Objectives

Causai Relatioi

Problem Statement

Causation v/s

Simple Regre

Model

Least Square

Regression

Data Eitting

Derivation of Lea

Square Estimates

Thankyou

Simple Regression Analysis

Parthasarathi Edupally¹

¹DataConscientious LLP, Mumbai

Elements of Econometrics, Russell Square International College, Mumbai

Outline

Objectives

Objectives

Drohlam Statamant

Problem Statemen

Correlation

Simple Regre

Disturbance te

Least Square

method of

Regression

Derivation of Lea

Derivation of Lea Square Estimates

Outline

Objectives

Causal Relation 1 Obje

Correlation
Simple Regression

Disturbance term

method of Regression

Data Fitting

Derivation of Leas Square Estimates

Chankwon

1 Objectives

2 Causal Relation

Problem Statement Causation v/s Correlation Simple Regression Model Disturbance term

Outline

Objectives

Problem Statement

Causation v/s Correlation

Simple Regressi Model

Disturbance

method of

Data Fitting

Derivation of Lea Square Estimates

Chankvo

Objectives

2 Causal Relation

Problem Statement Causation v/s Correlation Simple Regression Model Disturbance term

3 Least Squares method of Regression

Data Fitting
Derivation of Least Square Estimates

Outline

Objectives

Causal Relation

Problem Statement

Correlation Simple Regression

Disturbance to

Least Squar

Regression

Derivation of Lea

Chankyo

Objectives

2 Causal Relation

Problem Statement
Causation v/s Correlation
Simple Regression Model
Disturbance term

- 3 Least Squares method of Regression
 Data Fitting
 Derivation of Least Square Estimates
- 4 Thankyou

Objectives

Objectives

Causal Relation

Causation v/s

Simple Regress

Disturbance t

Least Square

method of

Regression

Data Fitting

Derivation of Les

Derivation of Lea Square Estimates

Γhankyou

• To understand motivation for Regression Model

Objectives

Objectives

Causal Relation

Problem Statement

Correlation Simple Regress

Disturbance to

Laget Sangra

method of

regression

Derivation of Lea

Γhankyou

- To understand motivation for Regression Model
- Procedure for finding Regression Coefficients

Objectives

Objectives

Causal Relation

Problem Statement

Simple Regressi

Disturbance to

Least Square

method of

Data Fitting

Derivation of Leas Square Estimates

- To understand motivation for Regression Model
- Procedure for finding Regression Coefficients
- Interpretation of Regression Coefficients

Objectives

Cancal Palatic

Problem Statement

Correlation

Simple Regres Model

Disturbanc

Least Squar

Regression

Data Cittian

Derivation of Lea

Thankyou

Problem Statement

• Given a population of interest, What causes a particular characteristic of its people ?

Objectives

Causal Relatio

Problem Statement

Correlation Simple Regression

Distuibance

method of

Regressio

Derivation of Leas

Thankwou

- Given a population of interest, What causes a particular characteristic of its people ?
- For example,
 - Population: All patients suffering with Diabetes in Mumbai

Objectives

Causal Relati

Problem Statement

Correlation
Simple Regressi

Disturbance ter

method of

D . F...

Derivation of Least Square Estimates

Thankyou

- Given a population of interest, What causes a particular characteristic of its people ?
- For example,
 - Population: All patients suffering with Diabetes in Mumbai
 - Characteristic: Average sugar levels (say Y, during early morning, averaged over a month)

Objectives

Causal Relation

Problem Statement

Correlation
Simple Regressio

Disturbance to

Least Squares method of

Data Fitting

Derivation of Leas Square Estimates

Thankyou

- Given a population of interest, What causes a particular characteristic of its people ?
- For example,
 - Population: All patients suffering with Diabetes in Mumbai
 - Characteristic: Average sugar levels (say Y, during early morning, averaged over a month)
 - Possible causes: Lifestyle (say X_1 , number of hours of walking, everyday averaged over a month), Diet (say X_2 , percentage of fibre in diet in any day, averaged over a month) etc

. . . .

Caucal Palati

Problem Statement

Correlation Simple Regressio Model

Disturbanc

Least Squares method of Regression

Data Fitting

Derivation of Leas Square Estimates

Thankyo

- Given a population of interest, What causes a particular characteristic of its people ?
- For example,
 - Population: All patients suffering with Diabetes in Mumbai
 - Characteristic: Average sugar levels (say Y, during early morning, averaged over a month)
 - Possible causes: Lifestyle (say X_1 , number of hours of walking, everyday averaged over a month), Diet (say X_2 , percentage of fibre in diet in any day, averaged over a month) etc
- How do we approach this problem ?

