STATISTICAL METHODS IN FISHERIES

Statistics is the study of the collection, organization, and interpretation of data. It deals with all aspects of this
including the planning of data collection in term of design and surveys and experiments. Statisticians improve the quality of data with the design of experiment and survey sampling .statistics also provides tools for prediction and forecasting using data and statistical models.

Statistical methods can be used to summarize or describe a collection of data, this is called descripture statistics.

This is useful in research, when communicating the results of experiments. In addition, patterns in the data may
be modeled in a way that accounts for randomness and uncertainty in the observations, and are then used to draw inference about the process or population being studied, this is called inferential statistics. Inference is a
vital element of scientific advance, since it provides a prediction (based in data) for where a theory logically leads.

Examples of descriptive statistical tools include:

1) Mean, median, mode, pie-chart, bar-chart, histogram standard deviation, standard error, confidence limit, etc

Meanwhile inferential statistical tools include:

- Correlation analysis
- Regression analysis
- Variance tests

In this course, emphasis will be on the applicability of some of these tools in data analysis and presentation in fisheries management and Aquaculture experiments.

SAMPLING STRATEGIES

- Random sampling e.g simple and systematic random samplings
- Proportion sampling
- Stratified sampling

DATA ANALYSIS AND PRESENTATION

Mean values, variance and confidence limits.

Assuming x = 2, 5, 6, ------- N

` X = Σx

N

If the data are collected with frequency, X becomes

Σfx

Σf

Where X = observed data n = number of observed data f = frequency of occurrence

The variance therefore for the ungrouped data in calculated as

S_2 = Σ x – x

N – 1

And standard deviation S is equal to

Σf x – x $_2$

√ Σf

Some authors in applied sciences present the mean of observed data in the form of:

X + S

This in the real sense is just a measure of dispersion from the mean value, but nowadays in biostatistics, the mean
value of any set of data should be presented to show the uncertainty around them using confidence limit at 95%

fractile. This is more scientific and allows for comparison with any other set of data. This parameter is therefore
called mean standard error, MSE

.

$$MSE = X + t_{n-1} * \frac{S}{\sqrt{n}}$$

Where X = mean value

t_{n-1} = fractile at 95% CL

S = standard deviation

n = number of observation

For example data collected from a feeding experiment in the laboratory in better represented as follows:

	Treatments			
Replicate	A	B	C	D
1				
2				
3				
4				
X	XA±SE	XB±BE	XC±BE	XD±BE

Ordinary linear regression Analysis

This method is used when we want to describe the variation of one quantity as a linear function of another quantity e.g the body depth of a fish as a function of its total length. The theory requires that the quantity of the
horizontal axis (independent variable) is measured with absolute precision. The method is often applied, however,
when the requirement is violated the effect of the inaccuracy of the values of the independent variable, the slope
of the line becomes flatter i.e. close to zero

$$Y = b * X$$

It illustrates that as X increases, Y increases. The line passes through the origin when we allow a deviation from a
proportionality between X and Y by introducing second parameter 'a' the model changes to

$$Y_{(i)} = a + b * X_{(i)}$$

Here a is the intercept on the Y- axis

Parameters a and b can be estimated graphically or by using least square method.

Variance about the regression line is

$$S^2 = \frac{1}{n-2} * \sum_{I=i}^{n} [Y_{(i)} a - b * X_{(i)}]^2$$

There are n– 2 degree of freedom

$$b = \frac{\sum_{i=I}^{n} X_{(1)} * Y_{(i)} - \sum_{i=I}^{n} X_{(1)} * \sum_{i=i}^{n} Y_{(1)}}{n(\sum_{i=I}^{n} x^2) - (\sum_{i=1}^{n} x)^2}$$

$$a = Y - X * b$$

Correlation coefficient, r, is a measure of the linear association between two quantities, both of which are subject
to random variable.

$$r = \frac{SXY}{SX * SY}$$

Where SXY is the covariance which is equal to:

1

n-1 * [ΣX $_{(i)}$*ΣY(i) – 1 *ΣX $_{(i)}$ *ΣY$_{(i)}$]

n

SX = ΣX$^2_{(i)}$ – 1 *[Σ X $_{(i)}$]^2

n

SY = ΣY$^2_{(i)}$ 1 * [ΣY$_{(i)}$]^2

n

the range of r is :- 1.0 ≤ r ≤ 1.0

Comparison of means

Students will be exposed to the methods used to analysis means comparison.

It should be recalled from STS201 methods used to compare pair of means in which one is known and the other is
unknown. This method shall not be dealt with here. These are referred to as t-test in STS201. Here we shall deal
with t- test that is used to compare two unknown means (2- tailed t- test)

S = S^2

$_1$ + S^2

$_2$

√ n_1 n_2

Where S_1 and S_2 are the standard deviations of the two means to be compared ($x_1$and x_2)

t = x $_1$– x_2

S

Here error degree of freedom is equal to $n_1 + n_2 – 2$

Where n_1 and n_2 are numbers of observations in the two samples of data for comparison.

Comparison for more than two means

When comparing more than two mean, it is better to use F- test rather than using several t-tests to compare the pairs
of means. This will then lead us to what is called experimental design.

Experimental design: this is the method of arranging treatments in order that their effects may be meaningfully tested.

Before choosing any of the designs, one must answer the following questions satisfactorily:

1. How many factors are involves?
2. What is the nature of the experimental materials? Is there any reason to suspect heterogeneity?

TYPES OF DESIGN

There various types of designs, but two are more relevant in aquaculture and fisheries management research. These
include:

a. Complete randomization design

b. Complete randomization block design

Layouts of the Designs

CRD: 4x3 designs

$T2R_1$ $T1R_2$ $T3R_2$ $T2R_3$

$T4R_1$ $T1R_2$ $T2R_2$ $T4R_2$

$T3R_1$ $T2R_1$ $T4R_3$ $T3R_3$

CRBD: 4X3 design

T_3 T_1 T_4 T_2

T_2 T_1 T_3 T_4

T_4 T_3 T_2 T_1

Analysis of variance (ANOVA) is then used to compare the means all together. When the means are found to be significantly different at any alpha- risk, then a follow - up tests are used to detect or declare which pair(s) of the means are actually different. Commonly used are Duncan Multiple Range Test and Least significant difference (LSD)