

evolve  
college of massage therapy

---

Sampling, Statistics and  
Causation

INSTRUCTOR: DAN BAILEY  
RESEARCH METHODS IN MASSAGE THERAPY

---

---

---

---

---


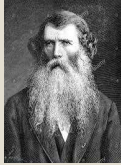
---

---

---

Why are numbers Important?

Gregor Mendel 1822-1884	Charles Naudin 1815-1899
----------------------------	-----------------------------

	
---	---

---

---

---

---

---

---

---

---

What Do We Need To Understand Research?

- ❖ We need to understand the data.
- ❖ We need to understand the type of research.
- ❖ We need to understand where the research comes from.
- ❖ We need to understand why the research was done
- ❖ We need to understand the research question
- ❖ **We need to understand how the data was collected.**
- ❖ We need to understand the data analysis.
- ❖ We need to understand if the process was done ethically.
- ❖ We need to understand how to formally present the information.

---

---

---

---

---

---

---

---

### Sampling

Allows researchers to select a sample of a large population that mirrors the characteristics of that population

- Representativeness

Two categories of Sampling

1. Probability Sampling
2. Non-probability Sampling

---

---

---

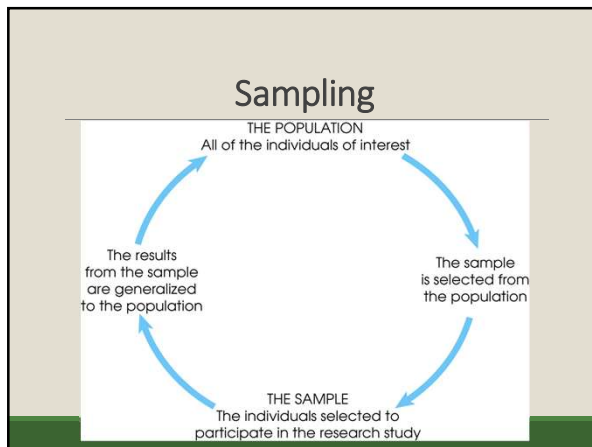
---

---

---

---

---



---

---

---

---

---

---

---

---

### Sampling

Target Pop. (N)

Sample (n)

- Effective Sampling produces a  $n$  which is representative of  $N$
- Note:  $n$  is only ever representative of the  $N$  it was drawn from, i.e. not necessarily the general population.

---

---

---

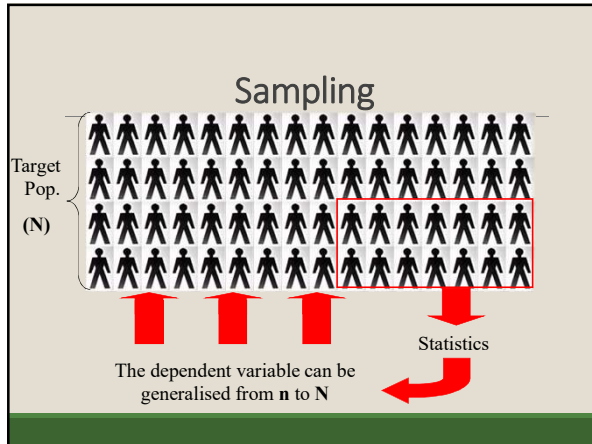
---

---

---

---

---



---

---

---

---

---

---

---

---

### Sampling

## Probability Sampling

Samples selected in accord with probability theory, typically involving some random selection method.

- ❖ Probability theory – branch of Mathematics that provides tools for sampling techniques to produce representative samples.

---

---

---

---

---

---

---

---

### Sampling

Probability Theory

THE MONTY HALL PROBLEM

LET'S MAKE A DEAL

---

---

---

---

---

---

---

---

Sampling

---

**Probability Sampling**

ESPEM – equal probability of selection method: The individuals in the population all have the same probability of being selected

- ❖ Simple Random Sampling
- ❖ Systematic Sampling
- ❖ Stratified Sampling

---

---

---

---

---

---

---

---

Sampling

---

**Non-Probability Sampling**

- ❖ Sampling not using probability theory
- ❖ Used where population parameters might not be available or when certain parameters are required for the study

---

---

---

---

---

---

---

---

Sampling

---

**Non-Probability Sampling**

- ❖ Purposive Sampling aka Judgmental Sampling
- ❖ Snowball Sampling
- ❖ Quota Sampling
- ❖ Convenience or haphazard sampling

---

---

---

---

---

---

---

---

## What Do We Need To Understand Research?

- ❖ We need to understand the data.
- ❖ We need to understand the type of research.
- ❖ We need to understand where the research comes from.
- ❖ We need to understand why the research was done
- ❖ We need to understand the research question
- ❖ We need to understand how the data was collected.
- ❖ **We need to understand the data analysis.**
- ❖ We need to understand if the process was done ethically.
- ❖ We need to understand how to formally present the information.

