Տնտեսական զարգացման և հետազոտությունների կենտրոն

ԷԿՈՆՈՄԵՏՐԻԿԱ
Փոքրագույն քառակուսիների եղանակը
Դասական գծային ռեգրեսիոն մոդելը

Տնտեսական զարգացման և հետազոտությունների կենտրոն

Ելենա Մանուկյան

ՀՀ Էկոնոմիկայի նախարարություն

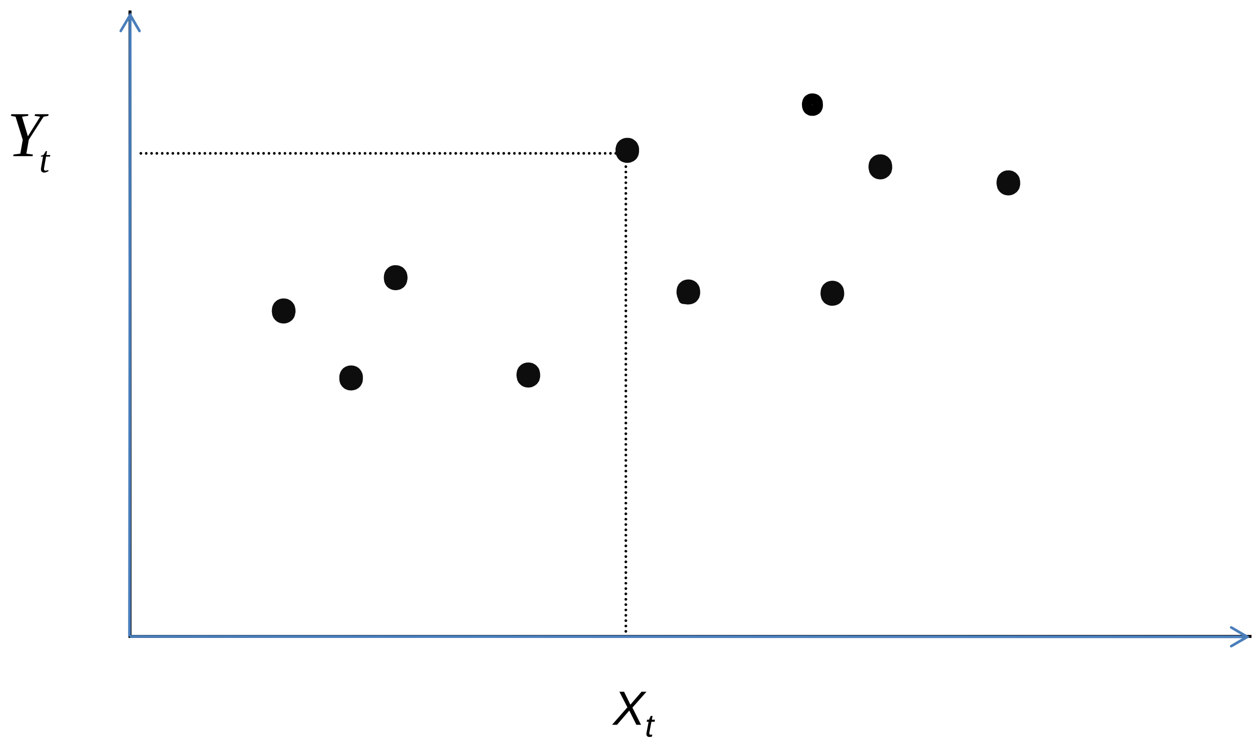
Երևան - 2010

Գերմանական տեխնիկական համագործակցության ընկերություն

