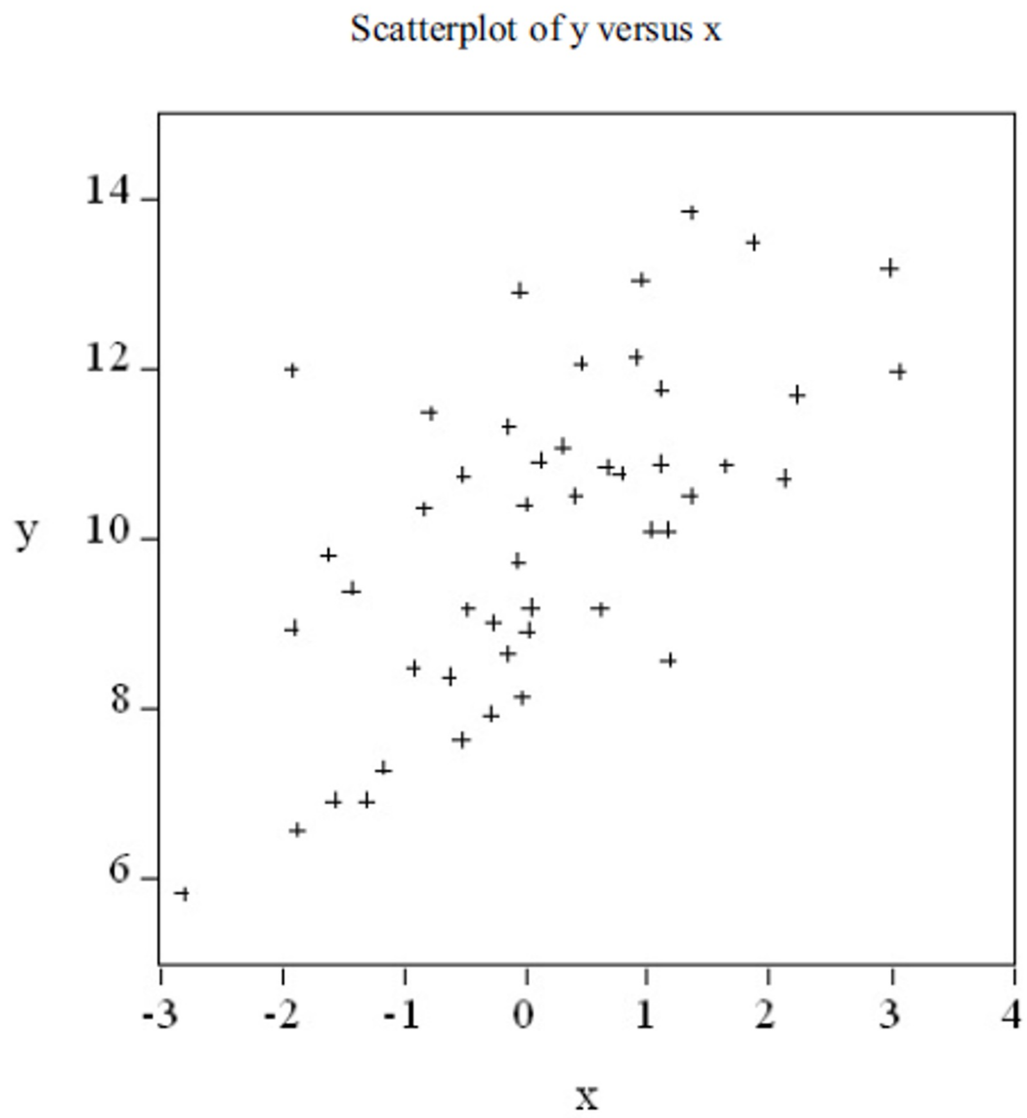
Regression

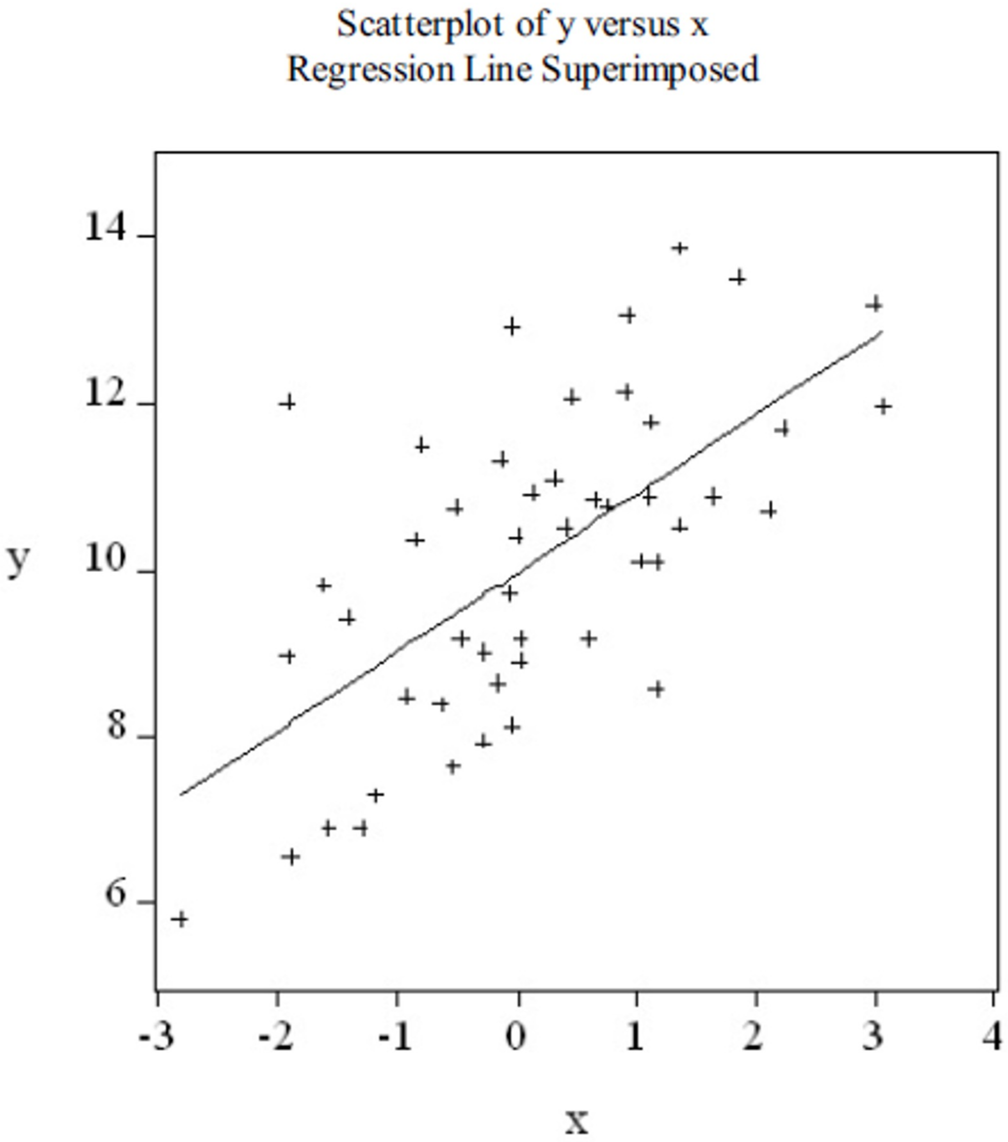
Overview

Regression Analysis

- One of the basic tools for forecasting
- A statistical technique to describe relationships among variables
- Consider two variables y and x
  - Describe y using x
  - y: dependent variable
  - x: independent variable (explanatory, exogenous)

Regression Analysis





Regression Analysis

- How to find the line that fits best?
  - Line: $y = \beta_0 + \beta_1 X$
  - How to find β_0 and β_1?
- Example

