E-Statistics as an instrument for reliable source of data

A A Gwani\textsuperscript{1,2,a} and A Aliyu\textsuperscript{1}

\textsuperscript{1}Department of Mathematical Sciences, Bauchi State University Gadau, Nigeria
\textsuperscript{2}Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Nigeria

E-mail: aaagwani1@gmail.com

Abstract. This paper identifies the modern technique of data collection, analysis and presentation using information technology facilities and this technique is termed as E-Statistics. We highlighted advantages of E-Statistics over other techniques for data collection including the economic advantages in terms of time, resources and money. This will assist in getting 100\% coverage of the targeted population and provide reliable and latest data. In addition, E-Statistics would eradicate the idea of sampling and its techniques, provided that the entire population information has been captured in the database and the data capturing centres have direct update access. However, computer scientist, computer engineers and technologists should design platforms with a centralized database which would give the National Bureau of Statistics, Statisticians and other organisations to access and update statistical information online. Based on this World data was collected online from statistical web site on internet users and two statistical techniques were used in presenting and analysing the data (descriptive and inferential). After the analysis we realized that there is significance difference between the regional mean values of the internet users in the world.

1. Introduction
Size of Population can be large and the data which are collected on the population could be larger than the population itself because of the number of variables that can be collected from a population member [1]. Statistician can assist decision makers in summarizing the data and this can be achieved by using Graphs and numerical techniques [1]. Using the statistical analysis makes it possible for us to test a numerical data for relevance, reliability and validity. The primitive way of collecting data consumed a lot of time and resources and which led to errors in collected data; and these errors might result from the respondents, selected sampling technique and procedure and so on [1]. In finding a measure to prevent these shortcomings, E-Statistics was introduced. However, developed countries are advanced in respect to statistics [2]. It is necessary for E-Statistics to preserve and provide the previous and current data respectively, and this should be carried out in manner that will be useful to the society for gathering relevant information which will enable them to make correct judgments.

2. E-Statistics
E-statistics is the deployment of internet and the World Wide Web for organisational and individuals data gathering for easy representation and analysis of data. It also involves the uses of Information and Communications Technology (ICT) to facilitate and improve on the efficiency and effectiveness of data gathering and delivery in both public and private sectors [11]. The E-Statistics demonstrates applications to facilitate the operation and the collection and disbursement of data or information. E-statistics relies heavily on the internet work of Information and Communications Technology (ICT) infrastructures, internet and non-internet applications to aid the data operations in order to facilities...
decision making [8]. The recent data big bang coupled with the development of the ICTs have raised
in providing us with a stream of fresh and digitized data related to how individuals, companies and
other organizations interact [12]. Following this task requires to carefully organised plan of the whole
process of data collection and analysis taking into account the particularities of the social, legal and
economic issues [10]. This includes the wide variety of heterogeneous sources of information and a
strict governance policy. Grounded on the data to provide a specifically designed system for
forecasting social and economic behaviours, trends and change [15]. The E-statistics system should
turn the data into knowledge [9,13]. E-Statistics must deal with large amount of unstructured and
heterogeneous data [11]. E-Statistics is the modern technique of data collection, analysis and
presentation using information and communication technology facilities in order to have advantages of
speed and accuracy. E-Statistics can also be seen as a system that deal with multiple data sources that
can be opened at the same time online, making it easier to compare the contents of different data
sources, create multiple subsets of cases or variables for analysis, merge multiple data sources from
various data formats for a specific purpose.

2.1. Data Sources
Several statisticians experiences difficulties in obtaining data for research and analysis [2]. The
sources of data can be classified in two including the primary source and the secondary source [3].

2.1.1. Primary Sources of Data
Primary data are the data from first hand source or data collected by the researcher(s). Researchers
gather these kinds of data in a situation where the required data cannot be gotten from outside sources.
The process for gathering this data may require that questions and collection should address the need
of the research questions [2]. This process can be expensive and requires permission and authorization
from the relevant organization before such data can be collected [13]. Also, issues concerning consent
and confidentiality are of great importance. An advantage of primary data is its capability to provide
researchers with current information which can be obtained from the primary source [11].

2.1.2. Secondary Sources of Data
The secondary data is historic in nature and provides information about what has already been
discovered on a specific research topic. Primary data can be collected online via monkey survey,
emails, Facebook, etc. Secondary source of data is a method for data collection which can be adopted
when the duration for data collection is extensive, or when the data collection has already been done
[3]. Secondary data is normally obtained from the studies performed by other institutions or
organizations. The process for data collection and its validation was discussed in detail [13]. It is easy
to obtain some secondary data online when it is requested for, because its freely available to everyone.
Online secondary data includes data from the database, webpage history, Facebook history,
Organisational web history etc.

3. E-Statistics as a Data Source
E-statistics serves as a source of data since both primary and secondary data could be obtained online
[13]. Example of E-statistic includes online survey, online interview, online observation and online
record or documents [10].

3.1. E-Statistics instruments for data collection
E-statistics instruments for data collection are online interview, Monkey survey, Online observation,
online record or documents.

3.1.1. Monkey Survey
Monkey surveys are standardized instruments that can be administered online. The primary advantage
of monkey surveys is the cost in relation to the amount of data that can be collected [2]. Generally,
Monkey surveying is considered efficient because, it can cover large population at a relatively low
cost [11]. Although it has some disadvantages: the rate of response to survey conducted online may be
low due to some factors which threatens the validity of the data collected [13]. Mechanisms should be adopted to increase response rates, however, adopting these mechanisms can increase the cost of survey [3]. Monkey survey is an online instrument for data collection that would enable the researcher to track grassroots of an organizations, it is categorised as online primary source of data. It is a tool used for performing E-statistics in data collection.