Փոքրագույն քառակուսիների եղանակը

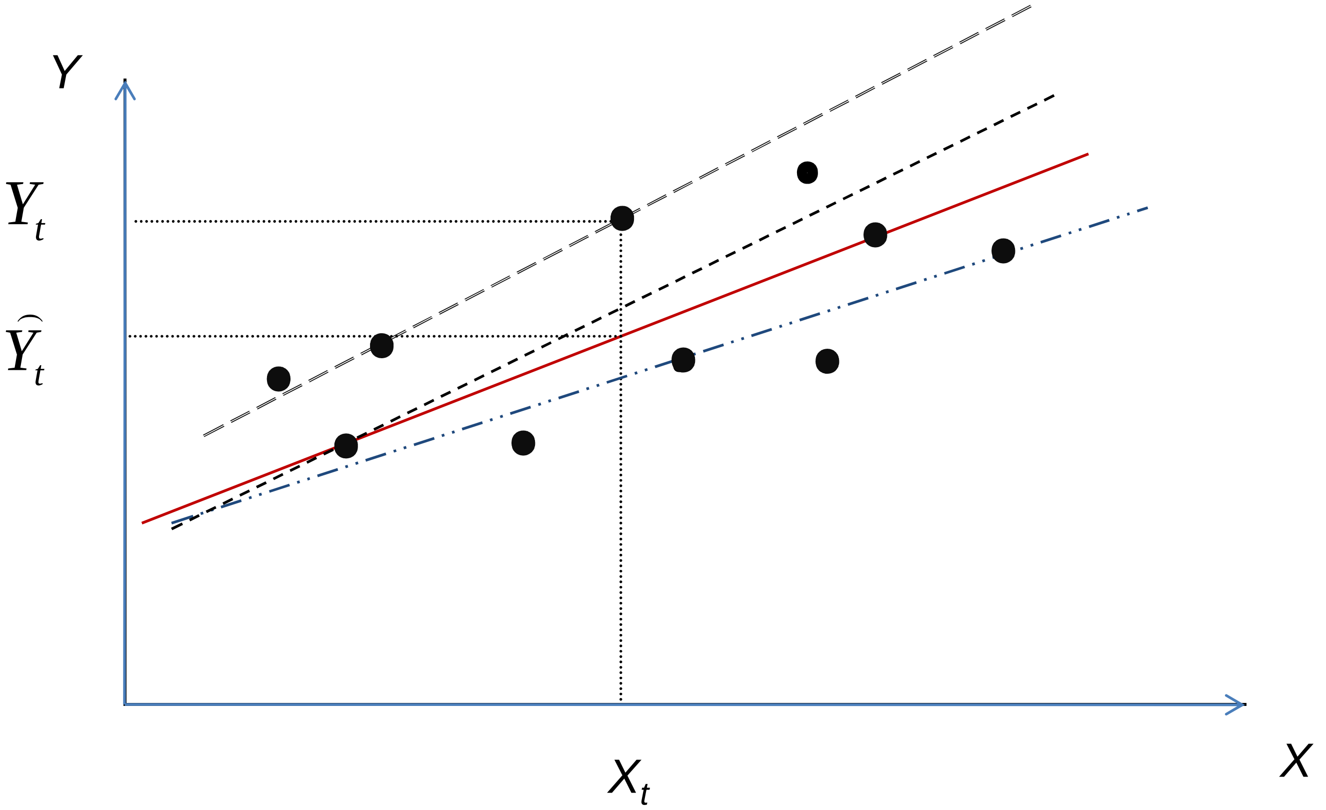
- Ենթադրենք ունենք որևէ X և Y ցուցանիշների փաստացի արժեքները, որոնք ներկայացնենք տվյալների ամպով:

  Դիցուք տվյալների ամպի ուսումնասիրությունը հուշում է, որ փոփոխականների միջև կապը կարելի է բնութագրել գծային ֆունկցիայով:

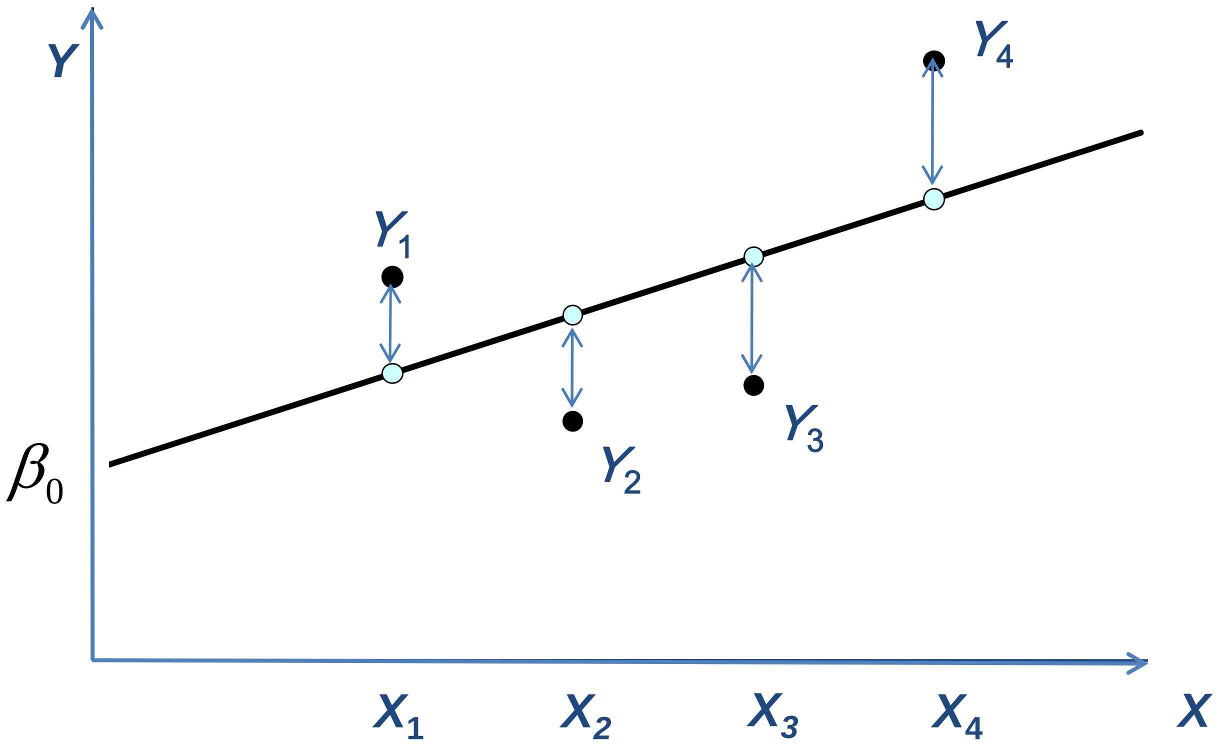


Փոքրագույն քառակուսիների եղանակը

- Ո՞ր ուղիղն է լավագույնս բնութագրում տվյալների ամպը:
  Ինչպիսի՞ հատկություններով պետք է օժտված լինի լավագույն ուղիղը:



Y_t - ով նշանակենք բացատրվող փոփոխականի փաստացի արժեքները, և դիտարկենք դրանց շեղումները առաջարկված գծից:

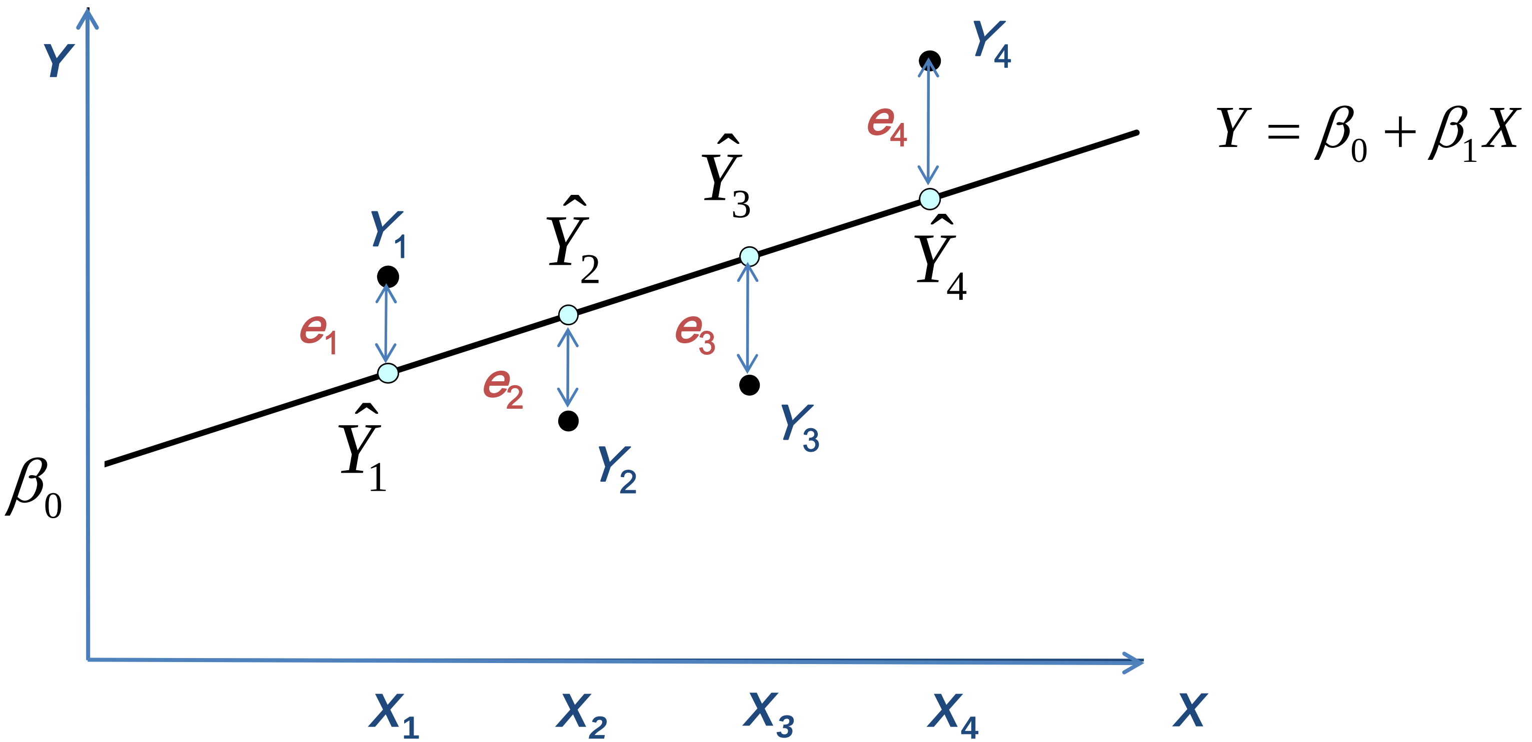


$$Y = \beta_0 + \beta_1 X$$

Նշանակենք

- Y_t - փաստացին
- $\hat{Y}_t$ - գնահատականը

$Y_t - \hat{Y}_t = e_t$ - մնացորդը

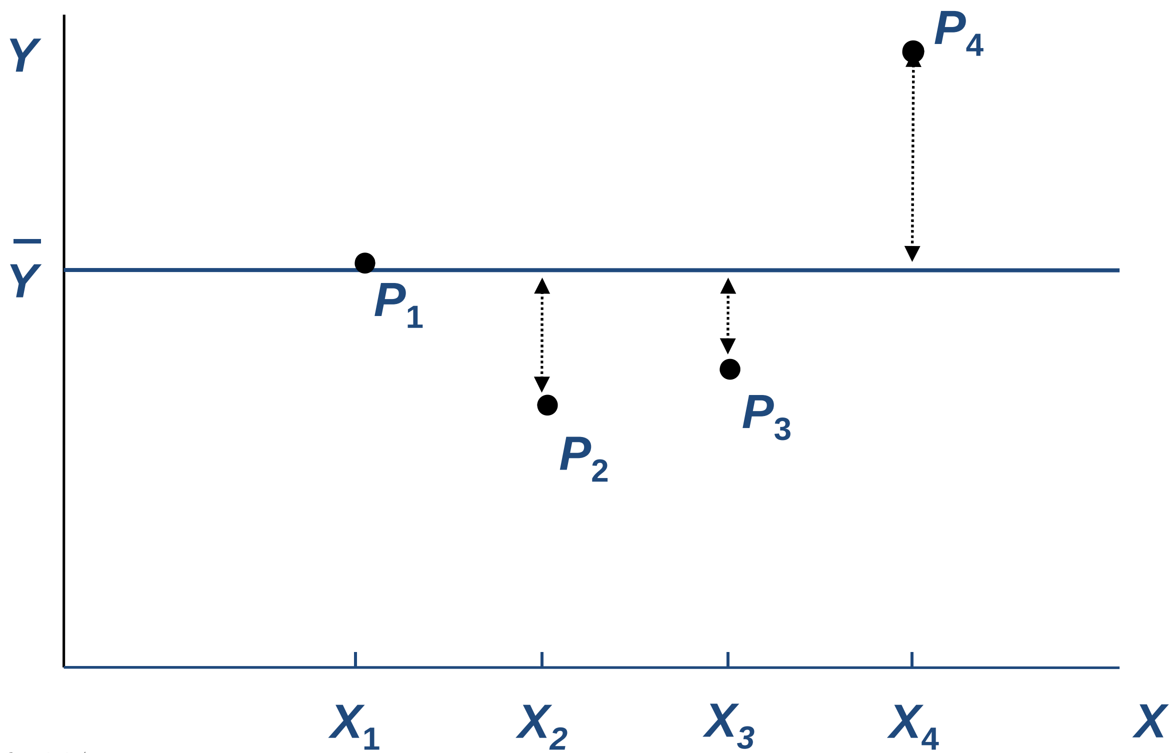


ՓՔԵ սկզբունքը

Նվազագույնի ծգտեցնել RSS (residual sum of squares),

որտեղ $RSS = \sum_{t=1}^{n} e_i^2 = e_1^2 + \ldots + e_n^2$

Գուցե բավական է նվազագույնի ծգտեցնել $\sum_{t=1}^{n} e_i = e_1 + \ldots + e_n$



- Հետազոտվող ուղղի β_0 և β_1 անհայտ պարամետրերի գնահատման համար պետք է հետազոտվի հետևյալ ֆունկցիան.

