



Socio-Economic Background of Households Spent on Medical Tourism

Manoj Kumar Srivastava^{1,*}, Namita Srivastava² and Sandeep³

1 Department of Statistics, Institute of Social Sciences, Dr. B.R. Ambedkar University, Agra, Uttar Pradesh, India.

2 Department of Statistics, St. John's College, Agra, Uttar Pradesh, India.

3 NITI AAYOG, Government of India, New Delhi, India.

Abstract: Linear regression is a factual technique for figuring the estimation of a reliant variable from an autonomous variable. Linear regression estimates the relationship between two factors. It is a demonstrating technique where a reliant variable is anticipated dependent on at least one autonomous factor. Linear regression investigation is the most generally utilized of every single factual technique. This article clarifies the essential ideas and clarifies how we can do linear regression counts in SPSS and exceed expectations. Statistics make it conceivable to examine true business issues with genuine information so you can decide whether an advertising procedure is truly working, how much an organization should charge for its items, or any of a million other reasonable inquiries. The study of statistics utilizes relapse examination, theory testing, sampling dispersions, and more to guarantee precise information investigation.

Keywords: Continuous variable test, excel and SPSS analysis, linear regression, statistics.

© JS Publication.

1. Introduction

The purpose of statistical evaluation of medical data is often to describe relationships between two variables or among several variables. For example, one would like to know not just whether patients have high blood pressure, but also whether the likelihood of having high blood pressure is influenced by factors such as age and weight. The variable to be explained (blood pressure) is called the dependent variable, or, alternatively, the response variable; the variables that explain it (age, weight) are called independent variables or predictor variables. Measures of association provide an initial impression of the extent of statistical dependence between variables. If the dependent and independent variables are continuous, as is the case for blood pressure and weight, then a correlation coefficient can be calculated as a measure of the strength of the relationship between them.

Sir Francis Galton first proposed the concept of linear regression in 1894. Linear regression is a statistical test applied to a data set to define and quantify the relation between the considered variables. Univariate statistical tests such as Chi-square, Fisher's exact test, *t*-test, and analysis of variance do not allow taking into account the effect of other covariates/confounders during analyses (Chang 2004). However, partial correlation and regression are the tests that allow the researcher to control the effect of confounders in the understanding of the relation between two variables.

In biomedical or clinical research, the analyst regularly attempts to comprehend or relate at least two autonomous (indicator) variables to anticipate a result or ward variable. This might be comprehended as how the hazard factors or the indicator

* E-mail: mkiss87@gmail.com

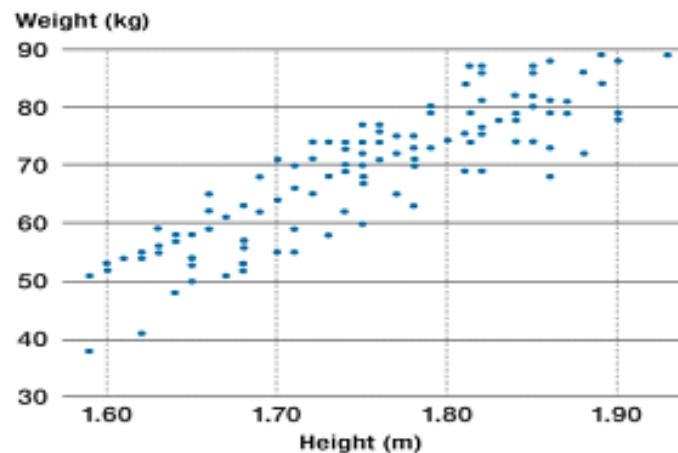
variables or autonomous variables represent the forecast of the opportunity of an infection event, i.e., subordinate variable. Hazard elements (or ward variables) partner with organic, (for example, age and sexual orientation), physical, (for example, weight record and pulse [BP]), or way of life, (for example, smoking and liquor utilization) variables with the malady. Both correlation and regression give this chance to comprehend the "hazard factors-malady" relationship. While correlation gives a quantitative method for estimating the degree or quality of a connection between two variables, regression examination scientifically portrays this relationship. Regression investigation permits foreseeing the estimation of a needy variable dependent on the estimation of in any event one autonomous variable.

