

Analiza bolne i bezbolne dijabetičke polineuropatije logističkom regresijom

Bošković, Paula

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:322430>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-11**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



Analiza bolne i bezbolne dijabetičke polineuropatije logističkom regresijom

Bošković, Paula

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:322430>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2022-10-25**



Repository / Repozitorij:

[Repository of Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Paula Bošković

ANALIZA BOLNE I BEZBOLNE
DIJABETIČKE POLINEUROPATIJE
LOGISTIČKOM REGRESIJOM

Diplomski rad

Voditelj rada:
Prof. dr. sc. Anamarija Jazbec

Zagreb, rujan 2022.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

*Hvala mentorici prof. dr. sc. Anamariji Jazbec na savjetima i pomoći.
Hvala svim mojim prijateljima i prijateljicama koji su bili uz mene.
Hvala svim mojim suigračima i suigračicama s kojima sam prošla kroz najljepše sportske
trenutke i uspjehe koji su unijeli neopisivu radost u moje studiranje.
I na kraju, najveća hvala mojoj obitelji na podršci i razumijevanju.*

Sadržaj

Sadržaj	iv
Uvod	1
1 Logistička regresija	2
1.1 Regresijska analiza	2
1.2 Logistički model	4
1.3 Procjena parametara	6
1.4 Testiranje adekvatnosti modela (eng. <i>Goodnes-of-fit</i>)	7
1.5 Interpretacija parametara	8
1.6 ROC krivulja	10
2 Analiza bolne i bezbolne dijabetičke polineuropatije logističkom regresijom	12
2.1 Opis varijabli i deskriptivna statistika	12
2.2 Rezultati univarijatnih logističkih regresija	17
2.3 Rezultati multivarijatnih logističkih regresija	30
2.4 Stepwise procedura	36
2.5 Zaključak	39
3 Dodatak	40
3.1 SAS kod	40
Bibliografija	45

Uvod

Dijabetička polineuropatija jedna je od najčešćih komplikacija šećerne bolesti. Već kod otkrića šećerne bolesti oko 14 – 30% bolesnika ima neki oblik oštećenja perifernog živčanog sustava, a oko 13% bolesnika ima dijabetičku polineuropatiju. Zahvaća motorne i senzorne živce, a najizraženija je na nogama. U načelu razlikujemo bolnu i bezbolnu (asimptomatsku) polineuropatiju. Simptomi su: distalne parestezije, bolovi noću, grčevi u mišićima, pečenje u nogama i stopalima, preosjetljivost na dodir, osjećaj boli od dodira pokrivača, neosjetljivost stopala na toplinu i hladnoću, suha koža i drugi. Sve tegobe pojavljuju se simetrično, najprije na prstima, potom na stopalima i zatim na donjim dijelovima potkoljenica. Kako bolest napreduje, tako zahvaća i ruke istim načinom širenja od prstiju prema podlaktici. Tegobe su izraženije u mirovanju i noću, pa remete san. Napredovanje bolesti na donjim dijelovima udova, osobito nogu, uzrokuje hipotrofiju malih mišića s posljedničnom mišićnom slabošću. Također, teži simptomi utječu na kvalitetu života uzrokujući komplikacije poput padova i otvorenih rana na stopalu, što može dovesti do prijeloma, amputacije, čak i smrti. Lakši simptomi očituju se samo kao nelagoda i neugodnost. [1] [11]

U ovom ćemo radu koristeći logističku regresiju analizirati koje varijable statistički značajno utječu te kakav je njihov omjer šanse da dijabetička polineuropatija bude bolna s obzirom na bezbolnu. Originalna baza sadrži podatke 160 pacijenata te je korištena prilikom istraživanja "The impact of neuropathic pain and other comorbidities on the quality of life in patients with diabetes". [2] Za potrebe ovog rada korišteni su podaci 100 pacijenata. Analizu podataka provest ćemo koristeći statistički program SAS.

Poglavlje 1

Logistička regresija

1.1 Regresijska analiza

Regresijska analiza je metoda koja proučava ovisnost jedne zavisne varijable (varijable odaziva) o jednoj ili više drugih nezavisnih varijabli (varijabli poticaja). Glavni cilj je pronaći regresijski model koji najbolje objašnjava odnos zavisne i nezavisnih varijabli, tj. matematičku jednadžbu koja kvantificira tu povezanost. Ako imamo samo jednu nezavisnu varijablu, tada se radi o univarijatnom regresijskom modelu, inače imamo multivarijatni regresijski model. Također, razlikujemo linearnu i nelinearnu povezanost između zavisne varijable i jedne ili više nezavisnih varijabli. Kod univarijatne regresije rezultat tada može biti jednadžba pravca, odnosno krivulje dok je kod multivarijatne regresije rezultat jednadžba ravnine, odnosno plohe. [10]

Linearni model dan je jednadžbom

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon, \quad (1.1)$$

gdje je

- Y - zavisna varijabla (varijabla odaziva)
- X_i - nezavisne varijable (varijable poticaja), $i = 1, \dots, n$
- β_i - parametri modela, $i = 1, \dots, n$
- ϵ - slučajna greška.

U praksi često imamo više opažanja pa linearni model zapisujemo na sljedeći način

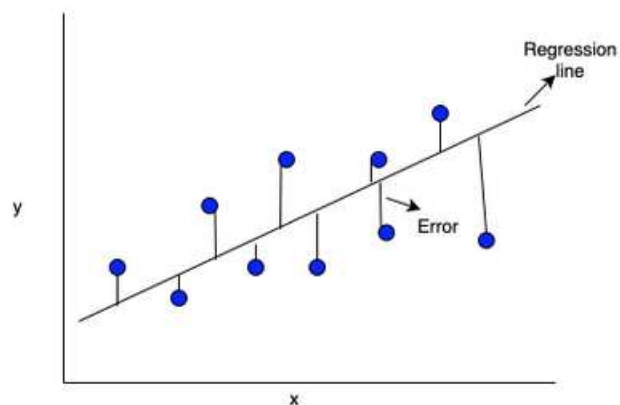
$$Y_j = \beta_0 + \sum_{i=1}^n \beta_i X_{ij} + \epsilon_j, \quad j = 1, \dots, m. \quad (1.2)$$

Četiri glavne pretpostavke koje opravdavaju korištenje linearnog regresijskog modela su:

- linearni odnos između varijabli poticaja i odaziva
- nezavisnost grešaka
- homogenost grešaka
- normalna distribuiranost grešaka.

Ako neka od pretpostavki nije opravdana, naša predviđanja mogu biti pogrešna. [7]

Kao što smo već spomenuli, kod univarijatne linearne regresije povezanost između zavisne i nezavisne varijable opisana je jednadžbom pravca. Taj pravac određujemo metodom najmanjih kvadrata. Iz skupa svih pravaca odaberemo onaj čija je suma odstupanja svake točke od pravca najmanja.



