African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)
www.abjournals.org

ASSESSING THE PREVALENCE OF TUBERCULOSIS IN TARABA STATE

Nwaokolo Martin Afam and Okorie Charity Ebelechukwu

Department of Mathematics and Statistics, Federal University, Wukari

Corresponding author's email: okoriec636@gmail.com

Cite this article:

Nwaokolo M.A., Okorie C.E. (2023), Assessing the Prevalence of Tuberculosis in Taraba State. African Journal of Mathematics and Statistics Studies 6(4), 81-90. DOI: 10.52589/AJMSS-GEGK0SGR

Manuscript History

Received: 3 Sept 2023

Accepted: 24 Oct 2023

Published: 23 Nov 2023

ABSTRACT: *Tuberculosis is an old disease that poses a new threat and remains a public health issue in Nigeria, having been ranked as the second biggest cause of death from single infectious disease after HIV/AIDS. The world health organization initiated the stop TB programs and directly observed treatment short course centers to eliminate tuberculosis, yet tuberculosis in developing countries is still on the high side. Therefore, this work is primarily targeted at providing reliable and concrete information on the rate of occurrence and prevalent rate of TB in Taraba State. The aim of this research is to analyze the cases of TB in Taraba State. In order to carry out this research, data were collected from Specialist Hospital, Jalingo. The data collected were analyzed using the SPSS (Statistical Package for Social Science) software 16.0. Ordinary Least Square and Analysis of Variance (ANOVA) were used in carrying out the analysis. The analysis with the least square method yields the linear trend equation as* $y = 551.917 – 14.582x + e$, *where* $y = a + bx + e$. $X = 2(t-2013.5)$. *This trend indicates that there is a falling trend. For instance, using the Linear Trend to forecast for the future, we discovered that there will be a decrease in tuberculosis from 2023 with the value of 413.388, 2024=399.806, 2025=384.228.*

KEYWORDS: Tuberculosis, Disease, Significant, Least Square, Variance.

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)

www.abjournals.org

INTRODUCTION

Background of the Study

Tuberculosis (TB) is an infectious disease usually caused by Mycobacterium tuberculosis (MTB) bacteria. TB may infect any part of the body, but most commonly occurs in the lungs (known as pulmonary tuberculosis). Extra-pulmonary TB occurs when tuberculosis develops outside of the lungs. It occurs more commonly in people with a weakened immune system and young children. In those with HIV, this occurs in more than 50% of cases (Vikram & Golden, 2005). Notable extra-pulmonary infection sites include the pleura (in tuberculous pleurisy), the central nervous system (in tuberculous meningitis), the lymphatic system (in scrofula of the neck), the genitourinary system (in urogenital tuberculosis), and the bones and joints (in pott disease of the spine), among others. Most infections do not have symptoms in which case is known as latent tuberculosis; about 10% of latent infections progress to active disease which, if left untreated, kills about half of those affected. The classic symptoms of active TB are chronic cough with blood-containing mucus, fever, night sweats and weight loss (WHO, 2015).

Those at high risk include household, workplace and social contacts of people with active TB (WHO, 2008). Treatment requires the use of multiple antibiotics over a long period of time. Antibiotic resistance is a growing problem with the increasing rate of multi drug-resistant tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) (WHO, 2015). It was first observed in 2003 in Italy (Migliori G.B. et al., 2007), but not widely reported until 2012 (McKenna, 2012), and has also been found in Iran and India (Paul, 2014; Farnia et al., 2009).

Caminero et al. (2010) further categorizes the risks of acquiring drug-resistant TB into three sets of vulnerable groups. The first group are patients who based on bacteriological results are classified as being at high risk of drug-resistant TB. In a review conducted by (Faustini et al., 2006) in Europe, the most significant risk factor for drug resistant TB was unsuccessful previous treatment. The second group of patients are those who are at high risk of drug-resistant TB based on close contact with drug-resistant TB patients or patients classified as category 1 TB regimen failure. Close contacts with MDR-TB cases have been reported to have different resistant strains from the index case (WHO, 2011; Kritski, 1996). Another group of patients at moderate risk of acquiring MDR-TB are those who remain smear positive at the second or third month of treatment initiation (Chavez & Blank, 2004).