3.1.2. Online interview
Interview is an instrument of data collection conducted either in person or over platforms such as Skype, Facebook video call, GSM, and telephone. This instrument allows complex questions to be asked during interview since there would be an opportunity to clarify any confusion. Unlike the online survey, the interview allows respondents to elaborate on their answers and eliciting more in-depth information [2]. The disadvantage of interviews is the high cost and time consumption [1] (and therefore costs more money or internet data) involved in conducting telephone calls or video calls. The use of information and communication technology interview is very efficient since the information would be collected from fewer people (Sample). However, the reliability of online interview can be an issue if interviewers are not experts because responses can be bias when questions can be asked in different ways [1].

3.1.3. Online observation
Observations can capture the data behaviours, interactions between groups or individuals, events, or physical site conditions [1]. Observations require well-trained observers who follow details such as what data to observe, when, where, for how long, and by what online method of recording. The main advantage of this instrument is observation validity. When observation is performed accurately, it is being considered as a strong data collection method because it generates first hand, unbiased information which is retrieved from the system [1]. Online data observation requires a time frame for design and development of the observations tool/system, training of the observers, and data collection or implementation of the tool/system [5]. An example of online observations includes: The number of Nigerian citizen that are currently online on Facebook. Based on this information, we can generate the Statistics.

3.1.4. Online Record or Documents
Online Record or Document is the current or instant record in the database (web sites) or document review [8]. Online record is a systematic and organised method for collecting data from existing records. These records include financial documents, employee record, Clinical or hospital record, monthly reports, activity logs, purchase orders, etc.[5,11]. An advantage of using online records in an organization is to ease the process of data collection [2]. The data already exists in the database and no extra effort needs to be made to collect it (assuming the specific data needed is really available and up-to-date) [2,15]. In a situation where the data is readily available and timely, record review will be an efficient and cost-effective method for data collection [15].

3.2. Data Reliability
If a measure is repeatable, we should get the same result when we measured the same thing twice at same time [6,13]. Reliability refers to the consistency of data while the validity is the accuracy of the information generated. Reliability involves the extent to which data is reproducible [3,15]. Does the system repeatedly produce the instantaneous results? Inconsistencies in the sample, undermined sample size, or bias in the data collection instrument, all these are the primary threat to reliability and they can be avoided or reduced using E-statistics [3,12]. Reliability must give a quantitative value for repeatability, [6,15]. Validity is highly important and often overlook aspect of any study whether inferential, descriptive or quantitative [6,12]. Validity simply refers to whether the E-statistics system measures, records, observes or accesses the appropriate items [6,15]. Reliable data are collected at a specific time such as current data and data with interval. This time interval can range from 5 seconds up to 20 years or above of an area or place, institution or organisation.
3.2.1 Advantages of E-Statistics
Data can be saved or recorded automatically instantaneous data as well as recorded data can be generated with just a press of button thereby reduced cost. Data can be automatically coded and analysed by the system by just pressing a button. This saves time.
- Data can be automatically coded and analysed by the system by just pressing a button. This saves time.
- Comparison of groups with similar or the same characteristics can be done, and this is useful in demonstrating the success or failure of the data collected.
- Data could be collected a number of times which is more reliable
- E-statistics facilitates faster disseminations of information there by improving the communication system/task accomplishment.
- It simplifies data/information transaction process via electronic transactions.
- E-statistics reduces cost of survey through elimination of physical barriers
- E-statistics can prevent double entry by setting up a system to collect data more than once and then compare their saving, sorting, time and cost for discrepancies which ensures the reliability of the extracted data.
- It is easy to sort data to find missing, high, or low values in the database or spreadsheet function by the using formulas or sorting functions.
- Report can be generated easily provided the database has a standard format.

3.2.2 Challenges of E-Statistics
- It lacks steady and constant power at some rural data centres [5].
- Shortage of IT facilities for proper implementation [5].
- There exist data discrepancies from multi-data organization due to data quality assurance [7].
- The designation of integrated databased that would accommodate various data from various organisations [5].
- Data security issues [5].
- Geographical variations and the variations in national laws [7].
- Some organisations are not online, and some are online, but their website is not up-to-date [5].

3.3 E-Statistics Theory
Let n be a sample size from a population of size N, such that n can partition, as then sample size increases over time as a result of advancement in information technology, as \( n \rightarrow N \) there exist a point where the sample size \( n = N \). this is the what we term as exact point or point of Exactness. Exact point is the situation where by the sample size is equal to the size of the population of interest. This theory is based on insight world is changing or moving from analogue to digital, as some individuals and organisations are changing to embrace the technological evolution, we hope in future statistician would use the technology to get information about the entire population so that we can get exact statistics of a targeted population.