Regression Analysis

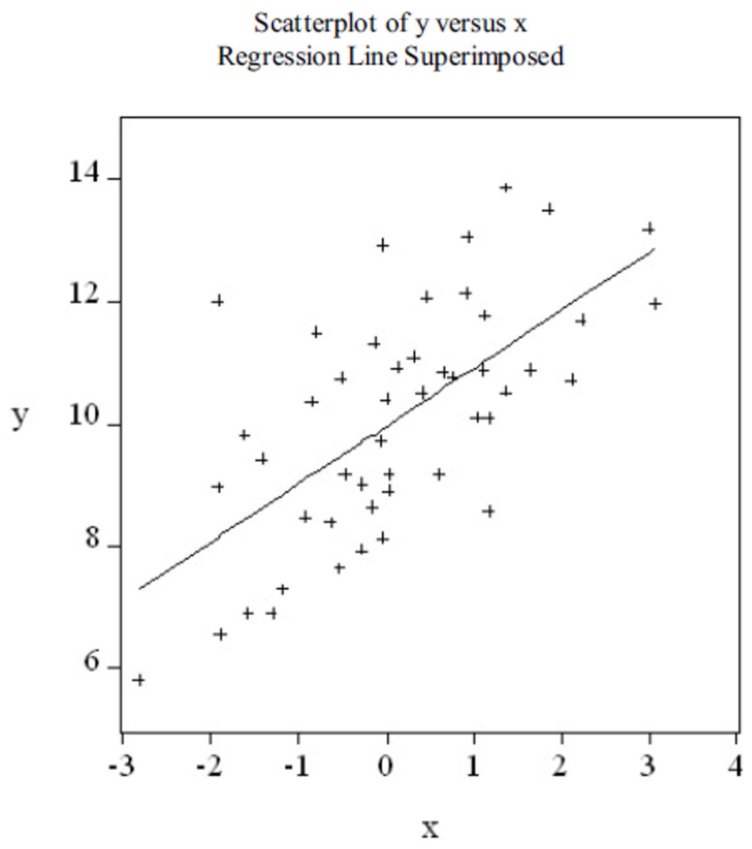
- Probabilistic Model
  - $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$
  - $\varepsilon_t \overset{iid}{\rightarrow} N(0, \sigma^2)$
  - Model parameters: $\beta_0, \beta_1, \sigma^2$
- If this model is correct:
  - Expected value of y conditional on $x = x^*$
    - $E(y|x^*) = \beta_0 + \beta_1 x^*$

