6. Gaia

Erregresio Lineal Orokorreko Eredua

Inferentzia eta aurresana

Pilar González eta Susan Orbe

Ekonomia Aplikatua III (Ekonometria eta Estatistika) Saila

Ikasketa-emaitzak

- Inferentziaren helburua ulertu.
- KTA koefizienteen estimatzaileen banaketa lortu.
- Erregresioko koefizienteen tartezko estimazioa lortu.
- Koefizienteen populazio-balioei buruzko hipotesien kontrasteak.
- Aldagai azalduaren aurresana lortu.

Edukia.

1. Zer da inferentzia?
2. KTA estimatzailearen banaketa.
3. Tartezko estimazioa.
4. Kontrasterako hipotesia.
   - Koefiziente bakarraren kontrasteak.
   - Murrizketa lineal orokorren kontrastea.
   - Baterako esangura kontrastea.
   - Alde bateko kontrasteak.
5. Aurresana.
   - Puntuzko aurresana.
   - Tartezko aurresana.
6. Proposatutako jarduera: J6.
7. Proposatutako ariketak: A6.1, A6.2 eta A6.3.

Edukia.

Zer da inferentzia?

Aurreko gaietan erregresio lineal orokorreko ereduaren zehaztapena eta KTA metodoaren estimazioa ikusi da.

Gai honetako helburua, ELOEko hipotesiak eta eskuragarri dagoen laginarekin lortutako estimazioak emanik, populazioko parametroei buruzko ondorioak ateratzea da esangura kontuan hartuz, hau da, merezi duten konfiantzarekin.

Adibidea. Pizzaren kontsumoa.

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i$$

Galdera interesgarriak:

Errentak pizza kontsumoan eragiten du? Emakumeek gizonek baino pizza gehiago kontsumitzen dute? Pizza kontsumoa bezeroaren hezkuntza mailaren mendekoa da? Bigarren mailako hezkuntza eta goi mailako hezkuntza duten bezeroek pizza kontsumo maila berdina dute?

Zer da inferentzia?

Inferentziarako tresnak.

a) **Konfiantza-tarteak:**

Datu-multzo bat lortzeko prozedura zeinetan parametroen benetako balioa izatearen probabilitatea handia den. Balio soil bat izan beharrean, balioen tarte bat izatean, benetako parametroei buruzko informazio gehiago dago, zeren parametroaren balioaz gain bere zehaztasuna ere ematen baitu.

b) **Kontrasterako hipotesiak**:

Parametroen benetako balioei buruzko susmoak eta laginarekin estimatutako balioak alderatzen uzten dituen estatistikako prozedurak dira. Hipotesien kontrasteek, lagineko datuak dauden susmoekin (hipotesiekin) bategarriak diren ala ez, erabakitzen laguntzen du.

Zer da inferentzia?

Inferentzia, populazio-parametroei buruzko susmoak baieztatzeko balio du.

Adibidea. Pizzaren kontsumoa.

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i$$

Errentak pizza kontsumoan eragiten du? $\Rightarrow$ $\beta_2 = 0$

Emakumeek gizonek baino pizza gehiago kontsumitzen dute? $\Rightarrow$ $\beta_4 > 0$

Pizza kontsumoa bezeroaren hezkuntza mailaren mendekoa da?
$\Rightarrow$ $\beta_5 = \beta_6 = \beta_7 = 0$

Bigarren mailako hezkuntza eta goi mailako hezkuntza duten bezeroek pizza kontsumo maila berdina dute? $\Rightarrow$ $\beta_5 = \beta_6$

Edukia.

KTA estimatzailearen banaketa.

Konfiantza-tarteen edota hipotesien kontrasteen prozedura, KTA estimatzailearen banaketaren mendekoak dira.

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u}$$

ELOEko H1-H6 hipotesien mendean,

$$\begin{array}{rcl} E(Y|X) & = & X\beta \\ Bar(Y|X) & = & E((Y - E(Y|X))(Y - E(Y|X))'|X) = E(uu'|X) = \sigma^2 \\ Y|X & \sim & \text{banaketa normala} \end{array}$$

$$\Rightarrow \quad \boxed{Y|X \ \sim \ N(X\beta, \sigma^2)}$$

KTA estimatzailearen banaketa

- $\hat{\beta}$ estimatzailea Xri baldintzaturik, Y-rekiko lineala da.