Slika 1.1: Grafički prikaz univarijatnog linearnog modela
izvor: <https://community.cloudera.com/t5/image/serverpage/image-id/25068iFF075A5AEC3B85284>

Ako je zavisna varijabla kategorijska, a ne kontinuirana, tada dolazi do problema. Narušava se pretpostavka o homogenosti i normalnosti grešaka te više nije opravdano koristiti linearni regresijski model, tj. koristeći taj model naša predviđanja mogu biti pogrešna. Kako bi se riješio taj problem, razvili su se različiti alternativni pristupi, a danas je najpopularniji logistička regresija. [6]

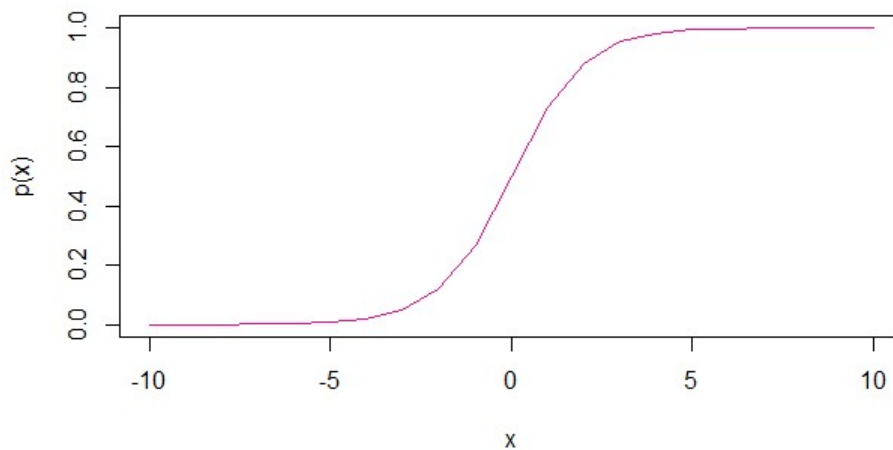
Logistička regresija je vrsta regresijske analize čija je zavisna varijabla kategorijska, najčešće dihotomna (može poprimiti samo dvije vrijednosti). Za razliku od linearne regresije, nema pretpostavki o distribuciji nezavisnih varijabli te za procjenu parametara koristimo metodu maksimalne vjerodostojnosti.

1.2 Logistički model

Logistička funkcija $p : (-\infty, +\infty) \rightarrow (0, 1)$ dana je s

$$p(x) = \frac{1}{1 + e^{-x}} \quad (1.3)$$

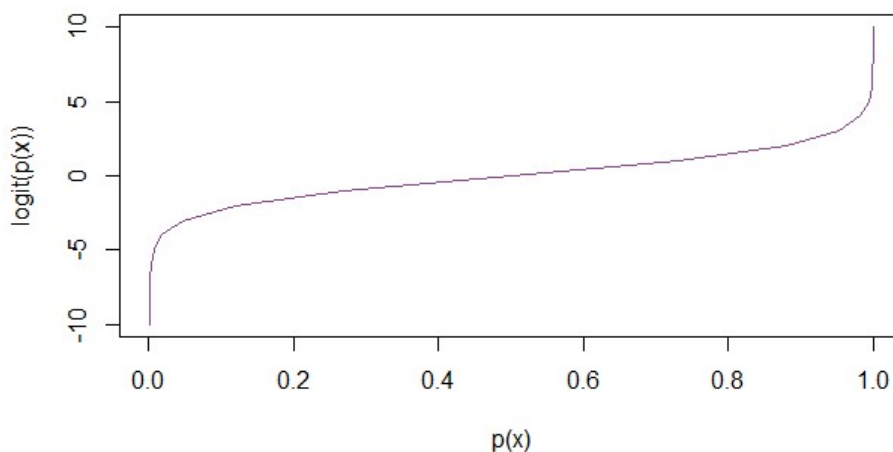
te je prepoznatljiva po svojem S-obliku.



Slika 1.2: Grafički prikaz logističke funkcije

Funkcija *logit* : $(0, 1) \rightarrow (-\infty, +\infty)$, njoj inverzna funkcija, dana je s

$$\text{logit}(p(x)) = \log \left[\frac{p(x)}{1 - p(x)} \right] = \log(p(x)) - \log(1 - p(x)). \quad (1.4)$$



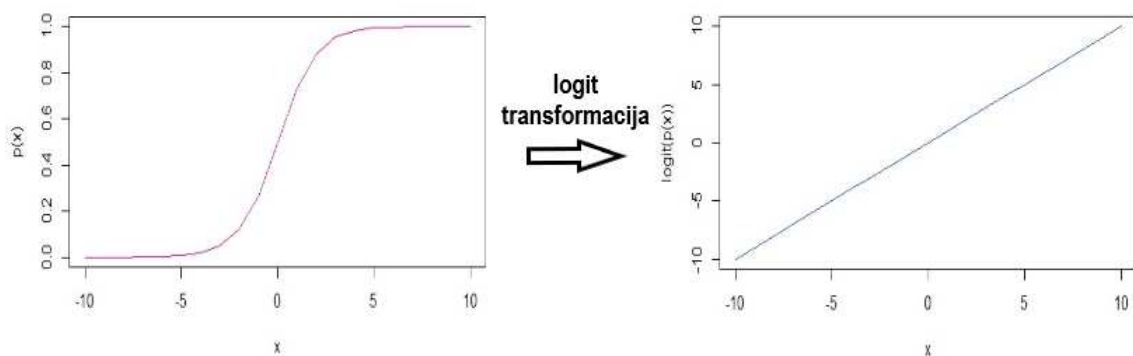
Slika 1.3: Grafički prikaz logit funkcije

Logistički model izražava vjerojatnost da će neka jedinica opažanja ući u jednu dihotomnu skupinu umjesto u drugu. S obzirom na to da je linearni model lakše za promatrati, naš cilj je linearizirati logistički model. Međutim, dolazi do problema jer je funkcija vjerojatnosti ograničena s 0 i 1, dok je linearna funkcija neograničena. Rješenje je transformirati vjerojatnost kako bismo maknuli donju i gornju granicu.

