|  |  |
| --- | --- |
|  | Zbornik radova Fakulteta tehničkih nauka, Novi Sad |

**UDK: 517.98**

**DOI:** [**https://doi.org/10.24867/22JV01Drinic**](https://doi.org/10.24867/22JV01Drinic)

**LINEARNA I NELINEARNA REGRESIJA**

**LINEAR AND NON-LINEAR REGRESSION**

Jelena Drinić, *Fakultet tehničkih nauka, Novi Sad*

**Oblast – MATEMATIKA U TEHNICI**

**Kratak sadržaj –** *Rad je podeljen u četiri poglavlja. U prvom poglavlju upoznajemo se sa pojmom regresione analize i vrstama regresije. U drugom poglavlju uvode se osnovni pojmovi linearne regresije, vrši se njena podela na prostu i višestruku linearnu regresiju i prikazana je metoda najmanjih kvadrata kojom dobijamo ocenu regresionih parametara. Definisani su intervali poverenja i odgovarajuće statistike za testiranje značajnosti. Treće poglavlje razmatra najjednostavnije slučajeve nelinearne regresije, kao i način kako u slučaju nelinarnog modela ocenjujemo parametre. Na kraju su izneseni zaključci.*

**Ključne reči:** *Linearna regresija, Nelinearna regresija, Ocena parametara*

**Abstract –** *The paper is divided into four chapters. In the first chapter, we introduce the concept of regression analysis and types of regression. In the second chapter, the basic concepts of linear regression are introduced, its division into simple and multiple linear regression, and we perform the method of least squares by which we obtain an estimate of the regression parameters. Confidence intervals, appropriate statistics for significance testing are defined. The third chapter discusses the simplest cases of non-linear regression, as well as the way we evaluate the parameters in the case of a non-linear model. Finally, the conclusions are presented.*

**Key words:** *Linear regression Non-linear regression, Parameter estimation*

1. **UVOD**

Jedan od ciljeva u velikom broju istraživanja jeste da se opišu veze među pojavama koje nas okružuju. Najpre, postavljamo pitanje može li se, i u kom obliku, uspostaviti povezanost između obeležja na matematički način. Regresija je statistička tehnika za modelovanje i ocenu veze između promenljivih.

Da bismo opisali ovu vezu potrebno je pronaći model koji povezuje jednu ili više zavisnih promenljivih *,,...,* sa jednom ili više nezavisnih promenljivih pomoću neke funkcionalne zavisnosti, čiji oblik najčešće nije poznat, pa ostaje na istraživaču da izabere onu koja je po nekom kriterijumu najbolja.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**NAPOMENA:**

**Ovaj rad proistekao je iz master rada čiji mentor je bio prof. dr Ljubo Nedović.**

Regresija može biti linearna i nelinearna, u zavisnosti od toga da li vezu opisujemo linearnom ili nelinearnom funkci­jom. Cilj regresione analize je ustanoviti kako se menja u zavisnosti od i šta još pored nezavisnih promenljivih utiče na ishod razmatranja.

1. **LINEARNA REGRESIJA**

Linearna regresija je koristan alat za predviđanje kvanti­tativnog odgovora, na osnovu jedne ili više nezavisnih promenljivih, koje još nazivamo prediktorima. Funkcija kojom se opisuje veza između promenljivih je u ovom slučaju približno linearna. Što je linearna povezanost iz­među odgovora i prediktora veća, to je predviđanje tač­nije. To je upravo zadatak linearne regresije, da pronađe takvu pravu kojom će se najtačnije opisati veza između promen­ljivih. Ukoliko imamo samo jednu nezavisnu promenljivu reč je o prostoj (jednostrukoj) linearnoj re­gresiji. Ako je broj prediktora bar dva, tada je reč o više­strukoj linearnoj regresiji.

