

# Basic Statistical Analysis and Interpretation

Concepts and Variables  
Structure of Statistical Analysis  
Statistical Packages/Software

<http://orcid.org/0000-0003-0006-8841>

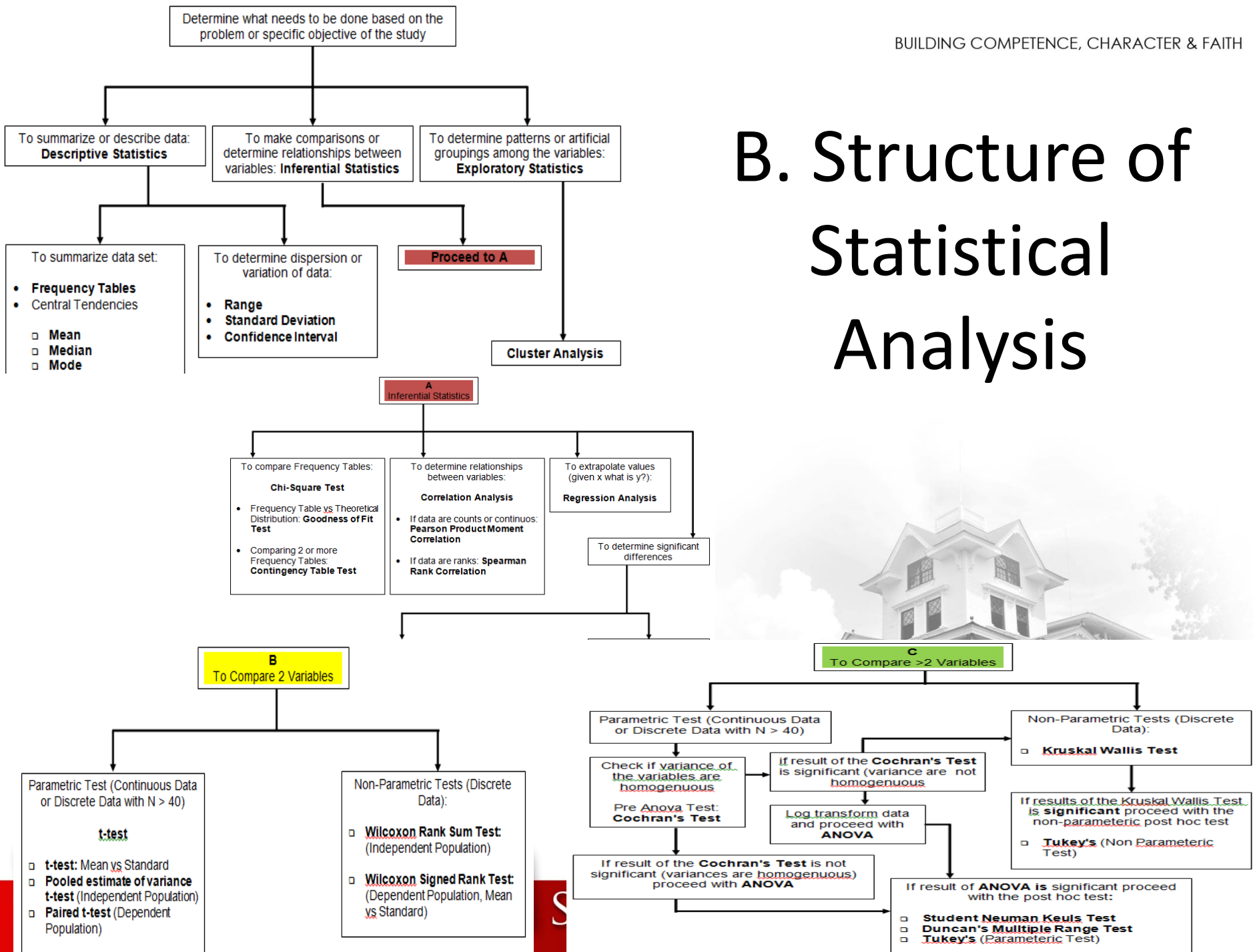
**Dave E. Marcial**  
[www.davemarcial.net](http://www.davemarcial.net)  
[demarcial@su.edu.ph](mailto:demarcial@su.edu.ph)



SILLIMAN UNIVERSITY

[www.su.edu.ph](http://www.su.edu.ph)

# B. Structure of Statistical Analysis



# Structure of Statistical Analysis

Determine what needs to be done based on the problem or specific objective of the study

```
graph TD; A[Determine what needs to be done based on the problem or specific objective of the study] --> B[Blue Segment]; A --> C[Red Segment]; A --> D[Green Segment]; C --> E[Black Square]; E --> F[Purple Square]; E --> G[Light Blue Rectangle]; E --> H[Orange Rectangle];
```

# Structure of Statistical Analysis

Determine what needs to be done based on the problem or specific objective of the study

To summarize or describe data:  
**Descriptive Statistics**

To summarize data set:

- **Frequency Tables**
- **Central Tendencies**
  - **Mean**
  - **Median**
  - **Mode**

## Descriptive Statistics

- **Summarizing Data**
  - **Frequency** (For discrete data sets usually but there are also instances wherein continuous data sets are summarized into frequency tables)
- **Central Tendencies**
  - Mean
  - Median
  - Mode



# Structure of Statistical Analysis

Determine what needs to be done based on the problem or specific objective of the study

To summarize or describe data:  
**Descriptive Statistics**

To summarize data set:

- **Frequency Tables**
- Central Tendencies
  - **Mean**
  - **Median**
  - **Mode**

## Descriptive Statistics

- **Summarizing Data**
  - Measures of Dispersion (variations among the data)
    - **Range** (minimum and maximum values)
    - **Standard Deviation** (measure of precision: “how close are your measurements”)
    - **Confidence Interval** (measure of accuracy: “how close are you to the true value”)

# Structure of Statistical Analysis

Determine what needs to be done based on the problem or specific objective of the study

To summarize or describe data:  
**Descriptive Statistics**

To summarize data set:

- **Frequency Tables**
- Central Tendencies
  - **Mean**
  - **Median**
  - **Mode**

To determine dispersion or variation of data:

- **Range**
- **Standard Deviation**
- **Confidence Interval**

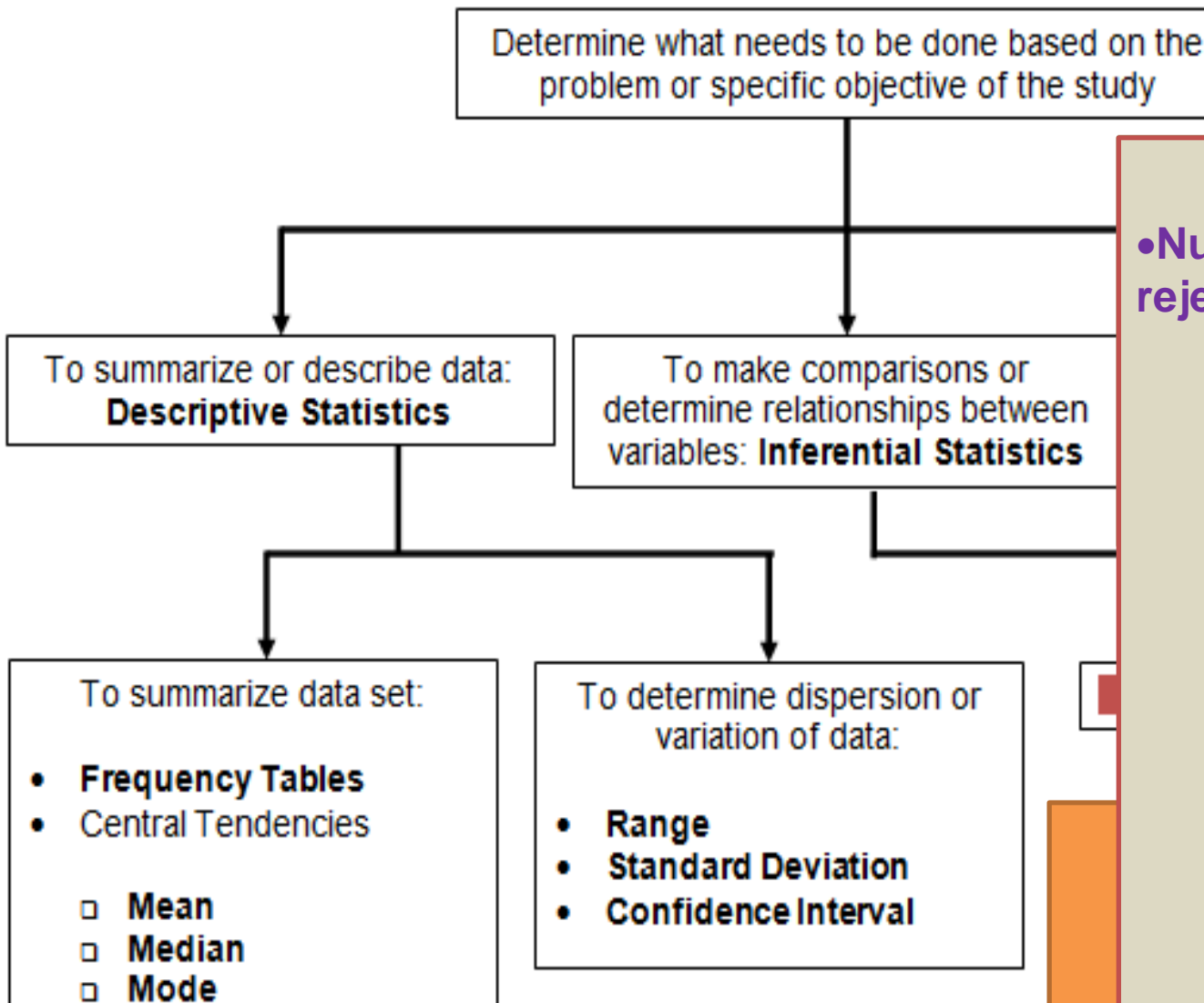
## Inferential Statistics

• **Significant relationships are determined by rejecting the null hypothesis and accepting the alternative hypothesis**

$\Rightarrow H_0$ : Variable A = Variable B

$\Rightarrow H_1$ : Variable A = Variable B

# Structure of Statistical Analysis



## Inferential Statistics

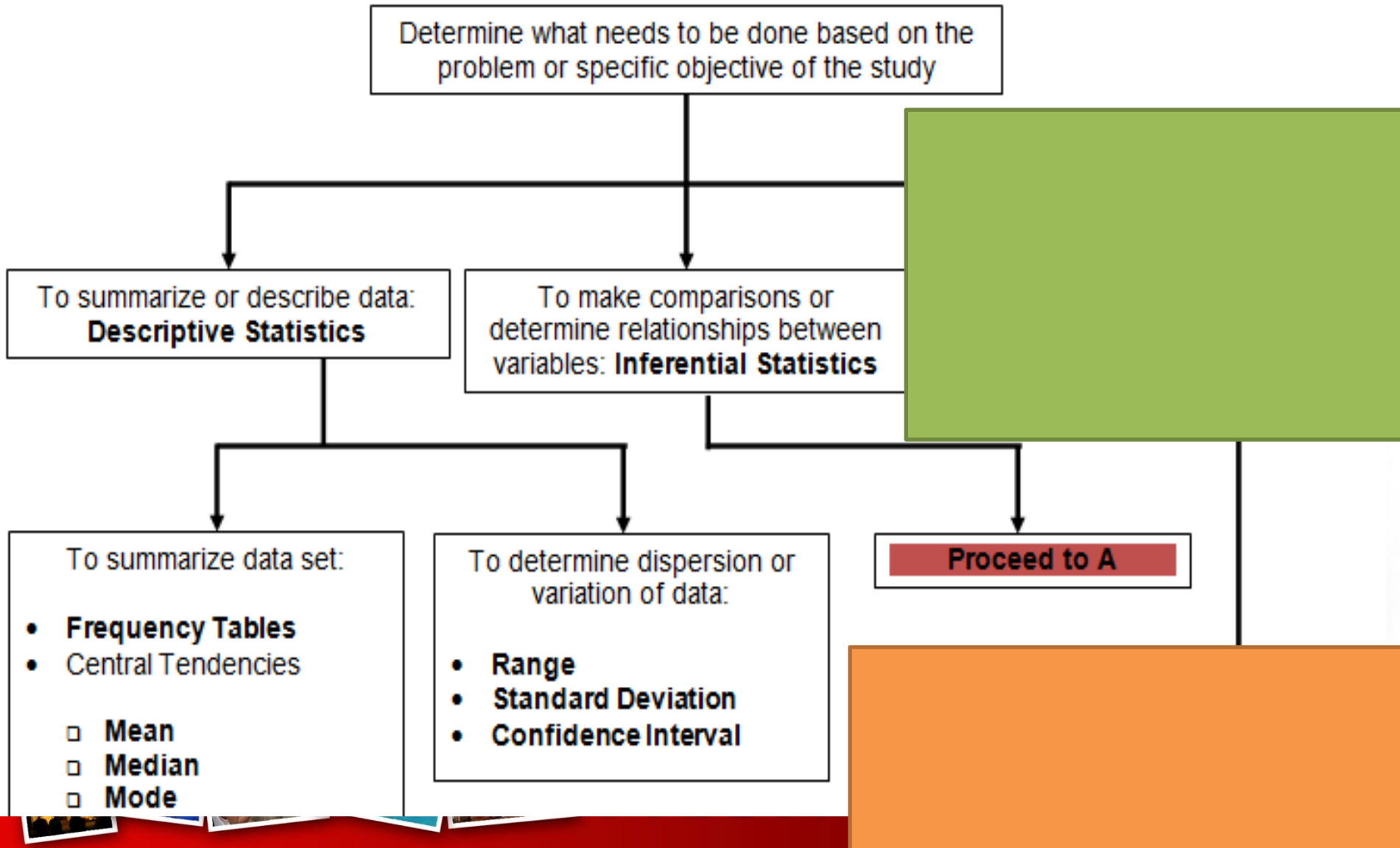
• **Null hypothesis are rejected if:**

⇒ computed statistics is greater than the table (critical) value at  $\alpha$  (for manual computation)

⇒ probability value is less than  $\alpha$  (computer generated)

$\alpha$  is the confidence level (usually set at 95% or 0.05)