3.3.1 Limitations of the theory
a. At the Exact point (n=N) the data may large but the advancement of technology would dealt with this.
b. Not all humans or groups of human, organisations and Government has internet access or are online but we hope in near future all humans or groups of humans, organisations and Government would be online.
c. It may take long-time to realise the imaginary point of exactness due to technological challenges in some portion of the world.
3.4. **E-Statistics versus Sampling**

E-Statistics as a source of data can be used in many ways in the production of official statistics. E-Statistics data are the sources of data used for the production of some statistic about a population of interest. To achieve this selection of appropriate method of inference in necessary [17] but in some instance the results remain biased if specific subpopulations are completely missing in the E-Statistics data set [16]. This could happen when the technology used is insufficient or deficient in coverage. E-Statistics data can be used as auxiliary data in a procedure primarily based on sample survey data. Statistics based on E-Statistics data could be simply as covariate in model-based estimation techniques applied to survey sample data [16], among the potential gains of this approach is cost reduction and reduction of burden on respondents [16]. At same note this methods has tendency of predicting parameter values for unobserved units, and are usually chance upon in data mining and machine learning contexts [19]. E-Statistics may enrich the official publications of Statistics although, E-Statistics data may be selective in the sense that not everybody in the world population uses the internet that is why we consider the population of internet users as our auxiliary variable as per as this article E-Statistics data as concerned. On the other hand among the reasons we sampled are: when the population of interest is large, when we are dealing with a limited resources, when in getting the sample we may destroy the part of the population or entire population [20]. Sampling is a process of obtaining the sample from the population of interest, to obtain the sample we most used some techniques depends on the nature of the population and the variable of interest [20]. For now since not everybody is on internet then we can sample online provided each unit of the population is represented. E-Statistics can be considered as non-probability source of data while sampling could be both probability and non-probability depends on population characteristics. E-Statistics may be helpful in research pertaining health statistics, Geographical locations, Biostatistics and others, and the result could be publish, its limitation must be clearly state as well as assumptions if any, and E-Statistics should be publish in a format that would avoids confusion with other Statistics. Result obtained from E-Statistics data through investigative data analysis or visualisation, could be amalgamated or further verified using sample survey. We hope in near future that everybody would be on internet, at this juncture no need of sample because we would have data or information on the whole population.

3.5. **E-Statistics data flow pattern**

The data pattern is show in the figure 1 below,
As the name implies, Figure 1 above shows the data flow pattern of the forecasted E-Statistics. The manual input units are the registration centres or any point that the internet user can enter the data, this would serve as a data sources to the regional databases, that is African, American, Asian, European, Oceania and Middle East. The six regional databases would serve as a data source to the central world database of which any client would use View to get the E-Statistics data or analysed result of the data. Although along the line there should be checks and coders/designers would take care of that.

4. Data
In assessing the reliability of the extracted data which will be considered for E-statistics, we obtained data from online source [4] containing the estimated World 2017 population regionally and internet users and other related statistics as shown in table 1–6 as presented in section 4.2 [4].

4.1. Methodology
In this paper, secondary source was used as the methodology for obtaining the data. Simple percentage was used to determine the amount of internet users in different parts of the world [4]. By having the high percentage of internet usage in a region, it indicates the suitability and reliability of the need for E-Statistics. Descriptive statistics tools were used in the data display and in the inferential part ANOVA was used to evaluate the null hypothesis in order to know whether or not there is significance difference among the regional means of the Internet Users Population [4]. SPSS and Excel would be used in the data analysis and data presentation respectively.

4.1.1. One way ANOVA null hypothesis
ANOVA is an abbreviation of the name “Analysis of Variance” it derived from a partitioning total variability in to its component parts [21]. Suppose we would have different level of a single factor that we would wish to compare [21], and our factor here is the number of internet users in a place. The different levels of a factor are often called treatments [21]. The null hypothesis for the auxiliary variable (Population of Internet Users) is based on our interest to test the equality of treatment effects (regions) on the grand mean, in other word whether the treatment means are equal. Mathematically, the appropriate hypotheses are equation (1) and (2) Let \( \mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \ldots \ldots \mu_i \) represents the means values of internet users in the sub-populations or region then,

\[
H_0: \mu_1 = \mu_2 = \mu_3 \ldots \ldots \mu_i \]  
(1)
\[ H_1: \mu_i \neq \mu_j \text{ for atleast one } i \quad (2) \]

Equation (1) and equation (2) above is the null hypothesis and its alternative hypothesis respectively, the first could be interpreted as there is no significant difference between the treatment means of the groups and the second one as an alternative interpreted as there is significant difference between the treatment means at this point further test should be done to know which among the treatment means are/is different.

4.2. Data Presentation

The data would be presented regionally as in the tables 1 and 2 contains European Population and Internet Users Statistics in December 2017, table 3 shows Asia Population and Internet Users Statistics in December 2017, table 4 indicated American Population and Internet Users Statistics in December 2017, table 5 revealed Middle East Population and Internet Users Statistics in December 2017 and table 6 shows Oceania Population and Internet Users Statistics in December 2017. Table 7 contains summarised tables 1 – 6. Also, figures 2 and 3 [4] illustrate the obtained data.