* 1. **Prosta linearna regresija**

Prosta linearna regresija predstavlja veoma jednostavan pristup koji se koristi da se objasni slučajna promenljiva pomoću jednog prediktora . Obično se daje u obliku:

|  |  |
| --- | --- |
| . | (1) |

gde smo u smestili sve ono što još utiče na promenu vrednosti u *,* ali nije obuhvaćeno posmatranjem, i njega nazivamo šum ili greška. predstavlja nagib prave, dok predstavlja odsečak na -osi [1].

Za svako istraživanje imamo dat uzorak za i , tj. neka je ) uzorak obima za prediktor i neka je ) uzorak obima za odgovor . Podatke možemo zapisati u obliku uređenih parova Svako može se predstaviti na sledeći način [1]:

|  |  |
| --- | --- |
|  | (2) |

Standardne pretpostavke za model:

* ima normalnu raspodelu,
* ,
* (homoskedastičnost),
* i su međusobno nezavisne za sve i
* je nestohastička promenljiva i za () važi:

i

gde je aritmetička sredina uzorka ().

Iz prva tri uslova zaključujemo da [1].

* + 1. **Ocena parametara**

U modelu (1) predstavljaju koeficijente ili parametre, koji su nam nepoznati. Korišćenjem datih trening podataka , ocenićemo tražene parametre, primenom metode najmanjih kvadrata. Posmatramo uređene parove podataka , . Pretpostavili smo da svako možemo predstaviti kao u modelu (2). Želimo da nađemo onu pravu za koju je greška najmanja tj. da je greška između stvarne vrednosti i vrednosti procenjene modelom minimalna. Neka je:

Tražimo rešenje optimizacionog problema [1]:

|  |  |
| --- | --- |
|  | (3) |

Nalaženje minimuma ove funkcije svodi se na rešavanje:

 ,

Dobijamo sistem normalnih jednačina čijim rešavanjem dobijamo:

 = ,

 .

gde su , aritmetičke sredine uzorka, , vidi [1].

Tako dobijamo uzoračku regresionu pravu = + .

Može se pokazati da je *E*()i , što znači da su ocene nepristrasne, kao i da imaju normalne raspodele tj. i [1].

Kvadratni koreni varijansi ocenjenih parametara su njihove standardne greške tj:

()=

 ()= .

Centrirana ocena za varijansu reziduala se može dobiti pomoću ocene za zbir kvadrata reziduala Detaljnije o ovome vidi u [1]. Centrirana ocena iznosi:

. (4)

* + 1. **Intervali poverenja i testiranje hipoteza**

Do sada smo našli tačkaste ocene za sva tri parametra. Pomoću standardnih grešaka se mogu izračunati intervali poverenja za njih. Interval poverenja se definiše kao opseg vrednosti koji sa određenom verovatnoćom sadrži pravu nepoznatu vrednost parametra.

Opseg je definisan u smislu donje i gornje granice.

Interval poverenja za :

.

Interval poverenja za :

.

Interval poverenja za :

.

Standardne greške se takođe mogu koristiti za testiranje hipoteza. Značajnost koeficijenata regresionog modela (1) vrši se pomoću odgovarajućih statistika. Testiramo hipotezu:

 naspram hipoteze

U praksi, izračunavamo -statistiku, datu sa:

 [1, 2].

* + 1. **Koeficijent determinacije**

Ukupno odstupanje jedne registrovane vrednosti promen­ljive od srednje vrednosti se može podeliti na: mode­lom objašnjeno odstupanje − i neobjašnjeno odstupa­nje registrovanih vrednosti od vrednosti određenih modelom. To možemo zapisati [1]:

.

Uvedimo oznake: , , .

Neka je sa označen količnik odstupanja objašnjenog modelom i ukupnog odstupanja [2]:

= .