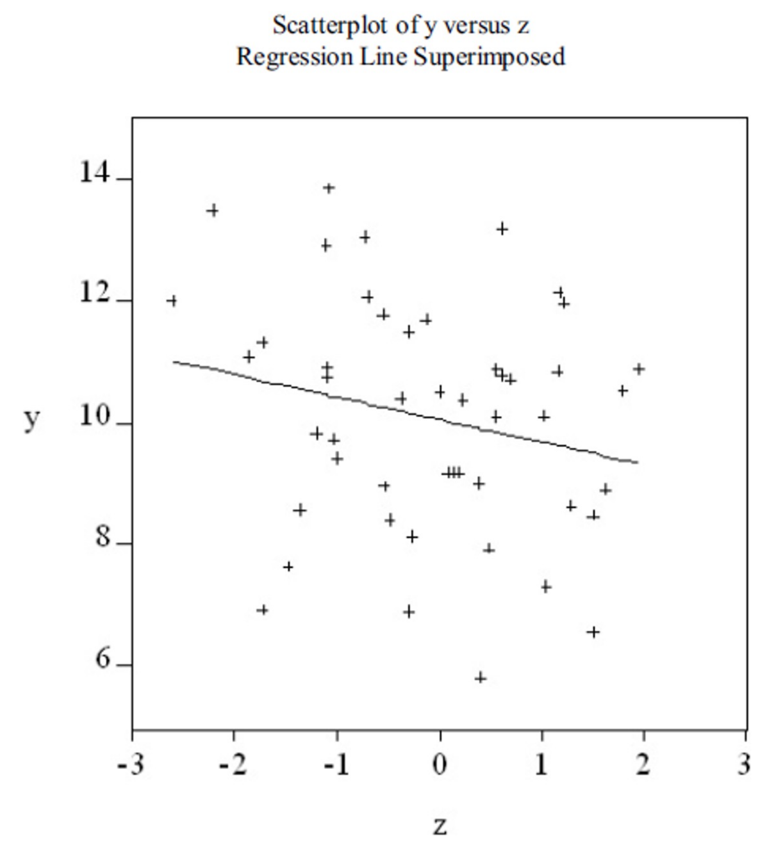
Regression Analysis

- Fitted values
  - $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$
- Minimize residuals
  - Residuals = in-sample forecast errors
  - $e_t = y_t - \hat{y}_t$
- Least-squares estimation
  - $\sum_{t=1}^{T} (y_t - \hat{y}_t)^2$

Regression Analysis

- Multiple linear regression
  - y is described by more than one explanatory variable





Regression Analysis

- Multiple linear regression model
  - $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t$
  - $\varepsilon_t \xrightarrow{iid} N(0, \sigma^2)$
- Fitted values
  - $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2\, z_t$
- Residuals
  - $e_t = y_t - \hat{y}_t$
- Least-squares estimation
  - $\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$

Goodness-of-fit Statistics

- Sum squared resid. $$SSR = \sum_{t=1}^{T} e_t^2$$

- Objective of the least-squares estimation
- Sum of squared residuals
- Not of much value in isolation
- Input to other diagnostics
- Useful for comparing models and testing hypotheses

Goodness-of-fit Statistics

- R-squared (R^2)
  - Indicates how much of var(y) can be explained by the variables included in the regression
  - Intuition: $\frac{var(y|x)}{var(y)}$
  - Measurement of in-sample success of the regression equation
  - If intercept is included, $0 < R^2 < 1$

$$R^2 = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T} e_t^2}{\frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y}_t)^2}$$

Goodness-of-fit Statistics

- Adjusted R-squared ($\bar{R}^2$)
  - Same interpretation as R^2 but formula is slightly different
  - Adjusted for degrees of freedom used in fitting the model
  - Adjustment: penalize the amount of right-hand-side variables

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k}\sum_{t=1}^{T} e_t^2}{\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

Goodness-of-fit Statistics

- Akaike info criterion (AIC)
  - Estimate of the out-of-sample forecast error variance
  - Similar to s^2 but penalizes degrees of freedom more harshly
  - Used to compare forecasting models

$$\mathrm{AIC} = e^{\left(\frac{2k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

Goodness-of-fit Statistics

- Schwarz criterion (SIC)
  - Alternative to AIC
  - Harsher penalty for degrees-of-freedom
  - Used to compare forecasting models

$$SIC = T^{\left(\frac{k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

Goodness-of-fit Statistics

- Durbin-Watson stat. (DW)
  - Errors from a good forecasting model should be unforecastable
    - Forcastable error -> room for improvement in the model
    - Correlation among errors -> forecastable information
    - DW tests if the regression disturbances over time are serially correlated.
  - $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, where $\varepsilon_t = \varphi \varepsilon_{t-1} + v_t$

    $$v_t \xrightarrow{iid} N(0, \sigma^2)$$

    - ε_t is serially correlated when $\varphi \neq 0$.
    - Ideal case $\varphi = 0$

Goodness-of-fit Statistics

- Durbin-Watson stat. (DW)
  - $H_0: \varphi = 0$
  - $0 \leq DW \leq 4$
    - **OK:** $DW \sim 2$
    - **Alarm:** $DW < 1.5$
      - Consult DW tables for significance level for rejecting the null hypothesis

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Goodness-of-fit Statistics

- Residual plot
  - Examine :
    - Actual data (y_t)
    - Fitted values ($\hat{y}_t$)
    - Residuals (e_t)