

# DA5020

COLLECT, STORE, RETRIEVE, ANALYZE DATA  
WEEK 11 – LIVE SESSION



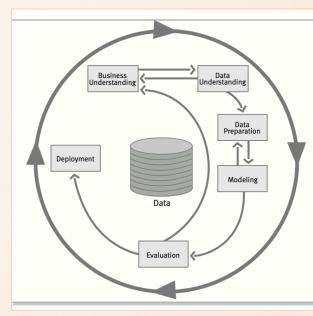
## OUTLINE

### Week 11: Forecasting Future Values

- ❖ Recap
- ❖ Topic: Forecasting Future Values
  - ❖ The modeling phase
  - ❖ Forecasting

## RECAP!

- ❖ CRISP-DM
- ❖ Where are we now?
  - ❖ Modeling
  - ❖ Evaluation



Phases of the CRISP-DM Model (Chapman et al., 2000)

## THE MODELING PHASE

- ❖ What is a model?
- ❖ Why do we need models?
- ❖ What is a good model?
- ❖ What does it mean to 'fit' the data?

Num_Rooms	Lot_Area	Condition	Sale_Price
7	8450	Normal	208500
6	9600	Normal	181500
7	11250	Normal	223500
7	9550	Abnorml	140000
8	14260	Normal	250000
5	14115	Normal	143000

## THE MODELING PHASE

- ❖ What is regression?
- ❖ Types of regression models:
  - ❖ Linear regression: simple & multiple
  - ❖ Logistic regression
  - ❖ Poisson regression
  - ❖ Polynomial regression, etc...

Num_Rooms	Lot_Area	Condition	Sale_Price
7	8450	Normal	208500
6	9600	Normal	181500
7	11250	Normal	223500
7	9550	Abnorml	140000
8	14260	Normal	250000
5	14115	Normal	143000

## THE MODELING PHASE

- ❖ Models are approximations
- ❖ Start with a baseline (a simple model)
- ❖ Evaluate the model
  - ❖ Make improvements:
    - ❖ to the same model or build a more advanced model
    - ❖ to the data

## THE MODELING PHASE

KEEP IN MIND THAT...

“ Essentially, all models are wrong, but some are useful.”

—George E.P. Box

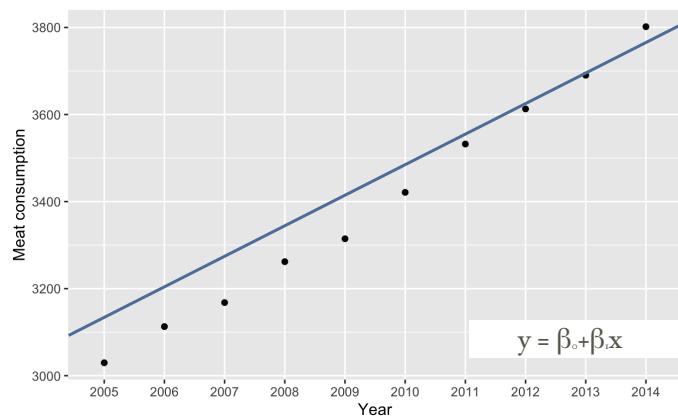
Source: <https://images.app.goo.gl/G4scytZkBBfLrvmyq>

## THE MODELING PHASE

- ❖ Assumptions of linear regression:
  - ❖ Linearity
  - ❖ Independence
  - ❖ Normality
  - ❖ Homoscedasticity

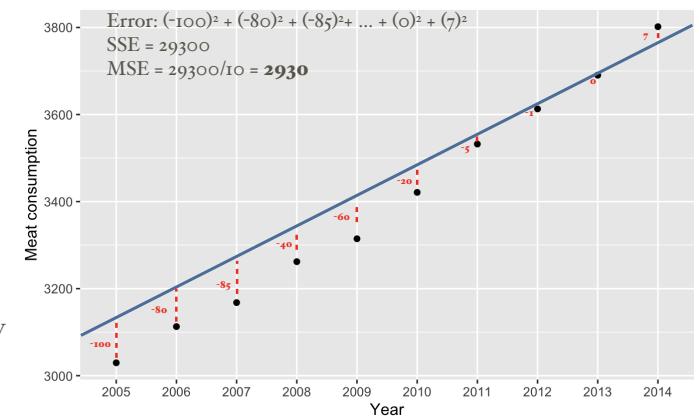
## MODELING

- ❖ Assumptions of linear regression:
- ❖ Linearity
- ❖ Independence
- ❖ Normality
- ❖ Homoscedasticity



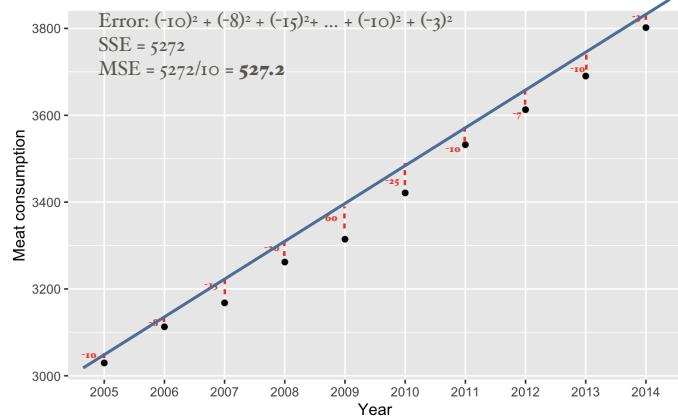
## MODELING

- ❖ Assumptions of linear regression:
- ❖ Linearity
- ❖ Independence
- ❖ Normality
- ❖ Homoscedasticity



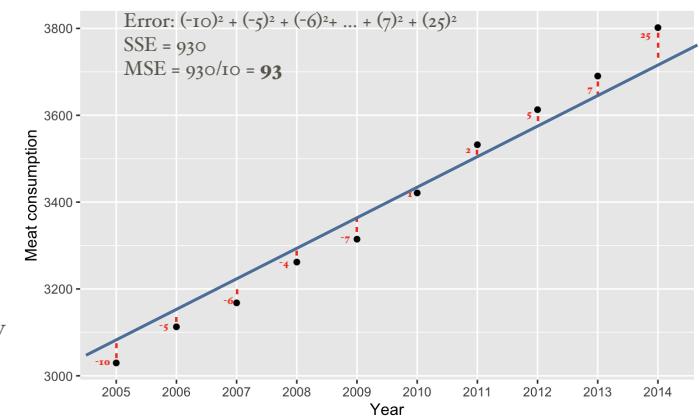
## MODELING

- ❖ Assumptions of linear regression:
- ❖ Linearity
- ❖ Independence
- ❖ Normality
- ❖ Homoscedasticity



## MODELING

- ❖ Assumptions of linear regression:
- ❖ Linearity
- ❖ Independence
- ❖ Normality
- ❖ Homoscedasticity



# DEMO

PREDICTING THE ANNUAL MEAT CONSUMPTION

BREAKOUT GROUPS!

CALCULATE THE MSE FOR THE MEAT CONSUMPTION MODEL