Ovaj količnik se zove koeficijent determinacije i pred­stavlja relativnu meru fitovanja. Uvek je između 0 i 1. Za slučaj kada je blizu 0, varijabilnost nije u potpunosti objašnjena modelom, već greškom i za ovakav model smatramo da je veoma loš. Kada je blizu 1, varija­bilnost u je uglavnom objašnjena nezavisnom promen­ljivom , što znači da je model dobar [1, 2].

* 1. **Višestruka linearna regresija**

Model višestruke linearne regresije u matričnom zapisu je oblika:

 (5)

gde su:

Y = , X= =,

Kod standardnih pretpostavki modela treba dodati da u slučaju višestruke regresije imamo i pretpostavku da su promenljive linearno nezavisne [3].

* + 1. **Ocena parametara**

Parametre možemo oceniti različitim metodama, ali najčešće korišćen je metod najmanjih kvadrata, kao i kod proste linearne regresije. Dakle, tražimo takve vrednosti vektora za koje funkcija:

ima minimalnu vrednost [3]. Rešavanje ovog sistema svodi se na sistem normalnih jednačina i njegovim rešavanjem dobijamo vrednosti ocenjenih parametara koje iznose:

Ove ocene su nepristrasne, jer se može pokazati da je =. ima raspodelu. Vidi [3].

Ocena varijanse reziduala je [3]:

. (8)

* + 1. **Intervali poverenja i testiranje hipoteza**

Interval poverenja za parametre :

Interval poverenja za :

Postoje dve grupe postupaka za testiranje hipoteza o parametrima regresije:

* Test o jednom regresijskom parametru (pojedinačan test, test) gde testiramo hipotezu

 naspram hipoteze

Korišćenjem statistike:

*.*

* Test o svim regresijskim parametrima (test o značajnosti regresije, test), gde testiramo hipotezu

 naspram hipoteze .

Test statistika je [3]:

.

* + 1. **Koeficijent determinacije i adjungovani koeficijent determinacije**

Pored koeficijenta determinacije koji računamo kao i u slučaju proste linearne regresije, u slučaju višestruke regresije imamo i adjungovani (korigovani) koeficijent determinacije koji računamo [3]:

.

Uopšteno govoreći, uvodeći dodatne nezavisne promen­ljive u model, povećava se broj ocenjenih parametara uz nepromenjen obim uzorka, čime se povećava broj stepeni slobode pa se smanjuje pouzdanost ocenjivanja. Stoga je prednost korigovanog koeficijenta determinacije to što on uzima u obzir odnos broja promenljivih i obim uzorka, pa ga je prikladno koristiti za modele koji sadrže različit broj nezavisnih promenljivih [3].

1. **NELINEARNA REGRESIJA**

U situacijama u kojima funkcionalan odnos između zavisne promenljive i nezavisne promenljive ne može biti adekvatno aproksimiran linearnim odnosom, koris­timo nelinearnu regresiju. Nelinearni regresioni modeli aproksimiraju vezu izmedu zavisne i nezavisnih promen­ljivih nelinearnom funkcijom. Opisaćemo najjednostav­nije slučajeve nelinearne regresije.

##  Polinomni regresijski model

Često je potrebno proširiti linearnu regresiju na nelinearnu regresiju kako bi povećali reprezentativnost modela. Standardan način da se to uradi je tako da se veza između promenljivih opiše polinomskom funkcijom tj [4]:

Ovaj pristup naziva se polinomna regresija. Isto kao i u slučaju linearne regresije, predstavlja grešku, a nepoznate koeficijente koje je potrebno oceniti. Ocene parametara dobijamo metodom najmanjih kvadrata. Neka je:

.

Tražimo minimum funkcije .

Minimum funkcije tražimo tako što nalazimo prve parcijalne izvode funkcije po parametrima i izjednačavamo ih sa nulom:

Tako dobijamo sistem normalnih jednačina:

Dobijamo jednačinu, a toliko imamo i nepoznatih parametara. Njihovim rešavanjem dobićemo ocenjene koeficijente. Vidi [4].

