Homework - 2

Linear Regression and Time Series

Tushar Raj

Clark University  master’s in science in Data Analytics

**1. What is the population regression function?**

Population regression function is the set of all the point where they satisfy the equation of the conditional mean of variable Y which is a dependent variable for a fixed independent variable X.

**2. What is the difference between the population and sample regression functions? Is this a distinction without difference?**

Population regression function is the set of all the point where they satisfy the equation of the conditional mean of variable Y which is a dependent variable for a fixed independent variable X.

Whereas the Sample regression function is showing the estimated relation between Independent variable X with respect to the dependent variable Y.

Equation: yi = βo + βixi + µi

y -> dependent variable.

x-> independent variable

µ-> error term

βo -> Intercept

βi -> Slope coefficient

whereas in the real life we do have to work with the sample data and the equation/logic derived is applied on the real data. And the regression function obtained using the sample data is called sample regression function.

Equation: yi (cap)= βo(cap) + βi(cap).xi + µi(cap)

Y(cap)-> estimator of Y.

x-> independent variable

µ(cap)-> error term

βo(cap) -> estimator of βo

βi(cap) -> estimator of βi

Diagram

Description automatically generated

Note : The above diagram is extracted from the google search

The population regression function is a hypothetical conjunction between the variable and explanatory variable. The parameter used in thein the model indicates the weight given to each predictor with respect to the linear combination of them. Whereas on the other hand sample regression function contains estimated values of those coefficient. The set are chosen to be best fit data from that sample.

|  |  |
| --- | --- |
| **Population Regression function** | **Sample regression function** |
| Whole data is considered | Sample data is considered |
| It is a conceptual analysis | It is a realistic analysis |
| From the given population data one PRF can be drawn | One SRF can be drawn from the sample data |

**3. What is the role of the error term 𝑈𝑖 in regression analysis? What is the difference between the 𝑈𝑖 and 𝑈𝑖(cap) ?**

**ANSWER:**

The error term can be defined as the margin of the error within a statistical model. It can also be referred as the deviations within the regression lines, which explains the difference between the theoretical value vs actual observed value of the model.

Statistical error or the error term is mostly theoretical since they are calculated using the parameters. whereas the residual or the Ui(cap) is calculated using the relevant statistic. In context to the linear regression the model which gives the that gives out based on the input provided and the difference between the expected and the actual value obtained is called the error but calculating true regression is not possible based out of sample data. Thus, the difference between observed and estimated value from regression called residual.

|  |  |
| --- | --- |
| **𝑈𝑖** | **𝑈𝑖(cap)** |
| It is the difference between the observed and expected value of dependent variable. (true regression equation) | It refers to the difference between observed and estimated value of independent variable. |
| Real-world concept | Theoretical concept |
| It is the error term denoted by Ui | It is considered as estimate of error term. |

**4. Why do we need regression analysis? Why not simply use the mean value of the regressand as its best value?**

**ANSWER:**

It helps us understand the magnitude and direction of the dependent and independent variables even though it falls to give precise explanation of the event but it helps in forecasting using linear regression.

Reason Why we need regression analysis:

1. We will generally have small number of samples. Average of dependent variable on the selected sample will have lot of variability in it will be too far from the actual average. The linear regression uses all the data so the estimate for mean will be better choice.
2. There might be the scenario where zero samples in the subset of data ,where we cannot take the straight average we can build the regression model to estimate that scenario.
3. In case to conduct inference, we need regression model.

In actual data we never get the mean of the dependent variable in a regression. We rather get the samples which can be close or far away from the mean so it can be uncertain as to the sample value.

We cannot derive mean of regressand from the population so we reach mean to the population by using the sample of population and regress to estimate it.

**5. What do we mean by a linear regression model?**

**ANSWER:**

“A linear regression model describes **the relationship between a dependent variable, y, and one or more independent variables, X**.”

Yi = β1+β2Xi + µi

In math, regression indicates going of something from the chaotic actual data of observations to the simpler estimate of a line fitted to the data. Thus, the linear regression model can also be described as the process of taking observed data and finding the 'best fit' line to describe the relationship of two variables.

Example:

There is mobile manufacturer, and they are try to estimate the cost to volume analysis for a particular phone model with the sample data we have to come up with the equation with which we can estimate the total cost given a certain proposed volume of production as accurate as possible. Few of the method for that is Ordinary Least square method etc.

**6> Determine whether the following models are linear in the parameters, or the variables, or both. Which of these models are linear regression models?**

Table

Description automatically generated with medium confidence

**ANSWER:**

**In Model (a),** The X is in reciprocal, so the model is **Not linear in variable.** But the coefficient (Beta) is with power 1 so, **the equation is Linear (in parameter) regression mode**l.

**In Model (b),** The Log function of X is in the model so **Not linear in variable.** But the coefficient (Beta) is with power 1 so, **the equation is Linear (in parameter) regression mode**l.

**In Model (c),** The Log function of Y is in the model so **Not linear in variable.** But the coefficients are with power 1 so, **the equation is Linear (in parameter) regression mode**l.

**In Model (d),** The Log function are there in the model and it is not completely possible to transform the equation into the linear form so **Not linear in variable. The equation is not linear in parameter** also, but it can be transformed into, **the equation is Linear (in parameter) regression mode**l.

If,

α= ln(β1)

Thus, the equation becomes,

ln(Y) = α+ β2ln(Xi) + Ui

**In Model (e),** The Log function of Y and reciprocal are in the model so **Not linear in variable.** But the coefficients are with power 1 so, **the equation is Linear (in parameter) regression mode**l.

**Solution for the classroom exercises:**

****

Thank You