Journal 11/13/18

If you are using a survey:

who will participate in your survey? Why did you decide on that? Explain

If you are not using a survey:

Where will you look for information? Why did you decide on that? Explain

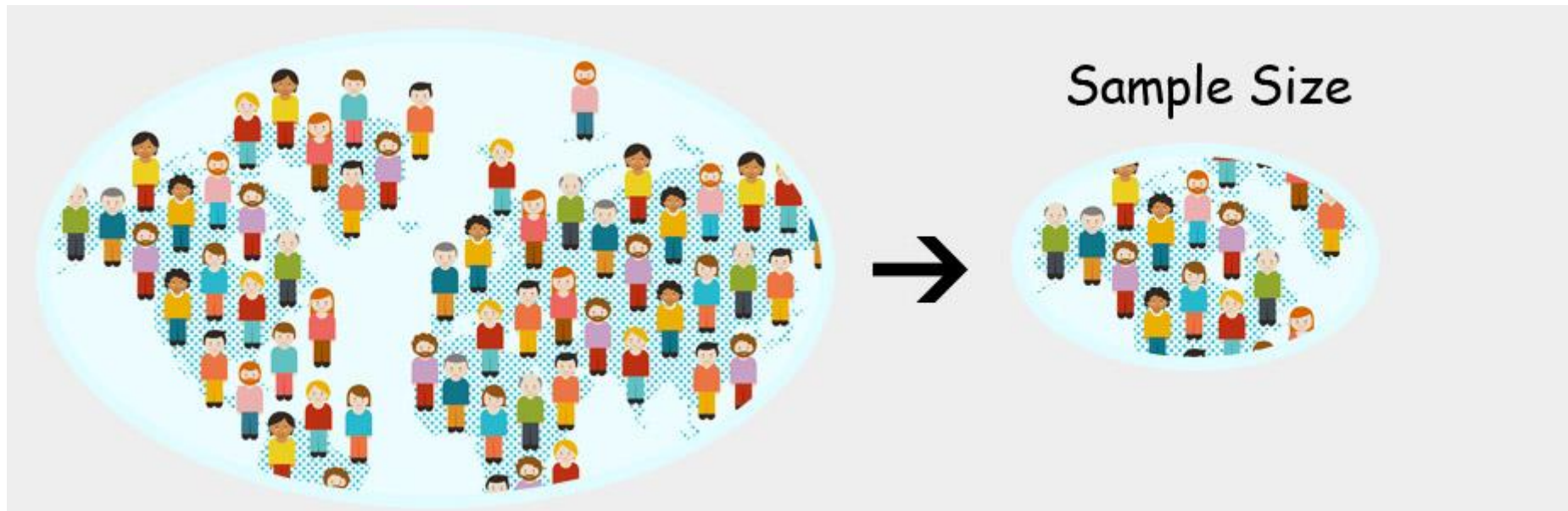# Unit 2: Research Methods
Table of Contents

Sampling

Determining Your Sample Size

Population

- Entire group of individuals that we want information from.

Sample

- A part of the population that we examine in order to gather information.

Sample Size

Sample Size

- *Sample size* is an important concept in statistics, and refers to the number of individual pieces of data collected in a survey.

- A survey or statistic's sample size is important in determining the **accuracy and reliability** of a survey's findings.

The Definition of Sample Size

Sample size measures the number of individual samples measured or observations used in a survey or experiment.

- For example,
  - if you observe 100 people to measure length of social interactions, your sample size is 100.
  - If an online survey returned 30,500 completed questionnaires, your sample size is 30,500.
- In statistics, sample size is generally represented by the variable "n".