Problem Statement

- Given a population of interest, What causes a particular characteristic of its people?
- For example,
 - Population: All patients suffering with Diabetes in Mumbai
 - Characteristic: Average sugar levels (say Y, during early morning, averaged over a month)
 - Possible causes: Lifestyle (say X_1 , number of hours of walking, everyday averaged over a month), Diet (say X_2 , percentage of fibre in diet in any day, averaged over a month) etc
- How do we approach this problem?
 - Correlation: to measure possible linear relation between variables Y and X_1, X_2

Problem Statement

Problem Statement

Given a population of interest, What causes a particular characteristic of its people?

- For example,
 - Population: All patients suffering with Diabetes in Mumbai
 - Characteristic: Average sugar levels (say Y, during early morning, averaged over a month)
 - Possible causes: Lifestyle (say X_1 , number of hours of walking, everyday averaged over a month), Diet (say X_2 , percentage of fibre in diet in any day, averaged over a month) etc
- How do we approach this problem?
 - Correlation: to measure possible linear relation between variables Y and X_1, X_2
 - Regression: Assuming (Contemplating) that linear relation exists, we try to fit regression model to data

Causation v/s Correlation

Causation v/s Correlation

• Correlation doesn't imply Causation, Why?

Objectives

Causal Relatio

Problem Statement

Causation v/s

Correlation
Simple Regressi

Disturbance

Least Square method of

Regression

Derivation of Leas Square Estimates

Thankyou

Causation v/s Correlation

- Correlation doesn't imply Causation, Why?
- Because there might be other variables, which we are not aware of, actually influencing both X and Y together

Objectives

Causal Relatio

Problem Statement

Causation v/s

Correlation Simple Regressi Model

Disturbance

Least Square

Regression

Data Fitting

Derivation of Least

Derivation of Least Square Estimates

Thankyou

Causation v/s Correlation

- Correlation doesn't imply Causation, Why?
- Because there might be other variables, which we are not aware of, actually influencing both X and Y together
- Medical researchers solve this probelm with curated experiments to get data - control and treatment groups

Causation v/s Correlation

Causation v/s Correlation

- Correlation doesn't imply Causation, Why?
- Because there might be other variables, which we are not aware of, actually influencing both X and Y together
- Medical researchers solve this probelm with curated experiments to get data - control and treatment groups
- In Economics, we have only observational data so difficult to establish causation
- Ceteris Paribus assumption is basis of all regression modelling in Economics

<Scatterplot Demo>

Simple Regression Model

Objectives

Causal Relatio

Problem Stateme

Simple Regression Model

man a second

Least Squan

method of

Regression

Derivation of Lea

Square Estimates

• Assumptions:

Objectives

Causal Relation

Problem Statemen

Simple Regression Model

method of

Regression

Data Eitting

Derivation of Leas

Thankyou

Simple Regression Model

- Assumptions:
 - There is linear relation between X and Y (Model is correctly specified)

Objectives

Causal Relation

Problem Stateme

Simple Regression Model

Disturbance

Least Square method of

Regression

Data Fitting

Derivation of Leas Square Estimates

Thankyou

Simple Regression Model

- Assumptions:
 - There is linear relation between X and Y (Model is correctly specified)
 - All the other variables influencing Y can be modelled with error term

Causal Relation

Problem Statement

Simple Regression Model

Disturbance

Least Squares method of

Regressio

Derivation of Leas Square Estimates

Thankvo

Simple Regression Model

- Assumptions:
 - There is linear relation between X and Y (Model is correctly specified)
 - All the other variables influencing Y can be modelled with error term
- Model: $Y_i = \beta_1 + \beta_2 X_i + u_i$
- Where Y_i , X_i and u_i are Y, X and u values of observation i; β_1 and β_2 are parameters of model

Simple Regression Model

Objective:

Causal Relation

Problem Stateme Causation v/s

Simple Regression Model

Disturbance t

method of

Data Fitting

Derivation of Leas Square Estimates

- Assumptions:
 - There is linear relation between X and Y (Model is correctly specified)
 - All the other variables influencing Y can be modelled with error term
- Model: $Y_i = \beta_1 + \beta_2 X_i + u_i$
- Where Y_i , X_i and u_i are Y, X and u values of observation i; β_1 and β_2 are parameters of model
- Y is called dependent variable
- X is called regressor or independent variable

Objectives

Causal Relation

Correlation
Simple Regress

Disturbance term

method of Regression

Data Fitting

Derivation of Lea

Thankyou

Disturbance or Error Term

• It is used to capture, mathematically, everything else not explained by this simple regression model

Objectives

Causal Relation Problem Statement

Correlation Simple Regressi Model

Disturbance term

Least Squares method of Regression

Data Fitting
Derivation of Least

Thombroom

- It is used to capture, mathematically, everything else not explained by this simple regression model
- Following are reasons for its inclusion:
 - Omission of explanatory variables: there might be other variables influencing Y

