Skewness

Skewness refers to the asymmetry or distortion in the distribution of data around its mean. A distribution is skewed if one tail is longer or fatter than the other.

Types of Skewness

1. **Symmetrical Distribution (Zero Skewness):**

- Mean = Median = Mode
- Bell-shaped curve (e.g., normal distribution)

2. **Positively Skewed (Right Skewed):**

- Tail on the right
- Mean > Median > Mode
- More values are concentrated on the lower side

3. **Negatively Skewed (Left Skewed):**

- Tail on the left

Symmetrical Distribution

A symmetrical distribution is a type of frequency distribution where the left and right sides are mirror images of each other. In such distributions:

Mean=Median=Mode

Characteristics :

1. Balanced shape — data is evenly spread around the center.
2. The curve is bell-shaped (like the normal distribution).
3. No skewness — Skewness = 0
4. Tails on both ends of the distribution are equal in length and shape.

Graphical Distribution

Example:

Inferential Statistics

While descriptive statistics summarize the characteristics of a data set, inferential statistics help you come to conclusions and make predictions based on your data.

- Making estimates about populations (for example, the mean SAT score of all 11th graders in the US).
- testing hypotheses to draw conclusions about populations (for example, the relationship between SAT scores and family income).

Descriptive Versus

Descriptive statistics :Allow you to *describe* a data set, while inferential statistics allow you to make *inferences* based on a data set.

- The **distribution** concerns the frequency of each value.
- The **central tendency** concerns the averages of the values.
- The **variability** concerns how spread out the values are.

Parameter vs Statistic

A parameter is a number describing a whole population (e.g., population mean), while a statistic is a number describing a sample (e.g., sample mean).

The goal of quantitative research is to understand characteristics of populations by finding parameters. In practice, it's often too difficult, time-consuming or unfeasible to collect data from every member of a population. Instead, data is collected from samples.

Population vs Sample

A sample is a smaller group taken from the population. The sample is the group of elements that you will actually collect data from.

A population is the entire group that you're interested in studying. This may be a group of people (e.g., all adults in the US or all employees of a company), but it can also mean a group containing other kinds of elements: objects, events, organizations, countries, species, organisms, etc.

Standard Error

- The standard error of the mean, or simply **standard error**, indicates how different the population mean is likely to be from a sample mean. It tells you how much the sample mean would vary if you were to repeat a study using new samples from within a single population .
- The standard error of the mean (SE or SEM) is the most commonly reported type of standard error. But you can also find the standard error for other statistics, like medians or proportions. The standard error is a common measure of sampling error—the difference between a population parameter and a sample statistics.

Hypothesis Testing

Hypothesis testing is a formal procedure for investigating our ideas about the world using statistics .It is most often used by scientists to test specific predictions, called hypotheses, that arise from theories.

There are 5 main steps in hypothesis testing:

1. State your research hypothesis as a null hypothesis and alternate hypothesis (H_o) and (Ha or H1).
2. Perform an appropriate statistical test.
3. Decide whether to reject or fail to reject your null hypothesis.
4. Present the findings in your results and discussion section.

Null & Alternative Hypotheses

The null and alternative hypotheses are two competing claims that researchers weigh evidence for and against using a statistical test:

- **Null hypothesis (H_0):** There's no effect in the population.
- **Alternative hypothesis (H_a or H_1):** There's an effect in the population.

The effect is usually the effect of the independent variable on the dependent variable .

Statistical Significance

Statistically significant, that means it's unlikely to be explained solely by chance or random factors. A statistically significant result has a very low chance of occurring if there were no true effect in a research study.

Type I & Type II Errors

The probability of making a Type I error is the significance level, or alpha (α), while the probability of making a Type II error is beta (ß). These risks can be minimized through careful planning in your study design.

Example: Type I vs Type II error

- **Type I error (false positive):** the test result says you have coronavirus, but you actually don't.
- **Type II error (false negative):** the test result says you don't have coronavirus, but you actually do.

Statistical Power

Statistical power, or sensitivity, is the likelihood of a significance test detecting an effect when there actually is one.

A true effect is a real, non-zero relationship between variables in a population. An effect is usually indicated by a real difference between groups or a correlation between variables.

Power is mainly influenced by sample size, effect size, and significance level. A power analysis can be used to determine the necessary sample size for a study.

Statistical Test

- **Reporting Statistics:**

Statistical analysis involves gathering and testing quantitative data to make inferences about the world. A statistic is any number that describes a sample: it can be a proportion, a range, or a measurement, among other things.

- **Numbers and Measurements :**

In general, APA advises using words for numbers under 10 and numerals for 10 and greater. However, always spell out a number that appears at the start of a sentence (or rephrase).

- Exact numbers before units of measurement or time
- Mathematical equations
- Percentages and percentiles
- Ratios, decimals, and uncommon fractions

Units of measurements and time

Measurements using numerals, and use symbols or abbreviations for common units of measurement when they accompany exact measurements. Include a space between the number and the abbreviation.

Measurements should be reported in metric units. If you recorded measurements in non-metric units, include metric equivalents in your report as well as the original units.

Percentage:

- Use numerals for percentages along with the percent symbol (%). Don't insert a space between the number and the symbol.
- Words for "percent" or "percentage" should only be used in text when numbers aren't used, or when a percentage appears at the start of a sentence.

Decimal places and leading zeroes

The number of decimal places to report depends on what you're reporting. Generally, you should aim to round numbers while retaining precision. It's best to present fewer decimal digits to aid easy understanding.

Use two or three decimal places and report exact values for all *p* values greater than .001. For p values smaller than .001, report them as $p < .001$.

Leading zeroes:

A leading zero is zero before the decimal point for numbers less than one. In APA Style, it's only used in some cases.

Use a leading zero only when the statistic you're describing can be greater than one. If it can never exceed one, omit the leading zero.