Mikroekonometrija

Modeli zavisne promenljive sa ograničenjem

Master studije

Predavač: Aleksandra Nojković

Beograd, školska 2021/22

Struktura predavanja

Modeli zavisne promenljive sa ograničenjem:

- Modeli odsečene raspodele zavisne promenljive
- Modeli cenzurisane zavisne promenljive
- Modeli korekcije uzoračkog izbora
- Modeli trajanja

Modeli ograničene distribucije

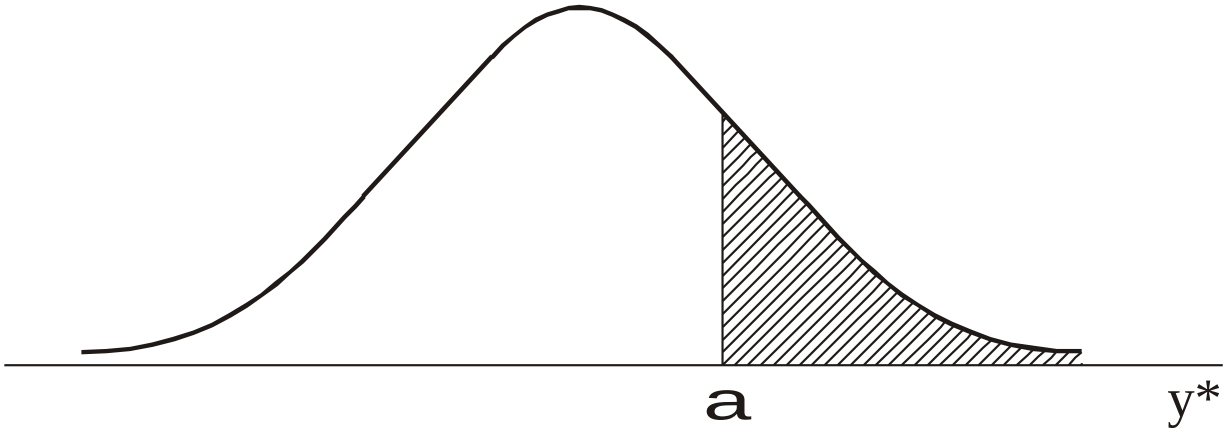
- Ograničene distribucije se koriste pri izučavanju neke kontinuirane pojave koja se javlja u ograničenom intervalu vrednosti.

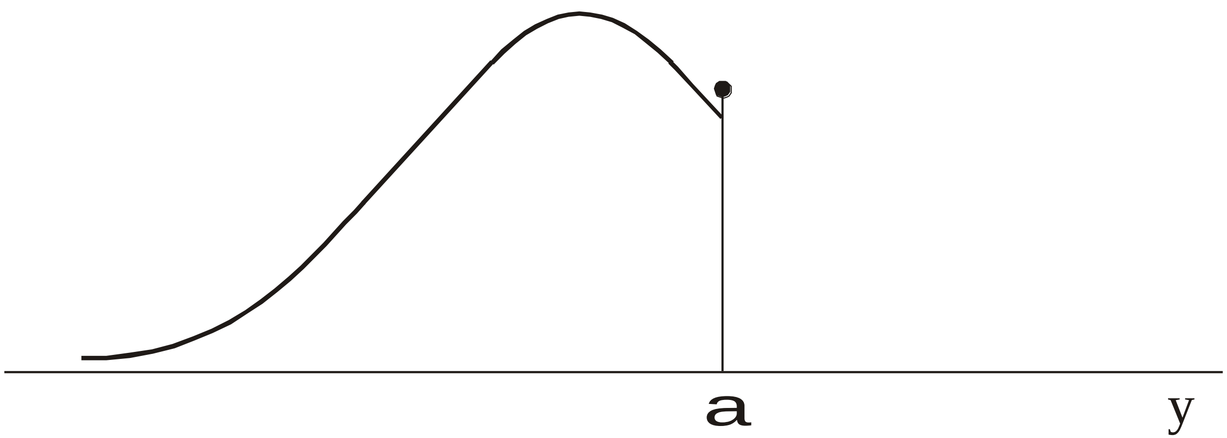
- Podaci sa ograničenjem na vrednosti zavisne promenljive dele se na:

1) <u>Cenzurisane podatke</u>, kada se neke opservacije **zavisne promenljive ne mere** iako su za njih **poznati podaci nezavisnih promenljivih**. Npr. za sve porodice koje su ispod linije siromaštva uzima se da su na samoj liniji.
2) <u>Odsečene raspodele</u>, kada su vrednosti nezavisnih promenljivih poznate **jedino ako se vrednost zavisne promenljive opaža**, odnosno meri. Npr. iznos plaćenog poreza je poznat samo ako su prihodi pojedinca viši od nekog nivoa iznad koga se plaća porez, i slično.

