Central Limit Theorem and Its Applications

Chanoknath Sutanapong★ & Louangrath, P.I. ★★

About the author

★Chanoknath Sutanapong is an independent researcher. She may be reached by email at: Chanoknath.sutanapong@gmail.com

★★ Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT

The purpose of this paper is to explain the central limit theorem and its application in research. Two concepts are constant companions in statistics: Central Limit theorem and distribution. The central limit theorem states that the arithmetic mean of sufficiently large number of iterations of independently random variable is the expected value of the iterations, and it is normally distributed with the mean equal to the expected value. Distribution is the probability of occurrence of a certain value within a defined range of values. The distribution type that describes the central limit theorem is the normal distribution curve. A normal distribution curve describes the probability distribution of continuous data. A normal distribution curve has the following properties: (i) it is symmetric around the point where $x = \mu$; (ii) unimodal; (iii) it has two inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$; (iv) it is log-concave in shape; and (v) it is infinitely differentiable.

Keywords: central limit theorem, distribution

CITATION:

Sutanapong, C. and Louangrath, P.I. (2015). "Central Limit Theorem and Its Applications– Part 3 of 4." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 1, No. 3: pp. 30-41. (Jul. – Sep. 2015).

1.0 INTRODUCTION

1.1 Central limit theorem

In this paper, we *shall* review basic statistics with the single purpose of gaining a better understanding of statistics so that we can carry our research in any field of social science. In this attempt we must begin with the normal distribution. The normal distribution is derived from the probability density of a continuous distribution between dependent variable *X* and independent variable *Y*. "Normal distribution" is a character of distribution that takes a shape of a normal curve. A normal curve is a bell shape curve with perfect symmetry bowing downward from the top of the bell to the right and left at equal rate of decline. If the relationship resulted from a single *variate* (determinant) in *X*, it is called a Univariate Normal Distribution. If the relationship is expressed from two variates of *X*, that is x_1 and x_2 matching to one resulting value of *Y*, a probability density resulted is called Bivariate Normal Distribution. Finally, if the relationship resulted from more than two *x* variates, it is called Multivariate Normal Distribution. In all three cases, the curve obeys a normal shape, that is, a symmetrical bell shape curve. In addition, this paper also covers other types of distributions: *chi-square, t,* and *F*. All these distribution serves one purpose: testing for normality of the density of distribution fitting into a perfect bell. A normal distribution curve has the following properties: (i) it is symmetric around the point where $x = \mu$; (ii) unimodal; (iii) it has two inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$; (iv) it is log-concave in shape (Patel et al., 1996); and (v) it is infinitely differentiable (Fan, 1991; pp. 1257-1272).

1.2 Random variable and probability distribution

Random means happening by chance. When we speak of randomness, we must also consider the concept of probability distribution. Two definitions are necessary. "Probability" is defined as a strong likelihood of an event to happen. "Distribution" is the dispersion of the events that happened. Let assume that *x* is a random variable and a function $Y = f(x)$ is a probability density of *x*. The probability of *x* is defined as: $\Pr(x = x_j) = f(x_j)$.

The probability function in (1) has the following characteristics or properties:

(1) $0 \le \Pr(x \le x_j) \le 1$;

(2) $\Pr(-\infty \le x \le \infty) = 1$; and

(3) For $x_k > x_j$, $\Pr(x_j \le x \le x_k) = \Pr(x \le x_k) - \Pr(x \le x_j)$.

Property (1) means that the probability of *x* is between zero and 1. Property (2) means that the probability of *x* occurring between negative and positive infinity will also add up to 1. It means that the probability is express in percentage. When expressed the percentage in decimal points, it is less than zero. However, the sum of all points will add up to 1. The third property (3) if *k* is larger than *j* , the probability of *x* lies between *j* and *k* which is the same saying the probability of *x* equal to or less than *k* minus the probability of *x* less than *j*.