In correlation examination, the correlation coefficient "r" is a dimensionless number whose worth reaches from -1 to $+1$. An incentive toward -1 shows reverse or negative relationship, while towards $+1$ demonstrates a positive connection. When there is a typical circulation, the Pearson's correlation is utilized, though, in no normally conveyed information, Spearman's rank correlation is utilized.

The linear regression examination utilizes the scientific condition, i.e., $y = mx + c$ that portray the line of best fit for the connection between y (subordinate variable) and x (free factor). The regression coefficient, i.e., r^2 infers the level of fluctuation of y because of x.

Linear regression models are utilized to appear or foresee the connection between two variables or components. The factor that is being anticipated (the factor that the condition unravels for) is known as the reliant variable. The components that are utilized to foresee the estimation of the reliant variable are known as the autonomous variables.

Great information doesn't generally recount to the total story. Regression investigation is usually utilized in inquire about as it builds up that a correlation exists between variables. In any case, correlation isn't equivalent to causation. Indeed, even a line in a straightforward linear regression that fits the information focuses well may not say something conclusive regarding a circumstances and logical results relationship.



In simple linear regression, every perception comprises of two qualities. One worth is for the needy variable and one worth is for the free factor. Simple Linear Regression Analysis The simplest type of a regression investigation utilizes on subordinate variable and one autonomous variable. In this simple model, a straight line approximates the connection between the reliant variable and the autonomous variable. Different Regression Analysis When at least two autonomous variables are utilized in regression investigation, the model is never again a simple linear one.

1.1. Simple Linear Regression Model

The simple linear regression model is spoken to like this: $y = (\beta_0 + \beta_1 + E$. By numerical show, the two factors that are engaged with a simple linear regression investigation are assigned x and y. The condition that depicts how y is identified with x is known as the regression model. The linear regression model likewise contains a mistake term that is spoken to by E , or the Greek letter epsilon. The blunder term is utilized to represent the fluctuation in y that can't be clarified by the linear connection among x and y. There likewise parameters that speak to the populace being contemplated. These parameters of the model that are spoken to by $(\beta_0 + \beta_1 x)$.

The simple linear regression condition is spoken to like this: $E(y) = (\beta_0 + \beta_1 x)$.

The simple linear regression condition is diagrammed as a straight line.

β_0 is the y catch of the regression line.

β_1 is the slant.

$E(y)$ is the mean or anticipated estimation of y for a given estimation of x.

A regression line can show a positive linear relationship, a negative linear relationship, or no relationship. On the off chance that the charted line in a simple linear regression is level (not inclined), there is no connection between the two variables. On the off chance that the regression line slants upward with the lower stopping point at the y capture (pivot) of the diagram, and the upper finish of line broadening upward into the chart field, away from the x catch (hub) a positive linear relationship exists. In the event that the regression line slants descending with the upper stopping point at the y catch (hub) of the chart, and the lower end of line expanding descending into the diagram field, close to the x block (hub) a negative linear relationship exists.

2. Business Statistics

At the point when you're working with populaces and tests (a subset of a populace) in business statistics, you can utilize three normal kinds of measures to portray the informational collection: focal propensity, scattering, and affiliation.

2.1. Measures of central tendency

In statistics, the mean, median, and mode are known as measures of *central tendency*; they are used to identify the center of a data set:

- **Mean:** The value between the largest and smallest values of a data set, obtained by a prescribed method.
- **Mode:** The most commonly observed value in a data set

Samples are arbitrarily looked over populaces. In the event that this procedure is completed effectively, each example ought to precisely mirror the attributes of the populace. Along these lines, an example measure, for example, the mean, ought to be a decent gauge of the relating populace measure. Think about the accompanying instances of mean:

1. Population mean
2. This formula simply tells you to add up all the elements in the population and divide by the size of the population.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

The procedure for figuring this is the very same; you include every one of the components in the example and gap by the size of the example. Notwithstanding proportions of focal inclination, two other key sorts of measures will be proportions of scattering (spread) and proportions of affiliation.