| AFRICA       | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) | Facebook Subscribers 31-Dec-2017 |
|--------------|------------------------|-----------------------------|-----------------------------|----------------------------------|
| Algeria      | 42,008,054             | 18,580,000                  | 44.20%                      | 19,000,000                      |
| Angola       | 30,774,205             | 5,951,453                   | 19.30%                      | 3,800,000                       |
| Benin        | 11,458,674             | 3,801,758                   | 33.10%                      | 920,000                          |
| Botswana     | 2,333,201              | 923,528                     | 39.60%                      | 830,000                          |
| Burkina Faso | 19,751,651             | 3,704,265                   | 18.80%                      | 840,000                          |
| Burundi      | 11,216,450             | 617,116                     | 5.50%                       | 450,000                          |
| Cabo Verde   | 553,335                | 265,972                     | 48.10%                      | 240,000                          |
| Cameroon     | 24,678,234             | 6,128,422                   | 24.80%                      | 2,700,000                        |
| Central African | 4,737,423        | 256,432                     | 5.40%                       | 96,000                           |
| Chad         | 15,553,184             | 768,274                     | 5.00%                       | 260,000                          |
| Comoros      | 832,347                | 130,578                     | 15.70%                      | 120,000                          |
| Congo        | 5,399,895              | 650,000                     | 12.00%                      | 600,000                          |
| Congo, Dem.  | 84,004,989             | 5,137,271                   | 6.10%                       | 2,100,000                        |
| Cote d'Ivoire| 24,905,843             | 6,318,355                   | 26.30%                      | 3,800,000                        |
| Djibouti     | 971,408                | 180,000                     | 18.50%                      | 180,000                          |
| Egypt        | 99,375,741             | 49,231,493                  | 49.50%                      | 35,000,000                       |
| Equatorial Guinea | 1,313,894       | 312,704                     | 23.80%                      | 67,000                           |
| Eritrea      | 5,187,948              | 71,000                      | 1.40%                       | 63,000                           |
| Ethiopia     | 107,534,882            | 16,437,811                  | 15.30%                      | 4,500,000                        |
| Gabon        | 2,067,561              | 985,492                     | 47.70%                      | 620,000                          |
| Gambia       | 2,163,765              | 392,277                     | 18.10%                      | 310,000                          |
| Ghana        | 29,463,643             | 10,110,000                  | 34.30%                      | 4,900,000                        |
| Guinea       | 13,052,608             | 1,602,485                   | 12.30%                      | 1,500,000                        |
| Guinea-Bissau| 1,907,268              | 120,000                     | 6.30%                       | 110,000                          |
| Kenya        | 50,950,879             | 43,329,434                  | 85.00%                      | 7,000,000                        |
| Lesotho      | 2,263,010              | 627,860                     | 27.70%                      | 310,000                          |
| Liberia      | 4,853,516              | 395,063                     | 8.10%                       | 330,000                          |
| Libya        | 6,470,956              | 3,800,000                   | 58.70%                      | 3,500,000                        |
### Table 1

| Country          | Population  | Internet Users | Penetration (%) | Facebook Subscribers |
|------------------|-------------|----------------|-----------------|----------------------|
| Madagascar       | 26,262,810  | 1,900,000      | 7.20%           | 1,700,000            |
| Malawi           | 19,164,728  | 1,828,503      | 9.50%           | 720,000              |
| Mali             | 19,107,706  | 12,480,176     | 65.30%          | 1,500,000            |
| Mauritania       | 4,540,068   | 810,000        | 17.80%          | 770,000              |
| Mauritius        | 1,268,315   | 803,896        | 63.40%          | 700,000              |
| Mayotte (FR)     | 259,682     | 107,940        | 41.60%          | 71,000               |
| Morocco          | 36,191,805  | 22,567,154     | 62.40%          | 15,000,000           |
| Mozambique       | 30,528,673  | 5,279,135      | 17.30%          | 1,800,000            |
| Namibia          | 2,587,801   | 797,027        | 30.80%          | 570,000              |
| Niger            | 22,311,375  | 951,548        | 4.30%           | 440,000              |
| Nigeria          | 195,875,237 | 98,391,456     | 50.20%          | 17,000,000           |
| Reunion (FR)     | 883,247     | 480,000        | 54.30%          | 420,000              |
| Rwanda           | 12,501,156  | 3,724,678      | 29.80%          | 490,000              |
| Saint Helena     | 4,049       | 2,200          | 54.30%          | 1,700                |
| Sao Tome & Principe | 208,818  | 57,875         | 27.70%          | 170,000              |
| Senegal          | 1,391,385   | 446,051        | 32.10%          | 52,000               |
| Seychelles       | 95,235      | 67,119         | 70.50%          | 61,000               |
| Sierra Leone     | 7,719,729   | 902,462        | 11.70%          | 450,000              |
| Somalia          | 15,181,925  | 1,200,000      | 7.90%           | 1,100,000            |
| South Africa     | 57,398,421  | 30,815,634     | 53.70%          | 16,000,000           |
| South Sudan      | 12,919,053  | 2,229,963      | 17.30%          | 180,000              |
| Sudan            | 41,511,526  | 11,816,570     | 28.50%          | 2,600,000            |
| Swaziland        | 1,391,385   | 446,051        | 32.10%          | 170,000              |
| Tanzania         | 59,091,392  | 23,000,000     | 38.90%          | 6,100,000            |
| Togo             | 7,990,926   | 899,956        | 11.30%          | 560,000              |
| Tunisia          | 11,659,174  | 7,898,534      | 67.70%          | 6,400,000            |
| Uganda           | 44,270,563  | 19,000,000     | 42.90%          | 2,600,000            |
| Western Sahara   | 561,257     | 28,000         | 5.00%           | 24,000               |
| Zambia           | 17,609,178  | 7,248,773      | 41.20%          | 1,600,000            |
| Zimbabwe         | 16,913,261  | 6,796,314      | 40.20%          | 880,000              |
| **TOTAL AFRICA** | **1,287,914,329** | **453,329,534** | **35.20%**   | **177,005,700** |
| **WORLD TOTAL** | **7,634,758,428** | **4,156,932,140** | **54.40%**   | **2,119,060,152** |

Table 1 above indicated the column of population of African countries, the population of Internet Users in each country by December 2017, the rate of internet penetration in percentage and the population of Facebook Users June 2017. It also shows the total population of African, African internet and Facebook Users, as well as the rate of internet penetration which was up to 35% of its population.