---

---

---

---

---

---

---

---

## Statistics

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write"

H.G. Wells



---

---

---

---

---

---

---

---

## What is Statistics?

- **Statistics:** The mathematics of collecting, organizing, and interpreting numerical data. Statistics especially concerns methods for figuring out population characteristics by looking at sample characteristics.

---

---

---

---

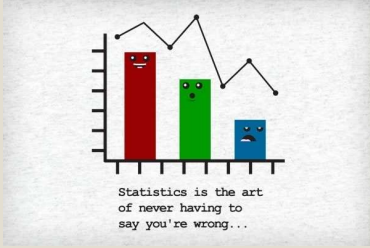
---

---

---

---

## Statistically speaking...



---

---

---

---

---

---

---

---

## Statistics

The most common and important question in statistics for research is this:

Are these two groups the same or are they different?

---

---

---

---

---

---

---

---

## Types of Statistics

### Research begins when there is a question

Different kinds of questions:

**Descriptive:**

How many men work at Health Sciences?  
How many hours a week do employees spend at their desks?

**Inferential:**

Does having a science degree help students learn statistical concepts?  
What risk factors most predict heart disease?

**Descriptive Statistics:** describe the relationship between variables.

**Inferential Statistics:** make inferences about the population, based on a random sample.

---

---

---

---

---

---

---

---

## Measures Of Central Tendency

How do you describe a group of numbers?

The goal of central tendency is to describe the entire distribution of variables with a single measurement.

**Mean:** The average. Add all the numbers up, divide them by the size of the sample. That's the mean.

**Median:** If you arrange the numbers in sequence, the median is the one in the middle (if there are two in the middle, add them & and divide that number by 2).

**Mode:** The most commonly observed number.

---

---

---

---

---

---

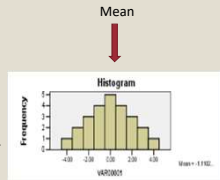
---

---

## Measures Of Central Tendency

"A statistician is someone who can have his head in an oven and his feet in ice, and say that on the average he feels great."

- The mean can be misleading because it can be greatly influenced by extreme scores (very high, or very low scores).
- For example, the average length of stay at a hospital could be greatly influenced by one patient that stays for 5 years.
- Extreme cases or values are called outliers.
- Sometimes the median may yield more information when your distribution contains outliers, or is skewed (not normally distributed).




---

---

---

---

---

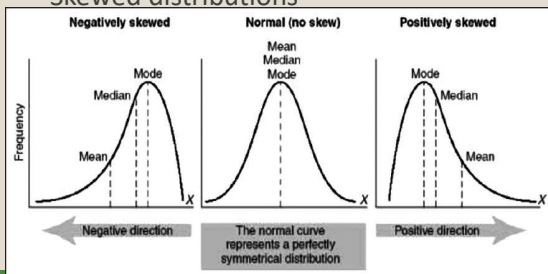
---

---

---

## Measures Of Central Tendency

### Skewed distributions




---

---

---

---

---

---

---

---

### Measures Of Central Tendency

Skewed distributions

Test #1 Results: Are these results “positively” or “negatively” skewed?

Mean	85.1%
medium	89.8%
Mode	94.9%
Frequency of mode	3

---

---

---

---

---

---

---

---

### Measures Of Central Tendency

**Range:** The difference between the highest and lowest number.

**Frequency of the mode:** How often the most common number appears. What does this tell us?

**Dispersion (also called variability, scatter, or spread) :** How spread out a data set is about the mean.

- ❖ Variance and Standard Deviation are the two main measures of dispersion within a data set.