2.2. Evaluated Linear Regression Equation

In the event that the parameters of the populace were known, the simple linear regression condition (demonstrated as follows) could be utilized to figure the mean estimation of y for a known estimation of x .

$$E(y) = (\beta_0 + \beta_1 x).$$

Be that as it may, by and by, the parameter esteems are not known so they should be evaluated by utilizing information from an example of the populace. The populace parameters are evaluated by utilizing test measurements. The example measurements are spoken to by $b_0 + b_1$. At the point when the example insights are substituted for the populace parameters, the evaluated regression condition is shaped. The evaluated regression condition is demonstrated as follows.

$$(\hat{y}) = \beta_0 + \beta + 1x$$

(\hat{y}) is articulated y cap. The chart of the evaluated simple regression condition is known as the assessed regression line. The b_0 is the y block. The b_1 is the incline. The (\hat{y}) is the evaluated estimation of y for a given estimation of x . Regression examination isn't utilized to translate circumstances and logical results connections between variables. Regression examination can, in any case, show how variables are connected or to what degree variables are related with one another. In this manner, regression investigation will in general make striking connections that warrant an educated scientist investigating. Regression examination is a kind of measurable assessment that empowers three things:

	Application	Dependent Variables	Independent Variables
Linear Regression	Description of the linear relationship	Continuous (weight, blood pressure)	Continuous and/ or categorical
Logistic Regression	Prediction of the Probability of	Dichotomous (Success of treatment: yes/no)	
Proportional hazard regression (Cox regression)	Belonging to groups (outcome: yes/no) Modelling of survival data Modelling of counting processes	Survival time (time from diagnosis event) Counting data: whole numbers representing event in temporal sequence (e.g. a no. of times of women gave birth over a certain of time)	

- Description: Relationships among the reliant variables and the free variables can be measurably depicted by methods for regression investigation.
- Estimation: The estimations of the needy variables can be assessed from the watched estimations of the autonomous variables.
- Prognostication: Risk factors that impact the result can be distinguished, and singular Regression investigation utilizes a model that depicts the connections between the needy

variables and the free variables in a streamlined numerical structure. There might be organic motivations to expect from the earlier that a specific sort of numerical capacity will best portray such a relationship, or simple suppositions must be made this is the situation (e.g., that circulatory strain rises linearly with age).

2.3. Estimating Using Linear Regression

Backwoods items, for example, paper, boards and sawn wood, showing up in countless end use items, influence day by day lives. Moreover, in a few nations the woodland area contributes in a huge manner to the general financial and social advancement. The expanding volume of exchange and specifically the import of woods items into Europe, assumes a significant job likewise for worldwide backwoods segment improvements. A depiction of an econometric examination of the timberland area in Europe and how the subsequent models, together with presumptions in regards to financial development and cost and cost improvements, used to create nation explicit projections of utilization, generation and exchange of wood items are contemplated. The impacts of emptying and its effects on net incomes and all out expenses in various beneficial lines of Mazandaran Wood and Paper Company are researched. Adopted a model choice strategy to the topic of whether a class of versatile expectation models like Artificial Neural Networks is helpful for foreseeing future estimations of nine macroeconomic variables. An assortment of out-of-test estimate based model determination criteria including conjecture mistake measures and gauge bearing exactness is utilized. So as to contrast the proposed expectations with expertly accessible overview forecasts, a constant anticipating method is actualized.

The examination is an essential econometric investigation to investigate the elements clarifying the adjustments in the mechanical round wood request in Turkey. The investigation has included interest figures dependent on the econometric models proposed in this. Two separate econometric models are developed: one for national interest for locally delivered sawlog and the other for national interest for locally created non-sawlog mechanical round wood. Models are initially planned in multiplicative structure.