Prvo transformiramo vjerojatnost u šansu čime mičemo gornju granicu. Šansa (eng. *odds*) nekog događaja je omjer očekivanog broja puta kada će se on dogoditi naspram očekivanog broja puta kada se on neće dogoditi. Ako je $p(x)$ vjerojatnost nekog događaja i *odds* šansa, tada je

$$odds = \frac{p(x)}{1 - p(x)} = \frac{\text{vjerojatnost da se događaj dogodio}}{\text{vjerojatnost da se događaj nije dogodio}}. \quad (1.5)$$

Zatim logaritmiramo šansu čime mičemo donju granicu. Pripadna transformacija naziva se logit transformacija logističke funkcije, tj. logističkom regresijom modeliramo logit transformiranu vjerojatnost kao linearnu vezu s nezavisnim varijablama (slika 1.4). [10]



Slika 1.4: Logit transformacija

Dakle, univarijatni logistički model dan je s

$$\text{logit}(p(x)) = \log(odds) = \log \left[\frac{p(x)}{1 - p(x)} \right] = \beta_0 + \beta_1 x. \quad (1.6)$$

Slijedi da je

$$\frac{p(x)}{1 - p(x)} = e^{\beta_0 + \beta_1 x}. \quad (1.7)$$

Ako izrazimo $p(x)$ iz prethodne jednakosti, dobivamo

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}. \quad (1.8)$$

Slično imamo i za multivarijantni logistički model. Ako je $\mathbf{x} = (x_1, \dots, x_m)$, tada je

$$\text{logit}(p(\mathbf{x})) = \log(\text{odds}) = \log \left[\frac{p(\mathbf{x})}{1 - p(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m. \quad (1.9)$$

Slijedi da je

$$p(\mathbf{x}) = p(x_1, \dots, x_m) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}. \quad (1.10)$$

1.3 Procjena parametara

Metoda maksimalne vjerodostojnosti (eng. *maximum likelihood*, oznaka ML) standardna je metoda za procjenu parametara logističkog modela. Kod ML tražimo najmanje moguće odstupanje između opaženih (y) i prediktivnih (\hat{y}) vrijednosti koristeći iterativne matematičke metode. Jednom kada nađemo najbolje rješenje, tj. najmanje moguće odstupanje, zovemo ga *Deviance*, -2LogLikelihood ili *Likelihood ratio*. [10]

Neka je $\mathbf{X} = (X_1, \dots, X_n)$ slučajni uzorak s gustoćom $f(\cdot|\theta)$ gdje je $\theta \in \Theta \subseteq \mathbb{R}^d$. Neka je $\mathbf{x} = (x_1, \dots, x_n)$ realizacija tog uzorka. Funkcija vjerodostojnosti (eng. *likelihood*) $L : \Theta \rightarrow \mathbb{R}$ definirana je s

$$L(\theta|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (1.11)$$

Tada se procjena maksimalne vjerodostojnosti definira kao

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} L(\theta|\mathbf{x}), \quad (1.12)$$

a procjenitelj maksimalne vjerodostojnosti je statistika $\hat{\theta}(\mathbf{X})$.

Često se ne traži maksimum vjerodostojnosti nego njenog logaritma, tzv. log-vjerodostojnosti koja je definirana s

$$l(\theta|\mathbf{x}) = \ln L(\theta|\mathbf{x}) = \sum_{i=1}^n \ln f(x_i|\theta). \quad (1.13)$$

Budući da je logaritam rastuća funkcija, maksimumi tih dviju funkcija postižu se u istoj vrijednosti za θ , a log-vjerodostojnost je u pravilu jednostavnija funkcija za maksimiziranje. [9] [4]

Primijenimo ovo na naš primjer s dihotomnom zavisnom varijablom. Pretpostavimo da imamo n nezavisnih opažanja (x_i, y_i) , $i = 1, \dots, n$. X_i su nezavisne varijable, a Y_i su zavisne varijable s vrijednostima u skupu $\{0, 1\}$. Označimo s \hat{y} ML procjenu od y te sa \hat{p} ML procjenu od $p(x)$. Želimo procijeniti nepoznati parametar $\boldsymbol{\beta} = (\beta_0, \beta_1)$, pri čemu su β_0 i β_1

parametri univarijatnog logističkog modela (analogno vrijedi i za multivarijatni logistički model). [8][10]

Funkcija vjerodostojnosti dana je s

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[p(x_i)^{y_i} \cdot (1 - p(x_i))^{1-y_i} \right]. \quad (1.14)$$

Funkcija log-vjerodostojnosti dana je s

$$l(\boldsymbol{\beta}) = \ln(L(\boldsymbol{\beta})) = \sum_{i=1}^n [y_i \ln(p(x_i)) + (1 - y_i) \ln(1 - p(x_i))]. \quad (1.15)$$

Kako bismo pronašli procjenitelja maksimalne vjerodostojnosti, odnosno vrijednosti za parametar $\boldsymbol{\beta}$ koji maksimiziraju funkciju $l(\boldsymbol{\beta})$, potrebno je parcijalno derivirati funkciju $l(\boldsymbol{\beta})$ po β_0 i β_1 te izjednačiti s nulom.

$$\begin{aligned} \sum_{i=1}^n [y_i - p(x_i)] &= 0 \\ \sum_{i=1}^n x_i [y_i - p(x_i)] &= 0 \end{aligned} \quad (1.16)$$

Jednadžbe se rješavaju iterativnim metodama. Dobiveno rješenje naziva se ML procjenitelj parametra $\boldsymbol{\beta}$ i označava se s $\hat{\boldsymbol{\beta}}$. [10]

1.4 Testiranje adekvatnosti modela (eng. *Goodnes-of-fit*)

Želimo pronaći najbolji model, tj. model koji najbolje opisuje naše podatke. Osnovna mjera adekvatnosti modela je devijanca (eng. *deviance*). Ona je ekvivalentna sumi kvadrata reziduala kod linearne regresije. [8]

Devijancu (oznaka D) računamo kao omjer vjerodostojnosti

$$\begin{aligned} D &= -2 \ln \left[\frac{\text{likelihood modela}}{\text{likelihood saturiranog modela}} \right] \\ &= -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{p}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{p}(x_i)}{1 - y_i} \right) \right] \\ &\approx \chi^2. \end{aligned} \quad (1.17)$$

Saturirani model je model koji sadrži onoliko parametara koliko ima podataka.

Za testiranje razlike između modela s i bez nezavisnih varijabli koristimo statistiku G . Ideja je da testiramo pospješuju li nezavisne varijable model, tj. je li model bolji ako su one uključene. U slučaju da se model pospješuje, očekujemo da se D smanji. [10]

$$\begin{aligned} G &= D(\text{model bez varijabli, samo intercept}) - D(\text{model s } k \text{ varijabli}) \\ &= -2l(0) - (-2l(k)) \\ &= -2 \ln \left[\frac{L(0)}{L(k)} \right] \\ &\approx \chi^2(k) \end{aligned} \tag{1.18}$$

Testiranje značajnosti parametara provodi se pomoću Waldovog testa, gdje su kod univarijantnog logističkog modela hipoteze:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0. \end{aligned} \tag{1.19}$$

Kod multivarijantnog logističkog modela hipoteze su:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 &: \text{barem jedan } \beta_i \neq 0, i = 1, \dots, k. \end{aligned} \tag{1.20}$$

Kažemo da je parametar statistički značajan ako se statistički značajno razlikuje od nule. U tom slučaju pripadna nezavisna varijabla statistički značajno pospješuje model. Ako je parametar približno jednak nuli, onda pripadna nezavisna varijabla nema statistički značajan utjecaj na zavisnu varijablu.