- U literaturi se naješće posmatra uzorak izvučen iz normalne raspodele pod ograničenjem.

Funkcije gustine originalne i cenzurisane normalne raspodele





F-ja gustine cenzurisane normalne raspodele

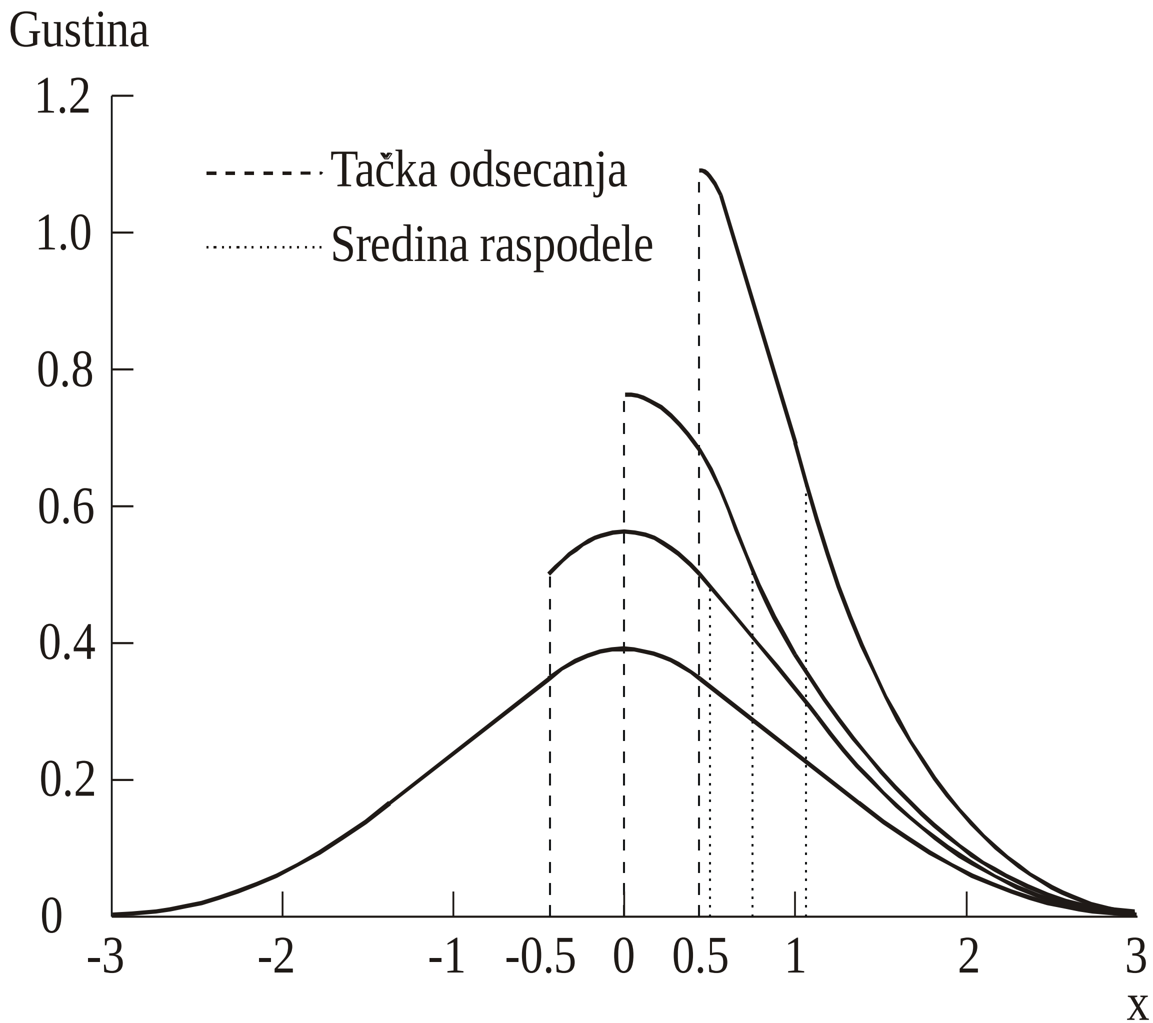
- Odgovarajuća funkcija raspodele cenzurisane slučajne promenljive y (za $y^* \sim N(\mu, \sigma^2)$) za $a = 0$ jeste: Prob(y = 0)