### Table 2

European population and internet users statistics in December 2017.

| EUROPEAN UNION      | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (%) Population | Facebook Subscribers 31-Dec-2017 |
|---------------------|-------------------------|----------------------------|----------------------------|----------------------------------|
| Austria             | 8,592,400               | 7,273,168                  | 84.60%                     | 3,600,000                        |
| Belgium             | 11,443,830              | 10,060,745                 | 87.90%                     | 6,500,000                        |
| Bulgaria            | 7,045,259               | 4,213,065                  | 59.80%                     | 3,300,000                        |
| Croatia             | 4,209,815               | 3,133,485                  | 74.40%                     | 1,800,000                        |
| Cyprus              | 1,187,575               | 901,369                    | 75.90%                     | 800,000                          |
| Czech Republic      | 10,555,130              | 9,323,428                  | 88.30%                     | 4,600,000                        |
| Denmark             | 5,711,837               | 5,534,770                  | 96.90%                     | 3,700,000                        |
| Estonia             | 1,305,755               | 1,196,521                  | 91.60%                     | 620,000                          |
| Finland             | 5,541,274               | 5,125,678                  | 92.50%                     | 2,700,000                        |
Table 2 above shows the estimated population of European countries, their population of Internet Users by December 2017, the rate of internet penetration in percentage and the population of Facebook Users June 2017. It also shows the total population of Europe, European Internet Users and Facebook Users, it also revealed the rate of internet penetration which was up to 86% of its population, this much more higher than that of African internet penetration as presented in Table 1.

Table 3. Asian population and internet users statistics in December 2017.
The above (table 3) shows the estimated population of Asian countries, their respective Internet Users population in December 2017, the percentage rate of internet penetration and their respective population of Facebook Users in June 2017. It also shows the total population of Asia, total number of Internet Users in Asia, as well as total population of Facebook Users in Asia, it also revealed the rate of internet penetration which was 48% of its total population; this is higher than that of African internet penetration as presented in Table 1 but less than that of Europe.

Table 4. America population and internet users statistics in December 2017.

| AMERICA REGIONS      | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) | Facebook Subscribers 31-Dec-2017 |
|----------------------|------------------------|-----------------------------|----------------------------|----------------------------------|
| North America        | 363,224,006            | 320,059,368                 | 88.10%                     | 263,081,200                     |
| South America        | 426,548,298            | 278,596,721                 | 65.30%                     | 257,242,500                     |
| Central America      | 177,249,493            | 105,771,952                 | 59.70%                     | 102,760,000                     |
| The Caribbean        | 43,806,854             | 19,900,490                  | 45.40%                     | 10,972,840                      |
| TOTAL THE AMERICAS   | 1,010,828,651          | 724,328,531                 | 71.70%                     | 634,056,540                     |
| WORLD TOTAL          | 7,634,758,428          | 4,156,932,140               | 54.40%                     | 2,119,060,152                   |

Table 4 above outline the estimated population of American regions together with their population of Internet Users by December 2017, it also indicated the rate of internet penetration in percentage and the population of Facebook Users June 2017. It also shows the total population of America, total number of Internet Users, Facebook Users and rate of internet penetration in each region.

Table 5. Middle East population and internet users.

| MIDDLE EAST        | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) | Facebook Subscribers 31-Dec-2017 |
|--------------------|------------------------|-----------------------------|----------------------------|----------------------------------|
| Bahrain            | 1,566,993              | 1,535,653                   | 98.00%                     | 1,100,000                        |
| Iran               | 82,011,735             | 56,700,000                  | 69.10%                     | 40,000,000                       |
| Iraq               | 39,339,753             | 19,000,000                  | 48.30%                     | 17,000,000                       |
| Israel             | 8,452,841              | 6,740,287                   | 79.70%                     | 5,800,000                        |
| Jordan             | 9,903,802              | 8,700,000                   | 87.80%                     | 5,300,000                        |
| Kuwait             | 4,197,128              | 4,104,347                   | 97.80%                     | 3,100,000                        |
| Lebanon            | 6,093,509              | 5,546,494                   | 91.00%                     | 3,600,000                        |
Table 5. Population in year 2017, population of Internet Users in the area in December 2017, the number of Facebook Users, internet population penetration in percentage and their totals. Middle East has internet penetration of 64.5% in December 2017.

| Country           | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) | Facebook Subscribers 31-Dec-2017 |
|-------------------|------------------------|----------------------------|----------------------------|-----------------------------------|
| Oman              | 4,829,946              | 3,310,260                  | 68.50%                     | 2,630,000                         |
| Palestine (State of) | 5,052,776              | 3,055,088                  | 60.50%                     | 1,700,000                         |
| Qatar             | 2,694,849              | 2,644,580                  | 98.10%                     | 2,300,000                         |
| Saudi Arabia      | 33,554,343             | 30,257,715                 | 90.20%                     | 18,000,000                        |
| Syria             | 18,284,407             | 6,025,631                  | 33.00%                     | 4,900,000                         |
| United Arab Emirates | 9,541,615             | 9,385,420                  | 98.40%                     | 8,700,000                         |
| Yemen             | 28,915,284             | 7,031,784                  | 24.30%                     | 2,352,942                         |
| TOTAL MIDDLE EAST | 254,438,981            | 164,037,259                | 64.50%                     | 116,482,942                       |

