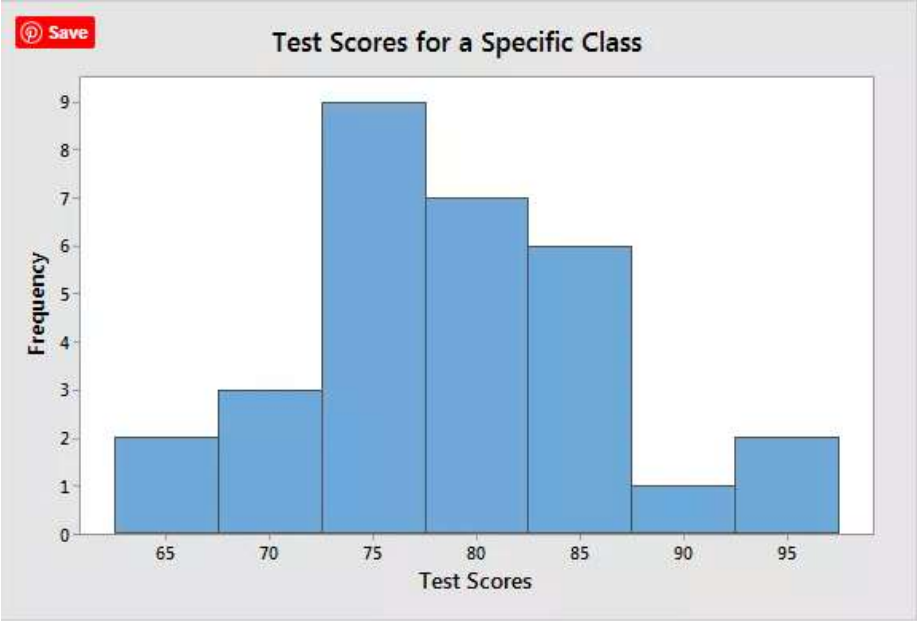
- **Central tendency**: Use the mean or the median to locate the center of the dataset. This measure tells you where most values fall.

- **Dispersion**: How far out from the center do the data extend? You can use the range or standard deviation to measure the dispersion. A low dispersion indicates that the values cluster more tightly around the center. Higher dispersion signifies that data points fall further away from the center. We can also graph the frequency distribution.



- **Skewness**: The measure tells you whether the distribution of values is symmetric or skewed.

Example of descriptive statistics



Statistic	Class value
Mean	79.18
Range	66.21 – 96.53
Proportion >= 70	86.7%

- Suppose we want to describe the test scores in a specific class of 30 students. We record all of the test scores and calculate the summary statistics and produce graphs.
- These results indicate that the mean score of this class is 79.18, distribution is symmetrically centered around the mean.
- The scores range from 66.21 to 96.53.
- A score of at least 70 on the test is acceptable.
- The data show that 86.7% of the students have acceptable scores.