- Banaketa baldintzatuaren batezbestekoa:

$$E(\hat{\beta}|X) = \beta + E[(X'X)^{-1}X'u|X] = \beta \qquad (E(\hat{\beta}_j|X) = \beta_j, \; j = 1, 2, \ldots, k)$$

- Banaketa baldintzatuaren bariantza:

$$Bar(\hat{\beta}|X) = \sigma^2 (X'X)^{-1} \qquad (Bar(\hat{\beta}_j|X) = \sigma^2_{\hat{\beta}_j} = \sigma^2 \, a_{jj} \; j = 1, 2, \ldots, k)$$

Bariantzak, estimatzaile bati dagokion zehaztasunari buruzko informazioa ematen du. KTA estimatzaileak alboragabeak direnez, bariantza zenbat eta txikiagoa izan, hainbat eta zehatzagoa izango da.

$$\boxed{\hat{\beta}|X \sim N[\beta, \sigma^2 (X'X)^{-1}]} \qquad \boxed{\hat{\beta}_j|X \sim N[\beta_j, \sigma^2_{\hat{\beta}_j}] \quad j = 1, 2, \ldots, k}$$

Edukia.

Tartezko estimazioa.

Koefizienteen estimazioa.

A. Puntuzko koefizientea: $\hat{\beta} = (X'X)^{-1}X'Y.$

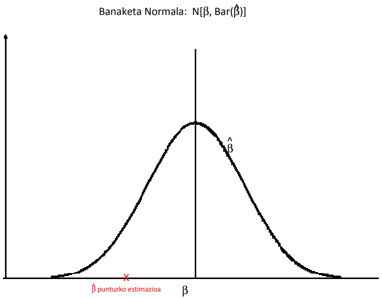
B. Tartezko estimazioa:

Konfiantza-tarteak: β_j benetako balioaren, balio posible guztien multzoa da.
$[A \quad B]$, non $P[A < \beta_j < B] = \% \, (1-\alpha)$ den.

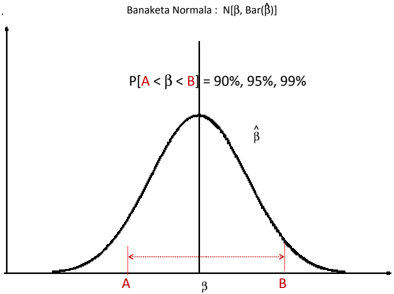
Interpretazioa: Populazio berdineko lagin aleatorio desberdinak izango balira, eta, lagin bakoitzarentzat $[A \ \ B]$ tartea kalkulatuko bagenu, orduan lortutako tarteen % $(1-\alpha)$, β_j benetako balioa izango lukete.

Tartezko estimazioa

Puntuzko estimazioa



Tartezko estimazioa



Tartezko estimazioa

Oharra: X_j aldagaiak estokastikoak izan daitezkenez, estimatzailearen analisi guztia baldintzatua da nahiz eta batzutan espreski ez adierazi.

$$\hat{\beta}_j \sim N[\beta_j, Bar(\hat{\beta}_j)] \qquad \longrightarrow_{Estandarizatuz} \quad \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N[0,1]$$

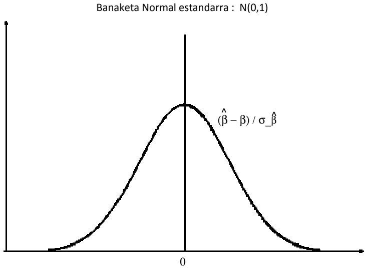
Probabilitate maila bat aukeratu: $(1-\alpha)\,\%$

