**Journal of Physical Science and Innovation** **Volume 7, No. 2, 2015**
**ISSN: 2277-0119**

# STATISTICAL ANALYSIS OF BIOLOGICAL DATA

[1]Aboko,Simon Igboye, [2]Amadi, Innocent Uchenna, [3]Mazi,Yellow Dimkpa, and [4]Anyamele Bethel Azunna

*[1&2] Department of Mathematics/Statistics, Rivers State College of Arts Science, Rumuola, Port Harcourt.*
*[3]Department of Mathematics/Statistics, Rivers State Polytechnic, Bori, Port Harcourt.*
*[4]Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Port-Harcourt.*
***Email**:amadiuchennainno@gmail.com.*

## ABSTRACT

Mann-Whitney test is a nonparametric counterpart of the t-test for the equality of two means for normal distributions. This work developed a practical approach of investigating infant Male and infant Female deaths in University of Port Harcourt teaching hospital for the period of twenty years. We discovered that there is no significance yearly difference in Mortality of infants at $\alpha$=0.05 and 0.01. This Model is found to be adequate

**Keywords: Nonparametric Test, Parametric, Mann-Whitney Test, Infant Mortality.**

## INTRODUCTION

In Statistical analysis parametric tests such as z,t,F, etc are used in testing hypothesis about population parameters. The outcome of their application will be valid if certain conditions are satisfied. One of the conditions is that measurements should be normally distributed. In most cases we know very little about the population from which samples are drawn. When available clues indicate radical departures from these conditions we have recourse to tests which utilize only the ordinal features of the data and which do not specify the form of the parent distribution, Nwobi(2003). Statisticians have developed alternative techniques which have become known as parametric methods. This is called distribution-free methods where we make no assumptions about the populations, except perhaps that they are continuous. It also includes methods which are nonparametric only in the sense that we are not concerned with parameters of populations of a given kind.

However, nonparametric methods, is very easy to handle , its computational burden is so light. Because of this reasons nonparametric methods have become more popular. Examples of nonparametric tests are the sign test, the signed-rank test ,the rank sum or Mann-whitney u-test, median test and kruskal- wallis H test. When we cannot measure outcomes of an experiment directly, they are ordered or ranked and nonparametric methods are used to analyze them. The aim of this research is to investigate infant Male and infant Female death using Mann Witney test as a tool.

### Advantages of Non-Parametric Methods

- They are easy to explain and understand.
- They are simple and less complex.
- -They require few assumptions about the shape of the population and their desirable properties remain unchanged under a wide variety of population types.

### Disadvantages of Nonparametric Methods

- They are not in position to infer as much as test that make use of known properties of a distribution.
- They avoid much sample information.
- They are less efficient than the parametric methods.

## MATERIALS AND METHODS

The data for this research is a record for infant males and infant females from University teaching hospital for the period of twenty years.

### Mann-Whitney U-test

This test was developed based on ranks in testing the equality of two populations .Since then a lot of literature on the optimal properties of his tests have been investigated.
Some of these results can be found in Bradley(1964) and Wilcoxon(1964).
They Mann-whitney U-test is is also known as the Wilcoxon rank sum test.

### Model Formulation

- Rank all $n_1+n_2$ observations from the smallest to the largest observations where $n_1 \leq n_2$. Note the sizes need not to be equal.
- Determine $T_1$ and $T_2$ , the sums respectively of ranks of the observations in samples 1 and 2
- For $n_1$ and $n_2 = n \geq 12$ , as is the case in practical problems, the test statistic becomes Z ,such that