Kako bismo mogli odlučiti odbacujemo li ili ne hipotezu H_0 u korist alternativne hipoteze H_1 , prvo moramo odrediti razinu značajnosti α (najčešće uzimamo da je $\alpha = 5\%$). Zatim dobivenu p -vrijednost uspoređujemo s α . Ako je p -vrijednost $< \alpha$, tada odbacujemo hipotezu H_0 u korist alternativne hipoteze H_1 te kažemo da je pripadna nezavisna varijabla statistički značajna na razini značajnosti α . U suprotnom ne možemo odbaciti hipotezu H_0 u korist alternativne hipoteze H_1 te kažemo da pripadna nezavisna varijabla nije statistički značajna na razini značajnosti α .

1.5 Interpretacija parametara

Za interpretaciju moramo definirati i omjer šanse (eng. *odds ratio*)

$$g(x) := \text{logit}(p(x)) = \log(\text{odds}(p(x))) = \log \left[\frac{p(x)}{1 - p(x)} \right] = \beta_0 + \beta_1 x \tag{1.21}$$

$$g(x + 1) = \beta_0 + \beta_1(x + 1)$$

$$\begin{aligned} g(x + 1) - g(x) &= \text{logit}(p(x + 1)) - \text{logit}(p(x)) \\ &= \log(\text{odds}(p(x + 1))) - \log(\text{odds}(p(x))) \\ &= \log \left[\frac{\text{odds}(p(x + 1))}{\text{odds}(p(x))} \right] = \beta_1 \end{aligned} \quad (1.22)$$

$$\text{odds ratio} = \frac{\text{odds}(p(x + 1))}{\text{odds}(p(x))} \quad (1.23)$$

Parametar β_0 (*intercept*) koji nužno postoji u svakom modelu očekivana je vrijednost zavisne varijable kada je nezavisna varijabla jednaka 0.

(i) Nezavisna varijabla je dihotomna s vrijednostima 0 ili 1, imamo:

$$\begin{aligned} x = 1 &\Rightarrow \text{odds}(p(1)) = \frac{p(1)}{1 - p(1)} \\ x = 0 &\Rightarrow \text{odds}(p(0)) = \frac{p(0)}{1 - p(0)} \end{aligned} \quad (1.24)$$

$$\begin{aligned} g(1) - g(0) &= \log \left[\frac{\text{odds}(p(1))}{\text{odds}(p(0))} \right] = \log \left[\frac{\frac{p(1)}{1 - p(1)}}{\frac{p(0)}{1 - p(0)}} \right] = \log(\text{odds ratio}) = \beta_1 \\ &\Rightarrow \text{odds ratio}(1, 0) = e^{\beta_1} \end{aligned} \quad (1.25)$$

Interpretacija: prijelaz nezavisne varijable iz niže u višu kategoriju povećava omjer šanse da se događaj dogodio ($p(x) = 1$) za e^{β_1} . [10]

Primjer: neka y označava prisutnost ($y = 1$) ili odsutnost ($y = 0$) raka pluća te neka x označava je li osoba pušač ($x = 1$ ako je osoba pušač i $x = 0$ ako osoba nije pušač), tada $\text{odds ratio} = 2$ procjenjuje da postoji dva puta veća vjerojatnost da će se rak pluća pojaviti među pušačima, nego među nepušačima. [8]

(ii) Nezavisna varijabla je kontinuirana, imamo:

$$g(x + 1) - g(x) = \beta_1 \quad (1.26)$$

Parametar β_1 pokazuje promjenu u logaritmiranoj šansi kada se nezavisna varijabla x pomakne za 1. No, često nam pomak za 1 kontinuirane varijable nije značajan te nam je zanimljiviji nešto veći pomak.

$$\begin{aligned} g(x + c) - g(x) &= c\beta_1, \quad c \text{ konstanta} \\ &\Rightarrow \text{odds ratio}(x + c, x) = e^{c\beta_1} \end{aligned} \quad (1.27)$$

Interpretacija: povećanjem nezavisne varijable za c povećava se omjer šanse da se događaj dogodio ($p(x) = 1$) za $e^{c\beta_1}$. [10]

Primjer: neka y označava prisutnost ($y = 1$) i odsutnost ($y = 0$) dijabetesa te neka x označava vrijednost $HbA1c$, ako se $HbA1c$ poveća za konstantu c , tada $oddsratio = 2$ procjenjuje da postoji dva puta veća šansa da će osoba oboljeti od dijabetesa.

Upravo ova jednostavna veza između parametara i omjera šansi jedan je od glavnih razloga zašto se logistička regresija pokazala tako snažnim analitičkim alatom. [8]

1.6 ROC krivulja

ROC krivulja (eng. *Receiver Operating Characteristic curve*), također poznata kao krivulja osjetljivosti, grafički je prikaz valjanosti dijagnostičkog testa. Valjanost dijagnostičkog testa ima dvije mjere: osjetljivost i specifičnost. Osjetljivost testa je udio bolesnih osoba (tj. onih koje je test pravilno prepoznao kao "pozitivne") u ukupnom broju bolesnih. Ako test ima visoku osjetljivost, tada ima nisku lažno negativnu stopu (eng. *false-negative rate*), tj. test neće pokazati lažan negativan rezultat kod mnogo oboljelih osoba. Specifičnost testa je udio osoba koje nisu oboljele (tj. onih koje je test pravilno prepoznao kao "negativne") u ukupnom broju zdravih. Ako test ima visoku specifičnost, tada ima nisku lažno pozitivnu stopu (eng. *false-positive rate*), tj. test neće pokazati lažno pozitivan rezultat kod mnogo neoboljelih osoba. Dakle, osjetljivost testa dobro detektira bolest dok specifičnost dobro detektira osobe koje nisu oboljele. ROC krivulja prikazuje omjer stvarno pozitivnih (osjetljivost) i lažno pozitivnih (1-specifičnost) osoba. Jedna od standardnih mjera za ocjenu dobrote prilagodbe modela je *c-statistika* ili *AUC* (eng. *area under curve*) koja predstavlja površinu ispod ROC krivulje, a čije se vrijednosti kreću između 0 i 1. Ona nam pokazuje koliko je naš model dobar klasifikator. [8] [3] Računa se prema sljedećoj formuli

$$c = \frac{nc + 0.5(t - nc - nd)}{t}, \quad (1.28)$$

gdje je

- nc - broj podudarnih (eng. *concordant*) parova
- nd - broj nepodudarnih (eng. *discordant*) parova
- t - broj parova s različitim vrijednostima odgovora.