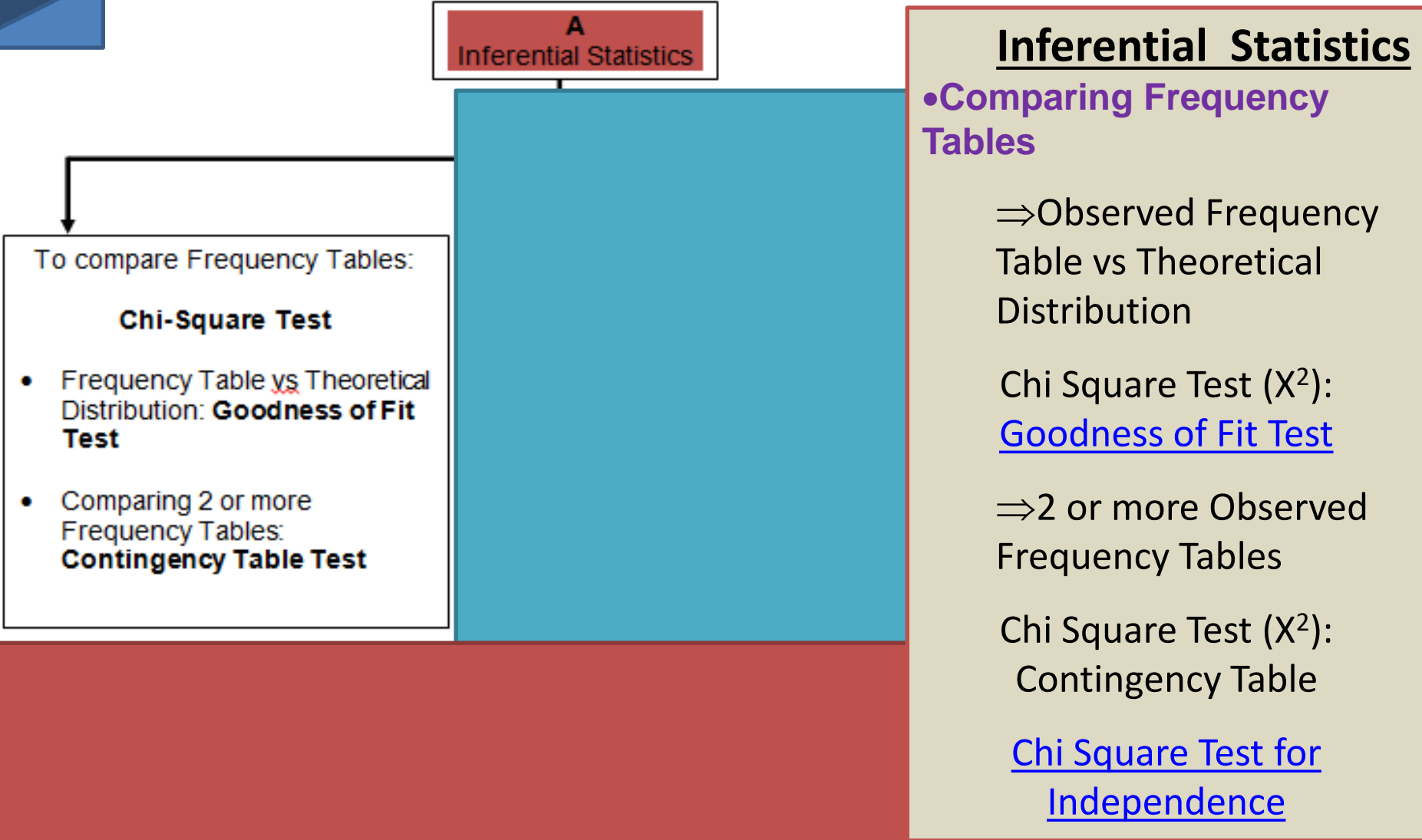
# Structure of Statistical Analysis







# Structure of Statistical Analysis



## A Inferential Statistics

To compare Frequency Tables:

### Chi-Square Test

- Frequency Table ~~vs~~ Theoretical Distribution: **Goodness of Fit Test**
- Comparing 2 or more Frequency Tables: **Contingency Table Test**

## Inferential Statistics

### •Comparing Frequency Tables

⇒ Observed Frequency Table vs Theoretical Distribution

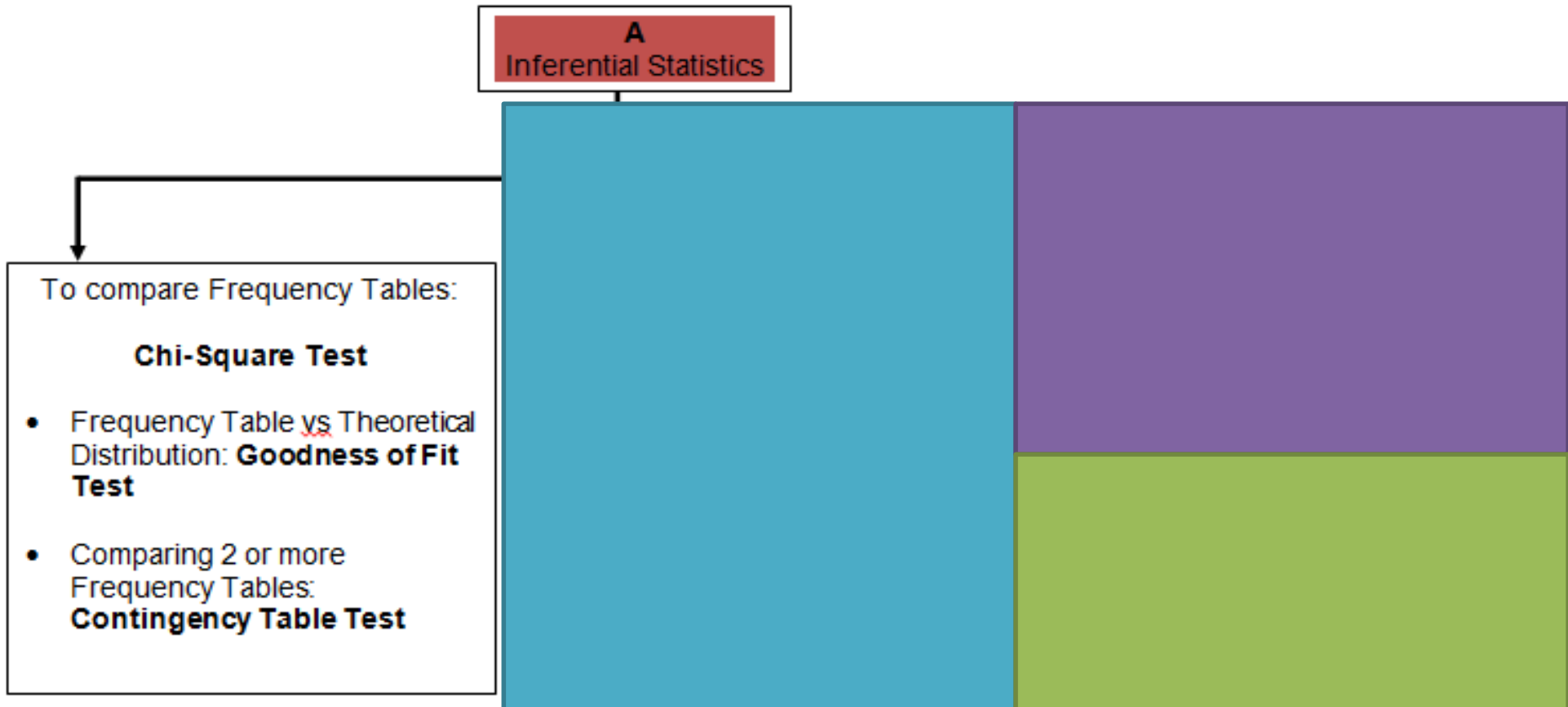
Chi Square Test ( $X^2$ ):  
[Goodness of Fit Test](#)

⇒ 2 or more Observed Frequency Tables

Chi Square Test ( $X^2$ ):  
Contingency Table

[Chi Square Test for Independence](#)

# Structure of Statistical Analysis



# Structure of Statistical Analysis

## A Inferential Statistics

Remember that a Pearson product-moment correlation is an index of the degree of linear relationship between two variables. That is, the correlation gives an indication of how closely the points in a scatter plot cluster around a straight line. But the relationship between two variables is not always linear.