Calculation of Sample Size

To determine the sample size needed for an experiment or survey, researchers take a number of desired factors into account.
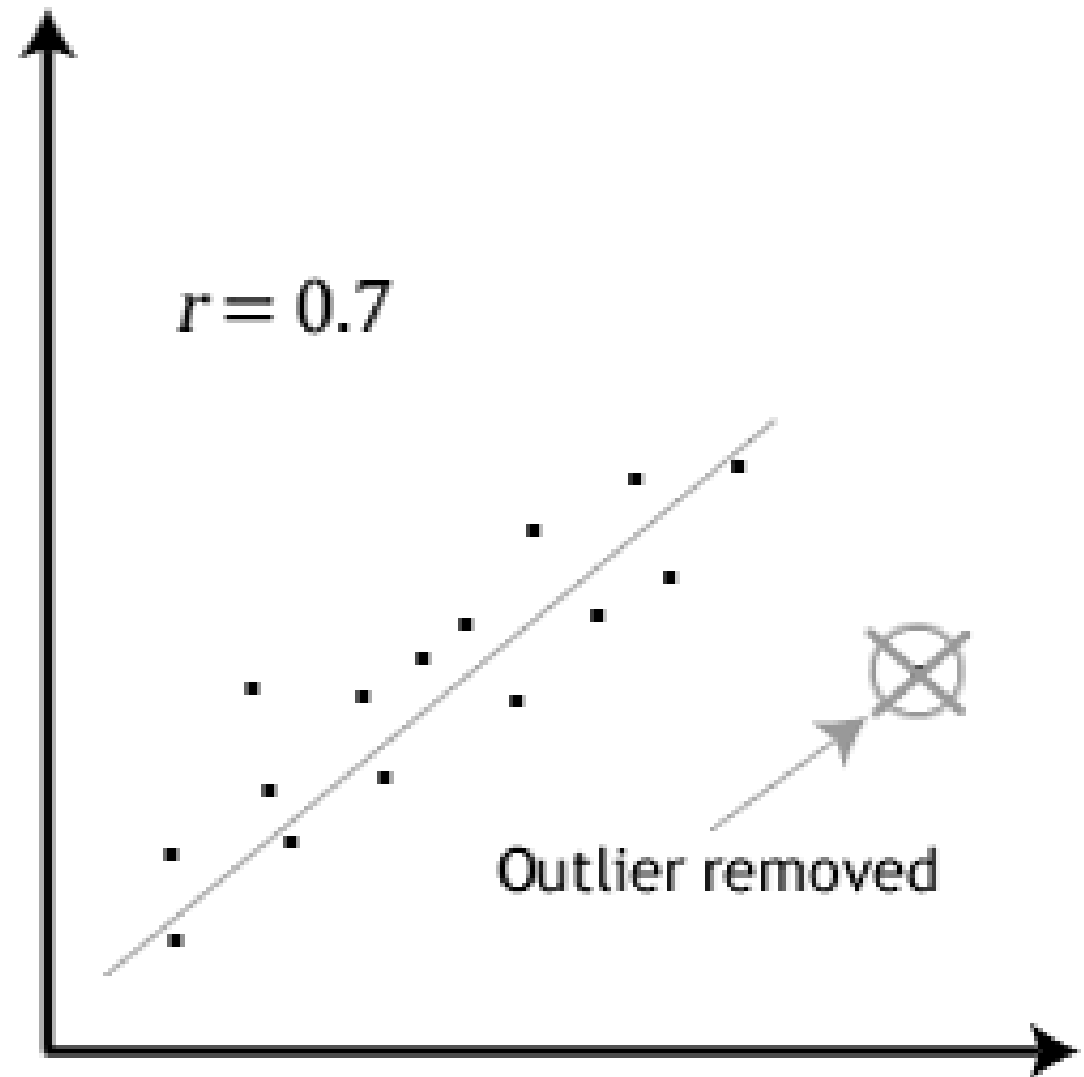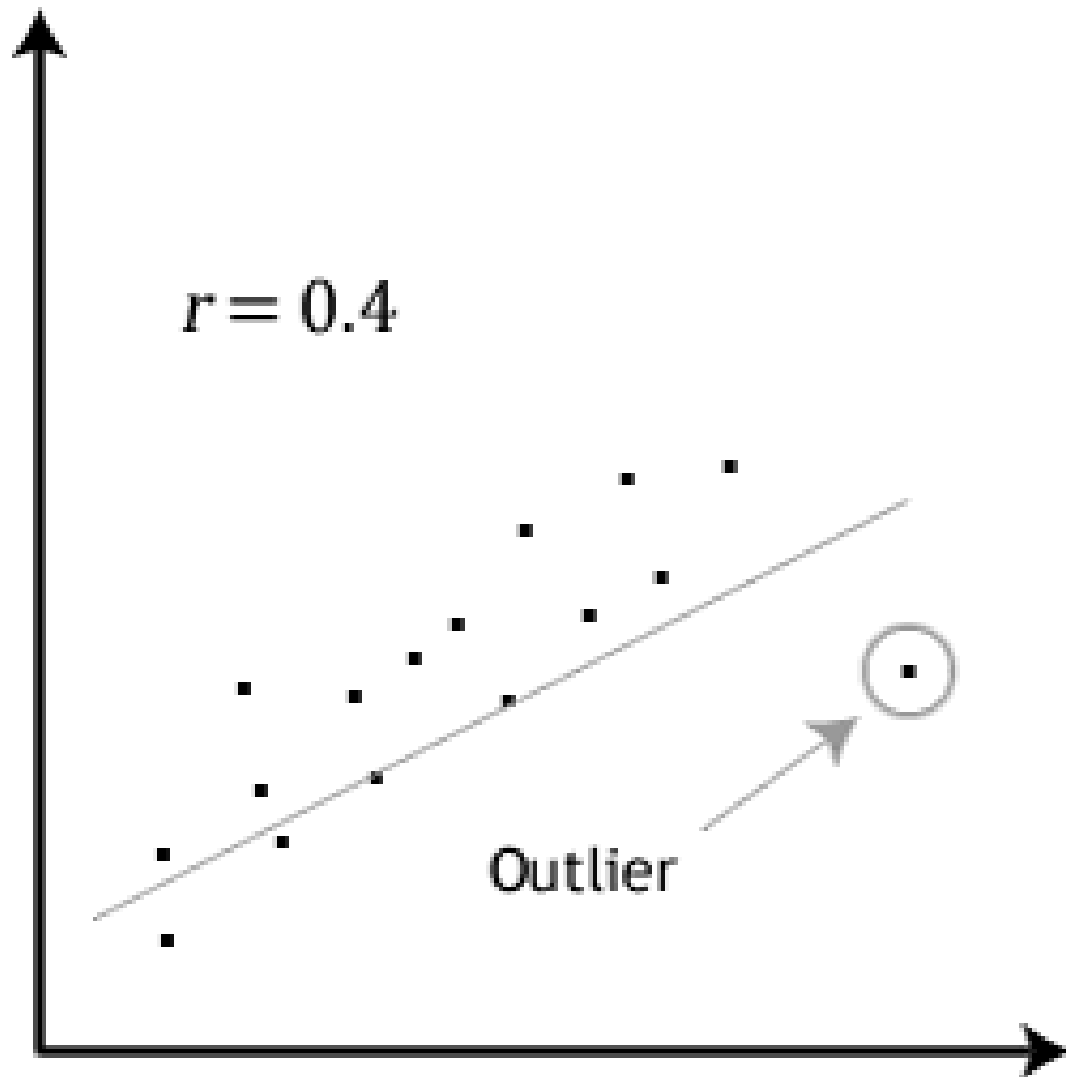
- Population Size
  - a survey that is looking to draw conclusions about all of California, for example, will need a much larger sample size than one specifically focused on Brentwood.