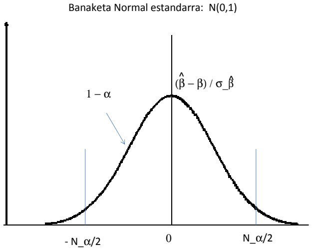
$$P\left[-N_{\alpha/2} \;\leq\; \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq N_{\alpha/2}\right] = (1-\alpha)\,\%$$

$$P\left[\hat{\beta}_j - N_{\alpha/2}\;\sigma_{\hat{\beta}_j} \;\leq\; \beta_j \;\leq\; \hat{\beta}_j + N_{\alpha/2}\;\sigma_{\hat{\beta}_j}\right] = (1-\alpha)\,\%$$

$$KT(\beta_j)_{(1-\alpha)\,\%} = \left[\hat{\beta}_j - N_{\alpha/2}\;\sigma_{\hat{\beta}_j} \qquad \hat{\beta}_j + N_{\alpha/2}\;\sigma_{\hat{\beta}_j}\right]$$

Tartezko estimazioa.





$P[- N_{\alpha/2} < (\hat{\beta} - \beta) / \sigma_{\hat{\beta}} < N_{\alpha/2}] = (1-\alpha)\%$

Tartezko estimazioa.

ARAZOA: Bariantza ezezaguna izaten da:

$$Bar(\hat{\beta}_j) = \sigma^2_{\hat{\beta}_j} = \sigma^2 \, a_{jj}$$

Bariantza estimatu daiteke: $\hat{Bar}(\hat{\beta}_j) = \hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 \, a_{jj}$

Zein da $\dfrac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}}$ estatistikoaren banaketa?

Banaketa hori ateratzea posiblea da zeren:

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N[0,1] \qquad \text{eta} \qquad \frac{\sum \hat{u}_i^2}{\sigma^2} \sim \chi^2(N-k)$$

delako eta gainera izandatzailean eta zatitzailean dauden aldagai aleatorioak independenteak direlako.

Tartezko estimazioa.

Emaitza hau aplikatuz:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2(N-k)}{N-k}}} \; = \; t(N-k)$$

hau lortzen da:

$$\frac{\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}}}{\sqrt{\frac{\frac{\sum \hat{u}_i^2}{\sigma^2}}{N-k}}} \; = \; \frac{\frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{a_{jj}}}}{\sqrt{\frac{\frac{\sum \hat{u}_i^2}{\sigma^2}}{N-k}}} \; = \; \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{a_{jj}}} \; = \; \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}}$$

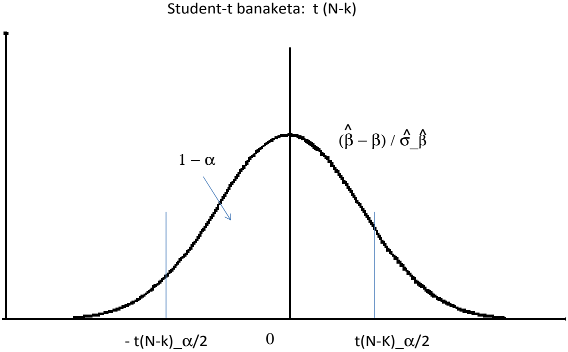
$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \; \sim \; t(N-k)$$

Gogoratu Student-t estatistikoa zeroan zentratuta dagoela baina banaketa normalak baino kola zabalagoak dituela.

Tartezko estimazioa.

Orokorrean, konfiantza-tarteak honako banaketa honetan oinarritzen dira:

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(N-k)$$



P[- t(N-k)_α/2 < (β̂ – β) / σ̂_β̂ < t(N-k)_α/2] = (1-α)%

Tartezko estimazioa.

Konfiantza-tartea.

$$P\left[-t_{\alpha/2}(N-k) \quad \leq \quad \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \quad \leq \quad t_{\alpha/2}(N-k)\right] = (1-\alpha)$$

$$P\left[\hat{\beta}_j - t_{\alpha/2}(N-k)\,\hat{\sigma}_{\hat{\beta}_j} \quad \leq \quad \beta_j \quad \leq \quad \hat{\beta}_j + t_{\alpha/2}(N-k)\,\hat{\sigma}_{\hat{\beta}_j}\right] = (1-\alpha)$$

$$KT(\beta_j)_{(1-\alpha)\,\%} = \left[\hat{\beta}_j - t_{\alpha/2}(N-k)\,\hat{\sigma}_{\hat{\beta}_j} \qquad \hat{\beta}_j + t_{\alpha/2}(N-k)\,\hat{\sigma}_{\hat{\beta}_j}\right]$$

Interpretazioa: Populazio bereko lagin askorentzat tarte hau kalkulatzen bada, orduan konfiantza-tarteen $\%\ (1-\alpha)$, β_j benetako balioa izango dute barnean.