2.4. Estimating Using Multiple Linear Regression

The impact of monetary and statistic variables on the yearly power utilization in Italy has been researched with the aim to build up a long haul utilization determining model. The timespan considered for the chronicled information is from 1970 to 2007. Distinctive regression models are created, utilizing recorded power utilization, Gross Domestic Product (GDP), GDP per capita and populace. An initial segment of the examination has thought about the estimation of GDP, cost and GDP per capita flexibilities of household and non-residential power utilization. explored the present moment (up 24 hours) load guaging of the interest for the South's (Sulewesi Island – Indonesia) Power System, utilizing a Multiple Linear Regression (MLR) technique. After a concise diagnostic exchange of the technique, the utilization of polynomial terms and the means to create the MLR model is clarified. Report on execution of MLR calculation utilizing financially accessible device, for example, Microsoft EXCEL™ is additionally talked about. As a contextual analysis, authentic information comprising of hourly burden request and temperatures of South Sulawesi electrical framework is utilized, to conjecture the transient burden.

3. Objective of the Study

1. To prepare statistical model
2. To forecast average cell phone bill.

3. To analyse the existing regression techniques to predict software effort estimation and the demand of pulpwood

3.1. Research Methodology

In this paper we have utilized linear regression model to discover the connection between month to month mobile phone bill regarding give autonomous variable . Linear regression model attempts to connection between two variable, coordinating a linear condition to the apparent information. One variable is to be considered, as a needy variable, and the second is considered as a free factor. Here the reliant variable is month to month PDA bill and the outcome will show how it relies upon same significant variable.

Number	Variable	Description
01	Age	One's age in years
02	Household Size	Number of people live in one's home
03	Shoes Pair	Number of pair of shoes one has
04	ActiveSIM1	1 if one has 1 SIM, 0 otherwise
05	DailyPocket1Exp500more	1 if one's out-of-pocket expenditure is more than 500tk, 0 otherwise
06	FavCoffee	1 if one like coffee, 0 otherwise
07	Fav Tea Coffee	1 if one like Tea,0 Coffee
08	Fav Drink Coke Pep	1 if one like Coke,0 Pepsi

Table 1. variable description

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std.Deviation
Age	134	20	58	28.39	8.359
Household size	132	2	22	4.73	2.536
Shoes Pair	134	1	60	8.85	9.310
ActiveSIM1	134	0	1	.34	.477
DailyPocket1Exp500more	134	0	1	.21	.408
FavCoffee	134	0	1	.22	.418
Fav Tea Coffee	134	0	1	.34	.477
Fav Drink Coke Pep	134	0	1	.18	.385
Valid n (list wise)	132				

Table 2. Descriptive Statistics

Variable	coefficient	Standard error	T state	p-value
Intercept	-862.435	320.39	-269	0.08
Age	38.318	8.83	4.34	0.00
Household size	90.55	26.54	3.141	0.01
Shoes pair	30.83	7.43	4.14	0.00
activeSIM1	-313.49	143.68	-2.18	0.03
Daily pocketExp500more	457.31	181.95	2.51	0.01
Favcoffee	716.47	174.24	4.11	0.00
Fav tea coffee	611.50	153.89	3.97	0.00
Fav Drink Cokepep	409.36	178.33	-2.29	0.23

Table 3. Modal specification

3.2. Random Variables and Probability Distributions in Business Statistics

Random variables and likelihood dispersions are two of the most significant ideas in statistics. An irregular variable doles out one of a kind numerical quality to the results of an arbitrary examination; this is a procedure that creates unsure results. A likelihood appropriation doles out probabilities to every conceivable estimation of an arbitrary variable. The two essential kinds of likelihood disseminations are discrete and nonstop. A discrete likelihood conveyance can just accept a limited number of various esteems.

3.3. Understand Sampling Distributions in Business Statistics

In statistics, sampling distributions are the probability distributions of some random measurement dependent on an irregular example, and are significant in light of the fact that they give a significant disentanglement on the course to factual induction. All the more explicitly, they enable investigative contemplations to be founded on the sampling dispersion of a measurement, as opposed to on the joint probability conveyance of all the individual example estimates. The estimation of an example measurement, for example, the example mean (\bar{X}) is probably going to be diverse for each example that is drawn from a populace. It can, along these lines, be thought of as an arbitrary variable, whose properties can be depicted with a probability distribution. The probability dispersion of an example measurement is known as a sampling distribution.