Objectives

Causal Relation Problem Statement

Correlation Simple Regression Model

Disturbance term

method of Regression Data Fitting

Derivation of Least Square Estimates

Thankyou

- It is used to capture, mathematically, everything else not explained by this simple regression model
- Following are reasons for its inclusion:
 - Omission of explanatory variables: there might be other variables influencing Y
 - Aggregation of variables: we are trying to find average relation

Objectives

Causal Relation

Correlation
Simple Regression
Model

Disturbance term

Regression

Data Fitting

Derivation of Least Square Estimates

Thankyou

- It is used to capture, mathematically, everything else not explained by this simple regression model
- Following are reasons for its inclusion:
 - Omission of explanatory variables: there might be other variables influencing Y
 - Aggregation of variables: we are trying to find average relation
 - Model misspecification: X chosen might be not appropriate

Objectives

Causal Relation

Correlation
Simple Regression
Model

Disturbance term

Least Squares method of Regression

Data Fitting
Derivation of Leas

Square Estimates

Thankyou

- It is used to capture, mathematically, everything else not explained by this simple regression model
- Following are reasons for its inclusion:
 - Omission of explanatory variables: there might be other variables influencing Y
 - Aggregation of variables: we are trying to find average relation
 - Model misspecification: X chosen might be not appropriate
 - Functional misspecification: there might be nonlinear relation

Objectives

Causal Relatio Problem Statement Causation v/s

Correlation Simple Regression Model

Disturbance term

Regression
Data Fitting

Derivation of Least Square Estimates

Thankyou

- It is used to capture, mathematically, everything else not explained by this simple regression model
- Following are reasons for its inclusion:
 - Omission of explanatory variables: there might be other variables influencing Y
 - Aggregation of variables: we are trying to find average relation
 - Model misspecification: X chosen might be not appropriate
 - Functional misspecification: there might be nonlinear relation
 - Measurement error: wouldnt have been possible to measure X and Y without errors

Data Fitting

Objectives

Causal Relation

Problem Statement

Simple Regress Model

Disturbance ter

Least Square

Data Fitting

Derivation of Leas Square Estimates

Thankyou

• This whole business of finding a causal relation between X and Y boils down to fitting a line

Objectives

Causal Relation

Problem Statement

Correlation Simple Regressi Model

method of Regression

Data Fitting

Derivation of Leas Square Estimates

Thankyou

Data Fitting

- This whole business of finding a causal relation between X and Y boils down to fitting a line
- Problem: We have to fit a "best" possible line (Model) to data

Data Fitting

Objectives

Causai Relation

Causation v/s Correlation Simple Regressio

Least Squares method of

Data Fitting

Derivation of Leas Square Estimates

- This whole business of finding a causal relation between X and Y boils down to fitting a line
- Problem: We have to fit a "best" possible line (Model) to data
- Means we have to find (estimate) best possible values of parameters β_1 and β_2

Data Fitting

Objective

Causal Relation

Causation v/s
Correlation
Simple Regression

method of Regression

Data Fitting

Derivation of Lea Square Estimates

- This whole business of finding a causal relation between X and Y boils down to fitting a line
- Problem: We have to fit a "best" possible line (Model) to data
- Means we have to find (estimate) best possible values of parameters β_1 and β_2
- Procedure to find "best" values:
 - Define best line: the one which has least "Residual Error"
 - Residual error: $e_i = Y_i \hat{Y}_i$
 - Fitted Y: $\hat{Y}_i = b_1 + b_2 X_i$

Data Fitting

Objective:

Causal Relation

Causation v/s
Correlation
Simple Regression

Least Squares method of Regression

Data Fitting

Derivation of Lea Square Estimates

- This whole business of finding a causal relation between X and Y boils down to fitting a line
- Problem: We have to fit a "best" possible line (Model) to data
- Means we have to find (estimate) best possible values of parameters β_1 and β_2
- Procedure to find "best" values:
 - Define best line: the one which has least "Residual Error"
 - Residual error: $e_i = Y_i \hat{Y}_i$
 - Fitted Y: $\hat{Y}_i = b_1 + b_2 X_i$
 - So we use standard calculus to minimize residual error to find parameters

Derivation of Least Square Estimates

Objectives

Causal Relation

Drohlam Statament

Correlation

Model

pro t

Least Squar

method of

Derivation of Least

Square Estimates

Γhankyou

Derive estimates of simple regression model using calculus

Interpretation of Coefficients

Objectives

Causal Relation

Problem Statement

Conciation

Model Model

Dieturbanea ta

Least Squar

method of

Data Eistina

Derivation of Lea

Square Estimates

From textbook!!

Objectives

Causal Relation

Problem Statement

Causation v/s
Correlation

Model

method of

Data Fitting

Derivation of Leas Square Estimates

Thankyou

"In God we trust, all others bring data."

William Edwards Deming (1900 - 1993).