* 1. **Jednostavni eksponencijalni regresijski model**

Drugi prostiji slučaj nelinearnog modela jeste jednostavan eksponencijalni regresijski model. On je nelinearan po parametrima i ima oblik [4]:

Najpre transformišemo izraz (10) da bude linearan u parametrima pomoću logaritamskih transformacija:

Sada ocenjujemo parametre metodom najmanjih kvadrata. Tražimo minimum funkcije:

.

Rešavanjem ovog optimizacionog problema dobijamo

,

.

Logaritamskom transformacijom model postaje linearan u parametrima (iako nelinearan u promenljivama), a nezavisna promenljiva ostaje u nepromenjenom obliku. Konačno, logaritmovane vrednosti parametara antilogarit­mujemo po bazi 10 da bismo dobili parametre u početnoj jednačini eksponencijalne regresije.

Kao što smo videli, polinomske i eksponencijalne funk­cije mogu se linearizovati određenim transformacijama promenljivih. Takve funkcije nazivamo linearizovane funkcije, jer ih možemo transformisati u funkciju linearnu po nepoznatim parametrima. Međutim, postoje i oni slu­čajevi kada nije moguće uraditi linearizaciju, ali oni neće biti opisani u ovom radu.

1. **ZAKLJUČAK**

Ovaj rad bavi se važnom temom statistike – regresijom. Iako je ova tema veoma opširna, odabrani su i opisani neki od njenih najvažnijih aspekata. Najpre je uveden pojam regresije tj. regresione analize, koja se koristi za utvrđivanje međusobnih odnosa između dve ili više pojava. Ustanovljeno je da u zavisnosti od vrste funkcije koju koristimo da aproksimiramo vezu između promen­ljivih, mogu postojati linearna i nelinarna regresija.

Line­arna regresija koristi linearnu vezu da objasni odnos izme­đu promenljivih, a u zavisnosti od broja nezavisnih pro­menljivih razlikujemo prostu i višestruku linearnu regre­siju. Objašnjeno je kako se pomoću metode najmanjih kvadrata mogu dobiti ocene nepoznatih parametara u modelu, pokazali smo da su ove ocene nepristrasne i normalno raspoređene. Koeficijent determinacije je mera repre­zentativnosti modela, pa je cilj da on bude što bliži 1, a pokazano je i da je kod višestruke regresije bolje za meru valid­nosti modela koristiti adjungovani koeficijent determina­cije. Dalje, u poglavlju nelinearne regresije pokazano je da je nekad umesto linearne funkcije vezu između promen­ljivih bolje opisati nelinearnom funkcijom. Navedeni su neki primeri nelinearnih modela koji se smenom mogu svesti na linearni model.

 **5. LITERATURA**

[1] Z. Lozanov-Crvenković, “Statistika”, *Prirodno- matematički fakultet,* Univerzitet u Novom Sadu, 2012.

[2] G. James, D. Witten, T. Hastie, R. Tibshirani, “An Introduction to Statistical Learning with Applications in R”*,* *Springer*, Njujork, 2013.

[3] Z. Lozanov-Crvenković, “Višestruka regresija*”*, materijal sa predavanja iz predmeta Statističko modeliranje, *Prirodno-matematički fakultet*, Univerzitet u Novom Sadu, 2012.

[4] I. Lulić, “Uporaba metode regresijske analize u rješavanju problema vezanih za inženjersku praksu”, završni rad, *Fakultet strojarstva i brodogradnje*, Univerzitet u Zagrebu, Hrvatska.

**Kratka biografija:**

|  |
| --- |
| **Jelena Drinić** rođena je u Novom Sadu 1998. god. Diplomirala na Prirodno-matematičkom fakultetu 2020. godine. Master rad piše na Fakul­tetu tehničkih nauka iz oblasti Linearne i nelinearne regresije.kontakt: drinicj5@gmail.com |