= Prob($y^* \leq 0$) = $\Phi\left(-\frac{\mu}{\sigma}\right)$ = 1 - $\Phi\left(\frac{\mu}{\sigma}\right)$,

a za $y^* > 0$ slučajna promenljiva y ima funkciju gustine neprekidne slučajne promenljive y^*.

- Očekivana vrednost i varijansa cenzurisane normalne slučajne promenljive y definišu se kao funkcija parametara raspodele $y^* \sim N(\mu, \sigma^2)$ i tačke cenzurisanja a.

Funkcija gustine odsečene normalne raspodele



Modeli odsečene distribucije

- Očekivana vrednost i varijansa slučajne promenljive x sa odsečenom normalnom raspodelom definisane su kao:

$$E\,[x \mid \text{odsečeno}] = \mu + \sigma\lambda(\alpha)$$

$$\text{var}\,[x \mid \text{odsečeno}] = \sigma^2[1 - \delta(\alpha)],$$

gde je $\alpha = (a-\mu) / \sigma$, *a* je tačka odsecanja (npr. -0,5; 0 ili 0,5 na slici), μ i σ^2 su parametri originalne normalne raspodele.

- $\boldsymbol{\lambda(\alpha) = \phi(\alpha) / [1 - \Phi(\alpha)] = \phi(\alpha) / \Phi(\alpha)}$ ako je odsečen deo funkcije gustine ispod tačke a (x >a), pri čemu su ϕ *i* Φ funkcija gustine i kumulativne raspodele verovatnoće normalne stan. raspodele.

- Funkcija $\lambda(\alpha)$ se naziva inverzni Mills-ov količnik (*inverse Mills ratio*).

Ocenjivanje modeli ograničene distribucije

- Izostavljanje dela opservacija (odsečena raspodela) ili njihovo izjednačavanje sa vrednošću neke definisane granice dovodi do <u>pristrasnosti.</u>

- Zato je potrebno primeniti metod maksimalne verodostojnosti, **ocenjujući prvo verovatnoću** da se dobiju opservacije koje su iz intervala odsečenih vrednosti.

- Modeli slučajnog odsecanja (modeliranje izbora uzorka).

Modeli korekcije uzoračkog izbora

- Model ponude ženske radne snage (da li žene sa više dece u proseku zarađuju više?)

- Dvostepeni metod ocenjivanja funkcije zarada (Heckman-ova procedura):

1) U prvom stepenu se ocenjije verovatnoća da je žena zaposlena, koja se ocenjuje za svaku osobu u uzorku.
2) U drugom stepenu se model determinisanja zarada koriguje, ugrađujući transformaciju **ovih predviđenih individualnih verovatnoća** kao **dodatnu objašnjavajuću promenljivu.**

- Pristrasnost usled izbora uzorka iz samo jednog dela populacije u suštini je rezultat uticaja promenljive koja nije merena, pa je sličan pristrasnosti usled izostavljanja relevantnog regresora.

Dve relevantne jednačine

1) Prvom jednačinom predstavljen je izbor uzorka:

$$z_i^* = \gamma' w_i + u_i ,$$

gde je: z* latentna promenljiva (koja se ne opaža), w vektor egzogenih promenljivih, a *u* slučajna greška koja ima *normalnu standardizovanu* funkciju raspodele (PROBIT model).

2) Druga jednačina predstavlja regresiju koja je u našem istraživanju od primarnog interesa:

$$y_i = \beta' x_i + \varepsilon_i$$

gde je: x vektor egzogenih promenljivih (od kojih neke mogu biti prisutne i u vektoru w) , a *ε* slučajna greška koja ima *normalnu* raspodelu.

- Uslov identifikovanosti je da postoji **bar jedna objašnjavajuća** promenljiva sadržana u vektoru **x koja ne ulazi u vektor w** (Maddala, 1983).

Dvostepena procedura

- Ako pretpostavimo da greške u_i i ε_i poseduju dvodimenzionalnu normalnu raspodelu sa sredinom 0 i *korelacijom ρ* **$((u_i, \varepsilon_i) \sim N_2\ [0, 0, 1, \sigma_\varepsilon, \rho]$**) i to uvrstimo u izraze za izračunavanje momenata, dobićemo regresioni model koji se može oceniti **na osnovu podataka iz uzorka kojim raspolažemo:**

$$
\begin{aligned}
E\,[y_i \mid y_i \text{ ulazi u uzorak}] &= E\,[y_i \mid z_i^* > 0] \\
&= E\,[y_i \mid u_i > -\gamma' w_i] \\
&= \beta' x_i + E\,[\varepsilon_i \mid u_i > -\gamma' w_i] \\
&= \beta' x_i + \rho\sigma_\varepsilon \lambda_i(\alpha_u) \\
&= \beta' x_i + \beta_\lambda \lambda_i(\alpha_u),
\end{aligned}
$$

pri čemu je **$\lambda_i(\alpha_u)$ – Mills-ov količnik:** $\lambda(\alpha_u) = \mathit{\Phi}(\gamma' w_i / \sigma_u) / \Phi(\gamma' w_i / \sigma_u$

- Dalje sledi da je: $y_i \mid z_i^* > 0 = E\,[y_i \mid z_i^* > 0] + v_i$

$$= \beta' x_i + \beta_\lambda \lambda_i(\alpha_u) + v_i.$$

Heckit model

1. U prvom stepenu metodom maksimalne verodostojnosti ocenjuje se probit model i dobijaju se ocene $\hat{\gamma}$ na osnovu kojih se računa $\hat{\lambda}_i$ za svaku opservaciju.

2. U drugom koraku metodom ONK (odnosno, ukoliko se u obzir uzme problem heteroskedastičnosti greške v_i, metodom ponderisanih NK) ocenjuje se regresija y na x i $\hat{\lambda}_i$ i dobijaju se ocene parametara β i $\beta_\lambda = \rho\sigma_\varepsilon$ (oznaka b i b_λ):

$$E [y_i \mid z_i = 1] = \beta' x + \beta_\lambda \hat{\lambda} (\gamma' w).$$

- Ocene dobijene primenom Heckman-ovog dvostepenog metoda su *konzistentne i asimptotski normalne*, a moguće je dobiti odvojene konzistentne ocene za parametre ρ i σ_ε.

- Primer ARS 2002: Uzorak zaposlenih žena **nije slučajan?** (odluku o učešću u radnoj snazi žene donose na osnovu razlike između rezervacione i tržišne zarade).

Heckman-ov model (napomena)

- Primenom programskog paketa Eviews/Stata Heckman-ovo postupak (primenom MV) ne ocenjuje direktno parametar ρ, već sledeću transformaciju:

$$\text{a tanh}\,\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

- Ovaj rezultat je u tabeli predstavljen kao */athrho*, a zatim je data i ocena za ρ. Kako je atanh (0) = 0, testiranje hipoteze da je atanh ρ = 0 ekvivalentno je testiranju hipoteze da je ρ = 0.
- Slično tome, standardna greška polazne jednačine zarada ne ocenjuje se direktno u proceduri (već se ocenjuje */lnsigma*), ali se u tabeli može pročitati i ocena za netransformisano σ (σ_ε u modelu).

Modeli trajanja

- Predmet je analiza vremena između dva događaja tj. trajanje određenog stanja. Npr. koliko traje i od čega zavisi dužina nezaposlenosti, vreme između dva kvara mašine ili dve porudžbine i sl.

- U ekonomiji se pitanje dužine trajanja nekog stanja vezuje za **verovatnoću da se to stanje promeni**.

- Podaci se odnose na vremenske intervale (**vreme se ne meri kontinuirano)**, pa se analiza svodi na upotrebu funkcije rizika, što nije moguće kada podaci posmatrane pojave nisu diskretni.

- Dve vrste podataka: ili se zna dužina nezaposlenosti ili je ispitanik još uvek nezaposlen.
- Funkcija rizika daje verovatnoću da nezaposlenost prestaje u vremenu t, uz uslov da je trajala sve do tog vremena.
- Reč je **o uslovnoj**, a ne bezuslovnoj **funkciji gusine.**
- **Pogodne funkcije rizika**: nepromenjiv rizik – eksponencijalna funkcija verovatnoće, za opadajući/rastući rizik – Weibull-ova raspodela.