While we use $f(x)$ as the function for the probability density for *x*, we shall denote the *distribution function* for random variable X by $F(x)$. This probability distribution function is written as:

$$\Pr(x \le x_j) = F(x_j) = \int_{-\infty}^{xj} f(x)dx \tag{1}$$

Equation (1) shows the *cumulative probability density* of a random variable *x*. For all non-negative numbers, the entire range of the probability in the function expressed in (1) is summed to 1.00. Probability is express in a ratio of observation over the entire observation, i.e. 3/5 means the probability of event 3 to occur over the total possibility of 5 is 3 divided 5 or 0.60. Take all these expressed decimal variables and add them up; they will be equal to one. There is no need to be technical about it. A collection of every possible outcome is added up to all outcomes of 100%. One hundred percent (100%) is the same as 1.00 if each observation data is expressed in a form *n/100* or observed event divided all possible events if:

$$\int_{-\infty}^{xj} f(x)dx = \Pr(x \le x_j) \tag{2}$$

Let us summarize the previous two arguments about random variable before going any further. We said that (i) Let there be some random variable *x* and that this *x* may occur according to a random variable probability function expressed in equation (1); (ii) each event of random variable occurring is a single event which such an event is a ratio to the whole probability. If all these events are summed together, they will add up to 1.00. All these events are called *cumulative probability density*; and (iii) from statements (i) and (ii), let *x* be an observed event and x_j be the number of total outcome of event, then the probability of *x* is equal to or les then the probability of x_j. This means that the probability of *x* occurring is a *fraction* of x_j; therefore must range from 0.01 to 1.00 or x/x_j where x_j = 1.00. For example, a fair coin has two sides: head and tail. The two sides is represented by x_j and head is x_{i1} and tail is x_{i2}. The probability of a coin flip for head to show is 1/2 or 0.50 and the probability of tail to show is also $1/2$ or 0.50. The *cumulative density of probability* here is: Pr(head) + Pr(tail) = 0.50 + 0.50 = 1.00.

2.0 NORMAL DISTRIBUTION

Let *x* be a random variable. Remember that random variable means that a variable with a random of probability that each *x* within the range, having an equal chance of occurring, may occur. The *normal distribution* of this random variable *x* with parameters μ and σ^2, where μ is the mean and σ^2 is the variance around the mean. These parameters are denoted as $N(\mu, \sigma^2)$, the density function of this random variable *x* is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (3)$$

The distribution function for this variable *x* is still the density function as expressed in equation (1):

$$\Pr(x \le x_j) = F(x_j) = \int_{-\infty}^{xj} f(x)dx$$

What is the relationship between distribution function $F(x)$ and the density function $f(x)$? The density function $f(x)$ represents the height or the level of *y* of the normal curve at point $x = x_j$. The distribution function $F(x)$ represents the area under the curve from points $-\infty$ to x_j. A block of density from points x_i to x_j may be calculated by taking the probability density of these two points; thus:

$$\Pr(x_i \le x \le x_j) = F(x_j) - F(x_i) \quad (4)$$

Earlier we introduced parameters μ and μ^2. At the on set of the experiment, we may not know the numbers of the mean or variance, but we can set up the expectation value for these parameters. Thus, the expected value for the mean of a set of random variable *x* is: $E(x) = \mu$.

Similarly, for parameter σ^2, the expected value for the variance of the set of data of random variable *x* is: $E(x-\mu)^2 = \sigma^2$.

Let us follow through with the idea of *normal distribution*. The normal distribution has a reproductive property. It means that it can be reproduced. By that we mean, for the parameters μ and σ^2 or $N(\mu,\sigma^2)$, there is a corresponding set of random distribution in the *y-axis*.