The risk of progression from latent infection to active disease is considerably higher in populations that therefore require special attention and targeted interventions (Getahun, 2015).

Drug resistance emerges quickly after the introduction of new anti-TB drugs (Bañuls et al., 2018). From a microbiological point of view, MDR-TB and XDR-TB are caused by genetic mutation of the M-Tuberculosis which renders anti-TB agents ineffective against the mutant tubercle bacilli (Migliori et al., 2010). However, Caminero et al. (2010) propose two categories of risk factors for drug resistant tuberculosis. The first category, they describe as those facilitating the selection of resistance in the community and the second as specific conditions that appear to increase some patients' vulnerability of resistance.

According to (CDC, 2016), there are five key components of a complete evaluation of tuberculosis disease. They are medical history, physical examination, test for M. Tuberculosis

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

infection, chest radiography and bacteriological examination of clinical specimens. The disease is usually chronic with varying clinical manifestations (Park, 2002).

Materials and Method

The methods used in order to meet the specific aim and objectives of the study are ANOVA, Percentage, Chart, and Ordinary least square method.

Least Squares Method

For the OLS method, linear trend is fitted to our data. Its variables are:

- The independent variable X, which is 2(t-2013.5)
- The dependent variable Y, which is cases.

The model describing the relationship between x and y can be express as:

$$Y = \beta_0 + \beta_1 X_i + e_i\,, \qquad for\ i = 1,2,3 \dots\dots\dots\dots n$$

where n = number of observations.

t = time in a year

X_i = known constant representing the i^{th} observation on X.

e_i = independent and normally distributed random error $\beta_o\ and\ \beta_i$ are parameters.

One of the methods usually employed to obtain the desired line of fit is known as the method of least square and the line obtained is called least square line. It is given by:

$$e_i = y_i - (\beta_o + \beta_1 x_i) \quad (1)$$

$$\sum_{i=1}^{n} \quad e_i^2 = \sum_{i=1}^{n} \quad (y_i - \beta_o - \beta_1 x_i)^2 \quad (2)$$

$$\frac{\partial}{\partial \beta_o}(e_i) = -2\sum_{i=1}^{n} \quad (y_i - \beta_o - \beta_1 x_i) \quad (3)$$

$$\sum_{i=1}^{n} \quad (y_i - \beta_o - \beta_1 x_i) = 0 \quad (4)$$

$$\sum_{i=1}^{n} \quad y_i = \sum_{i=1}^{n} \quad \widehat{\beta_o} + \widehat{\beta_1}\sum_{i=1}^{n} \quad x_i \quad (5)$$

$$\sum_{i=1}^{n} \quad \widehat{y_i} = n\widehat{\beta_o} + \widehat{\beta_1}\sum_{i=1}^{n} \quad x_i \quad (6)$$

$$\frac{1}{n}\sum_{i=1}^{n} \quad y_i = \widehat{\beta_o} + \frac{1}{n}\widehat{\beta_1}\sum_{i=1}^{n} \quad x_i \quad (7)$$

$$\underline{y} = \widehat{\beta_o} + \widehat{\beta_1}\underline{x} \quad (8)$$

$$\beta_o = \underline{y} - \beta_1\underline{x} \quad (9)$$

$$\frac{\partial \hat{\varepsilon}'\hat{\varepsilon}}{\partial \widehat{\beta_1}} = \sum_{i=1}^{n} \quad \left(y_i - \widehat{\beta_o} - \widehat{\beta_1} x_i\right) = 0 \quad (10)$$

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)

www.abjournals.org

$$\beta_1 = \frac{\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2} \quad (11)$$

The above formulas for estimating the value of β_0 *and* β_1 are used when t = 1,2,…,n given that t is use to number each point in the series.

In case of even number of years (for our research),

$$X = \frac{(t - arithmetic\ mean\ of\ two\ middle\ years)}{\frac{1}{2}}$$

To complete the model, we make the following additional assumptions:

1. $E(\mathcal{E}_i) = 0$ For all i = 1, 2,..., n, or, equivalently, $E(y_i) = \beta_o + \beta_1 x_i$.