## Inferential Statistics

### •Relationship between two variables

⇒Continuous Data

Pearson Product Moment Correlation (r)

Scatter plot

⇒Rank Data Set

Spearman Rank Correlation (r)

If  $r$  approaches 1 : the relationship is directly proportional

If  $r$  approaches 0 : there is no relationship

If  $r$  approaches -1: the relationship is inversely

### Test

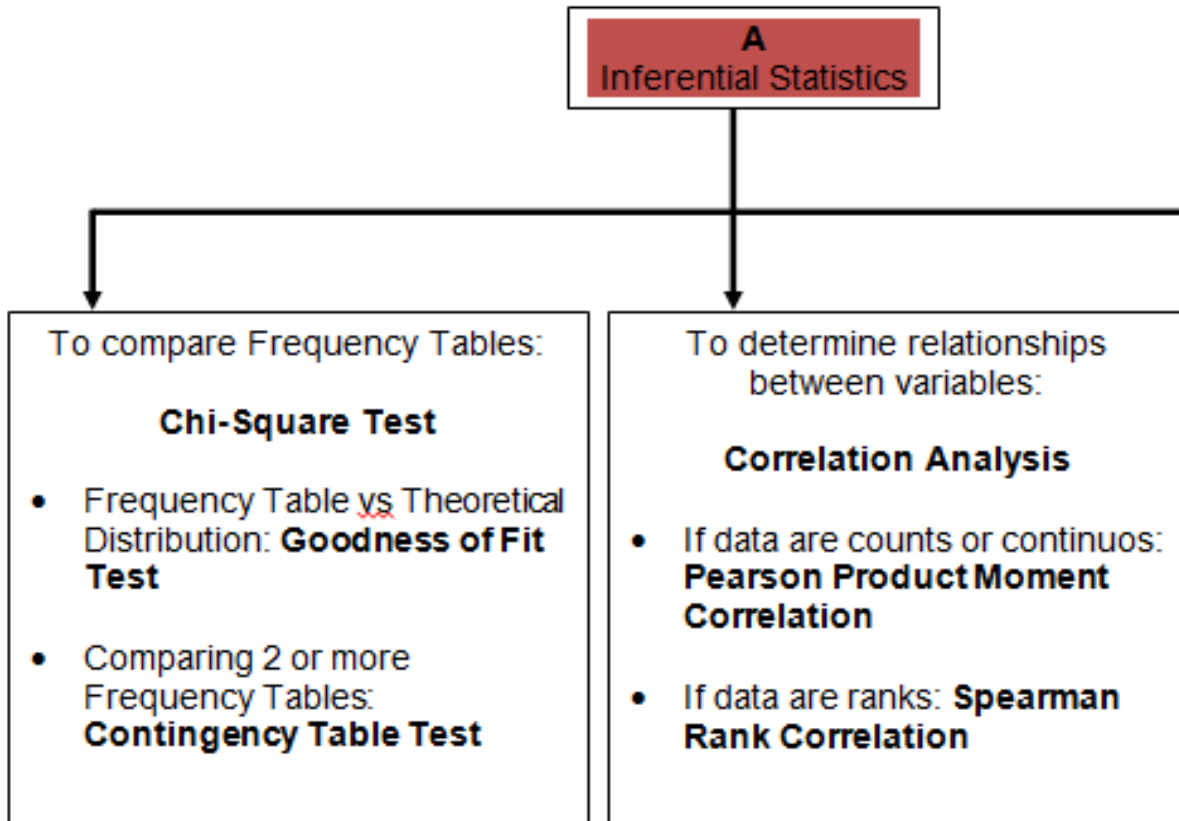
- Comparing 2 or more Frequency Tables:  
**Contingency Table Test**

### Pearson Product Moment Correlation

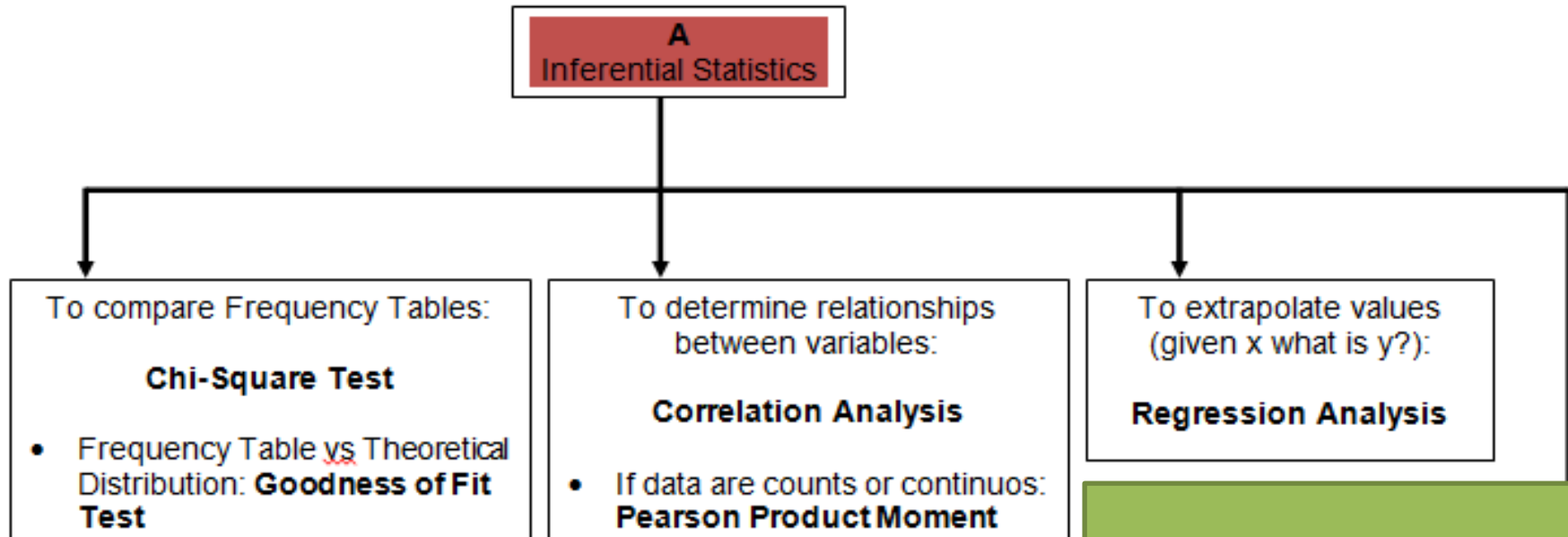
- If data are ranks: **Spearman Rank Correlation**

The Spearman rank-order correlation is used when both variables are at least ordinal scales of measurement, but one is not sure that both would qualify as interval or ratio scales of measurement.