An illustration is in order. Let $x_1,\cdots,x_n$ be independent distributed as $N(\mu,\sigma^2)$ where $i=1,\cdots,n$, then the random variable *y* is: $y = x_1 + \cdots + x_n$. These corresponding *y* values are distributed as $N\left(\mu = \sum \mu_i, \sigma^2 = \sum \sigma_i^2\right)$. This means that random variable *y* (which is dependent on random variable *x*) is distributed with parameters $\sum \mu_i$ and $\sum \sigma_i^2$. In order for this reproductive property to hold, the following conditions must exist: (i) the random variables in the data set $x_1, \cdots, x_n$ must be *independently distributed*; meaning each *x* occurs on its own and with equal chance and can be plotted in the *x*-*y* coordinate as individual points; (ii) the random variable data set $x_1,\cdots,x_n$ must be *identically distributed*; meaning each for each $x_1,\cdots,x_n$ there is a corresponding $y_1,\cdots,y_n$; and (iii) the cumulative probability density must be *normally distributed*; meaning distributed in a normal bell-shape curve fashion or all cumulative probability density add up to 1.00 under the set parameters of $N(\mu,\sigma^2)$. If these three conditions are met, then the random variable for *y* is: $y = \frac{x_1 + \cdots + x_n}{n}$. The distribution is equated as:

$$N\left(\mu, \frac{\sum \sigma^2}{n^2}\right) = N\left(\mu, \frac{n\sigma^2}{n^2}\right) = \left(\mu, \frac{\sigma^2}{n}\right) \qquad (5)$$

The condition above is known as Normal, Identically, Independently Distributed (NIID) variables. As for the mean for the random variable *y*, we can write its expected value as: $E(y) = \mu$ with a corresponding variance of *y* as:

$$E(y-\mu)^2 = \sigma_y^2 = \frac{\sigma^2}{n} \qquad (6)$$

Thus far, we have explain *normal distribution* through random variables *x* and *y*. It has been said that for a set of data called random variable *x*, there is a corresponding variable *y* in the *x*-*y Cartesian system*. The steps we took can be summarized as: (i) we define randomness, (ii) find the probability of the data set of random variable; we started with x as random variable and said for this random x, the there is a corresponding y which is also a random product of *x;* this is known as reproductive property of random variable and (iv) in order for random variable to shown reproductive property, the distribution must be independent, identical and normal. The character of *y* is identical to *x*. Our next objective is to relate the normal distribution concept to the Central Limit Theorem (CLT).

2.1 Central limit theorem and the normal curve

The central limit theorem is the foundation for statistics. An understanding of the central limit theorem is indispensable for the study of statistics. The procedural steps we must take in studying statistics may be summarized as:

In our study of the Central Limit Theorem, we will focus on the *classical form* of CLT. To be more specific, by classical form, we mean we will examine only the CLT as presented by the Lindeberg-Levy CLT, Lyapunov CLT, and Lindeberg CLT.

As an introductory note, we start with a random sample, hence the importance of introducing random number at the beginning of our discussion of normal distribution. Let these random variables be represented by a data set of $\{X_1, \cdots, X_n\}$ with the size or number of occurrence of *n*. Each event in the data set $\{X_1, \cdots, X_n\}$ is assumed to be independent and identical with the expected distribution value of the mean μ and the variance σ^2. Under these conditions, we can derive the mean for the data set or sample average as: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

As *n* gets larger and larger towards infinity, the sample average will converge towards the population mean. Here the term average is to have the same meaning as mean. From now on, we shall use the term mean instead of average. We can restate the relationship between the sample mean and the population mean as: $Ex(\mu) \cong \overline{X}$.

