

Chapter 07: Fundamentals of Inferential Statistics



Introduction to Statistics in Kinesiology

Furtado JR., O

Cal State Northridge

updated: 2022-02-16



```
xaringanExtra::use_webcam()
```

From sample to population

- Inferential statistics: the process of estimating population parameters based on sample statistics.

1. Difficult to access entire population
2. Population estimation is made based on sample

Statistic > Parameter

Parameter vs Statistic

Parameter is a characteristic of the population, while *statistic* is an estimate of a parameter based on the information obtained from a sample.

Measurement	statistics	parameter
Population Proportion	p	P
Data points	x	X
Population mean	\bar{x}	μ
Standard deviation	s	σ
Variance	s^2	σ^2
Number of persons/objects	n	N
Correlation coefficient	r	ρ

Also:

H_0 : for **null hypothesis**

H_1 : for **alternate hypothesis**

Sampling error

the amount of error in the estimate of a population parameter

Standard error of the mean (SE_m)

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{N}}$$

(SE_m) = estimate of the amount of error when a sample mean (\bar{x}) is used to estimate the population mean (μ).

Since the μ cannot be known, we use the \bar{x} and a estimate of the error (SE_m) we should to expect.

LOC and Probability error

Levels of confidence (LOC)

- A percentage value that establishes the probability that a statement is correct

Probability error

Common values:

- at 68% LOC, probability of error = 32%
- at 90% LOC, probability of error = 10%
- at 95% LOC, probability of error = 5%

Hypothesis testing

even

odd

id	scores
1	2
2	4
3	7
4	11
5	12
6	14

Median, cont.

- Represents the score at the **50th percentile**
- Divides the data set in two (visible in a boxplot)
- If N is odd, **median** is the middle score¹
- If N is even, do one of the following:
 - Use the higher of the two middle scores
 - Compute the average of the two middle scores

[1] data must be ranked first.

Median (cont.)

When to use the **median**?

- use with ordinal data or cases of highly skewed distributions

Important: not affected by extreme scores

Mean

The **mean**¹ is the most popular index of central tendency

Equation for the sample mean:

$$\bar{x} = \frac{\sum x}{n}$$

Equation for the population mean:

$$\mu = \frac{\sum x}{N}$$

[1] equal to arithmetic average

Mean (cont.)

The most sensitive of the central tendency indices

1. affected by every score in the
2. greatly affected by outliers

Play important role on statistical inference

Used with interval and ratio data¹

[1] In `jamovi`, interval and ratio = **continuous** data

Test

tab1

tab2

content tab1

The holy passion of Friendship is of so sweet and steady and loyal and enduring a nature that it will last through a whole lifetime...