$$Z = \frac{R_1 - \mu_{Ri}}{\sigma_{R1}}$$

Where the mean, $U_{R1} = \frac{n_1\,(n_1 + n_2 + 1)}{2}$ and standard deviation

$$\sigma_{R1} = \sqrt{\frac{n_1\,(n_1 + n_2 + 1)}{12}}$$

with this distribution we see that Z is approximately distribution.
The hypothesis is $H_0{:}n_1=n_2$ against $H_1{:}n_1\neq n_2$. $H_0$ is rejected if $Z<-Z_{\alpha/2}$ or $Z> Z_{\alpha/2}$.
If in a one –sided test as $H_0{:}n_1 = n_2$ versus $H_1{:}n_1>n_2$ ,if in a one sided test as $H_0{:}n_1\leq n_2$ against $H_1{:}n_1> n_2$.$H_0$ is rejected if $Z> Z_\alpha$.Note that $n_1= n_2$ is equivalent to $n_1\leq n_2$.if $H_0{:}n_1\geq n_2$ versus $H_1{:}n_1<n_2$,then we reject $H_0$ if $Z <Z_\alpha$.
The Mann-Whitney test is selected in preference to the t-test.

The table 1: below shows records of infant male and infant female deaths. Source of data: University of Port-Harcourt teaching Hospital. (1995-2014)

| years | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
|---|---|---|---|---|---|---|---|---|---|---|
| No of Males | 15 | 15 | 13 | 14 | 12 | 7 | 10 | 13 | 21 | 18 |
| No of Females | 13 | 14 | 16 | 22 | 9 | 12 | 8 | 14 | 18 | 16 |

| years | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No of Males | 10 | 20 | 21 | 15 | 16 | 12 | 11 | 15 | 9 | 16 | 283 |
| No of females | 8 | 17 | 20 | 12 | 16 | 13 | 12 | 13 | 11 | 10 | 274 |

$H_0: n_1 = n_2$ Vs $H_1: n_1 \neq n_2$

***Significance level***

$\alpha = 0.05$ and $0.01$

***Test Statistic***

$$\mu_{R1} = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma^2_{R1} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}, \qquad Z = \frac{R_1 - \mu_{Ri}}{\sigma_{R1}}$$

-Decision Criteria

$H_0$ is rejected if

$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$.

**Table 2: The Combine Observations and their Ranks where xi Represents the Individual Numbers Observed and Ri Represents their Corresponding Ranks.**

| S/N | Xi | Ri | S/N | Xi | Ri | S/N | Xi | Ri | S/N | Xi | Ri |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | (1) | 11 | 12 | (13) | 21 | 14 | (22) | 31 | 16 | 30 |
| 2 | 8 | 2.5 | 12 | 12 | (13) | 22 | 14 | 22 | 32 | 16 | 30 |
| 3 | 8 | 2.5 | 13 | 12 | 13 | 23 | 14 | 22 | 33 | 17 | 33 |
| 4 | 9 | (4.5) | 14 | 12 | 13 | 24 | 15 | (25.5) | 34 | 18 | (34.5) |
| 5 | 9 | 4.5 | 15 | 12 | 13 | 25 | 15 | (25.5) | 35 | 18 | 34.5 |
| 6 | 10 | (7) | 16 | 13 | (18) | 26 | 15 | (25.5) | 36 | 20 | (36.5) |
| 7 | 10 | (7) | 17 | 13 | (18) | 27 | 15 | (25.5) | 37 | 20 | 36.5 |
| 8 | 10 | 7 | 18 | 13 | 18 | 28 | 16 | (30) | 38 | 21 | (38.5) |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 11 | (9.5) | 19 | 13 | 18 | 29 | 16 | (30) | 39 | 21 | (38.5) |
| 10 | 11 | 9.5 | 20 | 13 | 18 | 30 | 16 | 30 | 40 | 22 | 40 |

The use Of the Z Statistic

$n_1$=20 and $n_2$=20,n=40>12

$R_1$=1+4.5+7+7+9.5+13+13+18+18+22+25.5+25.5+25.5+25.5+30+30+34.5+36.5+38.5+38.5 =423.

$R_2$ =2.5+2.5+4.5+7+9.5+13+13+13+18+18+18+22+22+30+30+30+33+34.5+36.5+40. =397.

$$\mu_{R1}= \frac{n_1\,(n_1+n_2+1)}{2}$$

$$= \frac{20(20+20+1)}{2}$$

$$= \frac{20(41)}{2}$$

$$= \frac{820}{2}$$

$\mu_{R1}$=410.

$$\sigma^2_{R1} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$= \frac{20\,x\,20(20+20+1)}{12}$$

$$= \frac{160400}{12}$$

=13366.67.