As the sample gets larger and larger $Ex(\mu) \cong \overline{X}$. This is the law of probability governing the Central Limit Theorem. As the number *n* gets larger, the variation or stochastic variation around the sample mean will tends to the center; that center is the population mean. This is what CLT tries to describe. What is the center? As *n* gets larger, say tending towards infinity, we say that $\overline{X} = \mu$ Or the sample mean equals the population mean. If that is the case, than the variance between the sample mean and population mean would be *zero*. It is the tendency for the difference between the sample mean and population mean becoming zero that we say the limit is central. This is the definition of the Central Limit Theorem: as *n* tends towards infinity, the variance between the population means and sample means becomes zero because $\overline{X} = \mu$. This common English is expressed formally, that means mathematically, in four ways: Lindeberg-Levy CLT, Lyapunov CLT, and Lindeberg CLT. Details of each statement follows:

2.1.1 LINDEBERG-LEVY CLT

Suppose random variable *x* in a data set of $\{X_1, \cdots, X_n\}$occurs independently and identically distributed. The expected value of *X* is: $E(X_i) = \mu$. If the event of *x* occurrence not matching exactly the mean of the population, this variance has the expected variance that is written as: $Var(X_i) = \sigma^2 < \infty$ as *i* approached ∞. We also know that as *n* approaches infinity, the distribution of the difference between the sample mean and its limit μ or population mean is multiplied by the factor *n*. This condition defines the distribution as: $\sqrt{n}(\overline{X} - \mu)$. As *n* approaches infinity, the random variable $\sqrt{n}(\overline{X} - \mu)$ approaches a normal distribution. 'Normal' is defined as: $N(0, \sigma^2)$ or the variance between the expected sample mean and the population mean is zero (Billingsley, 1995). This condition may be written as:

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0, \sigma^2) \qquad (7)$$

The condition above specifies a variance between the sample mean and population mean equal to zero. Can there be a situation where the population $\sigma > 0$? Yes, that situation occurs when there is a difference between the population mean and sample mean. In such a situation, the convergence in distribution as defined by the *cumulative distribution function* (cdf) of $\sqrt{n}(\overline{X} - \mu)$ occurs point-by-point, or we simply plot them pointwise. These points will converge to the limit at

$N(0, \sigma^2)$. The limit now may be expressed in a form of probability of the distribution among the variance between the sample mean and population mean, which can be written as:

$$\underset{n \to \infty}{Lim} \Pr\left[\sqrt{n}(\overline{X} - \mu) \leq Z\right] = \Phi\left(\frac{z}{\sigma}\right) \tag{8}$$

where $\Phi(x)$ is the standard normal cdf evaluated at x (hence, pointwise). The convergence in is uniform in Z because;

$$\underset{n \to \infty}{Lim} \underset{\approx \in R}{\sup} | \Pr\left[\sqrt{n}\left(\overline{X} - \sigma\right) \leq z\right] - \Phi\left(\frac{z}{\sigma}\right) = 0 \tag{9}$$

The term *sup* (*supremum*) is the upper bound of the set. We had used the term 'convergence' several times in our discussion above without defining it. 'Convergence' here means the tendency for all points to gravitate towards zero as *n* in the data set becomes larger and larger or approaches infinity. This is the tendency because the limit is set for zero or when the population mean equals the sample mean or $E(\mu) \cong \overline{X}$ in which case the variance or σ^2 between the population mean and sample mean tends towards zero in $N(0, \sigma^2)$.

2.1.2 LYAPUNOV CLT

A Russian mathematician named Alexandr Lyapunov also offers a proof for the Central Limit Theorem. According to Lyapunov, the random variable x in the set $\{X_i + \cdots + X_n\}$ must be independent, but not identically distributed. Furthermore, the random variable $\{X_i + \cdots + X_n\}$ must also have moments of the order $(2 + \delta)$ and the rate of growth of these moments is given by the Lyapunov condition (Ash and Doléans-Dade, 1999; Billingsley, 1996; Willey and Resnik, 1999) given as: 'If the $(2+\in)^{th}$ moment with $\in > 0$ exists for a statistical distribution of independent random variates x_i, the means μ_i and variances σ_i^2 are finite,' and:

$$r_n^{2+\in} = \sum_{i=1}^{n} \left\langle |x_i - \mu_i|^{2+\in} \right\rangle \tag{10}$$

then: $\lim\limits_{n \to \infty} \frac{r_n}{s_n} = 0$ where: $s_n^2 = \sum\limits_{i=1}^{n} \sigma_i^2$