$$F = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} \left(Y_t - \beta_0 - \beta_1 X_t\right)^2 \to \min$$

- Էքստրեմումի անհրաժեշտ պայմանների համաձայն առաջին ածանցյալներն ըստ յուրաքանչյուր անհայտի պետք է հավասարվեն 0 –ի.

$$\frac{\partial\left(\sum_{t=1}^{n} e_t^2\right)}{\partial\beta_0} = 0, \qquad \frac{\partial\left(\sum_{t=1}^{n} e_t^2\right)}{\partial\beta_1} = 0$$

- Արդյունքում կստանանք հետևյալ գնահատականները.

$$\widehat{\beta}_0 = \vec{\bar{Y}} - \widehat{\beta}_1 \overline{X}$$

$$\widehat{\beta}_1 = \frac{Cov(X,Y)}{V(X)} = \frac{\sum_{t=1}^{n}\left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2}$$

Ակնհայտ է, որ $V(X) \neq 0$:

Հասկանալու համար, թե ինչպե՞ս կարելի է ապահովել այս պայմանի տեղի ունենալը՝ վերհիշենք որոշ կարևոր սահմանումներ:

Հիշեցում

- Ընտրանքային միջին

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

- Ընտրանքային դիսպերսիա

$$V(X) = \frac{1}{n-1}\sum_{t=1}^{n} (X_t - \overline{X})^2 = s_x^2$$

որտեղ s_x-ը ստանդարտ շեղումն է:

- Ընտրանքային կովարիացիան

$$Cov(X,Y) = \frac{1}{n-1}\sum_{t=1}^{n} (X_t - \overline{X})(Y_t - \overline{Y}) = c_{xy}$$

- Ընտրանքային կորելյացիայի գործակիցը

$$r_{xy} = \frac{c_{xy}}{s_x s_y}$$

- Ենթ. ունենք հետևյալ գծային ռեգրեսիոն մոդելը, որի անհայտ պարամետրերը պետք է գնահատենք Փոքրագույն քառակուսիների եղանակով (ի դեպ, որո՞նք են այդ անհայտ պարամետրրերը).

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = \overline{1,\ n}$$

- Գծային ռեգրեսիոն մոդելի **լավագույն գնահատական** ասելով հասկանում ենք այն գհահատականները, որոնք

✓ անշեղելի են. $E(\widehat{\beta}_0) = \beta_0, \quad E(\widehat{\beta}_1) = \beta_1$

✓ կախյալ փոփոխականի նկատմամբ գծային են,

✓ ունեն նվազագույն դիսպերսիա:

- Գաուս-Մարկովի թեորեմի համաձայն ՓՔԵ-ով ստացված գնահատականները՝ լավագույն գնահատականներ են (BLUE – Best Linear Unbiased Estimator):

- Ռեգրեսիոն վերլուծության հիմքը կազմում է **դասական գծային ռեգրեսիոն մոդելը**, որի հիմքում ընկած վարկածները հետևյալն են.

1. մոդելը. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \quad t = \overline{1,n}$
2. անկախ փոփոխականները՝ $x_{t1}, x_{t2}, ..., x_{tk}$ ոչ պատահական մեծություններ են և $V(X) \neq 0$, ընդ որում $x_{t1} = 1, \quad t = \overline{1,n}$
3. ε պատահական մեծության համար իրավացի են հետևյալ պայմանները.

3.1 $E(\varepsilon_t) = 0$

3.2 $E(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ - հոմոսկեդաստիկության պայման,

3.3 $E(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t, s = \overline{1,n}$ - տարբեր ժամանակահատվածների սխալների միմյանց հետ կորելացված չլինելու պայման:

- Եթե ռեգրեսիոն մոդելը բավարարում է դասական գծային ռեգրեսիոն մոդելի 1-3 պայմաններին, համաձայն Գաուս-Մարկովի թեորեմի՝ փոքրագույն քառակուսիների եղանակով ստացված մոդելի անհայտ պարամետրերի գնահատականները կլինեն լավագույնը (նվազագույն դիսպերսիայի իմաստով) բոլոր գծային անշեղելի գնահատականների դասում:
- Հաճախ ավելացվում է նաև սխալների նորմալ բաշխվածության պայմանը.

4. $\varepsilon_t \sim N(0, \sigma^2)$

որի պարագայում 3.3 պայմանից $Cov(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t, s = \overline{1, n}$ հետևում է տարբեր ժամանակների սխալների անկախությունը:

Այս դեպքում մոդելը կոչվում է **դասական նորմալ գծային ռեգրեսիոն**: