

Analiza doživljenja: teorijske osnove i radionica

DUŠAN MUNDJAR

Sveučilište u Zagrebu, Fakultet organizacije i informatike

e-mail: dusan.mundjar@foi.hr

Prošireni sažetak

Analiza doživljenja obuhvaća različite teorije za analizu vremena koje prođe do nastanka nekog događaja. Standardni primjer je trajanje života, prema kojoj je metoda dobila ime. Analiza doživljenja primjenjuje se na različite vrste događaja, kao što su počinjenje kriminala, razvodi brakova, rađanje, nezaposlenost, završetak obrazovanja, ispravnost stroja i dr. Metoda se koristi u raznim disciplinama s manjim razlikama u metodi i notacijama. Npr. *failure-time* u tehničkim znanostima, *event modeling* u sociologiji itd.

U sklopu izlaganja biti će prikazan uvod u teoriju analize doživljenja. Uvodi se slučajna varijabla vremena do nastanka događaja (trajanje života), T . Za slučajnu varijablu definira se funkcija doživljenja, $S(t) = P(T > t)$. U nastavku se razmatraju izvedene slučajne varijable s uvjetovane na prošlo doživljeno razdoblje, T_x , te se izvode pripadne funkcije doživljenja, $S_t(t) = P(T > t + x \mid T > t)$. Tablice mortaliteta [1], služe kao primjer za slučaj slučajnog diskretnog vremena do nastanka događaja. Izračunava se očekivano vrijeme do nastanka događaja te varijanca slučajnog vremena do nastanka događaja, odnosno računa se očekivano trajanje života i varijanca trajanja života za pojedinu dob.

U nastavku se prelazi na neprekidni slučaj definiranjem intenziteta smrtnosti (hazarda), $\lambda(t)$. Intenzitet smrtnosti je trenutna mogućnost nastanka događaja pretpostavljajući da se događaj nije dogodio to trenutka t . Upotreba intenziteta smrtnosti korisna je iz više aspekata. Omogućuje lokalnu linearnu aproksimaciju funkcije doživljenja i način je dolaska do procjene funkcije doživljenja koja u pravilu nije unaprijed poznata.

Za procjenu funkcije doživljenja može se koristiti parametarski i neparametarski pristup. Najpoznatiji parametarski pristupi modeliranja su ekspanencijalni, Weibullov, lognormalni, log-logistički, Gompertzov, generalizirani gama. Ekspanencijalni model pretpostavlja vezu funkcije doživljenja i ekspanencijalne distribucije ($S(t) = 1 - F(t)$, gdje je $F(t)$ kumulativna funkcija distribucije ekspanencijalne slučajne varijable), a okarakteriziran funkcijom intenziteta $\lambda(t) = \lambda = const$. Prema ovom modelu uvjetna "vjerojatnost" nastanka događaja je konstantna kroz vrijeme. U Weibullovom modelu pretpostavlja se da je funkcija intenziteta oblika $\lambda(t) = \lambda p t^{p-1}$, a funkcija doživljenja tada poprima oblik $S(t) = e^{(-\lambda t)^p}$. Ostali modeli također pretpostavljaju vezanost oblika funkcije doživljenja sa slučajnom varijablom tipa navedenog u nazivu modela. Da bi se slučajna varijabla dobro modelirala rade se procjene parametara.

Ukoliko ni jedan od zakona nije prikladan za modeliranje funkcije doživljenja, do procjene se može doći *Kaplan-Meierovom* metodom. U sklopu izlaganja biti će prikazan problem, a potom definirane formule za modeliranje problema. Korišteni procjenitelj je procjenitelj maksimalne vjerodostojnosti (MLE). *Greenwoodova* formula omogućava nam izračun varijance za slučajne varijable vremena do nastanka događaja. U sklopu ovog dijela biti će ukratko komentiran problem cenzuriranih podataka.

Primjena *Kaplan-Meierove* procjene primarno je usmjerena modeliranju diskretnih slučajnih vremena do nastanka događaja. Za neprekidni slučaj u sklopu izlaganja prikazuje se *Nelson-Aalenova* procjena, također s pripadnim očekivanjima i varijancama.

Osim same procjene funkcije doživljenja, nekad primarni istraživački interes može biti u otkrivanju razlike između dviju ili više homogenih grupa. Funkciji vremena tada se pristupa kao funkciji više varijabli (kovarijati).

Da bi se odredila funkcija vremena radi se procjena modeliranja intenziteta smrtnosti $\lambda(t; z_i) = \lambda(t) \cdot e^{\beta \cdot z_i^T}$, gdje je $\lambda_0(t)$ osnovni, odnosno bazni hazard, a z_i kovarijate (druge varijable). Primjerice, kovarijate mogu biti indikatori spola, uzimanje određenih lijekova, pušenja i dr. β je vektor regresijskih koeficijenata koji pokazuje utjecaj kovarijata na intenzitet smrtnosti. Pošto ne postoji konkretan izračun varijance procjenitelja za testiranje β koeficijenata, uvijek može poslužiti asimptotska procjena, tzv. Cramer-Raova donja granica.

Alat prikazan za korištenje bit će programski jezik **R** s naglaskom na upotrebu paketa *RCmdr*, *survival* i *RExcel* paketa. Za praktični dio izlaganja koristiti će se podatci iz paketa *KMsurv* te će se izraditi simulirani podatci. Simulirat će se slučajno doživljenje dobi 65 godina osobe muškog ili ženskog spola koja trenutno ima 30 godina. Spol pojedine analizirane osobe je slučajno muškog ili ženskog spola s jednakom vjerojatnošću. Maksimalnu doživljenu dob modeliranom generiranjem diskretne slučajne varijable

$$\left(\begin{array}{cccccc} 30 & 31 & \cdots & 64 & & \geq 65 \\ q_{30} & p_{30}q_{31} & \cdots & p_{30} \cdots p_{63}q_{64} & p_{30} \cdots p_{63}p_{64} & \end{array} \right)$$

Vjerojatnosti doživljenja za pojedinu dob preuzete su iz [2]. Tablice mortaliteta razlikuju se za muško i žensko stanovništvo. Nakon izrade simulacija za 500 osoba provodi se test razlike stopa smrtnosti za muško i žensko stanovništvo, kako bi provjerili postojanje razlike na simuliranoj populaciji.

Ključne riječi: analiza doživljenja, R, hazard, Gompelz, Makeham, Kaplan-Meier, Greenwood, Coxov regresijski model

Literatura

- [1] Congdon, P., *Applied Bayesian Modelling*, John Wiley & Sons, Ltd. (2003)
- [2] DZS, *Tablice mortaliteta Republike Hrvatske 2000.-2002.* DZS, Zagreb, (2007)
- [3] German Rodriguez, *Survival Analysis*, Course materials, Princeton University, (2012), dostupno dana 27.7.2012. na <http://data.princeton.edu/pop509/>
- [4] Huzak, M., Bilješke s predavanja na predmetu “modeli doživljanja”, PMF-MO, Zagreb, (2006)
- [5] Macdonald, A.S., Hardy, M.R., *Survival Models: Core Reading–Subject 104*, Faculty and Institute of Actuaries, London, (2000)
- [6] Marloes Maathuis, *Seminar in Statistics: Survival Analysis*, Swiss Federal Institute of Technology Zurich, (2011), dostupno dana 27.7.2012. na <http://stat.ethz.ch/education/semesters/ss2011/seminar>
- [7] Matt Golder, *Introduction to Duration Models*, Course materials, Pennsylvania State University, (2012), dostupno dana 27.7.2012. na <https://files.nyu.edu/mrg217/public/essex.htm>