Let the random variable $\{X_i + \cdots + X_n\}$ be called $| X_i |$ be a sequence of independent random variable. Each of these random variables have the expected value as the population mean μ_i and variance σ^2. If $\delta > 0$, then the Lyapunov condition states that (Billingsley, p. 362):

$$\underset{n \to \infty}{Lim} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} E\left[| X_i - \mu_i |^{2+\delta}\right] = 0 \tag{11}$$

This is another way of saying the same thing that has been said by Lindeberg-Levy CLT. Here, Lyapunov uses *moments* as an exponential factor for the expected sum of variance between the population mean and sample mean. The larger the size of *delta* in the order of the moment, the faster the function approaches zero or the centrality of the distribution. Remember that δ is the term used to define the order of *moments* for the random variable $|X_i|$.

2.1.3 LINDEBERG CLT

The first CLT we discussed above came from the co-authorship between Lindeberg and Levy, but Lindeberg also developed CLT independently in 1920 (Lingberg, 1922; 211-225). Lindeberg' approach is the same as that found in Lyapunov, but Lindeberg uses the Greek letter *epsilon* ε istead of delta, and instead of using *moments* raised to some order $(2+\delta)$, Lindeberg uses the indicator function $1_{\{...\}}$. The condition starts with the statement: For every $\varepsilon > 0$, the limit of the variance between the population mean and sample mean approaches zero as $n \to \infty$. Thus, the condition is expressed as:

$$\underset{n\to\infty}{Lim}\frac{1}{s_n^2}\sum_{i=1}^{n}E\left[(X_i-\mu_i)^2\cdot 1_{\{|x_i-\mu_i|>\varepsilon s_n\}}\right]=0 \qquad (12)$$

As in approaches infinity, the standardized sum $\frac{1}{s_n^2}\sum_{i=1}^{n}(X_i-\mu_i)$ converges towards a normal distribution as $N(0,1)$.

In all three cases: Lindeberg-Levy CLT, Lyapunov CLT and Lindeberg CLT, we have three different approaches to demonstrate that for random variable *x* with the expected value of μ and σ^2 of the population and that of the sample, as the *n* approaches infinity----or with successive repetition of random selection---the cumulative density of the variance between sample means and population means will approach zero. Zero is the limit of the variation as *n* approaches infinity. This approaching *zero* is known as the *Central Limit Theorem*: zero comes from zero variance between the sample mean and population mean. Central means the distribution of the variance between the sample and population approaches zero; that zero is a point of reference with the variation approaching that zero from $-\infty$ and $+\infty$. This concludes of treatment and understanding of the Central Limit Theorem, i.e. the heart of statistics.

Research interests in the Central Limit Theorem could not be understated. It had generated interests among mathematicians through the history of mathematics. Tijms Henk wrote that:

> "The central limit theorem has an interesting history. The first version of this theorem was postulated by the French-born mathematician Abraham de Moivre who, in a remarkable article published in 1733, used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin. This finding was far ahead of its time, and was nearly forgotten until the famous French mathematician Pierre-Simon Laplace rescued it from obscurity in his monumental work *Théorie Analytique des Probabilités*, which was published in 1812. Laplace expanded De Moivre's finding by approximating the binomial distribution with the normal distribution. But as with De Moivre, Laplace's finding received little attention in his own time. It was not until the nineteenth century was at an end that the importance of the central limit theorem was discerned, when, in 1901, Russian mathematician Aleksandr Lyapunov defined it in general terms and proved precisely how it worked mathematically. Nowadays, the central limit theorem is considered to be the unofficial sovereign of probability theory." (Henk, 20014, p. 169).