# Structure of Statistical Analysis



# Structure of Statistical Analysis



## Inferential Statistics

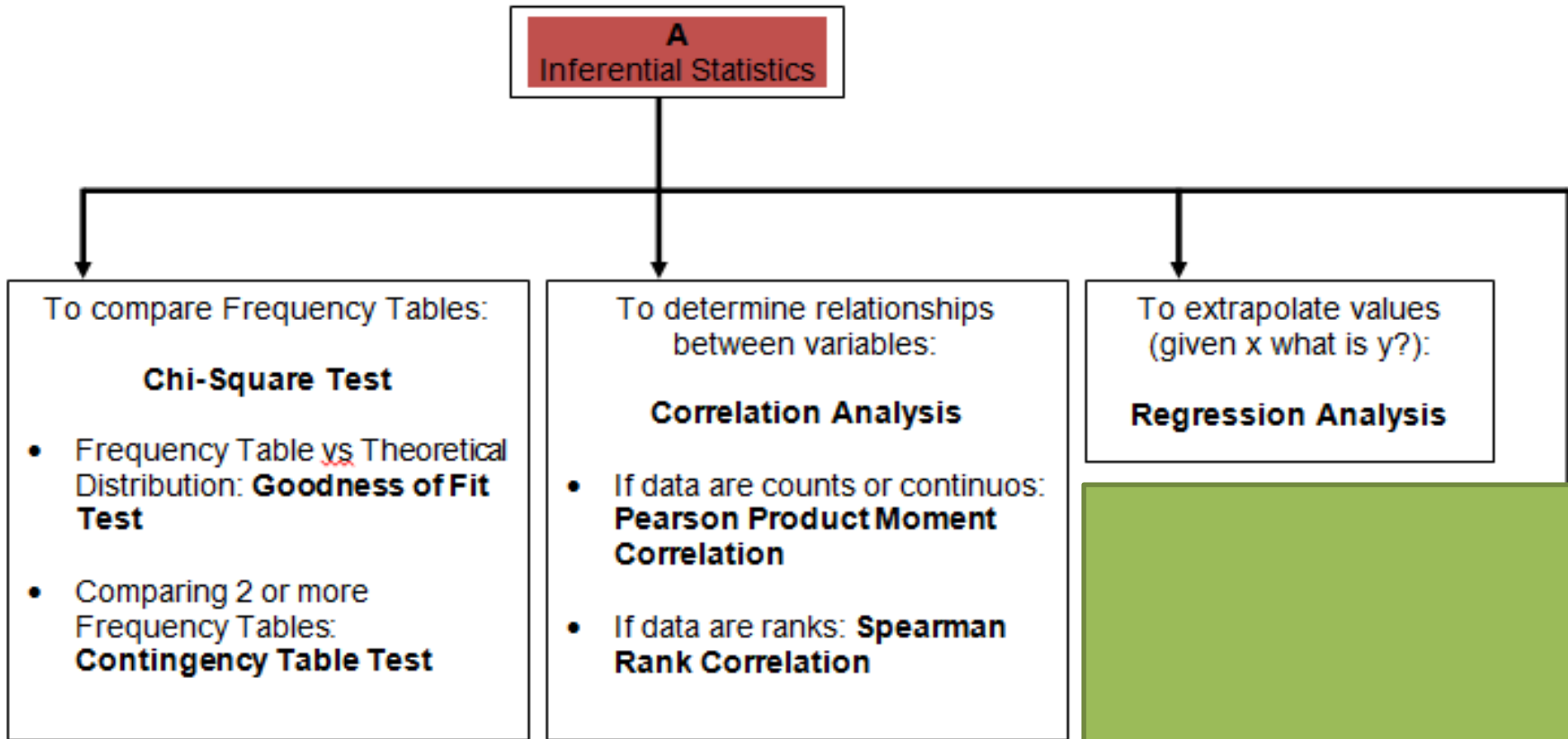
- To predict values for Y variable given a value for X variable

### Regression analysis

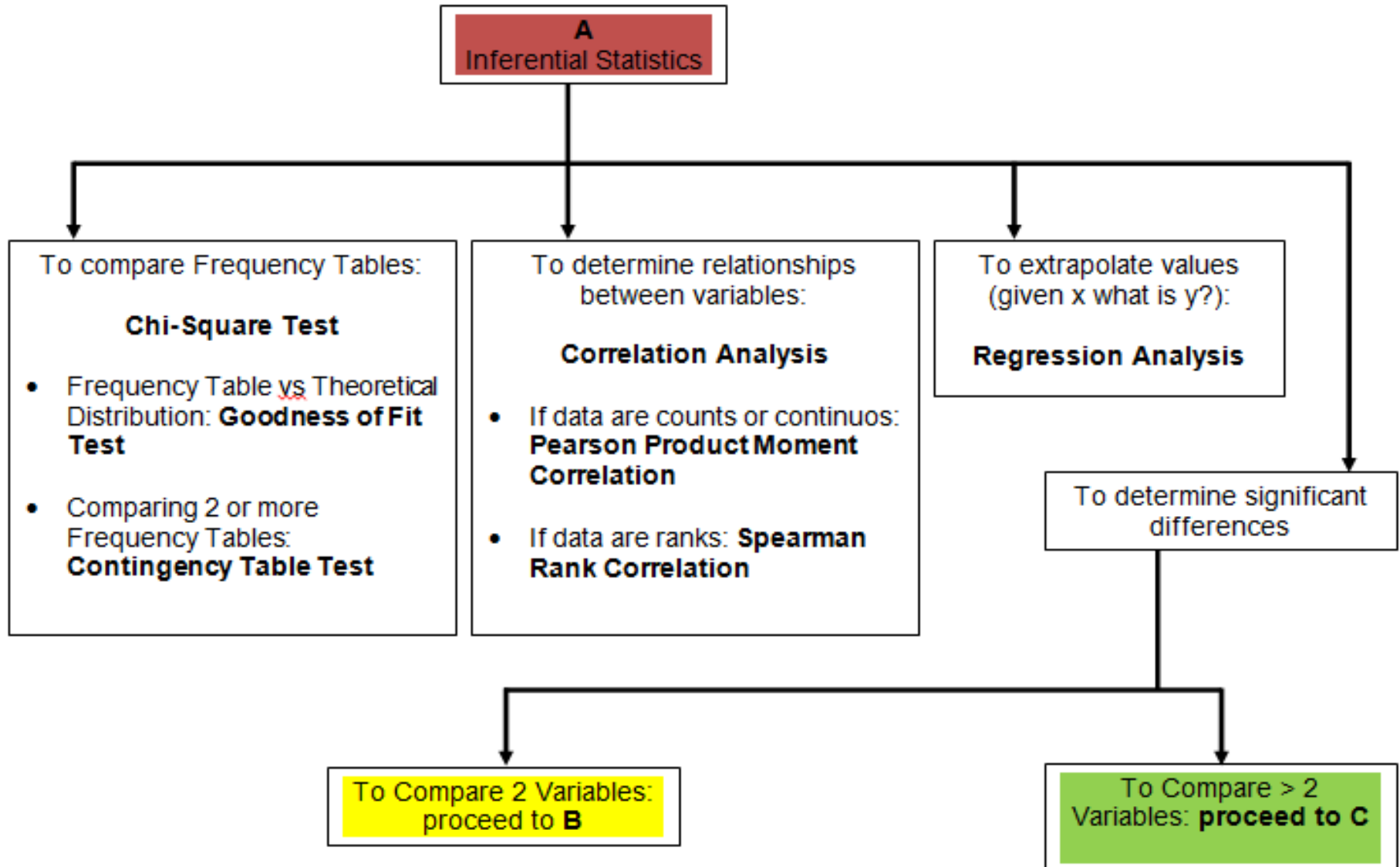
For a simple linear regression ( $y = a + bX$ ), the analysis will determine the a and b values in the equation

⇒ In principle, the regression analysis can only predict values with the range of the values of the samples used in the correlation.