According to a key result in statistics known as the Central Limit Theorem, the sampling distribution of the model mean is standard if one of two things is legitimate:

- The underlying population is normal
- The sample size is at least 30

Two minutes are expected to register probabilities for the example mean; the mean of the sampling distribution approaches:

$$\mu_{\bar{x}} = \mu$$

The standard deviation of the sampling distribution (also known as the *standard error*) can take on one of two possible values:

Explore Hypothesis Testing in Business Statistics: In statistics, speculation testing alludes to the way toward picking between contending theories about a probability dissemination, in light of watched information from the distribution. It's a centre theme and a basic piece of the language of statistics.

3.4. How Businesses Use Regression Analysis Statistics

Regression analysis is a measurable instrument utilized for the examination of connections between variables. Typically, the agent looks to discover the causal impact of one variable upon another—the impact of a cost increment upon request, for instance, or the impact of changes in the cash supply upon the swelling rate. Regression analysis is utilized to evaluate the quality and the heading of the connection between two straightly related variables: X and Y. X is the “autonomous” variable and Y is the “reliant” variable.

The two basic types of regression analysis are:

- **Simple regression analysis:** Used to estimate the relationship between a dependent variable and a single independent variable; for example, the relationship between crop yields and rainfall.
- **Multiple regression analysis:** Used to gauge the connection between a reliant variable and at least two autonomous variables; for instance, the connection between the pay rates of representatives and their experience and instruction.
- Different regression analysis presents a few extra complexities however may create more reasonable outcomes than basic regression analysis.
- Regression analysis depends on a few in number presumptions about the variables that are being assessed. A few key tests are utilized to guarantee that the outcomes are legitimate, including theory tests. These tests are utilized to guarantee that the regression results are not just because of irregular possibility yet show a genuine connection between at least two variables.

- An expected regression condition might be utilized for a wide assortment of business applications, for example,:
- Measuring the impact on a corporation's profits of an increase in profits
- Understanding how sensitive a corporation's sales are to changes in advertising expenditures
- Seeing how a stock price is affected by changes in interest rates

Regression analysis may likewise be utilized for anticipating purposes; for instance, a regression condition might be utilized to conjecture the future interest for an organization's items. Because of the extraordinary unpredictability of regression analysis, it is regularly executed using specific mini-computers or spreadsheet programs.

4. Conclusion

The presentation and understanding of linear regression investigation are dependent upon an assortment of entanglements, which are examined here in detail. The peruse is made mindful of normal mistakes of understanding through down to earth models. Both the open doors for applying linear regression examination and its constraints are displayed.

References

-
- [1] D. R. Anderson, D. J. Sweeney and T. A. Williams, *Essentials of Statistics for Business and Economics*, (3rd ed.) Mason, Ohio: Southwestern, Thompson Learning, (2013).
 - [2] S. Bekmeier-Feuerhahn, *Markorientierte Markenbewertung*, Wiebaden, (2013).
 - [3] G. Bergler, *Absatz: Materialien Fur EineAbsatzlehre*, Vol. 28, Spindler, Nurnberg, Germany (2010), Explained: Regression Analysis. MIT News, (2014).
 - [4] L. McIntyre, *Using Cigarette Data for An Introduction to Multiple Regression*, Journal of Statistics Education, 2(1)(2011).
 - [5] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*, (3rd ed.), New York, NY: Dellen Publishing Co, (2012).
 - [6] D. Panchenko, *D. 18.443 Statistics for Applications*, Fall 2006, Section 14, Simple Linear Regression. (Massachusetts Institute of Technology: MIT OpenCourseWare).
 - [7] David A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press, (2009).
 - [8] Alvin C. Rencher and William F. Christensen, *Methods of Multivariate Analysis*, Wiley Series in Probability and Statistics, (3rd ed.), John Wiley & Sons, (2012).
 - [9] Hilary L. Seal, *The historical development of the Gauss linear model*, Biometrika, 54(1/2)(2012), 1-24.
 - [10] Xin Yan, *Linear Regression Analysis: Theory and Computing*, World Scientific, (2009).
 - [11] Robert Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series B, 58(1)(1996), 267-288.
 - [12] Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani, *Least Angle Regression*, The Annals of Statistics, 32(2)(2004), 407-451.