Aplikazioa **6.1 Adibidean.**

Edukia.

Kontrasterako hipotesia.

Erregresio ereduko koefizientengan.

Kontrasterako urratsak:

Lehen urratsa. Erantzun nahi dugun galdera ereduko koefizienteekin erlazionatu. $\rightarrow$ Hipotesi hutsa eta aurkakoa definitu.

Bigarren urratsa. Kontrasterako estatistikoa zehaztu hipotesi hutsaren mendean jarraitzen duen banaketarekin batera.

Hirugarren urratsa. Estatistikoaren balioa kalkulatu eskuragarri dagoen laginarekin.

Laugarren urratsa. Erabaki-araua.

Aplikazioak **6.2 eta 6.3 Adibideetan.**

Kontrasterako hipotesiak.

Adibidea: pizzaren kontsumoa.

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 E_i + \beta_5 BA_i + \beta_6 GR_i + \beta_7 GO_i + u_i$$

A kasua. Errentak pizza kontsumoan eragiten du? $\Rightarrow \ \beta_2 = 0$

B kasua. Emakumeek gizonek baino pizza gehiago kontsumitzen dute?
$\Rightarrow \ \beta_4 > 0$

C kasua. Pizza kontsumoa bezeroaren hezkuntza mailaren mendekoa da?
$\Rightarrow \ \beta_5 = \beta_6 = \beta_7 = 0$

D kasua. Bigarren mailako hezkuntza eta goi mailako hezkuntza duten bezeroek pizza kontsumo maila berdina dute? $\Rightarrow \ \beta_5 = \beta_6$

E kasua. Ereduan barneratutako aldagaiek pizza kontsumoan eragiten dute?
$\Rightarrow \ \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$?

Kontrasterako hipotesiak.

Koefiziente bakarren kontrastea. [A kasua.]

Lehen urratsa. $\begin{cases} H_0: & \beta_j = c \\ H_a: & \beta_j \neq c \end{cases}$

Bigarren urratsa. Kontrasterako t-estatistikoa $t = \dfrac{\hat{\beta}_j - c}{\hat{\sigma}_{\hat{\beta}_j}} \overset{H_0}{\sim} t(N-k)$

Hirugarren urratsa. Estatistikoa kalkulatu.

Laugarren urratsa. Esangura-maila aukeratu: (% α).

α = P[Hipotesi hutsa baztertu | H_0 egia denean]

Erabaki-araua: Hipotesi hutsa baztertu aukeratutako esangura-mailarentzat honako hau ematen bada:

$$\left.\begin{array}{l} t > t_{\alpha/2}(N-k) \\ \\ t < -t_{\alpha/2}(N-k) \end{array}\right\} \Rightarrow \quad |t| > t_{\alpha/2}(N-k)$$

Kontrasterako hipotesia.

Murrizketa lineal orokorren kontrastea. [C kasua.]

Lehen urratsa.
$$\begin{cases} H_0: & \beta_5 = 0,\, \beta_6 = 0,\, \beta_7 = 0 \quad (\beta_5 = \beta_6 = \beta_7 = 0) \\ H_a: & \beta_5 \neq 0 \quad \text{edota} \quad \beta_6 \neq 0 \quad \text{edota} \quad \beta_7 \neq 0 \end{cases}$$

Murriztu Gabeko eredua:

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 E_i + \beta_5 BA_i + \beta_6 GR_i + \beta_7 GO_i + u_i$$

Eredu murriztua:

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 E_i + u_i$$

Kontrasterako hipotesia.

Murrizketa lineal orokorren kontrastea. [C kasua.]