Although de Moivre laid the foundation for the Central Limit Theorem in 1733, it was not until 1920 that the term "central limit theorem" was first used by George Polya in 1920. The original German term was: *zentraler Grentzwertsatz.* However, Polya's affixed a different meaning to the word 'central.' For Polya *central* meant the the theorem's importance is central to probability theory. Here, 'central' meant importance. It was made clearer by the French mathematician Lucien Le Cam who affixed the meaning of 'central' to literally mean that "it describes the behaviour of the centre of the distribution as opposed to its tails." (Le Cam, 1986; pp. 78-91).

3.0 Non-Normal distribution

3.1 Chi square distribution: χ^2

The Chi-square distribution is known as *gamma distribution.* It is a single variable distribution with a degree of freedom *k*. The mean, or expected value is equal to *k* and the variance is equal to *2k*. The probability function is not symmetrical. The mode is *k* – 2 and the 50 percentile point is approximately *k* – 0.7.

Percentile points of the chi-square may be expressed by approximation. The formula below approximates the percentile point:

$$\chi^2_{1-\alpha}(k) = \frac{1}{2}\left(\sqrt{2k-1} + z_{1-\alpha}\right)^2 \qquad (13)$$

Given 30 degree of freedom, the 90th percentile point of chi-square: $\chi^2_{0.90}(30) = \frac{1}{2}\left(\sqrt{2(30)-1} + 1.28\right)^2 = 40.20$. Where does 1.28 come from? Look at the Unit Normal Distribution Table, find 0.90 in the column marked $1-\alpha$, then find the corresponding number directly to the left of it in the column where marked $Z_{1-\alpha} = 1.28$.

The chi-square distribution may be generated by z. The variable z is a variable with a unit normal distribution. For single values of a random variable *X* selected from a normally distributed population, *z* may be expressed as:

$$z_i = \frac{X_i - \mu}{\sigma}, \qquad (14)$$

By squaring the *z*, we have:

$$z_i^2 = \frac{(X-\mu)^2}{\sigma^2} \qquad (15)$$

Is a random variable with a chi-square distribution with a single degree of freedom (df = 1.0) or $\chi_i^2(1)$. The range for $\chi_i^2(1)$ is between 1 and $+\infty$ and highly skewed with an expected value of 1 and a variance of 2.

Chi-square may be reproduced. The sum of all reproduced chi-squared is the sum of all chi-square with its form determined by the degree of freedom. This reproductive property of chi-square may be written as:

$$\chi^2(2) = z_1^2 + z_2^2 = \frac{(X_1-\mu)^2}{\sigma^2} + \frac{(X_2-\mu)^2}{\sigma^2} \qquad (16)$$

$$\chi^2(3) = z_1^2 + z_2^2 = \frac{(X_1-\mu)^2}{\sigma^2} + \frac{(X_2-\mu)^2}{\sigma^2} + \frac{(X_3-\mu)^2}{\sigma^2} \qquad (17)$$

The general expression for the reproductive sum property of chi-square may be written as:

$$\chi^2(k) = \sum_{i=1}^{k} z_i^2 = \frac{(X_1 - \mu)^2}{\sigma^2} + \frac{(X_2 - \mu)^2}{\sigma^2} + \cdots + \frac{(X_k - \mu)^2}{\sigma^2} \tag{18}$$

Generally, we do not know the population mean μ because in all experiments, we know only the sample mean $\overline{X}$. The equation above is expressed in terms of population mean μ; nevertheless, it is possible to substitute the sample mean $\overline{X}$ for μ, and thus rewrite as:

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \tag{19}$$

The chi-square distribution has $k = n - 1$ degrees of freedom. The term σ^2 is population variance, and since we do not generally known the population mean, we would also do not know the population variance; however, we can substitute sample variance for population variance σ^2 with s^2. Thus:

$$s^2 = \frac{\sum(X_1 - \overline{X})^2}{n-1} \tag{20}$$

Finally, we can also write that:

$$\frac{(n-1)s^2}{\sigma^2} \tag{21}$$

which has the chi-square distribution with n – 1 degree freedom. This ratio is useful for hypothesis testing and building confidence interval for population variance.