---

---

---

---

---

---

---

---

### Formulas

- ❖ Average  $\bar{X} = \frac{\sum X}{N}$
- ❖ Variance  $s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$
- ❖ Standard Deviation  $SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$

---

---

---

---

---

---

---

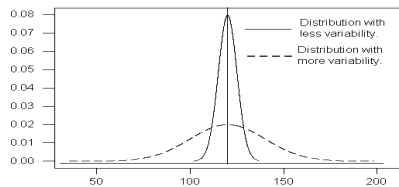
---



## Measures Of Central Tendency

**Variance:** The average of the squared differences from the mean.

- What is the variance within a group?
- What is the variance between groups?



---

---

---

---

---

---

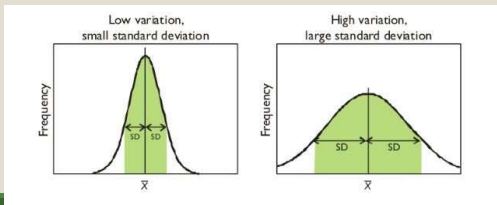
---

---

## Measures Of Central Tendency

**Standard Deviation:** Measure of how spread out the numbers are – square root of the Variance.

- The lower the variance the smaller the Std. Dev.
- The higher the variance the larger the Std. Dev.



---

---

---

---

---

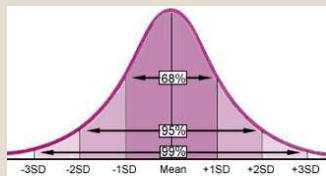
---

---

---

## Measures Of Central Tendency

### 68-95-99 Rule



68% of a sample will fall within one standard deviation

95% of a sample will fall within 2 standard deviations

99% of a sample will fall within 3 standard deviations

---

---

---

---

---

---

---

---

### Check Your Understanding: Question

The mean weight of stray dogs at a particular shelter is 70 lbs with a standard deviation of 2.5 lbs. With this information calculate the range of values for the first 3 Standard Deviations;

- 1<sup>st</sup> Standard Deviation = \_\_\_\_\_
- 2<sup>nd</sup> Standard Deviation = \_\_\_\_\_
- 3<sup>rd</sup> Standard Deviation = \_\_\_\_\_

---

---

---

---

---

---

---

---

### Check Your Understanding: Question

The mean weight of stray dogs at a particular shelter is 70 lbs with a standard deviation of 2.5 lbs.

1. What weight is 2 standard deviations below the mean?
2. What weight is 1 standard deviation above the mean?
3. The middle 68% of dogs weigh how much?

---

---

---

---

---

---

---

---

### Check Your Understanding: Question

The mean weight of stray dogs at a particular shelter is 70 lbs with a standard deviation of 2.5 lbs.

1. 2 standard deviations is  $2 * 2.5 = 5$  lbs. So if a dog is 2 standard deviations below the mean they weigh  $70 \text{ lbs} - 5 \text{ lbs} = 65 \text{ lbs}$
2. 1 standard deviation is 2.5 lbs, so a dog 1 standard deviation above the mean would weigh  $70 \text{ lbs} + 2.5 \text{ lbs} = 72.5 \text{ lbs}$ .
3. 1 standard deviation above (given in the answer to question 2) is 72.5 lbs; 1 standard deviation below is  $70 \text{ lbs} - 2.5 \text{ lbs}$  is 67.5 lbs. Therefore, 68% of dogs weigh between 67.5 and 72.5 lbs

---

---

---

---

---

---

---

---

### Measures Of Central Tendency

Basis for Comparison	Variance	Standard Deviation
<b>Meaning</b>	Variance is a numerical value that describes the variability of observations from its arithmetic mean.	Standard deviation is a measure of dispersion of observations within a data set.
<b>What is it?</b>	It is the average of squared deviations.	It is the root mean square deviation.
<b>Expressed in</b>	Squared units	Same units as the values in the set of data.
<b>Indicates</b>	How far individuals in a group are spread out.	How much observations of a data set differs from its mean.