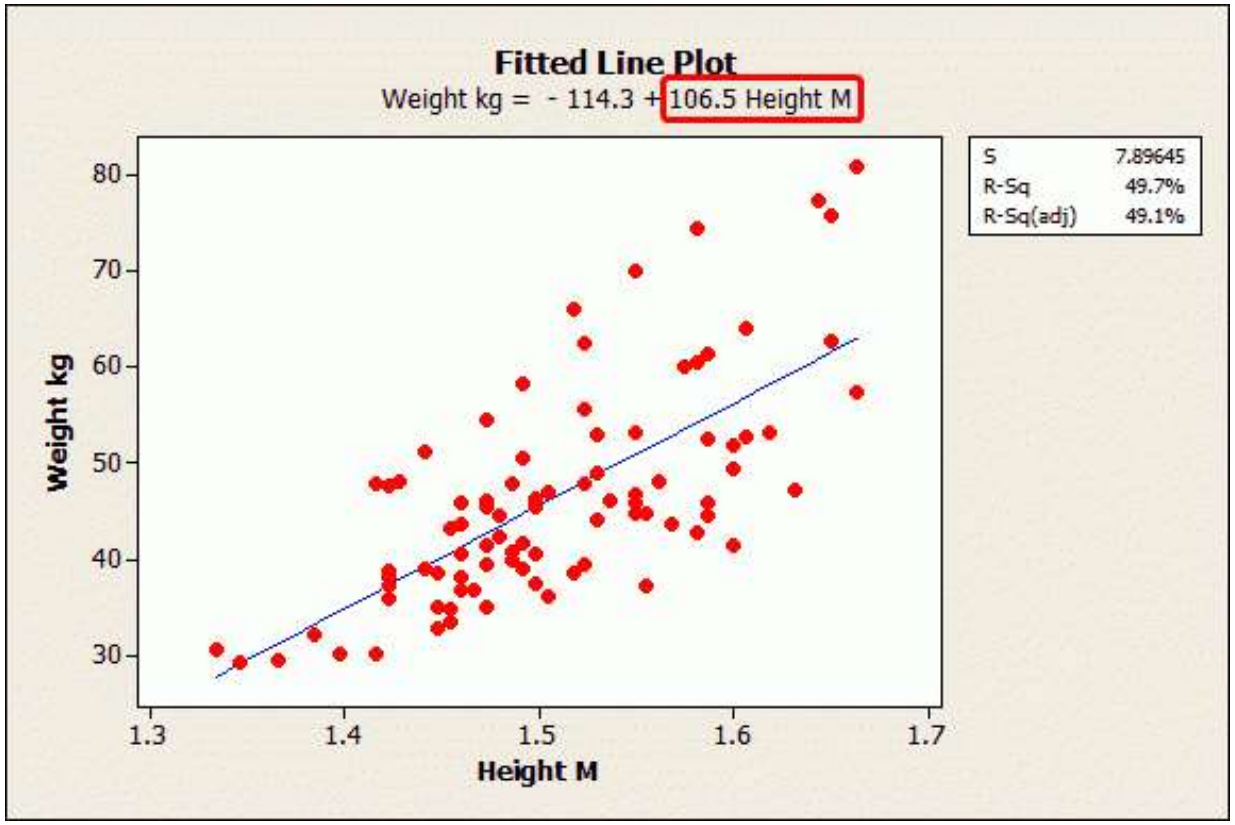
Inferential Statistics

- Inferential statistics takes data from a sample and makes inferences about the larger population from which the sample was drawn.
- Goal of inferential statistics is to draw conclusions from a sample and generalize them to a population, we need to have confidence that our sample accurately reflects the population.
- Random sampling allows us to have confidence that the sample represents the population.
- Consequently, when you estimate the properties of a population from a sample, the sample statistics are unlikely to equal the actual population value exactly.
- The difference between the sample statistic and the population value is the sampling error.

tools of inferential statistics

- Hypothesis tests use sample data answer questions like the following:
  - 1) Is the population mean greater than or less than a particular value?
  - 2) Are the means of two or more populations different from each other?
  - e.g. effectiveness of a new medication. After all, we don't want to use the medication if it is effective only in our specific sample. Instead, we need evidence that it'll be useful in the entire population of patients. Hypothesis tests allow us to draw these types of conclusions about entire populations.
- Confidence intervals incorporate the uncertainty and sample error to create a range of values the actual population value is like to fall within.

e.g. we draw a random sample from this population and calculate the mean height of 181 cm. Now, a confidence interval of [176 186] indicates that we can be confident that the real population mean falls within this range.



Regression analysis describes the relationship between a set of independent variables(height) and a dependent variable(weight).

We have sufficient evidence to conclude that this relationship exists in the population rather than just our sample.

Example of inferential statistics

Statistic	Population Parameter Estimate (CIs)
Mean	77.4 - 80.9
Standard deviation	7.7 - 10.1
Proportion scores >= 70	77% - 92%

- Assume that we are provided a list of names for the entire population and draw a random sample of 100 students from it and obtain their test scores. Given the uncertainty associated with these estimates, we can be 95% confident that:
- The population mean is between 77.4 and 80.9.
- The population standard deviation (a measure of dispersion) is likely to fall between 7.7 and 10.1.
- The population proportion of satisfactory scores is expected to be between 77% and 92%.

Some differences to remember!

S. No	Descriptive Statistics	Inferential Statistics
1	Concerned with the describing the target population	Make inferences from the sample and generalize them to the population.
2	Organize, analyze and present the data in a meaningful manner	Compares, test and predicts future outcomes.
3	Final results are shown in form of charts, tables and Graphs	Final result is the probability scores.
4	Describes the data which is already known	Tries to make conclusions about the population that is beyond the data available.
5	Tools- Measures of central tendency (mean/median/ mode), Spread of data (range, standard deviation etc.)	Tools- hypothesis tests, Analysis of variance etc.