Bigarren urratsa. Kontrasterako F-estatistikoa

$$F = \frac{(HKB_M \ - \ HKB_{MG})/q}{HKB_{MG}/(N-k)} \overset{H_0}{\sim} \mathcal{F}(q, N-k)$$

HKB_M: Murriztutako ereduaren hondar karratuen batura.
HKB_{MG}: Murriztu Gabeko ereduaren hondar karratuen batura.

Hirugarren urratsa. Estatistikoa kalkulatu.

Laugarren urratsa. Hipotesi hutsa baztertu α esangura-mailarentzat honako hau ematen bada:

$$F > \mathcal{F}_\alpha(q, N-k)$$

Kontrasterako hipotesia.

Baterako esangura kontrastea. [E KASUA.]

Aldagai guztien baterako eragina.

Lehen urratsa.

$$\begin{cases} H_0: \quad \beta_2 = \beta_3 = \beta_4 = \ldots = \beta_7 = 0 \\ H_a: \quad \beta_2 \neq 0 \quad \text{edota} \quad \beta_3 \neq 0 \text{ edota} \quad \beta_4 \neq 0 \ \ldots \ \beta_7 \neq 0 \end{cases}$$

Murriztu Gabeko eredua:

$$pizza_i = \beta_1 + \beta_2 errenta_i + \beta_3 adina_i + \beta_4 E_i + \beta_5 BA_i + \beta_6 GR_i + \beta_7 GO_i + u_i \quad (1)$$

Eredu Murriztua:

$$pizza_i = \beta_1 + u_i \quad (2)$$

Kontrasterako hipotesia.

Baterako esangura kontrastea. [E kasua.]

Bigarren urratsa. Kontraste horretako F-estatistiko zehatza:

$$F = \frac{(HKB_M - HKB_{MG})/q}{HKB_{MG}/(N-k)} = \frac{(HKB_M/KBT - HKB_{MG}/KBT)/q}{HKB_{MG}/KBT/(N-k)} =$$

$$= \frac{[(1-R_R^2) - (1-R_{MG}^2)]/q}{(1-R_{MG}^2)/(N-k)} = \frac{R_{MG}^2/q}{(1-R_{MG}^2)/(N-k)}$$

$$F = \frac{R^2/q}{(1-R^2)/(N-k)} \overset{H_0}{\sim} \mathcal{F}(q, N-k)$$

Hirugarren urratsa. Hipotesi hutsa baztertu α esangura-mailarentzat honako hau ematen bada:

$$F > \mathcal{F}_\alpha(q, N-k)$$

Alde bateko kontrastea.

Alde bateko kontrastea. [B kasua.]

Emakumeek gizonek baino pizza gehiago kontsumitzen dute?

Lehen urratsa.

$$\begin{cases} H_0 : \quad \beta_4 = 0 \ (\beta_4 \leq 0) \\ H_a : \quad \beta_4 > 0 \end{cases}$$

Bigarren urratsa. Kontrasterako t-estatistikoa $t = \dfrac{\hat{\beta}_4 - 0}{\hat{\sigma}_{\hat{\beta}_4}} \overset{H_0}{\sim} t(N-k)$

Hirugarren urratsa. Hipotesi hutsa baztertu α esangura-mailarentzat honako hau ematen bada:

$$t > t_\alpha(N-k)$$

Edukia.

Aurresana.

Zer aurresan?

Erregresio Lineal Orokorreko Eredua emanik:

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2 I)$$

Aurresanak.: Y_p balioa estimatu X_p balioak emanik.

Y_p aldagai aleatorioa aurresateko bi aukera daude:

- Puntuzko aurresana $(\hat{Y}_p)$: Y_p-rentzat balio bakar bat estimatzea.
- Tartezko aurresana$(KT(Y_p)_{1-\alpha})$: Y_p-rentzat balio posibleen tarte bat estimatzea, konfiantza-tartea hain zuzen.

Aurresana.