2. $var(\varepsilon_i) = \sigma^2$ For all i = 1, 2,…,n, or, equivalently, $var(y_i) = \varepsilon^2$. It is also known as the assumption of homoscedasticity, homogeneous variance or constant variance.

3. $cov(\varepsilon_i, \varepsilon_j) = 0$. For all i= j, or, equivalently, $cov(y_i, y_j) = 0$.

Analysis of Variance (Anova)

The ANOVA is a powerful statistical tool for tests of significance; the test of significance based on t-distribution is an adequate procedure only for testing the significance difference between two sample means. In a situation when we have three or more samples to be tested, ANOVA is used. The basic purpose of ANOVA is to test the homogeneity of several means.

We have the ANOVA for:

- One-way classification
- Two-way classification

For the sake of our study we will be considering only one-way classification.

Terminologies Related to ANOVA

- **Mean**

Mean is a simple or arithmetic average of a range of values. There are two kinds of mean that we use in ANOVA calculations, which are separate sample mean and the grand mean. The grand mean is the mean of sample means or the mean of all observations combined, irrespective of the sample.

$$\text{Grand mean } \underline{X} = \frac{1}{n}\sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} = \frac{\sum_{i=1}^{k} [\sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^{k} n_i} \quad (12)$$

$$= \frac{\sum_{i=1}^{k} n_i \underline{X_i}}{\sum_{i=1}^{k} n_i} = \frac{n_1\underline{X_1} + n_2\underline{X_2} + \cdots + n_k\underline{X_k}}{n_1 + n_2 + \cdots + n_k} \quad (13)$$

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)

www.abjournals.org

- **Hypothesis**

A hypothesis is an educated guess about something in the world around us. It should be testable either by experiment or observation. Just like any other kind of hypothesis that you might have studied in statistics, ANOVA also uses a null hypothesis and an alternate hypothesis. The null hypothesis in ANOVA is valid when all the sample means are equal, or they do not have any significant difference.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \textbf{\textit{Null hypothesis}}$$

$$H_1: \mu_1 \neq \mu_m \textbf{\textit{ Alternate hypothesis}}$$

where $\mu_l \neq \mu_m$ belong to any two sample means out of all the samples considered for the test.

- **Between Group Variability (Treatment Sum of Square)**

It refers to variations between the distributions of individual groups (or levels) as the values within each group are different. Each sample is looked at and the difference between its mean and grand mean is calculated to calculate the variability. If the distributions overlap or are close, the grand mean will be similar to the individual means, whereas if the distributions are far apart, the difference between the mean and grand mean would be large.

$$S_t^2 = \sum_{i=1}^{K} \quad \sum_{j=1}^{n_i} \quad \left(\underline{X}_i - \underline{X}\right)^2 = \sum_{i=1}^{k} \quad n_i\left(\underline{X}_i - \underline{X}\right)^2 \tag{14}$$

- **Within Group Variability (Sum of Square Error)**

It refers to variations caused by differences within individual groups (or levels) as not all the values within each group are the same. Each sample is looked at on its own and variability between the individual points in the sample is calculated. In other words, no interactions between samples are considered. We can measure within-group variability by looking at how much each value in each sample differs from its respective sample mean. So first, we will take the squared deviation of each value from its respective sample mean and add them up. This is the sum of squares for within-group variability.

$$S_e^2 = \sum_{i=1}^{k} \quad \sum_{j=1}^{n_i} \quad \left(X_{ij} - \underline{X}_i\right)^2 = \sum_{i=1}^{k} \quad nS_t^2 \tag{15}$$

- **Total Sum of Square**

This is the total sum of all the groups or treatments, and it is expressed as:

SST = Between group + Within group

$$S_T^2 = \sum_{i=1}^{k} \quad \sum_{j=1}^{n_i} \quad \left(X_{ij} + \underline{X}\right)^2 \tag{16}$$

- **Mean Sum of Square (MSS)**

The sum of squares divided by its degree of freedom gives the corresponding variance or mean sum of squares.

$$\text{MSS due to treatments} = \frac{SS\ due\ to\ treatment}{df} = \frac{S_t^2}{k-1} \tag{17}$$

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)

MSS due to error = $\frac{SS\ due\ to\ error}{df} = \frac{S_e^2}{n-k}$ (18)

- **F-Statistic**

The statistic which measures if the means of different samples are significantly different or not is called the F-Ratio. Lower the F-Ratio, more similar are the sample means. In that case, we cannot reject the null hypothesis.

