

# Research paper (predicting the number of internet users)



**ASSIGN  
BUSTER**

This paper is on internet adoption and penetration and as being influenced by socio-economic development and other demographics of a country (Mustafa2001). According to him, the research was aimed at establishing the variables which explain and predict internet growth in Africa.

### Hypothesis

To Mustafa (2001), the research sole purpose was to proof the below hypothesis;

H0: Internet adoption and penetration is influenced by socio-economic development and other demographics of a country

H1: Internet adoption and penetration is not influenced by socio-economic development and other demographics of a country

### Variables

The independent variables as Mustafa (2001) puts it are;

Population;

Gross domestic product per capita;

Gross national product per capita;

Telecommunication revenues;

Telecommunication investments;

Telephone lines;

Telephone lines per capita;

Cellular users;

Internet service providers;

Access charges; and

Internet hosts.

The researcher aimed at predicting the number of internet users (the dependent variable) using the above explanatory variables. Data for both <https://assignbuster.com/research-paper-predicting-the-number-of-internet-users/>

explanatory variables and the response are secondary. In this case, one misgiving is that there is no indication of the data being used for the sole purpose of the research as it should according to the tenets of research ethics. Ethically, sharing of data for other purposes other than the intended is un-ethical (Callahan 1998)

### Methodology

To get a best predictor of the number of internet users, the researcher used multivariate linear regression. In this type of methodology, each of the predictor variable is modelled against the response variable, in this case the number of internet users. This process is carried over with different combinations of the explanatory variables and the values of R, coefficient of correlation, and R<sup>2</sup>, coefficient of determination for the different models are calculated. The model with the highest value of R is normally selected as the best fitting model for the data (Bryman 1992). R<sup>2</sup> explains the variations in the response variable readings.

In this case, the researcher used all the explanatory variables in the initial model and used the backwards which eliminates the variables which are not better placed to explain the response variable as anticipated. The only problem with this technique is that it may result in the elimination of explanatory variables even before their effects on the entire model have been determined. As a best practice, I suggest individual simple regression equations to determine the individual effects on the response variable and then stepwise inclusion of the variables (Hinton 1995). This method leads to inclusive and exhaustive analysis of the explanatory variables resulting to the best ever model fit.

Adjusted R is not a good explainer of any model since it has the effects of <https://assignbuster.com/research-paper-predicting-the-number-of-internet-users/>

correlation, multi-collinearity and does not explain fully the model. R, the correlation coefficient is the best determinant used in the explanation of models. On the other hand, the researcher is also using R<sup>2</sup> not to explain the variations in the response variable as it should but also in the determination of the best model fit.

In the formulation of the regression linear line, the researcher used the un-standardized coefficients of the explanatory variables and the constant for the equation. This should not be the case and if possible the coefficients should be standardized since standardized values make it better to compared models. The use of Exploratory Data Analysis is another methodology well applied in this research study. EDA is used mostly in both qualitative and quantitative research since they give impressions of the data in a comprehensible way which can be understood even by non-statisticians. From the dissertation provided, the better methodology practices which can be applied in future research include;

The use of multivariate regression analysis for nominal and interval data;

Elimination of less significant explanatory variables in models to best predict the response variable;

The use of log values of predictors to test normality of the data;

Some areas which need changes in future include;

The use of R as opposed to R<sup>2</sup> and adjusted R to determine the best fitting model;

The use of step-wise regression as opposed to enter and backwards techniques;

More emphasis on EDA-charts and graphs analysis;

Omissions of variables with missing data in analysis.

<https://assignbuster.com/research-paper-predicting-the-number-of-internet-users/>

In conclusion, multivariate regression analysis is a good statistical methodology as it utilizes available information to model the future since predicted values always have no statistical difference from actual values (Diamond and Jefferies 2001)

#### References

- Bryman, A (1992). Quantity and quality in social research 2nd Ed.
- Callahan, T. C (1998). Research Ethics: Ethics in Medicine. [Online] Available from: <http://depts.washington.edu/bioethx/topics/resrch.html>
- Diamond, I. & Jefferies, J. (2001). Beginning Statistics. An Introduction for Social Scientists. CA: Sage Publications
- Hinton, P. R. (1995). Statistics Explained. A Guide for Social Science Students. NY: Routledge
- Mustafa, S. S. (2001). Factors That Explain Internet Growth in Africa: An Empirical Model