Introduction

- Historical Background:

Biostatics is the application of statics to a wide range of topics in biology, it is base in the design of biological experiments, is mostly use in the camp of medicine and agriculture; the collection, summarization, and analysis of data from those experiments; and the interpretation of, and inference from, the results.

Biostatistics is routinely used to drive biology experiments. Data is gathered and analyzed before, during, and after a biology experiment, with the intent to come to some form of logical conclusion about what might not be exactly empirical results. On the other hand, a biostatistics experiment can be entirely mathematical

- Static Value:

Average: According to the Thesaurus dictionary, average is: "Something, as a type, number, quantity, or degree, that represents a midpoint between extremes on a scale of valuation." The average is taken by the sum of all data and then dividing the result by the number of data. It is used for having a middle and sole number in a set of data that can represent the whole set.

Median: In statistics, the median is the numeric value that separates the higher half of a sample, population of probability distribution, from the lower half. When the data is even, the median can give us a good feel about the rank of the set of values.

Mode: In statistics, the mode is the most frequently occurring value of a data set. The mode gives us the value that is most likely to happen. However, the mode makes more sense with nominal values (non -numerical values)

Range: The statistic range (R) is the length of the smallest interval which contains all the data. It is the difference between the highest and the lowest values in the set and it provides an indication of statistical dispersion.

Variance: It is a measure that shows how spread out a distribution is. It describes how far the values of a set lie from the mean, in other words, it is a measure of dispersion. It is the mean of the squared deviation of a variable from its expected value or mean.

Standard Deviation: In statistics, the standard deviation is the square root of the variance (i.e. the square root of the square of the mean error). It is a measure of dispersion, for it shows how much variation there is from the average. When the standard deviation is low, it indicates that the data are close to the mean, while a high standard deviation indicates that the data is spread out a large range of values.

Standard Error: According to the Mosby's Dental Dictionary, standard error is "A measure or estimate of the sampling errors affecting a statistic; a measure of the amount the statistic may be expected to differ by chance from the true value of the statistic." The standard error is the result of the samples' standard deviation divided the square root of the number elements in the sample.

T-Test: According to the Encyclopædia Britannica, the student's t-test is "in statistics, a method of testing hypotheses about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown."

Data Colection:

The process of the data obtaining, was very simple, the objective was to make two data collections, one of some human anatomical part of two different populations, and the last one was of two different plants population that share the same environment.

The group decide to measure the large of the middle finger, we take the example of the female and the male population of one same grade, to analyze the relation between this two populations. .

Here is the data table of middle finger (right hand) size (PD The variable is in cm):

Bio-Statistics: Comparation of the middle finger length		
N	Male Population	Female Population
1	10,5	9,8
2	10,2	9
3	10	8,5
4	10	8,5
5	11,3	8,6
6	10,3	9
7	11,2	9
8	10	9
9	10,2	8,1
10	10	9,4
11	10,5	10,3
12	10,5	9,6
13	10,5	9,6
14	9,7	9
15	9	9

Mean	10,26	9,093333333
Mode	10,5	9
Range	2,2	2,2
Variance	0,312571429	0,324952381
Standard Deviation	0,559080878	0,570045946

T Value	4,60411E-06	
Correlation	0,131336392	

The second data collection was of the large of the leaves, of two different populations of plants, the Eugenia leave, and the Feijo leave.

Here is the data table of the sizes of the two types of plants:

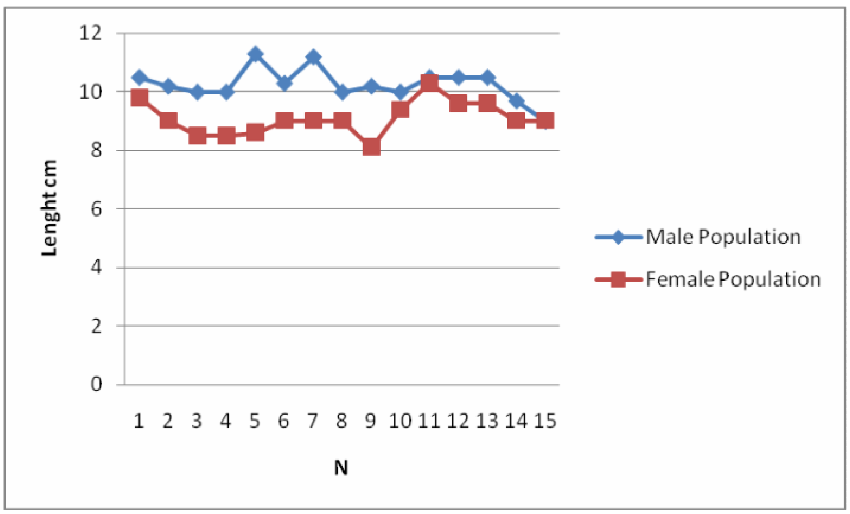
Bio-Statistics: Comparation of Eugenia &Feijo Leaves length		
N	Eugenia Population	Feijoa Population
1	6	7,3
2	6,1	6,2
3	5,6	5,9
4	5,7	8,9
5	6	7,1
6	4,6	6,5
7	6,8	6,3
8	5,4	5,4
9	5,2	6,9
10	6,6	7,3
11	6,6	6,3
12	5,1	6,6
13	4,4	6,7
14	4,6	7,3
15	5,6	6,8

Mean	5,62	6,766666667
Mode	5.7	7.3
Range	2,4	3,5
Variance	0,564571429	0,643809524
Standard Deviation	0,751379683	0,80237742
T Value	0,000379804	
Correlation	-0,011847694	