F = Between group variability / Within group variability

$F = \frac{s_t^2}{s_e^2}$ (18)

F follows F-distribution with $(V_1 = K - 1, V_2 = n - k)\ df$.

Anova Table for One-Way Classification Data

Sources of Variation	**Sum of Square**	**Df**	**Mean Sum of Square**	**F-Statistics**
Between group	S_t^2	k-1	$\frac{S_t^2}{k-1}$	$F = \frac{s_t^2}{s_e^2}$
Within group	S_e^2	n-k	$\frac{S_e^2}{n-k}$	
Total	S_T^2	n-1		

Assumption for the model:

i. All observation X_{ij} are independent and ij: N(μ_{ij}, σ_e^2).

ii. Different effects are additive in nature.

iii. E_{ij} are independent identically distributed with N(0,σ_e^2).

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 4, 2023 (pp. 81-90)

RESULTS

The analysis of data was performed using SPSS version 16.0. A linear trend was fitted to analyze the trend of the disease from 2008-2019. The age difference is analyzed using ANOVA and the trend within various age groups is recorded.

Table 1: Coefficients

	Unstandardized Coefficients		**Standardized Coefficients**		
	B	Std. Error	Beta	t	Sig.
time	-14.582	10.269	-.410	-1.420	.186
(Constant)	551.917	70.898		7.785	.000

Linear Trend Coefficients

The linear trend equation is y = 551.917 – 14.582x + e, where y = a + bx + e. X = 2(t-2013.5), the slope = b is -14.58 and the intercept, a = 551.92. The slope of the trend line represents the rate of decrease in TB per unit time, i.e., annually. Since the slope is negative (-14.58), the data exhibit a falling trend.

Table 2: TB Trend Values

	Cases (y)	**PREDICTED VALUE ($\hat{y}$)**	**RESIDUAL** $e = \hat{y} - y$
	172	712.3205	-540.3205
	901	683.1562	217.8438
	846	653.9918	192.0082
	568	624.8275	-56.8275
	898	595.6632	302.3368
	679	566.4988	112.5012
	352	537.3345	-185.334
	618	508.1702	109.8298
	665	479.0058	185.9942
	321	449.8415	-128.8415
	308	420.6772	-112.6772
	295	391.5128	-96.5128
Total	6623	6623	0.0005

Table 3 contains the trend values for all the given time periods in Graph 1 and also the residual (the vertical distance between any data point and the regression line). The negative values show

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

that the values are below the regression line while the positive shows that the values are above the regression line, and also the sum of our residual 0.0005 shows that predicted values does not differ from actual value (cases), i.e., there are no other factors influencing TB cases.