---

---

---

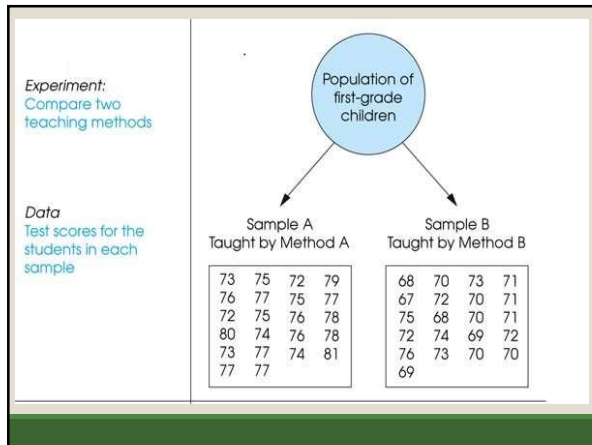
---

---

---

---

---




---

---

---

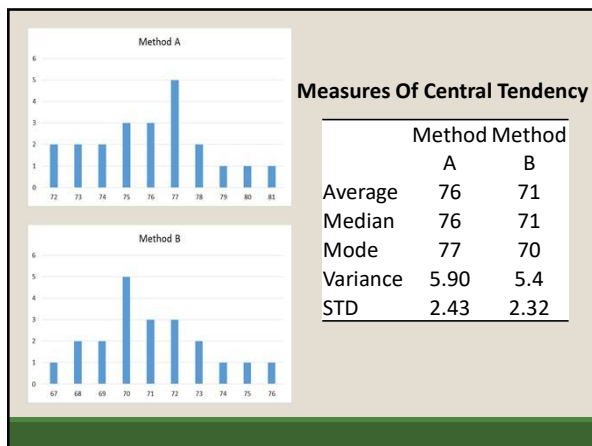
---

---

---

---

---




---

---

---

---

---

---

---

---

**Statistics Problem**

---

**Our Question:** The difference between teaching Method A and teaching Method B is 5 points. Is this difference significant or could it have occurred by chance?

**Step 1:** State the null and alternative hypotheses

**Step 2:** Calculate the value of the test statistic from your data.

**Step 3:** Develop a decision rule/rejection region.

---

---

---

---

---

---

---

---

**Statistics Problem**

---

Hypothesis Testing: is the art of testing if variation between two sample distributions can just be explained through random chance or not.

Secondary question: Are the differences between groups due to a systematic influence rather than chance significant or not?

- ❖ In order to statistically test this we must assume that all other factors that might contribute to differences are controlled

---

---

---

---

---

---

---

---

**Statistics Problem**

---

**Step 1:** State the null and alternative hypotheses

**Null Hypothesis:** There is no difference between the teaching methods. Both methods are equal.

**Alternative Hypothesis:** There is a difference between the teaching methods.

---

---

---

---

---

---

---

---

### Statistics Problem

---

**Step 2:** Choose the appropriate test statistic and calculate the value of the test statistic from your data.

- ❖ Chi-Square Test
- ❖ t-Test

---

---

---

---

---

---

---

---

### Statistics: Chi-Square Test

---

- ❖ Will test if two variables are independent of each other.
- ❖ Used to compare categorical variables.
- ❖ The result of a Chi-Square test is converted to a p-value in order to accept or reject the null.

---

---

---

---

---

---

---

---

### Statistics: Chi-Square Test

---

	Cat	Dog
Men	207	282
Women	231	242

- ❖ By doing the calculations, we come up with of P value = 0.043
- ❖ In this case  $p < 0.05$ , so this result is thought of as being "significant" meaning we think the variables are not independent.
- ❖ In other words, because  $0.043 < 0.05$  we think that Gender is linked to Pet Preference (Men and Women have different preferences for Cats and Dogs).

---

---

---

---

---

---

---

---

### Statistics: T-Test

- ❖ Widely used to compare the mean of two groups of samples.
- ❖ Used to evaluate whether the means of the two sets of data are statistically significantly different from each other.
- ❖ The result of the t-test is a 't' value; this value is then converted to a p-value
  - A large t-score tells you that the groups are different.
  - A small t-score tells you that the groups are similar.

---

---

---

---

---

---

---

---

### Statistics: T-Test

#### T-Test Types

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

1. **Independent samples t-test** which compares mean for two groups
2. **Paired sample t-test** which compares means from the same group at different times
3. **One sample t-test** which tests the mean of a single group against a known mean.

---

---

---

---

---

---

---

---

### Statistics: P-Value

- ❖ The P-Value or calculated probability, is the probability of finding the observed results when the **null hypothesis (H<sub>0</sub>)** of a study question is true.
- ❖ It is the value you use to determine if the difference between the means in your sample populations is significant.
- ❖ The P value is compared to the chosen significance level in order to accept or reject the null hypothesis

---

---

---

---

---

---

---

---

### Statistics: P-Value

$P > 0.10$	No evidence against the null hypothesis. The data appear to be consistent with the null hypothesis.
$0.05 < P < 0.10$	Weak evidence against the null hypothesis in favor of the alternative.
$0.01 < P < 0.05$	Moderate evidence against the null hypothesis in favor of the alternative.
$0.001 < P < 0.01$	Strong evidence against the null hypothesis in favor of the alternative.
$P < 0.001$	Very strong evidence against the null hypothesis in favor of the alternative.