Za par opservacija s različitim odgovorima kažemo da je podudaran ako opservacija koja ima više rangirani odgovor (npr. 2 - „događaj se ne dogodi“) ima nižu prediktivnu vjerojatnost da se događaj dogodi od opservacije s niže rangiranim odgovorom (npr. 1 - „događaj

se dogodi“), a kažemo da je nepodudaran ako opservacija koja ima više rangirani odgovor ima višu prediktivnu vjerojatnost da se događaj dogodi od opservacije s niže rangiranim odgovorom. Ako par opservacija nije ni podudaran, ni nepodudaran, kažemo da ima jednak odgovor (eng. *tie*). [10]

Ako je:

- $c < 0.5$, model nema dobru prediktivnu snagu
- $c = 0.5$, model ne predviđa ništa bolje od slučajnog izbora
- $c > 0.8$, model ima jako dobru prediktivnu snagu
- $c = 1$, model ima savršenu prediktivnu snagu. [8]

Poglavlje 2

Analiza bolne i bezbolne dijabetičke polineuropatije logističkom regresijom

2.1 Opis varijabli i deskriptivna statistika

Na pojavu simptoma dijabetičke polineuropatije utječe nekoliko čimbenika: starija životna dob, trajanje šećerne bolesti, lošija kontrola glukoze u krvi, poremećaj masnoća u krvi, pušenje, pretjerivanje u alkoholu, visoka tjelesna težina i drugi. [11] Podaci koje ćemo obraditi sastoje se od 100 opservacija, odnosno pacijenata, od kojih se svaka sastoji od 22 varijable (nezavisne varijable). Zavisna varijabla je dihotomna varijabla koja govori pati li pacijent od bolne ili bezbolne polineuropatije. Na temelju tih podataka želimo ispitati koje varijable i koliko statistički značajno utječu na to da polineuropatija bude bolna.

Nezavisne kontinuirane varijable koje koristimo su sljedeće:

- **Dob** pacijenta (izražena u godinama)
- **Dijabetes god** - koliko dugo pacijent boluje od dijabetesa (izraženo u godinama)
- **HbA1c** (glikozilirani hemoglobin A1c) - mjera količine šećera vezane uz hemoglobin u crvenim krvnim zrnima, pokazatelj prosječne razine šećera u krvi odnosno regulacije šećerne bolesti u posljednja tri mjeseca
- **HDL** - razina "dobrog kolesterola" u krvi (HDL-čestice skupljaju i prenose suvišan kolesterol iz raznih tkiva u jetru)
- **LDL** - razina "lošeg kolesterola" u krvi (LDL-čestice prenose kolesterol i odlažu ga u pojedina tkiva i organe)
- **Trigliceridi** - masnoće, spojevi alkohola glicerola i masnih kiselina.

Tablica 2.1: Deskriptivna statistika nezavisnih kontinuiranih varijabli (SAS ispis)

The MEANS Procedure						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Dob	Dob	100	63.3600000	9.2206619	22.0000000	85.0000000
Dijabetes_god	Dijabetes_god	100	18.7700000	9.6136004	2.0000000	51.0000000
HbA1c	HbA1c	100	7.2160000	1.0807517	5.1000000	10.9000000
HDL	HDL	100	1.3802000	0.3909170	0.6500000	2.6500000
LDL	LDL	100	2.6735000	0.9078417	0.8900000	5.6400000
Trigliceridi	Trigliceridi	100	1.7619000	1.2249483	0.4600000	7.4300000

Nezavisne kategorijske varijable koje koristimo su sljedeće:

- **Spol** pacijenta (0 - muškarac, 1 - žena)

Tablica 2.2: Deskriptivna statistika nezavisne kategorijske varijable Spol (SAS ispis)

The FREQ Procedure				
Spol				
Spol	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	55	55.00	55	55.00
1	45	45.00	100	100.00

- **Pušenje** - je li pacijent pušač (0 - nepušač, 1 - pušač)

Tablica 2.3: Deskriptivna statistika nezavisne kategorijske varijable Pušenje (SAS ispis)

Pušenje				
Pušenje	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	78	78.00	78	78.00
1	22	22.00	100	100.00

- **Alkohol** - konzumira li pacijent alkohol (0 - ne konzumira, 1 - konzumira)

Tablica 2.4: Deskriptivna statistika nezavisne kategorijske varijable Alkohol (SAS ispis)

Alkohol				
Alkohol	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	85	85.00	85	85.00
1	15	15.00	100	100.00

- **NSAR** (nesteroidni protuupalni lijekovi) - imaju analgetičke, antipiretičke i, u većim dozama, protuupalne osobine te ublažavaju bolove, simptome groznice i upale (0 - pacijent ih ne koristi, 1 - pacijent ih koristi)

Tablica 2.5: Deskriptivna statistika nezavisne kategorijske varijable NSAR (SAS ispis)

NSAR				
NSAR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	76	76.00	76	76.00
1	24	24.00	100	100.00

- **Tramadol** - analgetik koji pripada skupini opioda, lijekova protiv bolova koji djeluju na središnji živčani sustav, koristi se za liječenje umjerene do umjereno jake boli (0 - pacijent ga ne koristi, 1 - pacijent ga koristi)

Tablica 2.6: Deskriptivna statistika nezavisne kategorijske varijable Tramadol (SAS ispis)

Tramadol				
Tramadol	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	44	44.00	44	44.00
1	56	56.00	100	100.00

- **Antidepresivi** - koriste se za liječenje depresije, osim toga neki su indicirani i za liječenje kroničnih anksioznih poremećaja, neuropatske boli, bulimije i dr. (0 - pacijent ih ne koristi, 1 - pacijent ih koristi)

Tablica 2.7: Deskriptivna statistika nezavisne kategorijske varijable Antidepresivi (SAS ispis)

Antidepresivi				
Antidepresivi	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	66	66.00	66	66.00
1	34	34.00	100	100.00

- **BMIcat** - indeks tjelesne mase (eng. *Body Mass Index*) jedan je od načina procjene uhranjenosti, pacijenti su svrstani u 3 kategorije (1 - BMI manji od 25, 2 - BMI između 25 i 30, 3 - BMI veći od 30)

Tablica 2.8: Deskriptivna statistika nezavisne kategorijske varijable BMIcat (SAS ispis)

BMIcat				
BMIcat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	27	27.00	27	27.00
2	40	40.00	67	67.00
3	33	33.00	100	100.00

- **Polineuropatija_tip** - tip polineuropatije od koje pacijent boluje, svrstani su u 3 kategorije (1 - tip polineuropatije I, 2 - tip polineuropatije IIa i IIb, 3 - tip polineuropatije III).

Tablica 2.9: Deskriptivna statistika nezavisne kategorijske varijable Polineuropatija_tip (SAS ispis)

Polineuropatija_tip				
Polineuropatija_tip	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	18	18.95	18	18.95
2	69	72.63	87	91.58
3	8	8.42	95	100.00
Frequency Missing = 5				

Sljedeće varijable promatramo kao kontinuirane, ali su nam to zapravo faktori, tj. zbirni rezultati 8 dimenzija SF-36 upitnika (upitnik subjektivne ocjene zdravlja) [5]:

- **PF** - fizičko funkcioniranje
- **RP** - ograničenje zbog fizičkih teškoća
- **BP** - tjelesni bolovi
- **GH** - percepcija općeg zdravlja
- **VT** - vitalnost i energija
- **SF** - socijalno funkcioniranje
- **RE** - ograničenje zbog emocionalnih teškoća
- **MH** - mentalno zdravlje.