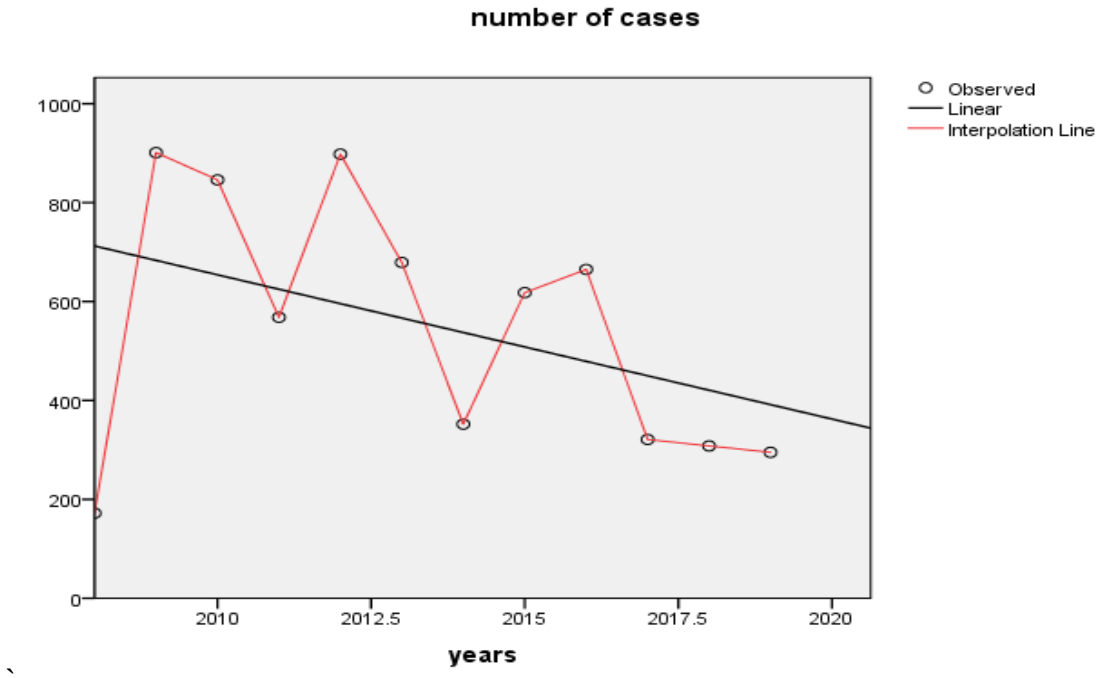

Figure 1: TB Trend over the Years

Figure 1 shows that there is a relative increase in TB cases from 2008 to 2009, then a slight decrease to 2010; it continues falling and rising till 2017. Which means that the data is fluctuating.

The descriptive table shows that the mean for male patients in 12 years is 322.67 and for females is 229.25, which is almost close with a difference of 93.42. The standard deviation for male patients is 179.242 with standard error mean of 49.145, and that of female patients is 115.366 with standard error mean of 33.303.

Decision Rule

Table 3: Descriptive Statistics with Respect to Gender

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
PATIENT	MALE	12	322.67	170.242	49.145
	FEMALE	12	229.25	115.366	33.303

If p-value > α-level at 95% level of significance (i.e., $= 0.05$), we do not reject H_0 and conclude that there is no significant difference in the prevalence of TB with respect to age.

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

Hypothesis

H_{o2}: There is no significant difference in the prevalence of TB according to age.

H_{12}: There is a significant difference in the prevalence of TB according to age.

Table 4: ANOVA Test for Prevalence of TB with Respect to Age

PATIENTS

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	149639.069	5	29927.814	9.808	.000
Within Groups	201399.917	66	3051.514		
Total	351038.986	71			

In Table 4, the F value, which is the between-group mean square, is large at 9.808 and is significant at P = 0.0001 which is less than α-level of significance (0.05). Therefore, the null hypothesis is rejected.

CONCLUSION

From our finding, there is a fluctuation in the rate of prevalence of TB, which means that there is a tendency that it can increase or decrease at any time. Using the Linear Trend to forecast for the future, we discovered that there will be a decrease in tuberculosis from 2023 with values of 413.388 (2023), 399.806 (2024), and 384.228 (2025).

REFERENCES

CDC. Tuberculin Skin Testing. CDC, 2016 [August 15 2016]. Available online: http://www.cdc.gov/tb/publications/factsheets/testing/skintesting.html

Centers for disease control and prevention (CDC).13 March 2012.Archived from the original on 6 February 2016.

Centers for Disease Control and Prevention. Managing tuberculosis patients and improving adherence. Atlanta, GA: CDC; 2014.

Vikram HR,Golden MP(November 2005). "Extrapulmonary Tuberculosis: an overview". American family physician.72(9): 1761-68.

Harris, Randall E. (2013). Epidemiology of chronic disease: global perspective Burlington, MA: Jones and Bartlett learning. P. 682. ISBN 978-0-7637-8047-0.

Hawn TR,Day TA, Scriba TJ, et al. (December 2014).″Tuberculosis vaccines and prevention of infection". Microbiology and molecular biology Reviews.78(4):71

World Health Organization. Global tuberculosis report 2016. Geneva: World Health Organization; 2016.

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR

World Health Organization. Global tuberculosis report 2021: World Health Organization; 2021.
World Health Organization. Global tuberculosis report 2016. Geneva: World Health Organization; 2016.

Article DOI: 10.52589/AJMSS-GEGK0SGR
DOI URL: https://doi.org/10.52589/AJMSS-GEGK0SGR