Table 6. Oceania population and internet users statistics in December 2017.

| OCEANIA          | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) | Facebook Subscribers 31-Dec-2017 |
|------------------|------------------------|----------------------------|----------------------------|-----------------------------------|
| American Samoa   | 55,653                 | 22,000                     | 39.50%                     |                                   |
| Antarctica       | 2,700                  | 4,400                      | 100.00%                    | 1,800                             |
| Australia        | 24,641,662             | 21,743,803                 | 88.20%                     | 15,000,000                        |
| Australia, Ext. Terr. | 1,651               | n/a                        | n/a                        | n/a                               |
| Christmas Island | 1,530                  | 790                        | 51.60%                     | 350                               |
| Cocos (Keeling) Is. | 596                   | 20                         | 3.40%                      | 20                                |
| Cook Islands     | 21,069                 | 11,377                     | 54.00%                     | 6,300                             |
| Fiji             | 902,547                | 421,958                    | 46.80%                     | 380,000                           |
| French Polynesia | 288,685                | 197,576                    | 68.40%                     | 140,000                           |
| Guam             | 174,214                | 124,145                    | 77.00%                     | 100,000                           |
| Kiribati         | 116,405                | 15,947                     | 13.70%                     | 10,000                            |
| Marshall Islands | 53,132                 | 20,000                     | 37.60%                     | 20,000                            |
| Micronesia       | 105,566                | 56,193                     | 53.20%                     | 21,000                            |
| Nauru            | 10,301                 | 5,152                      | 50.00%                     | 2,700                             |
| New Caledonia    | 269,736                | 240,000                    | 89.00%                     | 150,000                           |
| New Zealand      | 4,604,871              | 4,084,520                  | 88.70%                     | 3,000,000                         |
| Niue             | 1,614                  | 1,100                      | 68.20%                     | 700                               |
| Norfolk Island   | 2,210                  | 770                        | 34.80%                     | 420                               |
| Northern Marianas| 55,567                 | 29,000                     | 52.20%                     | 29,000                            |
| Palau            | 21,726                 | 7,700                      | 35.40%                     | 6,400                             |
| Papau New Guinea | 7,933,841              | 906,695                    | 11.40%                     | 380,000                           |
| Pitcairn Islands | 49                     | 170                        | n/a                        | 150                               |
| Samoa            | 195,743                | 68,000                     | 34.70%                     | 68,000                            |
| Smaller Territories (4) | 3,902               | n/a                        | n/a                        | n/a                               |
| Solomon Islands  | 606,215                | 66,684                     | 11.00%                     | 41,000                            |
| Terres Australes | n/a                   | n/a                        | n/a                        | n/a                               |
| Tokelau          | 1,300                  | 800                        | 61.50%                     | 150                               |
| Tonga            | 107,797                | 49,822                     | 46.20%                     | 43,000                            |
| Tuvalu           | 9,975                  | 5,000                      | 50.10%                     | 2,300                             |
| Vanuatu          | 276,331                | 82,764                     | 30.00%                     | 34,000                            |
Table 6 above expresses the total population of Oceania in 2017, the total population of Internet Users in Oceania, the total number of Facebook Users in the area and the percentage of internet penetration in the region. The presented tables above can be summarised by considering the six regions population, we consider internet users and population of the six regions, since Facebook Users are internet users.

Table 7. World population and internet users (Summarised Table 1 – 6).

| World Regions | Population (2017 Est.) | Internet Users 31-Dec-2017 | Penetration (% Population) |
|---------------|------------------------|-----------------------------|----------------------------|
| Africa        | 1,287,914,329          | 453,329,534                 | 35.20%                     |
| Asia          | 4,207,588,157          | 2,023,630,194               | 48.10%                     |
| Europe        | 827,650,849            | 704,833,752                 | 85.70%                     |
| America       | 1,015,892,658          | 782,662,124                 | 71.70%                     |
| Middle East   | 254,438,981            | 164,037,259                 | 64.50%                     |
| Oceania       | 41,273,454             | 28,439,277                  | 69.60%                     |
| WORLD TOTAL   | 7,634,758,428          | 4,156,932,140               | 54.40%                     |

The above table is a summary of the data presented as in tables 1 – 5, from the table we can calculate the six regions internet penetration rate in percentage which is closed to 50% and table 6 and the data would be more meaningful if we used some pictures to represent it.

The portion with the highest population is Asia which covered more than 50% of the whole world population, followed by Africa with 17%, America 13%, Europe 11%, Middle East 3% and the smallest portion of Oceania with 1%. This figure would not be interpreted in relation to internet users without considering figure 3.
Figure 3. Pie chart representing the percentage of regional internet users all over the world, updated on 31 December 2017.

Oceania is the portion with the smallest population which has only 1% of the Internet Users population in the world, followed by Middle East with 4%, America 19%, Europe 17%, Africa 11% and the largest portion of Asia with 48%. An estimate of 388,376,491 internet users in Africa was obtained in 31 December 2017 which represents 11.0% of the total world internet users. And the remaining 89.0% represents the rest of the world. Asia covered almost approximately 50% of the population of the world internet users.

Figure 4. Bar chart representing the world population (series1) and that of internet users (series2) by regions.
Figures 4 and 5 above almost have same meaning only that different chart were used, from the charts Asia has the highest number of internet users as well as population and Oceania has the smallest number of internet users and the Population. This occurs as a result of Asia population is almost 50% of the world population see figure 2 above. America has closed equal population of internet users and its population. This means that most of the people residing in America used internet. Same thing happen in European Union but in Africa the between the internet users and the population gap existed see figures 4 and 5. This means closed to half of the African population are not using internet due to some reasons discussed earlier in the paper.