- Margin of Error
  - the reliability that the data collected is generally accurate; and the confidence level, the probability that your margin of error is accurate.

- Standard Deviation
  - Standard deviation measures how much individual pieces of data vary from the average data measured. For instance, families from the same neighborhood will likely have a much smaller standard deviation in their average income than family incomes collected from across a whole county.

Dangers of Small Sample Size

- Large sample sizes are needed for a statistic to be accurate and reliable, especially if its findings are to be extrapolated to a larger population or group of data.

- Say you were conducting a survey about exercise and interviewed five people, two of whom said they run a marathon annually. If you take this survey to represent the population of the country as a whole, then according to your research, 40 percent of people run at least one marathon annually -- an unexpectedly high percentage.

- The smaller your sample size, the more likely outliers -- unusual pieces of data -- are to skew your findings.

Linear Regression measures the relationship between variables

99.7% of the data are within 3 standard deviations of the mean

95% within 2 standard deviations

68% within 1 standard deviation

$\mu - 3\sigma$ $\mu - 2\sigma$ $\mu - \sigma$ μ $\mu + \sigma$ $\mu + 2\sigma$ $\mu + 3\sigma$

(a) Negatively skewed

Mode

Median

Mean

Frequency

X

Negative direction

(b) Normal (no skew)

Mean
Median
Mode

X

The normal curve
represents a perfectly
symmetrical distribution

(c) Positively skewed

Mode

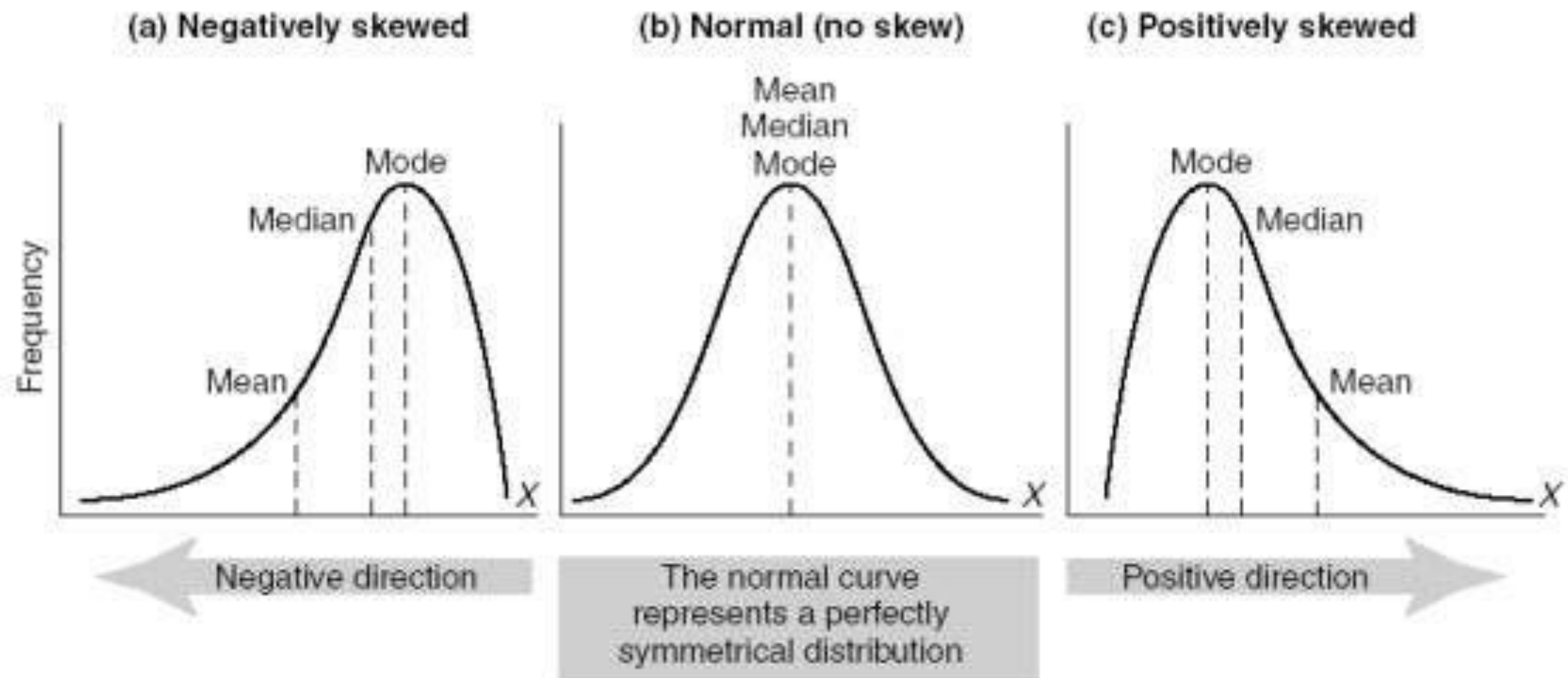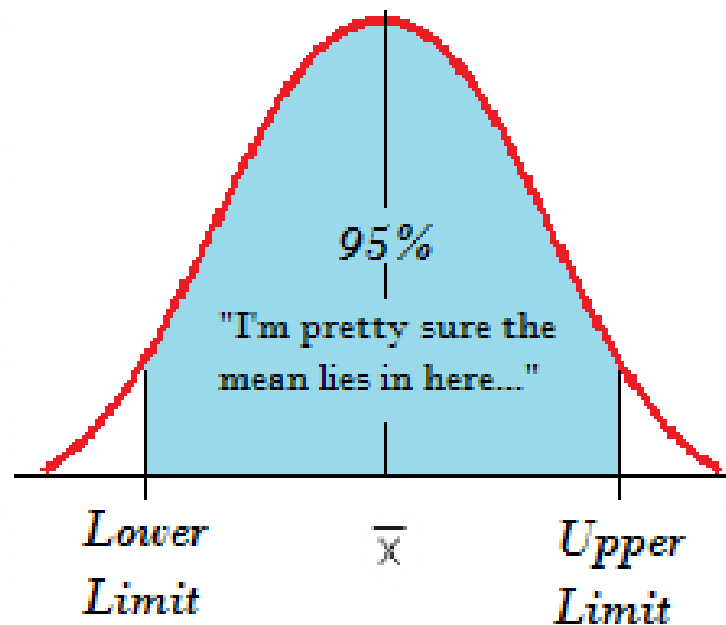Median