---

---

---

---

---

---

---

---

### Statistics Problem

**Step 3:** Develop a decision rule/rejection region.

- ❖ The **significance level** is a measure of the statistical strength of the hypothesis test.
- ❖ Typically the 0.05 or the 0.01 level is used. Or put another way, 95% or 99% sure the results are not due to chance.
- ❖ For this example we will select 95%.

---

---

---

---

---

---

---

---

		Variable 1	Variable 2
	Mean	76	71
	Variance	5.9	5.4
t-Test: Two-Sample	Observations	22	21
Assuming equal	Hypothesized		
Variations	Mean Difference	0	
	df	41	
	P(T<=t) two-tail	0.000000230	
	t Critical two-tail	2.01954097	
Conclusion	Reject the Null Hypothesis because p-value is less than 0.05 or 5%.		
In English	There is enough evidence to conclude the difference between the teaching methods is significant.		

---

---

---

---

---

---

---

---

### Statistical vs. Clinical Significance

This study says massage improves sleep compared to the control group!

That's awesome! How much longer did the massage group sleep?

... 47 seconds per night.



---

---

---

---

---

---

---

---

### Causation vs. Correlation

- ❖ Correlation: used to test relationships between quantitative variables or categorical variables.
- ❖ Measure of how things are related.

---

---

---

---

---

---

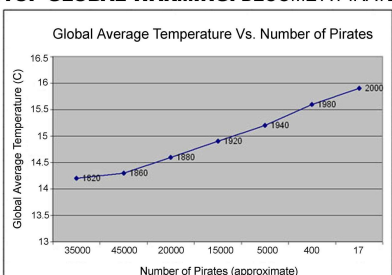
---

---

### Causation vs. Correlation

**STOP GLOBAL WARMING: BECOME A PIRATE**

Global Average Temperature Vs. Number of Pirates



Year	Number of Pirates (approximate)	Global Average Temperature (C)
1820	35000	14.0
1860	45000	14.2
1890	20000	14.5
1920	15000	14.8
1940	5000	15.2
1980	400	15.6
2000	17	16.0

---

---

---

---

---

---

---

---



### Causation vs. Correlation

**The Correlation Coefficient.**  
A correlation coefficient is a way to put a value to the relationship.

The figure shows three scatter plots side-by-side. The first plot, labeled 'Positive Correlation' with  $r = 0.4$ , shows a cluster of points with a blue arrow pointing upwards and to the right. The second plot, labeled 'No correlation' with  $r = 0$ , shows a random scatter of points with a horizontal blue line. The third plot, labeled 'Negative' with  $r = -0.4$ , shows a cluster of points with a blue arrow pointing downwards and to the right.

---

---

---

---

---

---

---

---

### Causation vs. Correlation

- value of between -1 and 1. A "0" means there is **no relationship** between the variables
- -1 or 1 means that there is a **perfect negative or positive correlation** (negative or positive correlation here refers to the type of graph the relationship will produce).

---

---

---

---


---

---

---

---

### Causation

 **causation**  
*noun* cau·sa·tion \kò-'zā-shən\

**Simple Definition of causation:** the act or process of causing something to happen or exist. The relationship between an event or situation and a possible reason or cause

---

---

---

---

---

---

---

---

### Causation

Three main criteria for causal relationship

- Correlation
- Time order
- Non-spurious

---

---

---

---

---

---

---

---

### Causation

Necessary and Sufficient Causes

- Necessary Cause – a condition that must be present for the effect to follow
- Sufficient Cause – a condition that, if it is present, guarantees the effect in question

---

---

---

---

---

---

---

---

### Assignment #5: Literature Review

❖ Drawing on the competencies you have developed prepare a research proposal investigating an important issue in massage therapy. The Research Proposal will include background information on the issue being investigated, a literature review of the issue, the methodology being employed, and what the possible outcome of the research might be. The Research Proposal will be double-spaced and typewritten and at least 1500 words in length.

❖ Email it to [da.bailey2019@gmail.com](mailto:da.bailey2019@gmail.com)

---

---

---

---

---

---

---

---