5. Data Discussion

The data presented in tables 1 – 6 indicated the estimated population of six regions in the world by countries, the number of internet users as from at 31 December 2017, the percentage of penetration for each region, and the number of Facebook users as of 2017 in each area. Table 7 summarised all the tables presented which shows the total number of internet users in each region in relation to its population. We also compare each region with the rest of the world example America which represents 18 per cent of the total world internet users. The remaining 82 per cent represents the rest of the world [4]. Table 1 also show that Nigeria as part of Africa has 47.7 percent of internet penetration with estimated population of 91,598,757 of internet users, [4] indicates that E-Statistics is suitable in respect to internet user in Nigeria. Table 7 and figure 2 summarised the whole world internet users in relation to the whole world population. The pie chart shows this in terms of percentage. Figures 4 and 5 almost have same interpretations where they differ is another charts were used. From the charts Asia has the highest number of internet users and population and Oceania has the smallest number of internet users as well as the Population. This occurs as a result of Asia population is almost 50% of the world population see figure 2 above. America has closed equal population of internet users and its population. This means that most of the people residing in America used internet. Same thing happen in European Union but in Africa the figures 4 and 5 showed the gap between the internet users and the population. This means closed to half of the African population are not using internet due to some reasons discussed earlier in the paper. Middle East has no much gap in the used of internet compared to its population. Based on this majority of the region in the world have been using internets.
5.1. Data Analysis and Discussion of the Result
The data were analysed using SPSS computer software, Version 16.0, one way Analysis of variance (ANOVA) used to compare variance between the six different groups of internet Users. That is between groups with variability between each of the groups. The result:

An F-ratio represents the variance between the groups divided by variance within the groups. A significance F-test enables us to reject the null hypothesis, which states that the Groups mean are equal but it does not indicated to us which of the group differ. To determine this we would conducted a post-hock analysis. We would use one way ANOVA with Post-hock tests at $\alpha=0.05$ to see whether there is any significance difference in the population of internet users in the six region of the world.

In the above table the number of sub-region is Oceania has 28 regions, Middle East has 14, America has the minimum number of region and Africa has the highest number of region. America is the region with the highest Mean value of 1.81E8 which is equal to 181,000,000 (One hundred and Eighty one million Average of Internet Users) and the region with smallest mean is Oceania, its mean value is 1.01E6 which is same as 1,010,000 (One Million Ten Thousand Average Internet Users).

From table 10, since the significance value (0.944) is greater than the level of significance 0.05 we did not violate the assumption of homogeneity of variance.
From the above ANOVA Table we have between groups and within groups’ sum of squares, degree of freedom (df). The significance value is 0.001 and it’s less than the level of significance (0.05), then we reject the null hypothesis and concluded that there is significance differences among the means of the groups (regions) and the regions’ Internet users Population are significantly difference from one another.

![Image of Mean plot of six region internet users’ population.](image)

**Figure 6.** Mean plot of six region internet users’ population.

Figure 6 above as the name implies Means Plot, this provide a graphical picture for the comparison of the means of the six regions in the world in which America has the highest mean value of the internet users while Oceania is the one with the lowest mean value.
The above Multiple Comparisons Table indicated that there is significant difference of the groups or regions. Then from the post-hoc test we will know exactly from where the difference among the groups occurs, and are those with sig values less than 0.05 or with asterisk (*) this shows that the means of the six regions being compared are significantly different from another at $p<0.05$.

5.1.1. Calculation of the effect size (Eta Value)
Eta-squared denoted by $\eta^2$ is a measure of effect size used in Analysis of variance (ANOVA). $\eta^2$ is equivalent to $R^2$ from multiple linear regression. In case of ANOVA we can calculate (Eta-square) $\eta^2$ is the value of Sum of Square between divided by Sum of square total [21].

Mathematically

$$\eta^2 = \frac{SSB}{SST}$$  \hspace{1cm} (3)

Where, SSB = Sum of Square between and SST = Sum of square total and this can be calculated traditionally using one way ANOVA formulas. In regression analysis $\eta^2$ is a proportion of variance in $Y$ explained by $X$ that is it measures the proportion of the total variance in a dependent variable that is associated with the membership of different groups defined by an independent variable. Eta-square can also be considered as Non-linear correlation coefficient. The eta-squared column in SPSS F-table output is actually partial eta-squared ($\eta^2$) in some versions of SPSS prior to V 11.0. [14] $\eta^2$ was not previously provided by SPSS but for the used version It can only be calculated manually: $\eta^2$ is the ration of Between-Groups Sum of Squares and Total Sum of Squares and these values can be obtained directly from Table 8 (ANOVA Table) $R^2$ is provided at the bottom of SPSS F-tables is the linear effect as per Multiple Linear Regression, however, in some instances this would not equate with $\eta^2$. In this case we would use Eta square value to measures effect of regions means on the grand mean of the World Population of Internet Users, the Values of Eta square are interpreted as follows: 0.01 – 0.59 indicates small effect, 0.06 – 0.13 indicates moderate effect and 0.14 – above 0.14 indicates large effect [14]. Using equation (3) above by substituting the values of SSB and SST from table 8 (ANOVA) the value of Eta-Square ($\eta^2$) can be calculated as $\eta^2 = 0.170$. This value of Eta-Square is categorised as above 0.14 which can be interpreted as large effect. Therefore, we can say that at least one of the treatment or sub group mean of six regions or sub-population of the internet users has a very large effect on the grand mean of the internet users’ population.