3.2 F distribution

The *F* distribution is known as the *beta distribution.* Another name for the *F* distribution is Snedecor's F distribution. The letter *F* is in honor of Fisher. Snedecor transformed a distribution obtained by Fisher. *F* is defined as:

$$\int_0^{F'} CF^{(k_{1/2})-1}\left(1 + \frac{k_1}{k_2}F\right)^{-(k_1+k_2)/2} dF\ , \tag{22}$$

where $C = \dfrac{(k_1 + k_2)^{k_{1/2}}}{\beta(k_1/2, k_2/2)}$

The denominator of *C* is called the *beta function*. *F* defines a family of distributions defined by k_1 and k_2 which are the degrees of freedom of the numerator and denominator.

F may also be defined as the ratio of two independent chi-square variables, each divided by its own degrees of freedom.

$$F = \frac{\chi_1^2 / k_1}{\chi_2^2 / k_2} \quad (23)$$

From earlier expression, we had: $\frac{(n-1)s^2}{\sigma^2}$

What is an F distribution? It is the ratio of two sample variance:

$$F = \frac{s_1^2}{s_2^2} \quad (24)$$

These two variances are taken from two separate samples from the same population. Therefore, it is an analysis of variance which can help determine whether there is bias in the sampling. The conclusion that *F* distribution is the ratio of two variance of the same population is derived from the following process:

$$X_1^2(n-1) = \frac{(n_1-1)s_1^2}{\sigma^2}, \quad (25)$$

$$X_2^2(n-1) = \frac{(n_1-1)s_2^2}{\sigma^2}, \quad (26)$$

and

$$F = \frac{\chi_1^2 / n_1 - 1}{\chi_2^2 / n_2 - 1} \quad (27)$$

By decomposing:

$$F = \frac{\left(\frac{(n_1-1)s_1^2}{\sigma^2 1}\right) \div n_1 - 1}{\left(\frac{(n_2-1)s_2^2}{\sigma^2}\right) \div n_2 - 1} \quad (28)$$

Or simply:

$$F = \frac{s_1^2}{s_2^2} \quad (29)$$

3.4 Relations among distributions

Each distribution is used in different situation. It is said that one type of distribution is used for a particular situation. A particular situation is determined by the type of data available, i.e. whether the data come from a sample, a series of sample, or an entire population. Nevertheless, in the final analysis we can also find a common ground because we are dealing with the *normality* of the distribution. In statistics, everything revolves around the *central limit theorem*. The relationships among all distribution, this far discussed, may be summarize as:

$$z_{1-\alpha}^2 = \chi_{1-2\alpha}^2(1) = F_{1-2a}(1,\infty) = t_{1-a}^2(\infty) \quad (30)$$

What does it all mean? The equation above states that *the square of Z value is equal to Chi-square with the degree of freedom (df) of 1; equal to the F value with the level of confidence of 1-2α and the degree of freedom of (1,∞); equal to the square of student t having degree of freedom (∞)* (Student, 1908). It means that if we found one, we can fund the others. All distributions are talking about the same thing but through different language.

The confusion among young researcher in what to use, i.e. what type of distribution test should be used, is a result of poor understanding of basic statistics. This lack of understanding finds its fault in the teaching of statistics. The fault does not lie with students, but with teachers who *failed* in teaching statistics in a useful and easily understandable manner. In social science research, statistics is an indispensable analytical tool. In order to engage a research topic either by doing research ourselves as researchers of reading the results of research done by others, we must have a good understanding of statistics. A mere rudimentary understanding is not enough to evaluate the research of others, and therefore could not be the basis for us to comment or evaluate the research of others; our understanding of statistics must be solid. *That* requires a good foundation in basic statistics.

The relationship among various distributions: *Z, chi-square, F-test and t-test,* is derived from: $t^2_{1-\alpha}(k) = F_{1-2\alpha}(1,k)$, then:

$$t^2 = \frac{\text{unit normal}^2}{\chi^2 / k} \tag{31}$$

$$t^2 = \frac{\chi^2 /1}{\chi^2 / k} \tag{32}$$

$t^2 = F(1,k)$ and $t(\infty) = z$.