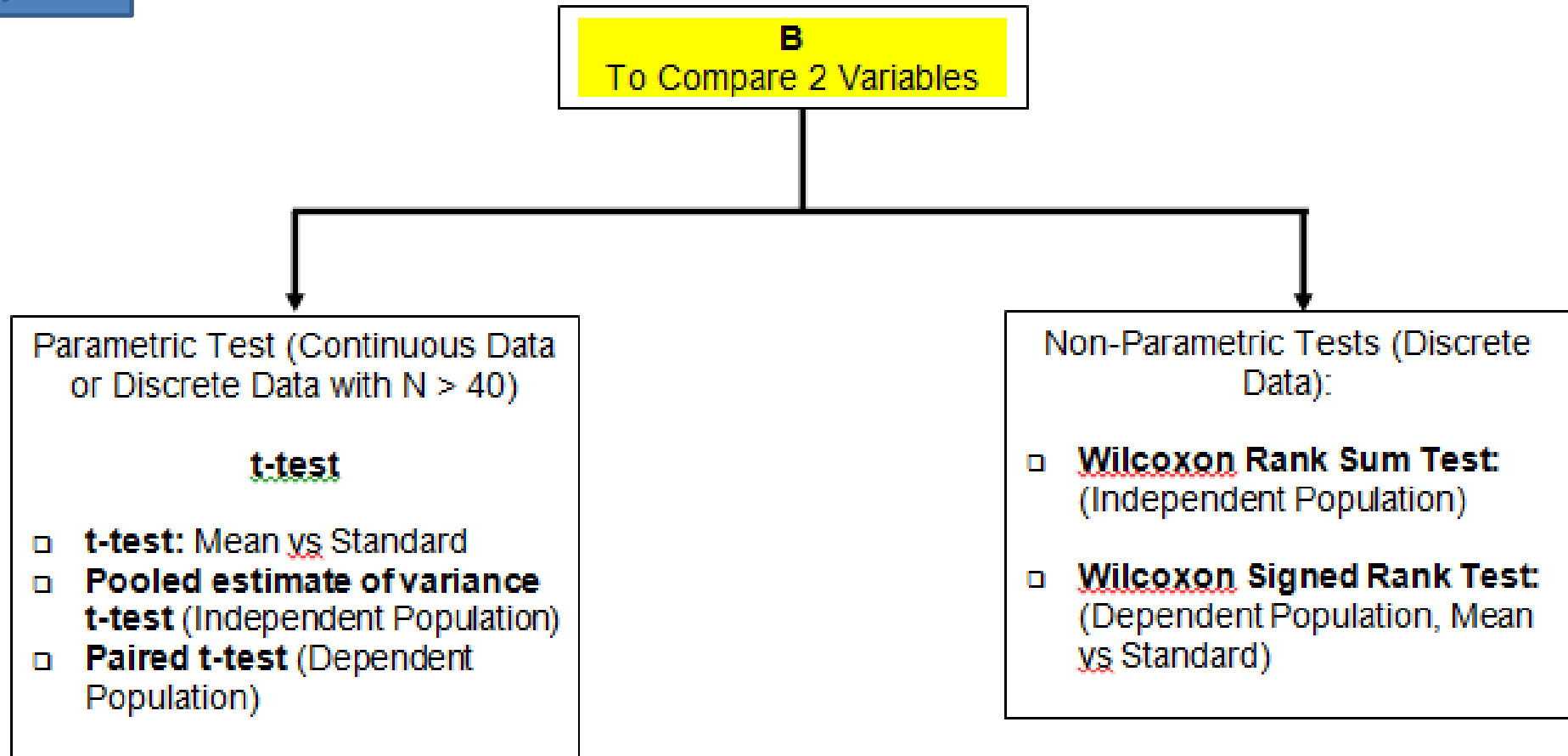
# Structure of Statistical Analysis



# Structure of Statistical Analysis



# Structure of Statistical Analysis

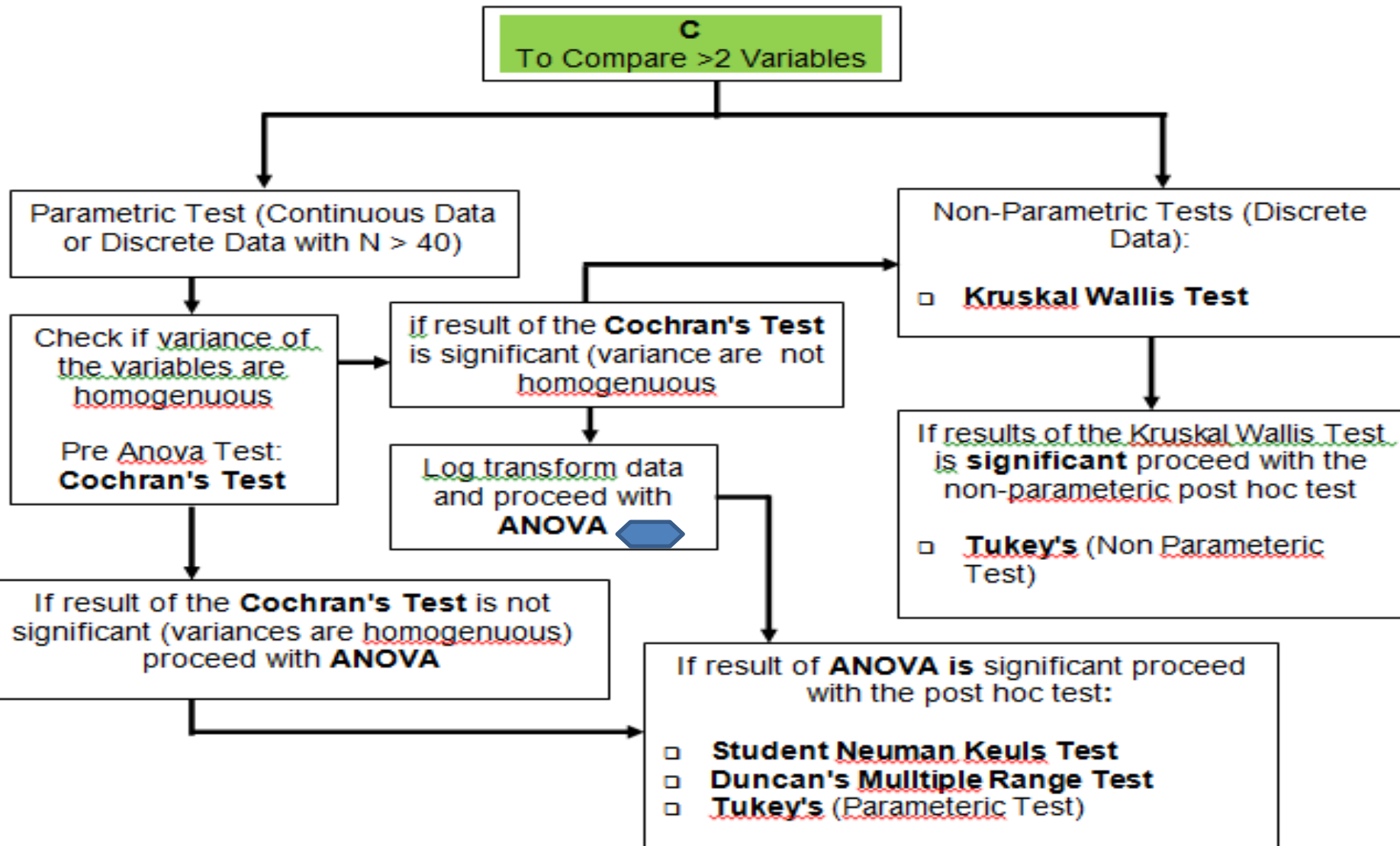


Independent - Correlated

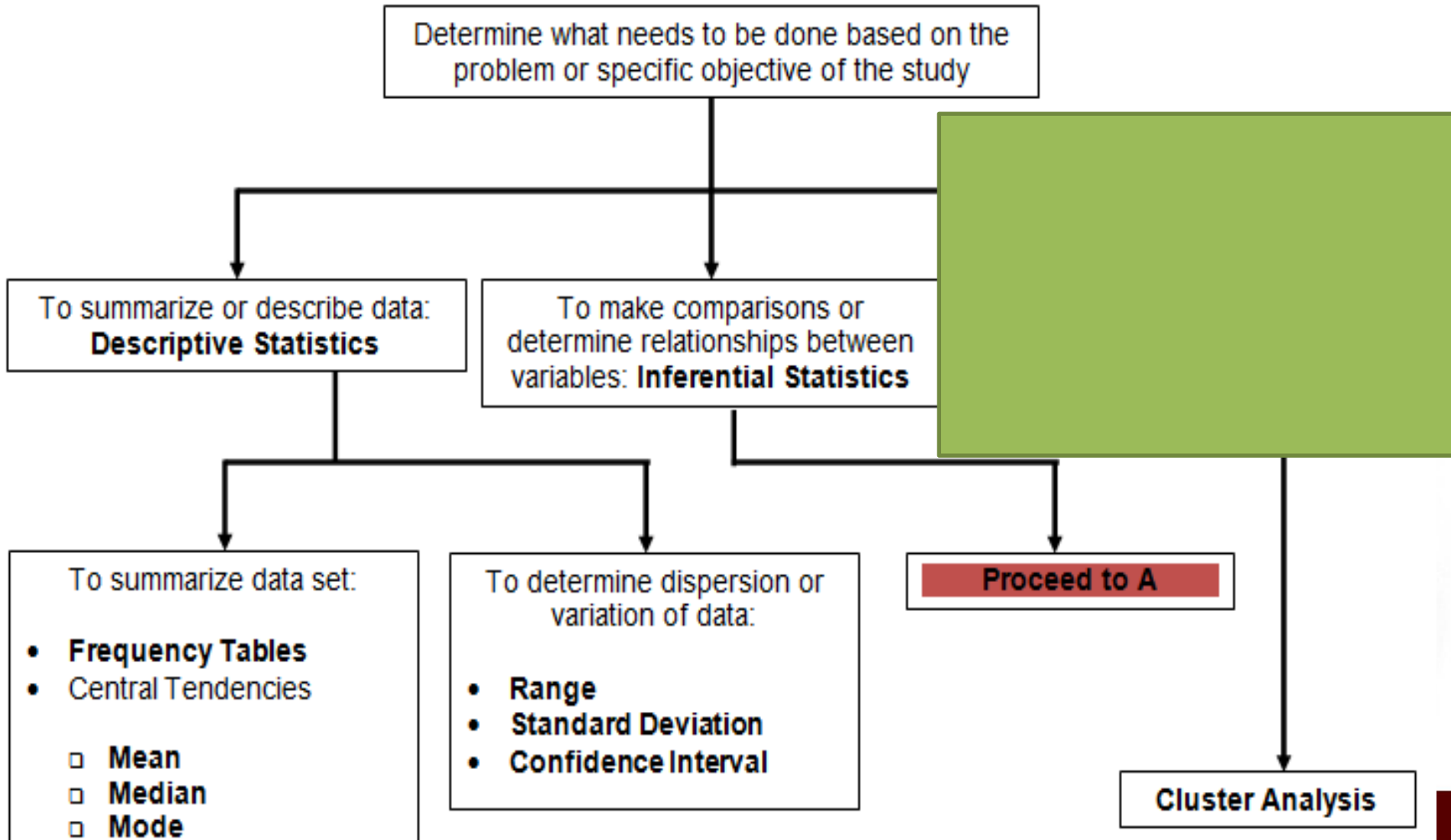




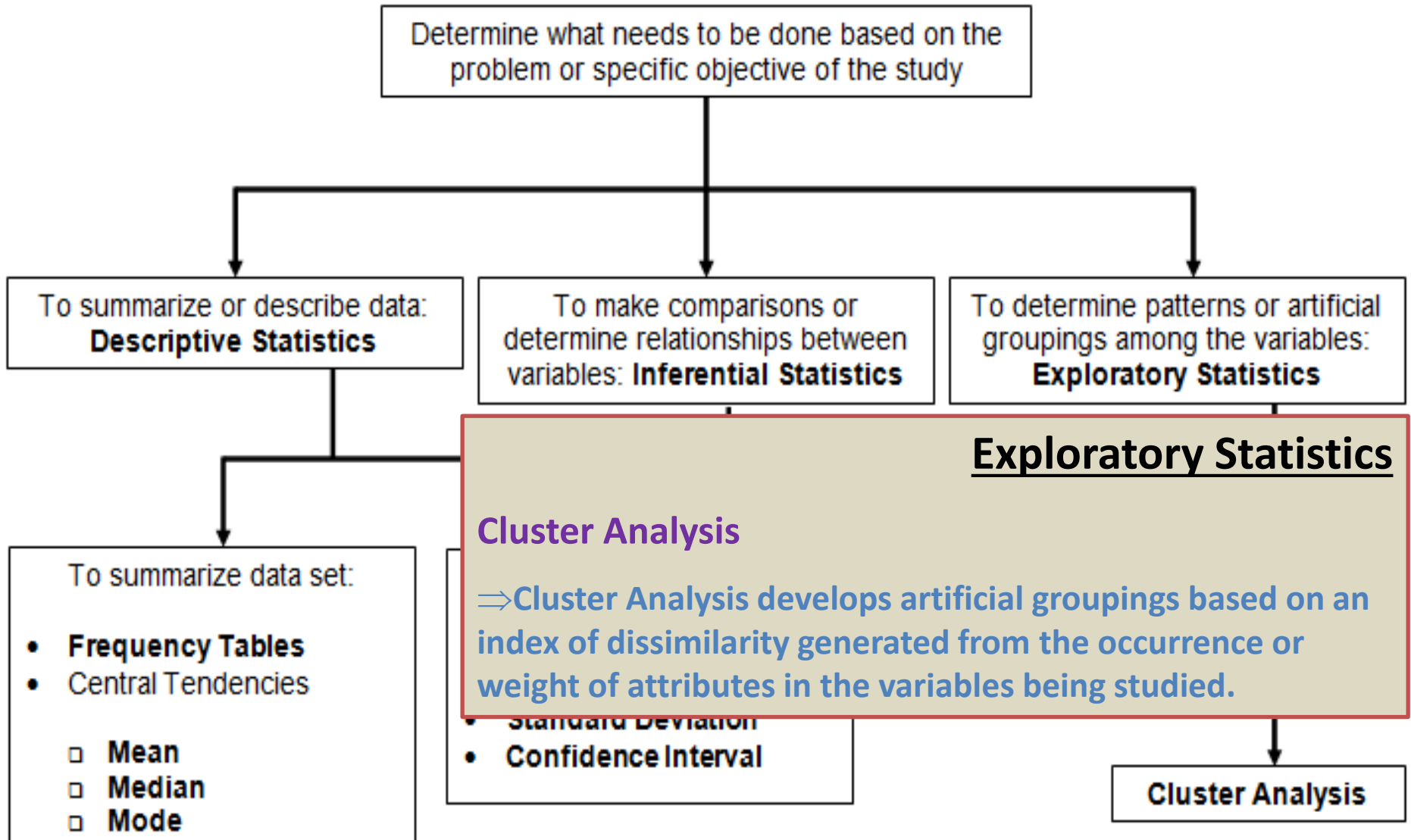
# Structure of Statistical Analysis



# Structure of Statistical Analysis



# Structure of Statistical Analysis



**Table 2. Fish larvae diversity in the three sampling areas**

Families	Northeast Monsoon		
	Bacong	Masaplod	Zamboangita
Acanthuridae	-	-	-
Apogonidae	-	+	+
Balistidae	-	-	-
Blenniidae	-	-	-
Carangidae *	-	+	+
Chaetodontidae	-	+	-
Clupeidae *	-	-	+
Engraulidae *	-	-	-
Gerreidae	-	+	+
Gobiesocidae	-	-	-
Haemulidae	-	+	-
Hoplichthyidae	-	-	+
Labridae	+	+	-
Lethrinidae	-	+	+
Lutjanidae	-	+	+
Mullidae	-	+	+