$\sigma_{R1}$=115.6.

$$Z = \frac{R_1 - \mu_{Ri}}{\sigma_{R1}} = \frac{423-410}{115.6} = \frac{13}{115.6} = 0.1125$$

$H_0$:$n_1$=$n_2$ against $H_1$:$n_1 \neq n_2$.

Z= 0.1125.

$\alpha$=0.05 and 0.01, $Z_{\alpha/2}$ =$Z_{0.025}$ =1.96 and $Z_{\alpha/2}$ =$Z_{0.01}$ =2.58

## RESULT AND DISCUSSIONS

The two random samples were combined to form a single array of $n_1$+$n_2$ and differentiate those observations belongs to $R_1$ and $R_2$ size n by underlining them, so smallest value taking rank 1. The result offers a flexible tool for the study of infant male and female deaths. We discovered that there is no significance yearly difference in mortality of infants at

$\alpha$=0.05.Even when we also tried at $\alpha$=0.01,it also gave us the same result. Hence the model is effective and reliable.

## CONCLUSION

FOR $\alpha$=0.05

Since $Z<Z_{\alpha/2}$ (0.1125<1.96),We accept the null hypothesis $H_o:n_1=n_2$.we therefore conclude that there is no significance yearly difference in the mortality of infant males and infant females at $\alpha$=0.05 .

FOR $\alpha$=0.01

Decision: Since $Z<Z_{\alpha/2}$(0.1125<2.58).The same result above

## REFERENCES

F. Wilcoxon (1964), Some Rapid Approximate Statistical Procedures, Ledarle Laboratories, Pear River, New York.

F.N Nwobi (2003), Statistics 2: Introductory Inference, Supreme Publisher, Owerri, Nigeria.

I.D Essi (2009), Fundamentals of Statistics, Geocelia Publisher, Diobu Port Harcourt.

J.E Freund and Ronald E. Walpole (1987), Mathematical Statistics. Fourth Edition, Prentice hall of India private limited New Delhi-110001.

K.E Okore(2009),College Statistics: a focus on hypothesis testing ,published by EL'Demak ,Enugu, Nigeria.

R.A. Bradley (1964) Application of the modified triangle test in sensory differences trials, *Journal of Food Science* 29,688-672.

S.A Abdulazeez and Lasisi A.R (2013),On the analytical efficiency of the extended Wilcoxon matched pairs signed rank test .journal of physical science and innovation. Volume 5,NO2,pp 120-127.

**Reference** to this paper should be made as follows: Aboko, Simon Igboye, *et al* (2015), Statistical Analysis of Biological Data. *J. of Physical Science and Innovation,* Vol. 7, No. 2, Pp. 1 – 5.

**Biographical Notes**

**Aboko Igboye Simon** had his B.Sc. in Pure and Applied Mathematics in 1998 and M.Sc. in Applied Statistics in 2013 both from Rivers State University of Science and Technology. He is a member of the Nigerian Statistical Association. His research interests are in Time Series Analysis, Demographic Measures, Sampling and Survey Methods.

**Amadi, I. U.** had his HND in Computer Science from Rivers State Polytechnic in 2001, POD in Statistics from Imo State University in 2008 and M.Sc. in Statistics from Rivers State University of Science and Technology in 2013. He is a member of the Science Teachers Association of Nigeria. His research interests are in Time Series Analysis and mathematics of Finance.

**Mazi Yellow Dimkpa** had B.Sc. Mathematics in 1989 from University of Port Harcourt, Nigeria and M.Sc. Applied Statistics in 2013. He is a member of the Nigerian Statistical Association. He is presently a Principal Lecturer of Statistics in the Rivers State Polytechnic Nigeria. His research interests are in Mathematical Modeling.

**Anyamele Bethel** had his B.Sc in Statistics in 2002 and M.Sc. in Mathematics (Applied Statistics option) in 2014. His research interests are Econometrics and Time series analysis.