Tablica 2.10: Deskriptivna statistika nezavisnih faktorskih varijabli (SAS ispis)

The MEANS Procedure						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PF	PF	100	46.9000000	29.0661037	0	100.0000000
RP	RP	100	36.0000000	42.2414895	0	100.0000000
BP	BP	100	47.4100000	21.0252830	0	90.0000000
GH	GH	100	39.0000000	20.8147190	0	89.0000000
VT	VT	100	52.6000000	21.7594749	0	100.0000000
SF	SF	100	67.5000000	28.2597083	0	100.0000000
RE	RE	100	61.6666667	45.0401990	0	100.0000000
MH	MH	100	63.8400000	21.5513106	4.0000000	100.0000000

Zavisna varijabla je:

- **Polineuropatija** - govori nam pati li pacijent od bolne ili bezbolne polineuropatije (0 - bezbolna, 1 - bolna).

Tablica 2.11: Deskriptivna statistika zavisne varijable Polineuropatija (SAS ispis)

The FREQ Procedure				
Polineuropatija				
Polineuropatija	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	49	49.00	49	49.00
1	51	51.00	100	100.00

2.2 Rezultati univarijatnih logističkih regresija

Za svaku od nezavisnih varijabli provodimo univarijatnu logističku regresiju kako bismo odredili značajnost svake od tih varijabli. S obzirom na to da imamo 22 nezavisne varijable, onda ćemo dobiti i 22 modela. U svakom modelu zavisna varijabla je Polineuropatija. Kod kontinuiranih nezavisnih varijabli nemamo problema za procjenu parametara β_0 i β_1 jer je linerana ovisnost između zavisne i nezavisne varijable, ali za kategorijske nezavisne varijable koje imaju $k \geq 3$ kategorija moramo koristiti pomoćne, tzv. *dummy varijable*. Za svaku od njih kreiramo k-1 pomoćnih varijabli i jednu od njih fiksiramo (tzv. *baseline*) te preostale modeliramo pomoću nje. Za razinu značajnosti uzimamo $\alpha = 5\%$.

Tablica 2.12: Rezultati adekvatnosti univarijatnih logističkih modela

Varijabla	-2LogL (Intercept only)	-2LogL (Intercept and Covariates)	Likelihood Ratio (χ^2)	p-vrijednost
Dob	138.589	135.878	2.7118	0.0996
Dijabetes_god	138.589	138.161	0.4285	0.5127
HbA1c	138.589	138.475	0.1142	0.7354
HDL	138.589	135.983	2.6066	0.1064
LDL	138.589	137.812	0.7777	0.3778
Trigliceridi	138.589	136.197	2.3920	0.1220
Spol	138.589	137.081	1.5085	0.2194
Pušenje	138.589	138.578	0.0113	0.9154
Alkohol	138.589	136.353	2.2361	0.1348
NSAR	138.589	138.252	0.3375	0.5613
Tramadol	138.589	123.742	14.8472	0.0001
Antidepresivi	138.589	134.672	3.9172	0.0478
BMI25 BMI30	138.589	137.732	0.8576	0.6513
tip1 tip3	131.603	129.368	2.2354	0.3270
PF	138.589	102.113	36.4767	< 0.0001
RP	138.589	116.355	22.2342	< 0.0001
BP	138.589	106.447	32.1421	< 0.0001
GH	138.589	115.332	23.2573	< 0.0001
VT	138.589	108.624	29.9652	< 0.0001
SF	138.589	119.941	18.6489	< 0.0001
RE	138.589	129.197	9.3926	0.0022
MH	138.589	119.469	19.1208	< 0.0001

Kriterij konvergencije je zadovoljen za sve varijable. Model je bolji što je devijanca manja, tj. što je *Likelihood ratio* veći. Gledajući p-vrijednosti u tablici 2.12, statistički značajni modeli su oni koji sadrže varijablu Tramadol, Antidepresivi te sve faktorske varijable.

Tablica 2.13: Rezultati procjene parametara univarijatnih logističkih modela metodom maksimalne vjerodostojnosti

Varijabla	Procjena intercepta	Procjena parametra	Standardna greška	Wald (χ^2)	p-vrijednost
Dob	-2.3180	0.0372	0.0234	2.5360	0.1113
Dijabetes_god	0.2979	-0.0137	0.0211	0.4253	0.5143
HbA1c	0.4935	-0.0628	0.1864	0.1137	0.7359
HDL	1.2073	-0.8462	0.542	2.5087	0.1132
LDL	-0.4856	0.1968	0.2248	0.7665	0.3813
Trigliceridi	-0.4292	0.2702	0.1849	2.1356	0.1439
Spol	-0.1823	0.4960	0.4055	1.4961	0.2213
Pušenje	0.0513	-0.0513	0.4828	0.0113	0.9154
Alkohol	0.1651	-0.8581	0.5894	2.1198	0.1454
NSAR	0.1053	-0.2720	0.4697	0.3355	0.5625
Tramadol	-0.8690	1.6163	0.4371	13.6731	0.0002
Antidepresivi	-0.2436	0.8497	0.4362	3.7942	0.0514
BMI25		0.0260	0.4986	0.0027	0.9585
BMI30	-0.1001	0.4055	0.536	0.7329	0.3919
tip1		-0.2521	0.520	0.2246	0.6355
tip3	0.0290	1.0696	0.8513	1.5788	0.2089
PF	2.4195	-0.0511	0.0103	24.5217	< 0.0001
RP	0.8903	-0.0248	0.00585	17.9979	< 0.0001
BP	3.2329	-0.0675	0.0147	21.1805	< 0.0001
GH	2.1341	-0.0538	0.0127	17.7994	< 0.0001
VT	3.2809	-0.0614	0.0134	20.9462	< 0.0001
SF	2.4792	-0.0354	0.00932	14.3968	0.0001
RE	0.9287	-0.0142	0.00480	8.7427	0.0031
MH	3.0777	-0.0469	0.0122	14.8335	0.0001

Na osnovu p-vrijednosti (tablica 2.13), možemo zaključiti da su varijable koje statistički značajno utječu na vjerojatnost da osoba pati od bolne polineuropatije Tramadol te sve faktorske varijable. Također, vidimo da je varijabla Antidepresivi statistički značajna na razini značajnosti 5.14%.