Another equivalence is:

$$\frac{1}{k}\chi^2_{1-\alpha} = F_{1-\alpha}(k,\infty) \tag{33}$$

which is the same as: $\chi^2_\alpha(k) = kF_\alpha(k,\infty) = \frac{1}{F_{1-\alpha(\infty,k)}}$ because $F_\alpha(k_1,k_2) = \frac{1}{F_{1-\alpha(\infty,k)}}$.

4.0 CONCLUSION

We learned that the Central Limit Theorem is the heart of statistics. Historically, there had been many attempts, by many people and in various manners, to illustrate and prove the Central limit Theorem. In words, the Central Limit Theorem states that in the long run all events that occur in a discrete time series will tend toward the center. That center is the population mean. When Sir Francis Galton work on the topic in the earlier period, he referred to then central tendency as *regression towards the mean*. The idea of centrality had been attempted by De Moivre, a French probability statistician, but De Moivre attached the meaning to centralism differently that what we understand centrality of today's statistics. For De Moivre, 'center' meant statistics based on analysis of the mean is the 'central idea.' Although de Moivre laid the foundation for the Central Limit Theorem in 1733, it was not until 1920 that the term "central limit theorem" was first used by George Polya in 1920. The original German term was: *zentraler Grentzwertsatz.* However, Polya's affixed a different meaning to the word 'central.' For Polya *central* meant the theorem's importance is central to probability theory. Here, 'central' meant importance. It was made clearer by the French mathematician Lucien Le Cam who affixed the meaning of 'central' to literally mean that "it describes the behaviour of the centre of the distribution as opposed to its tails" (Le Cam). The Central Limit Theorem is the key in for research methodology in social science. Much of social science research deals with randomness and variables that are not controllable. For the most part, in the natural sciences, experiments are observed in a controlled environment. Social science, by its nature and the nature of its subjects to be observed, does not have the luxury of laboratory observation. Therefore, experimental design in social science must find a tool that is scientific to render credibility to its claims. Statistics being a branch of mathematics and mathematics, being the claimant of universal truth, provides a solid foundation to social science research. The Central Limit Theorem, being the protagonist in probability and statistics, lay claims to primacy in experimental design for research in social science. Finally, at the heart of the Central Limit theorem is distribution. In a short and straight route, this paper led us through various types of distribution and the indispensable distribution density function.

REFERENCES

Ash, R. B. and Doléans-Dade, C. A. (1999). Probability & Measure Theory, 2nd ed. New York: Academic Press, p. 307.

Billingsley, P. (1986). Probability and Measure, 2nd ed. New York: p. 371.

Wiley and Resnik, S. (1999). A Probability Path. Boston, MA: Birkhäuser, p. 319.

Billingsley, Patrick (1995), Probability and Measure (3rd ed.). John Wiley & Sons, ISBN 0-471-00710-2 ; p. 357.

Fan, Jianqing (1991). "On the optimal rates of convergence for nonparametric deconvolution problems." The Annals of Statistics 19 (3): 1257–1272.

Le Cam, Lucien (1986). "The central limit theorem around 1935." Statistical Science 1 (1): 78–91.

Henk, Tijms (2004). Understanding Probability: Chance Rules in Everyday Life. Cambridge: Cambridge University Press. p. 169. ISBN 0-521-54036-4.

Lindeberg, K.W. (1922). "Eine Neue Herleitung des Exponentialgesetzes in der Warhscheinlickeitsrechnung." Mathematischer Zeitschrift 15 (1): 211-225.

Patel, Jagdish K.; Read, Campbell B. (1996). Handbook of the Normal Distribution (2nd ed.). CRC Press. ISBN 0-8247-9342-0.

Student (1908). "The Probable Error of a Mean." *Biometrika* **6**, 1-25.