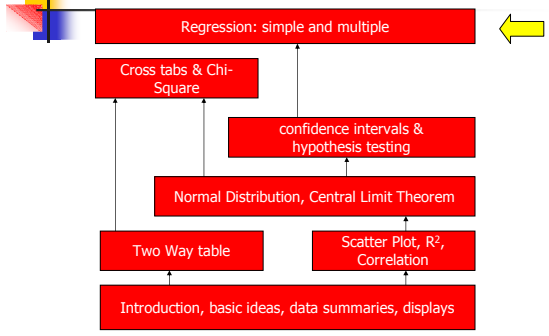
PUB – POS 316
Week 13b

Simple linear regression

Navid Ghaffarzadegan
navidg@gmail.com
Last updated – Jan 1, 10

Course Road Map



Agenda

- Introduction
  - Association
  - Scatter plots
- The linear regression model
- Tests for significance and CI
- ANOVA
- F-test

Introduction

- Association between variables:

Two variables are associated if knowing the value of one of them tells you something about the other one.

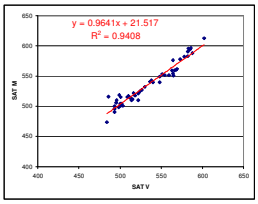
- Examples:
  - *Effort and grade*
    - *Positive association*
  - *Price and demand*
    - *Negative association*

Introduction

- Example:
- We have data on SAT average scores in different states.
  - Q1: Is there any association between SATM and SATV?
    - We can look at the data.
    - We can draw a graph that helps us to see if there is an association. → scatter plot

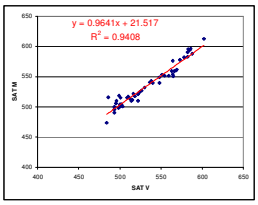
Introduction

- Example: what does the graph say?



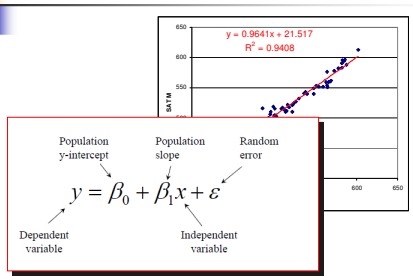
  - Q2: How can we predict SATM from SATV based on this data?
  - **If SATV=650 then SATM=?**

Least Square Regression



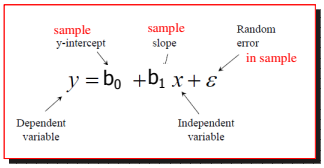
- Residual: $y_i - \hat{y}_i$
- The best fitting line minimizes the "sum of residual squared" $\sum(y_i - \hat{y}_i)^2$
- excel finds the line and we don't need to worry about it.

Least Square Regression



Least Square Regression

- What if we do not have the complete information about our population?



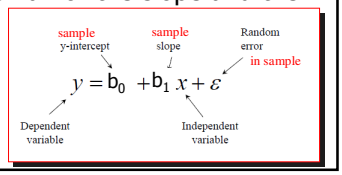
What does estimation of slope and intercept mean? (b estimation of β)

Least Square Regression

- Example
- Work with excel.

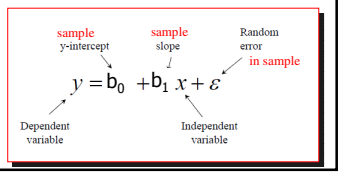
Tests for significance and CI

- So, if we are estimating the slope and the intercept of the line,…
- WE CAN BE WRONG
- → We need to report confidence intervals!
- Confidence interval for the slope and the intercept



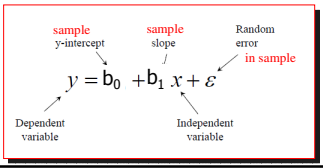
Tests for significance and CI

- Remember:
  - Margin of error = z. (proper standard deviation)
  - And if you do not have z, you use t.
  - The same here: (And the good thing is that excel gives you the proper standard deviation (standard error)
  - Margin of error = t. (SE) (df=n-2)
- Back to excel



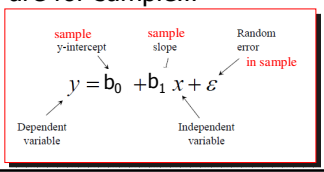
Tests for significance and CI

- What will happen for the slope and intercept if we conduct the study many times?
- The important question: <u>Are you confident enough that the slope is not zero? ($\beta_1 \neq 0$)</u>



Tests for significance and CI

- Hypotheses: H_0: $\beta_1 = 0$
- H_a: $\beta_1 \neq 0$
- Don't forget: β's are related to the population – b's are for sample…
- <u>Very simple:</u>
- $t = b_1 / SE_{b1}$



Tests for significance and CI

- So, 1. we should report confidence intervals for β's., or 2. We should test hypothesis that β's are different from zero.
- Back to excel.