Puntuzko aurresana.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t \qquad t = 1, 2, \ldots, N$$

$$Y_t = [1 \; X_{2t} \; X_{3t} \; \ldots \; X_{kt}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + u_t \qquad\qquad Y_t = X_t' \beta + u_t$$

Aurresanaren baldintzak:

- Y eta X aldagaien arteko erlazioa mantendu egiten da laginatik kanpo.
- $E(u_p|X) = 0, \quad Bar(u_p|X) = \sigma^2 \quad kob(u_p u_t|X) = 0 \qquad t = 1, 2, \ldots T.$

Aurresana.

ELOEn oinarrituz:

$$Y_p \;\; = \;\; X_p' \, \beta + u_p$$

Aldagai azaltzaileen balioak emanik, X_p, aldagai azalduaren puntuzko aurresana hau da:

$$\hat{Y}_p = X_p' \, \hat{\beta}$$

Aurresan errorea:

$$e_p \; = \; Y_p - \hat{Y}_p \; = \; Y_p - X_p' \, \hat{\beta} \; = \; X_p' \, \beta + u_p - X_p' \, \hat{\beta} \; = \; X_p' \, (\beta - \hat{\beta}) + u_p$$

Aldagai azalduaren tartezko aurresana, aurresan errorearen banaketatik eratortzen da.

Aurresana.

Aurresan errorearen banaketa:

$$e_p|X \sim N(0, \sigma_e^2)$$

1. Banaketa normala.
2. Batezbesteko baldintzatua: $E(e_p|X) = E(X_p'\,(\beta - \hat{\beta}) + u_p|X) = 0$
3. Bariantza baldintzatua:

$$\begin{aligned}
Bar(e_p|X) &= \sigma_e^2 = E[e_p - E(e_p)|X]^2 = E[(X_p'\,(\beta - \hat{\beta}) + u_p)\ (X_p'\,(\beta - \hat{\beta}) + u_p)'|X] = \\
&= E[(X_p'\,(\beta - \hat{\beta})(\beta - \hat{\beta})' X_p + u_p\,(\beta - \hat{\beta})' X_p + X_p'\,(\beta - \hat{\beta})\,u_p' + u_p u_p')|X] = \\
&= \sigma_u^2\ [1 + X_p'\,(X'X)^{-1}\,X_p]
\end{aligned}$$

Horrela: $E[X_p'\,(\beta - \hat{\beta})\,u_p'] = 0$

Aurresana.

Tartezko aurresana.

Aurresan-errorearen banaketarekin hasi behar da:

$$e_p|X \sim N(0, \sigma_e^2)$$

- Estandarizatu ondoren:

$$\frac{e_p}{\sigma_e}|X \sim N(0, 1)$$

- Aurresan-errorearen bariantza ezezaguna izaten denez,estimatua izan behar da:

$$\hat{\sigma}_e^2 = \hat{\sigma}^2 \, [1 + X_p' \, (X'X)^{-1} \, X_p]$$

Aurresana.

- Banaketa hau emanik :

$$(T-k) \quad \frac{\hat{\sigma}^2}{\sigma^2}|X \;=\; \frac{\hat{u}'\hat{u}}{\sigma^2}|X \;\sim\; \chi^2(T-k)$$

$$\frac{e_p}{\hat{\sigma}_e}|X \;\sim\; t(T-k) \qquad \Rightarrow \qquad \frac{Y_p-\hat{Y}_p}{\hat{\sigma}_e}|X \;\sim\; t(T-k)$$

Tartezko aurresana % (1- α)

$$P\left[-t_{\alpha/2}(T-k) \quad \leq \quad \frac{Y_p-\hat{Y}_p}{\hat{\sigma}_e} \quad \leq \quad t_{\alpha/2}(T-k)\right] = 1-\alpha$$

$$P\left[\hat{Y}_p \;-\; t_{\alpha/2}(T-k)\;\hat{\sigma}_e \;\leq\; Y_p \;\leq\; \hat{Y}_p \;+\; t_{\alpha/2}(T-k)\;\hat{\sigma}_e\right] = 1-\alpha$$

$$KT(Y_p)_{(1-\alpha)\,\%} \;=\; \left[\hat{Y}_p \;\pm\; t_{\alpha/2}(T-k)\;\hat{\sigma}_e\right]$$

Aplikazioa **6.4 Adibidean.**

Edukia.

Edukia.