Mean

X

Positive direction
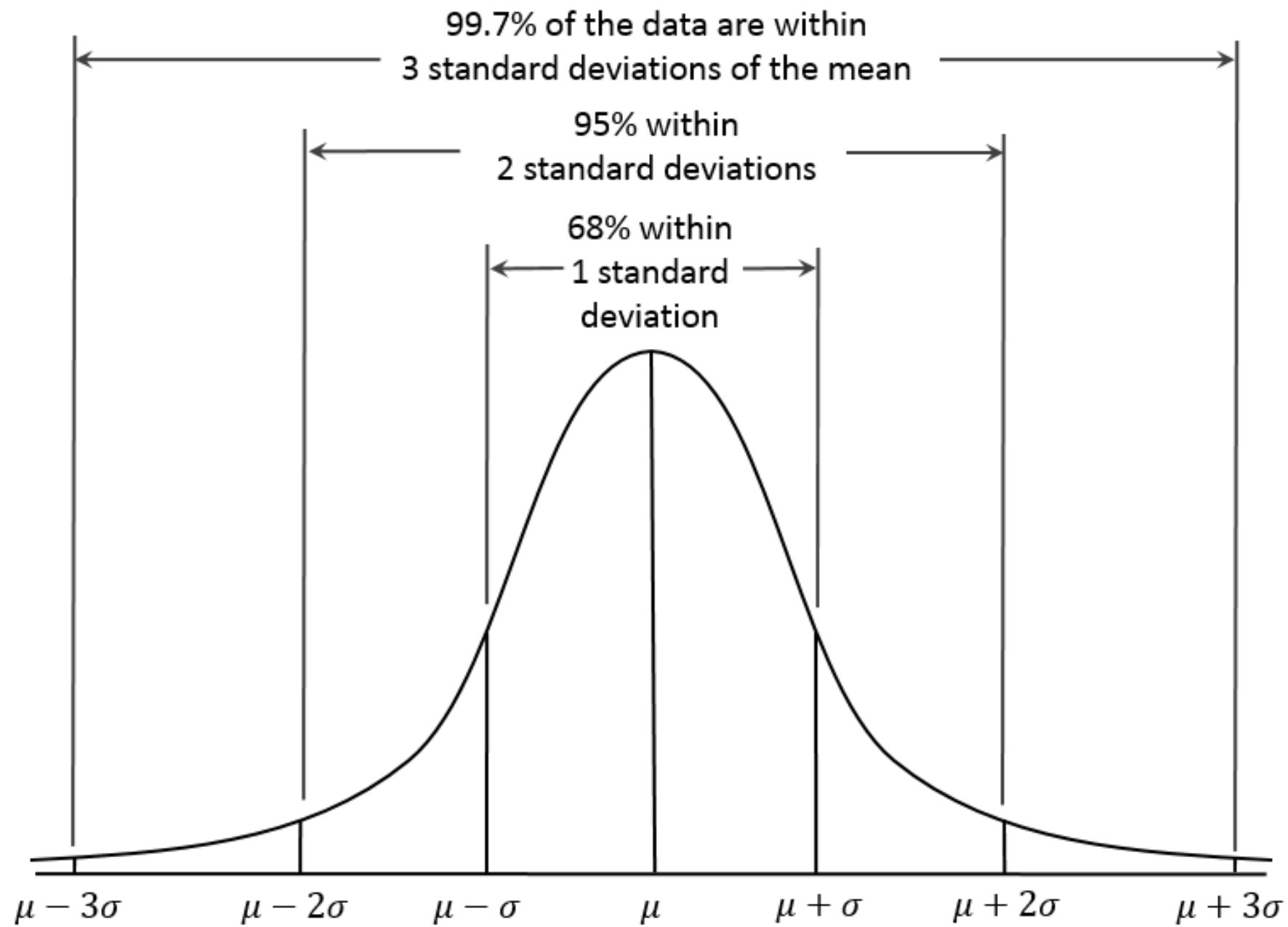
■ FIGURE 15.6   Examples of normal and skewed distributions

Sample Size and Margin of Error

- The sample size of a statistical survey is also directly related to the survey's margin of error.

- Margin of error is a percentage that expresses the probability that the data received is accurate.

- For example, in a survey about religious beliefs, the margin of error is the percentage of responders who can be expected to provide the same answer if the survey was repeated.

- To determine the margin of error, divide 1 by the square root of the sample size, and then multiply by 100 to get a percentage. For instance, a sample size of 2,400 will have a margin of error of 2.04 percent.

	Confidence level = 95%			Confidence level = 99%		
	Margin of error			Margin of error		
Population size	5%	2,5%	1%	5%	2,5%	1%
100	80	94	99	87	96	99
500	217	377	475	285	421	485
1.000	278	606	906	399	727	943
10.000	370	1.332	4.899	622	2.098	6.239
100.000	383	1.513	8.762	659	2.585	14.227
500.000	384	1.532	9.423	663	2.640	16.055
1.000.000	384	1.534	9.512	663	2.647	16.317

CONFIDENCE INTERVAL

95%

"I'm pretty sure the mean lies in here..."

Lower Limit $\overline{x}$ Upper Limit

TOLERANCE INTERVAL

95%

"I'm pretty sure a big % of values will fall in here..."

Lower Limit $\overline{x}$ Upper Limit
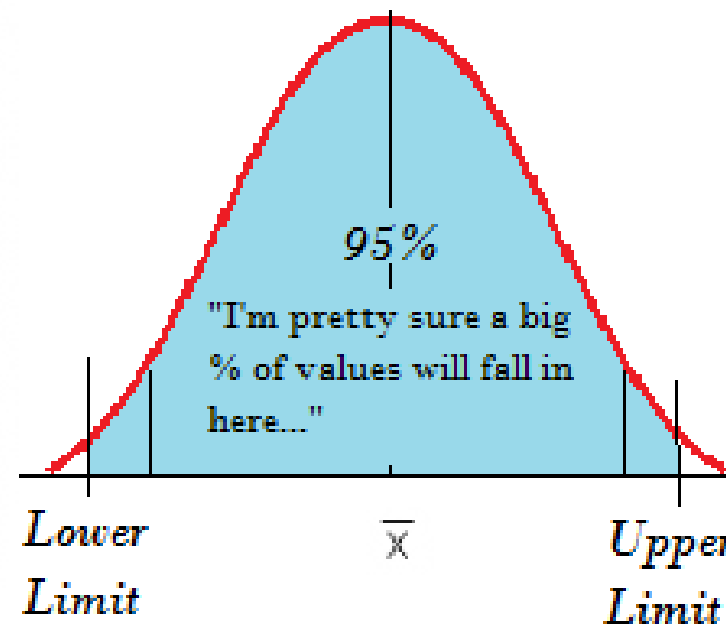
Simple Random Sampling

- **Each member of the population has an equal chance of being chosen for the study.**
- Simple random sampling is the most basic and common type of sampling method used in quantitative social science research and in scientific research generally.
- Guarantees that the sample chosen is representative of the population and that the sample is selected in an unbiased way.
- The statistical conclusions drawn from analysis of the sample will be valid.
- There are multiple ways of creating a simple random sample.
  - Lottery method, using a random number table, using a computer, and sampling with or without replacement.

Non Random Samples

- **Researchers sample based on hypotheses about the population of interest, known as selection criteria.**

- For example, if we're selecting our sample by stopping people on the street, attempting to stop an equal number of men and women (to coincide with the presumed gender distribution in the population) would be a criterion of nonrandom sampling.

- In these cases, since the selection of units for the sample isn't random, we shouldn't talk about error estimates. In other words, a nonrandom sample tells us about a population, but we don't know how precisely: we can't determine a margin of error or a confidence level.

- These types of sampling methods include convenience sampling, purposive sampling, quota sampling, snowball sampling

Convenience Sampling

- There is no evidence that they are representative of the populations we're interested in generalizing to -- and in many cases we would clearly suspect that they are not.

Purposive Sampling

- **Sample with a purpose in mind.**

- One or more specific predefined groups we are seeking.

- One of the first things they're likely to do is verify that the respondent does in fact meet the criteria for being in the sample.

- Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern.

- With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.

Quota Sampling

- **Select people non-randomly according to a fixed quota.**

- There are two types of quota sampling: proportional and non proportional.

- In proportional quota sampling you want to represent the major characteristics of the population by sampling a proportional amount of each.

- For instance, if you know the population has 40% women and 60% men, and that you want a total sample size of 100, you will continue sampling until you get those percentages and then you will stop. So, if you've already got the 40 women for your sample, but not the sixty men, you will continue to sample men but even if legitimate women respondents come along, you will not sample them because you have already "met your quota."

Snowball Sampling

- **Identify someone who meets the criteria for inclusion in your study. Ask them to recommend others who they may know who also meet the criteria.**

- Although this method would hardly lead to representative samples, there are times when it may be the best method available.

- Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find.

- For instance, if you are studying the homeless, you are not likely to be able to find good lists of homeless people within a specific geographical area. However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them

Reliability

- Reliability is another term for consistency.

- If one person takes the same personality test several times and always receives the same results, the test is reliable.

Validity

- A test is valid if it measures what it is supposed to measure.

- If the results of the personality test claimed that a very shy person was in fact out going, the test would be invalid.

Reliable Not Valid	Valid Not Reliable	Neither Reliable Nor Valid	Both Reliable And Valid