Iz tablice 2.13 možemo iščitati i jednadžbe univarijatnih logističkih modela:

- za Dob

$$\text{logit}(p) = -2.3180 + 0.0372 \cdot \text{Dob}$$

- za Dijabetes_god

$$\text{logit}(p) = 0.2979 - 0.0137 \cdot \text{Dijabetes_god}$$

- za HbA1c

$$\text{logit}(p) = 0.4935 - 0.0628 \cdot \text{HbA1c}$$

- za HDL

$$\text{logit}(p) = 1.2073 - 0.8462 \cdot \text{HDL}$$

- za LDL

$$\text{logit}(p) = -0.4856 + 0.1968 \cdot \text{LDL}$$

- za Trigliceridi

$$\text{logit}(p) = -0.4292 + 0.2702 \cdot \text{Trigliceridi}$$

- za Spol

$$\text{logit}(p) = -0.1823 + 0.4960 \cdot \text{Spol}$$

- za Pušenje

$$\text{logit}(p) = 0.0513 - 0.0513 \cdot \text{Pušenje}$$

- za Alkohol

$$\text{logit}(p) = 0.1651 - 0.8581 \cdot \text{Alkohol}$$

- za NSAR

$$\text{logit}(p) = 0.1053 - 0.2720 \cdot \text{NSAR}$$

- za Tramadol

$$\text{logit}(p) = -0.8690 + 1.6163 \cdot \text{Tramadol}$$

- za Antidepresivi

$$\text{logit}(p) = -0.2436 + 0.8497 \cdot \text{Antidepresivi}$$

- za BMIcat

$$\text{logit}(p) = -0.1001 + 0.0260 \cdot \text{BMI25} + 0.4055 \cdot \text{BMI30}$$

- za Polineuropatija_tip:

$$\text{logit}(p) = 0.0290 - 0.252 \cdot \text{tip1} + 1.0696 \cdot \text{tip3}$$

- za PF

$$\text{logit}(p) = 2.4195 - 0.0511 \cdot PF$$

- za RP

$$\text{logit}(p) = 0.8903 - 0.0248 \cdot RP$$

- za BP

$$\text{logit}(p) = 3.2329 - 0.0675 \cdot BP$$

- za GH

$$\text{logit}(p) = 2.1341 - 0.0538 \cdot GH$$

- za VT

$$\text{logit}(p) = 3.2809 - 0.0614 \cdot VT$$

- za SF

$$\text{logit}(p) = 2.4792 - 0.0354 \cdot SF$$

- za RE

$$\text{logit}(p) = 0.9287 - 0.0142 \cdot RE$$

- za MH

$$\text{logit}(p) = 3.0777 - 0.0469 \cdot MH.$$

Iz prethodnih jednadžbi možemo izračunati i p-vrijednost da osoba s određenim karakteristikama pati od bolne polineuropatije koristeći sljedeću formulu

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}. \quad (2.1)$$

Pogledajmo nekoliko primjera:

- vjerojatnost da osoba čiji je HbA1c jednak 6.4 pati od bolne polineuropatije je

$$p(6.4) = \frac{e^{0.4935 - 0.0628 \cdot 6.4}}{1 + e^{0.4935 - 0.0628 \cdot 6.4}} = 0.00118$$

- vjerojatnost da ženska osoba pati od bolne polineuropatije je

$$p(1) = \frac{e^{-0.1823 + 0.4960 \cdot 1}}{1 + e^{-0.1823 + 0.4960 \cdot 1}} = 0.5777881$$

- vjerojatnost da osoba koja spada u kategoriju BMI30 pati od bolne polineuropatije je

$$p(BMI30) = \frac{e^{-0.1001+0.0260 \cdot 0+0.4055 \cdot 1}}{1 + e^{-0.1001+0.0260 \cdot 0+0.4055 \cdot 1}} = 0.57576$$

- vjerojatnost da osoba čiji je VT score jednak 35 pati od bolne polineuropatije je

$$p(35) = \frac{e^{3.2809-0.0614 \cdot 35}}{1 + e^{3.2809-0.0614 \cdot 35}} = 0.756189.$$

Tablica 2.14: Rezultati procjene omjera šansi univarijatnih logističkih modela

Varijabla	Odds Ratio procjena	95% pouzdani interval
Dob	1.038	0.991 - 1.087
Dijabetes_god	0.986	0.946 - 1.028
HbA1c	0.939	0.652 - 1.353
HDL	0.429	0.151 - 1.223
LDL	1.218	0.784 - 1.892
Trigliceridi	1.310	0.912 - 1.883
Spol	1.642	0.742 - 3.635
Pušenje	0.950	0.369 - 2.447
Alkohol	0.424	0.134 - 1.346
NSAR	0.762	0.303 - 1.913
Tramadol	5.034	2.137 - 11.857
Antidepresivi	2.339	0.995 - 5.500
BMI25	1.026	0.386 - 2.727
BMI30	1.500	0.593 - 3.795
tip1	0.777	0.274 - 2.204
tip3	2.914	0.59 - 15.457
PF	0.950	0.931 - 0.970
RP	0.976	0.964 - 0.987
BP	0.935	0.908 - 0.962
GH	0.948	0.924 - 0.972
VT	0.940	0.916 - 0.965
SF	0.965	0.948 - 0.983
RE	0.986	0.977 - 0.995
MH	0.954	0.932 - 0.977

Drugi način na koji možemo provjeriti statističku značajnost varijable je 95% pouzdani interval za procjenu omjera šanse. Ako on sadrži jedinicu, onda varijabla nije statistički značajna, a ako ju ne sadrži, onda je varijabla statistički značajna. Iz tablice 2.14 možemo vidjeti da su statistički značajne varijable Tramadol te sve faktorske varijable.

Odds ratio procjena, tj. procjena omjera šanse, predstavlja omjere šansi prelaska zavisne varijable iz kategorije da osoba pati od bezbolne polineuropatije u kategoriju da osoba pati od bolne polineuropatije, uz prelazak nezavisne varijable iz niže u višu kategoriju (ako se radi o kategorijskoj varijabli), odnosno pomak za 1 nezavisne varijable (ako se radi o kontinuiranoj varijabli).

tablica 2.14 nam daje sljedeću interpretaciju:

- Povećanjem dobi za 1 godinu povećava se omjer šanse da osoba pati od bolne polineuropatije za 1.038 puta (3.8%).
- Povećanjem trajanja dijabetesa za 1 godinu smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 1.3%.
- Povećanjem HbA1c-a za 1 mjernu jedinicu smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 6.4%.
- Povećanjem HDL-a za 1 mjernu jedinicu povećava se omjer šanse da osoba pati od bolne polineuropatije za 1.218 puta (21.8%).
- Povećanjem triglicerida za 1 mjernu jedinicu povećava se omjer šanse da osoba pati od bolne polineuropatije za 1.310 puta (31%).
- Žene, u odnosu na muškarce, imaju 1.6412 puta (64.12%) veći omjer šanse da pate od bolne polineuropatije.
- Pušači imaju manji omjer šanse da pate od bolne polineuropatije u odnosu na nepušače za 5.2%.
- Osobe koje konzumiraju alkohol imaju manji omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne konzumiraju alkohol za 135.8%.
- Osobe koje koriste nesteroidne protuupalne lijekove imaju manji omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste nesteroidne protuupalne lijekove za 31.3%.
- Osobe koje koriste lijek Tramadol imaju 5.034 puta (403.4%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste lijek Tramadol.