6. Conclusions and future research

E-Statistics is a compendium; an elementary introduction to a growing field of statistics. In this paper we emphasised on defining the concept of E-statistic and identified some online techniques as the most appropriate, faster and useful methods for data collection, with no doubt in mind E-Statistic is considered efficient and effective due to its ability to ensure the integrity of the data collected and therefore the most reliable data source. It is a reliable data source because about 50% of the whole world population in 2017 are internet users [4] whom can be reached online. This is a good number as compared to sample proportion which is less than half of the sample population. The standard statistical inference procedures were used in the data analysis. The result of the analysis shows that there is significance difference among the population of the internet users from different regions and sub regions of the World. Eta-square value proved that the difference is great, from an indication of size of the effect is large. We learned that some countries advanced in using internet while others are still catching up. Some challenges were identified, and recommendations were made which when achieved or implemented will be useful in advancing the area of E-Statistics in the future, this serves as future research direction in the field of statistics. Other personal data could be considered in future analysis like age, gender etc. We also learned that based on the collected and analysed data, E-Statistics could be used as a reliable source of data more than that of the traditional sampling techniques.

7. Recommendation

- The creation of valid and up to date E-statistics web site under National Statistics Department or National Bureau of Statistics should be compulsory for every country including Nigeria.
- There is a need for intelligent system in future that will code text and voice information in the E-Statistics for easy statistical data analysis.
- All data should be centralised in a world database for easy access.
- The sampling technique is still valid since the e-data does not cover the whole population, this should also be included by the development agencies to cover the population.
• Experts in the field of information engineers, software engineers, computer scientists, data scientists, and statisticians should put more effort in the design and implantations of E-statistics.
• More features of data can be added in future design

Acknowledgments
We would like to thank the reviewers of this paper for taking their time to go through the manuscript and makes the necessary corrections and observations which shaped this writing to be considered as more statistically relevant, special thanks goes to our sponsor TETFUND, Nigeria for financing our conference through our Institution (BASUG). We appreciated the commitments of UPM and its members of the organising committee of this conference for the work well-done. Thank you all.

References
[1] Singapurwalla D 2013 A Hand Book of Statistic: An Overview of Statistical Method 1st Edition (online textbook, bookboon.com ISBN 978-87-403-0542-5) 8-65
[2] Shayib M A 2013 Applied statistics 1st Edition (online textbook, bookboon.com. ISBN 978-87-403-0493-0) 8-40
[3] Sullivan M 2013 Statistics, Informed Decision Using Data 4th Edition (Pearson, New York)
[4] De Argaez E 2017 Internet Usage And World Population Statistics (retrieved from http://www.internetworldstats.com)
[5] Adeyemo A B 2011 E-government Implementation In Nigeria – An Assessment Of Nigeria’s Global E-Gov Ranking (University of Ibadan, Nigeria, Electronic Journal of e-Government)
[6] Denis A 2011 Introduction to Statistics for Health, Life and Social Sciences 1st Edition (online textbook bookboon.com, ISBN 978-87-7681-740-4) 8-65
[7] Schwester R 2009 Examining the Barriers to e-Governance Adoption (Electronic Journal of e-Government) 7(1)
[8] United Nations 2008 United Nations e-governance Survey 2008 – From e-governance to Connected Governance (New York: Department of Economic and Social Affairs, Division for Public Administration and Development Management)
[9] Arrigo et al 2016 A multivariate approach to Facebook data for marketing communication Proceedings of the 1st International Conference on Advanced Research Methods and Analytics (CARMA 2016) (UPV Press, Universitat Politécnica de València, Valencia, Spain 10.4995/CARMA2016.2016.2974)
[10] Askitas and Zimmermann 2015 The internet as a data source for advancement in social sciences (Int. J. Manpow, 10.1108/IJM-02-2015-0029) 36(1) 2-12
[11] Blazquez and Domenech 2017 Web data mining for monitoring business export orientation Technol. Econ. Dev. Econ., (Online, 10.3846/20294913.2016.1213193) 1-23
[12] Chen M et al 2014 Big data: a survey Mob. Netw. Appl. 19(2) 171-209 (10.1007/s11036-013-0489-0)
[13] Corti L et al 2014 Managing and Sharing Research Data: A Guide to Good Practice 1st Ed (Sage Publications)
[14] John T E and Richardson 2011 Eta squared and partial eta squared as measures of effect size in educational research Educational Research Review 6(2) 135-147
[15] Einav L and Levin J 2014 The data revolution and economic analysis Innov. Policy Econ. 14(1) 1-24 (10.1086/674019)
[16] Baker R et al 2013 Report on the AAPOR Task Force on Non-Probability Sampling (AAPOR Report May)
[17] Bethlehem J 2011 Handbook of Nonresponse in Household Surveys (Wiley)
[18] Buelens B 2012 Shifting Paradigms in Official Statistics: From Design-Based to Model-Based to Algorithmic Inference (Discussion paper 201218, Statistics Netherlands)
[19] Breiman L 2001 Statistical Modelling: the two cultures *Statistical Science* **16**(3) 199-231
[20] Cochran W G 1977 *Sampling Techniques Third Edition* (New York: Wiley)
[21] Montgomery D C 2012 *Design and Analysis of Experiments 8th Edition* (ISBN: 978-1-118-14692-7: Wiley) 5-40