**Regression Model:**

* It is used when we want to predict a continuous dependent variable from several independent variables (can be dichotomous)
* If the dependent variable is dichotomous, then logistic regression should be used
* Independent variables with more than 2 levels can also be used. Like, X "predicts" Y, not X "causes" Y
* For example, it is used to determine what is the strength of the relationship between dose and effect, sales and marketing expenses, or age and income
* Used to forecast the effects or impact of changes. For example, how much additional sales income do I get for each additional $2000 spent on marketing?
* Predicts trends and future values. Like, what will the price of gold be in 6 months?
* Sample Correlation Coefficient (r), ranges between -1 and +1
* Measures the direction (the sign of the correlation) and strength (magnitude of the coefficient) of the linear association between the two variables

**R-Squared: Effectiveness of the Regression Model:**

* It is a goodness-of-fit measure for linear regression models
* Indicates the % of the variance in the dependent variable
* Measures the strength of the relationship between the model and the dependent variable on an appropriate 0 - 100% scale
* Usually, the larger the R2, the better the regression model fits

**References:**

1. Deborah R. Abrams, Tabachnick & Fidell (1989), Using multivariate statistics. (2nd edition). New York: HarperCollins was retrieved from <https://dss.princeton.edu/online_help/analysis/regression_intro.htm>
2. Jim Frost (2018), How to Interpret R-squared in Regression Analysis was retrieved from <https://statisticsbyjim.com/regression/interpret-r-squared-regression/>

I like the way you have jotted down the key points and findings. I would like to add a few points to this.

**Multiple Regression Analysis**

* It is used to see if there is a statistically significant relationship between sets of variables. It’s used to find trends in those sets of data
* The only difference between simple linear regression and multiple regression is in the number of predictors (“x” variables) used in the regression. Simple regression analysis uses a single x variable for each dependent “y” variable
* Normal linear regression usually isn’t enough to take into account all of the real-life factors that affect an outcome

Thanks & Regards,

Sunil Raj Thota

I like the way you have jotted down the key points and findings. I would like to add a few points to this.

A good value for R-squared are:

* It depends on the variable to which you measure it and the units in which that variable is measured
* It depends on the decision-making context
* If the dependent variable is a nonstationary time series, an R-squared value very close to 1 may not be very impressive
* If the dependent variable is a properly stationarized series then an R-squared of 25% may be quite good

In general, the important criteria for a good regression model are

* To make the smallest possible errors when predicting what will happen in the future
* To derive useful conclusions from the structure of the model and the estimated values of its parameters

Thanks & Regards,

Sunil Raj Thota