Data processing:

The first data to process is the human one:

- Graphic:



- Data Process:

Male-Mean: $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = 10.26$

Female-Mean: $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = 9.092$

Male-Mode: $= 10.5$

Female-Mode: = 9

Female/Male-Range: $R = x_{max} - X_{min}$

$R = 2,2$

Male-Variance: $\sigma^2 = \frac{\sum(\bar{x} - x)^2}{n-1}$

$\sigma^2 = 0.312$

Female-Variance: $\sigma^2 = \frac{\sum(\bar{x} - x)^2}{n-1}$

$\sigma^2 = 0.324$

Male-Standard Deviation: $\sigma = \sqrt{\frac{\sum(\bar{x} - x)^2}{n-1}}$

$\sigma = 0,5590$

Female-Standard Deviation: $\sigma = \sqrt{\frac{\sum(\bar{x} - x)^2}{n-1}}$

$\sigma = 0,5700$

- Final Static Value:

T test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{var_1}{n_1} + \frac{var_2}{n_2}}}$$

$$t = 4{,}6041E-06$$

Correlation coefficient:

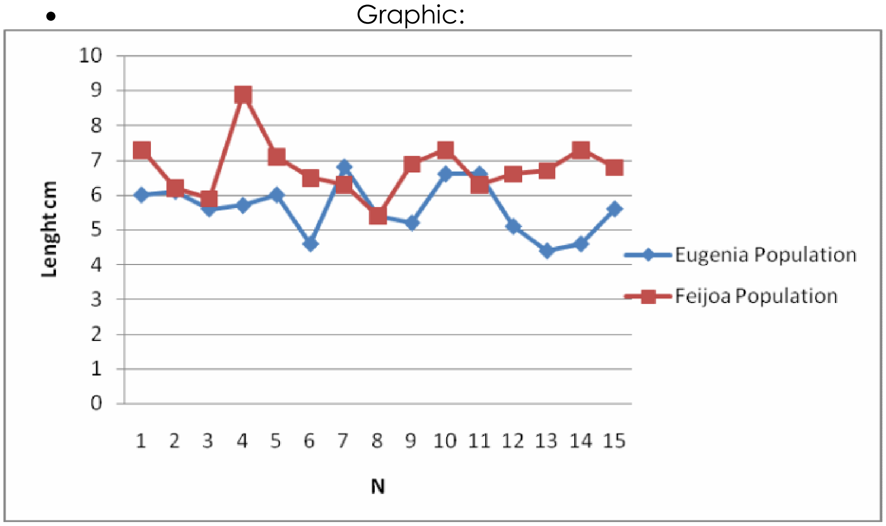
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2] \times [(\sum y^2) - (\sum y)^2]}}$$

$$r = 0{,}133692$$

Data processing:

The second data to process is the leaves size:

- Graphic:



- Data Process:

Eugenia-Mean: $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = 5.62$

Feijoa-Mean: $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = 6.766$

Eugenia -Mode: =5.7

Feijoa -Mode: = 7.3

Eugenia-Range: $R = x_{max} - X_{min}$

$R = 2{,}4$

Feijoa-Range: $R = x_{max} - X_{min}$

$R = 3.5$

Eugenia -Variance: $\sigma^2 = \frac{\sum(\bar{x} - x)^2}{n-1}$

$\sigma^2 = 0564$

Feijoa -Variance: $\sigma^2 = \frac{\sum(\bar{x} - x)^2}{n-1}$

$\sigma^2 = 0.643$

Eugenia -Standard Deviation: $\sigma = \sqrt{\frac{\sum(\bar{x} - x)^2}{n-1}}$

$\sigma = 0.7513$

Feijoa -Standard Deviation: $\sigma = \sqrt{\frac{\sum(\bar{x} - x)^2}{n-1}}$

$\sigma = 0.8023$

- Final Static Value:

T test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{var_1}{n_1} + \frac{var_2}{n_2}}}$$

$$t = 0{,}0003798 \quad 4$$

Correlation coefficient:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2] \times [(\sum y^2) - (\sum y)^2]}}$$

$$r = -0{,}011847 \quad 94$$

Discussion:

- What does the value mean?

  In the first case, the one of the middle fingers, the correlation was of 0.13, that means that this to populations are comparable, the correlation level is near to 0, by the other side the T value notes, that there is more than a 50% of probability, that this compatibility is not coincidence, so we can conclude that this two population are similar, and comparable, this is a result expected because, this two populations are of the same specie and age.

  In the second case, the one of the length of the leaves, the correlation is of -0.0118, they are not exactly the same, but they have a low level of compatibility, but the T value expose that it is a huge percent of probability that his values are causality. And is because actually, this two populations are two different species, so they are not comparable at all.

- Is Biostatistics important for science

The biostatics give the possibility to any camp of science, to make deeper and exact studies of population or phenomena, the biostatics give the chance to find similarities, and to make comparations of populations, is a necessary tool to analyze nature. The statistics part involves the accumulation, tracking, analysis, and application of data. Biostatistics is the use of statistics procedures and analysis in the study and practice of biology. As such, it has many real-world and scientific applications.

- What was bad in the practice

During the practice, maybe was committed a mistake of interpretation in the T value table, so the analysis is not 100% trustable .

- Who it can be improved

For the next time, the solution is jut in consulting and discus this type of question during class, beside the rest of the laboratory was take in a satisfactory way.