Nemipteridae	-	+	-
Pempheridae	-	+	-
Pomacanthidae	-	-	+
Pomacentridae	-	+	+
Priacanthidae	-	-	+
Scaridae	-	-	+
Schindleriidae	-	+	+
Serranidae	+	+	+
Sphyraenidae *	+	-	+
Teraponidae	+	+	+
<b>Total Families<sup>#</sup></b>	<b>3</b>	<b>14</b>	<b>13</b>

Legend: (+) present; (-) absent; (\*) pelagic fishes, <sup>#</sup> = reef fishes only

# Statistical Analysis

be done based on the  
objective of the study

comparisons or  
relationships between  
**Inferential Statistics**

To determine patterns or artificial  
groupings among the variables:  
**Exploratory Statistics**

## Exploratory Statistics

Analysis develops artificial groupings based on an  
arbitrary generated from the occurrence or  
rates in the variables being studied.

**Cluster Analysis**

Table 2. Fish larvae diversity in the three sampling areas

Northeast Monsoon			
Family	BACONG	MASAP	ZAMBO
Acanthuridae			
Apogonidae			
Balistidae			
Blenniidae			
Carangidae			
Chaetodontidae			
Clupeidae			
Engraulidae			
Gerreidae			
Gobiesocidae			
Haemulidae			
Hoplichthys			
Labridae			
Lethrinidae			
Lutjanidae			
Mullidae			
Nemipteridae			
Pempheridae			
Pomacanthidae			
Pomacentridae			
Priacanthidae			
Scaridae			
Schindleridae			
Serranidae			
Sphyraenidae			
Teraponidae	+	+	+
Total Families <sup>#</sup>	3	14	13

Legend: (+) present; (-) absent; (\*) pelagic fishes, <sup>#</sup> reef fishes only

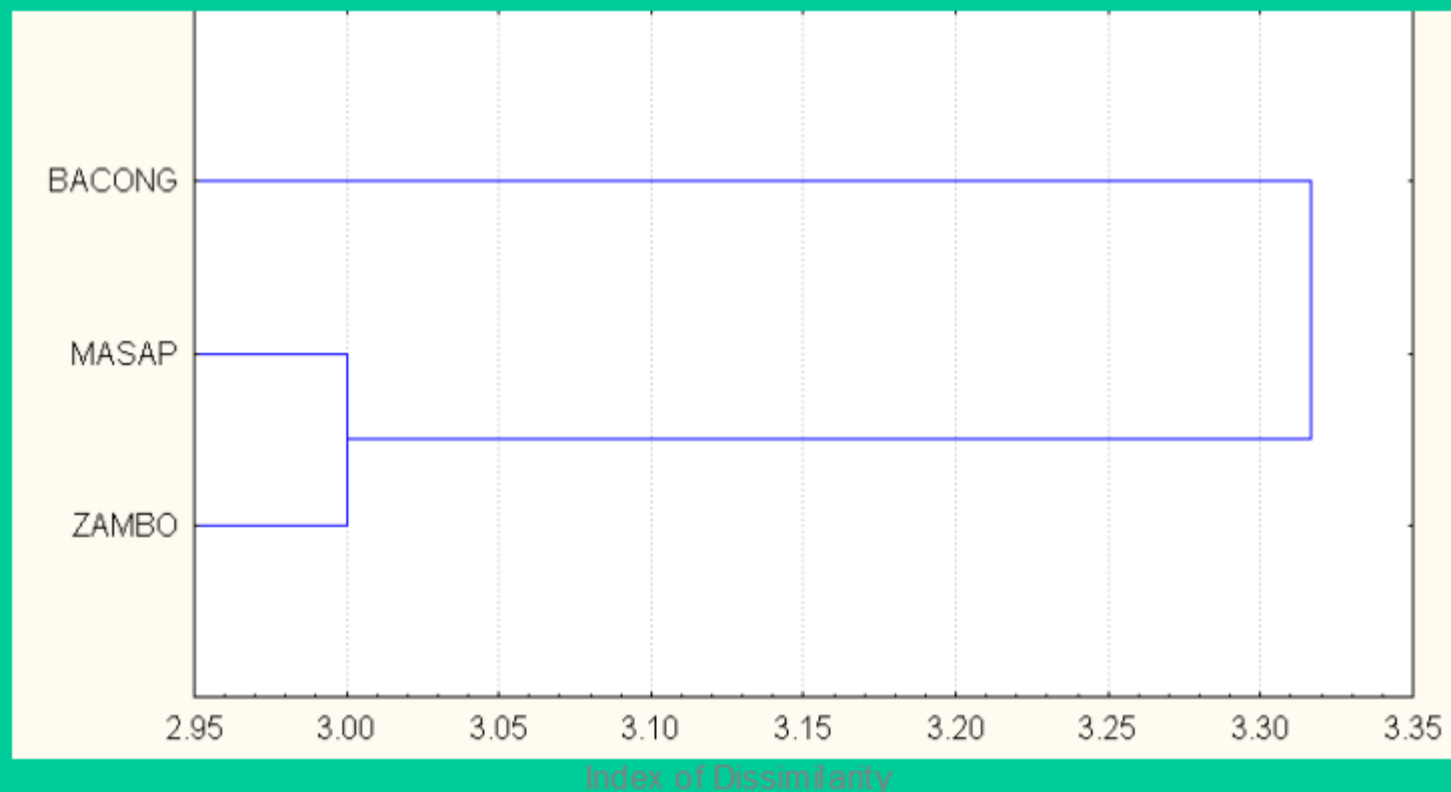


Figure 9. Cluster analysis for diversity of fish families encountered in the three sampling sites during the Northeast monsoon. *Bacong*–Bacong area; *Masap*–marine reserve in Masaplod Norte; *Zambo*–Zamboangita area

Cluster Analysis

# Structure of Statistical Analysis