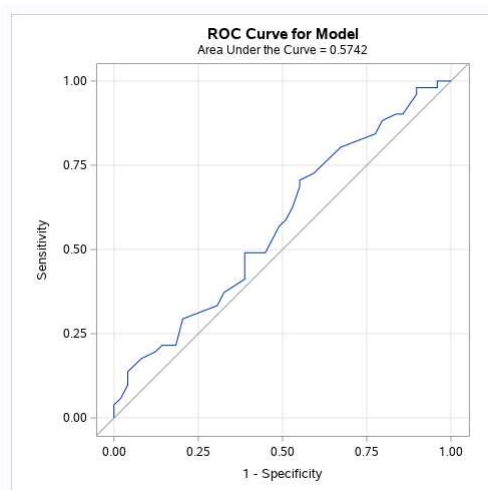
- Osobe koje koriste antidepresive imaju 2.339 puta (133.9%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste antidepresive.
- Osobe koje se nalaze u kategoriji BMI25 imaju 1.026 puta (2.6%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje se nalaze u kategoriji BMI25-30.
- Osobe koje se nalaze u kategoriji BMI30 imaju 1.500 puta (50%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje se nalaze u kategoriji BMI25-30.
- Osobe koje boluju od polineuropatije tipa I imaju manji omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje boluju od polineuropatije tipa II za 28.67%.
- Osobe koje boluju od polineuropatije tipa III imaju 2.914 puta (191.4%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje boluju od polineuropatije tipa II.
- Povećanjem PF *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 5.2%.
- Povećanjem RP *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 2.5%.
- Povećanjem BP *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 6.98%.
- Povećanjem GH *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 5.5%.
- Povećanjem VT *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 6.3%.
- Povećanjem SF *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 3.6%.
- Povećanjem RE *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 1.4%.
- Povećanjem MH *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 4.8%.

Tablica 2.15: Rezultati *c*-statistike za univarijatne logističke modele

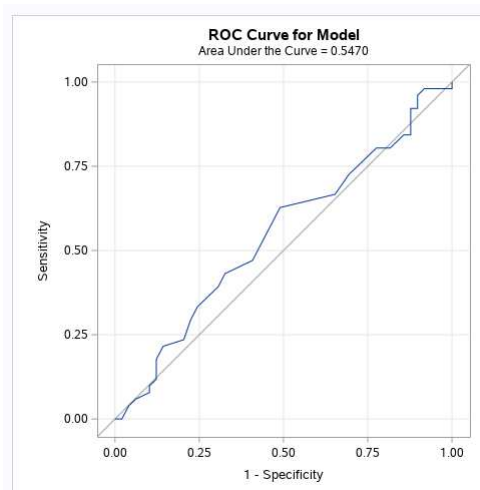
Varijabla	<i>c</i> -statistika	Varijabla	<i>c</i> -statistika
Dob	0.574	Tramadol	0.689
Dijabetes_god	0.57	Antidepresivi	0.593
HbA1c	0.52	PF	0.815
HDL	0.1132	RP	0.762
LDL	0.53	BP	0.789
Trigliceridi	0.575	GH	0.767
Spol	0.561	VT	0.799
Pušenje	0.504	SF	0.732
Alkohol	0.53	RE	0.659
NSAR	0.55	MH	0.745
BMI25 BMI30	0.55	tip1 tip3	0.57

Treći način na koji možemo provjeriti koliko je naš model dobar je površina ispod ROC krivulje koju predstavlja *c*-statistika. Iz tablice 2.15 vidimo da model s varijablom PF (fizičko funkcioniranje) ima najveću prediktivnu snagu ($c = 0.815$). Model s varijablom HDL nema dobru prediktivnu snagu jer je $c = 0.1132$. Također, za modele s varijablama Pušenje, HbA1c, LDL i NSAR možemo reći da ne predviđaju ništa bolje od slučajnog izbora jer im je *c*-statistika dosta blizu 0.5.

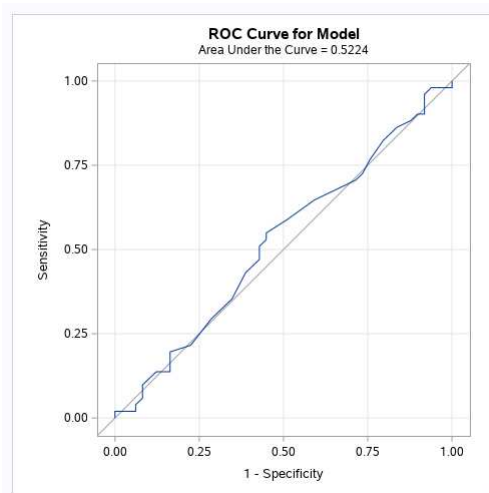
Grafički prikaz ROC krivulja za sve univarijatne logističke modele:



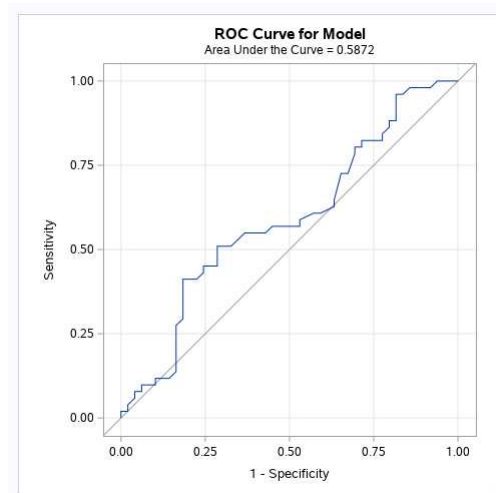
Slika 2.1: ROC krivulja za Dob (SAS ispis)



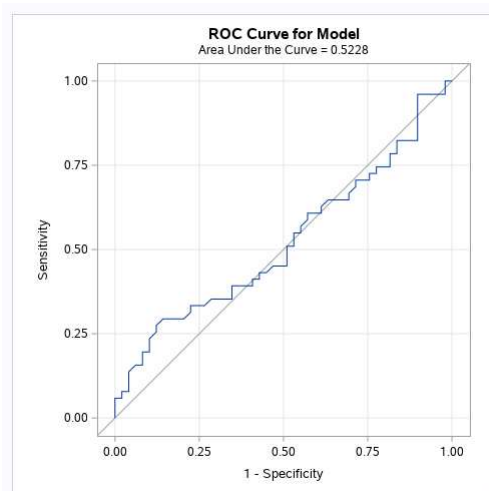
Slika 2.2: ROC krivulja za Dijabetes_god (SAS ispis)



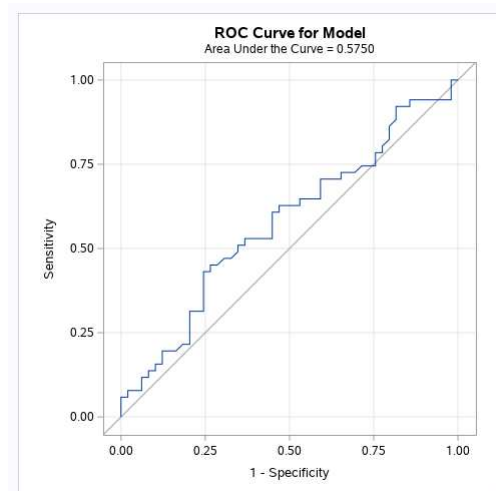
Slika 2.3: ROC krivulja za HbA1c (SAS ispis)



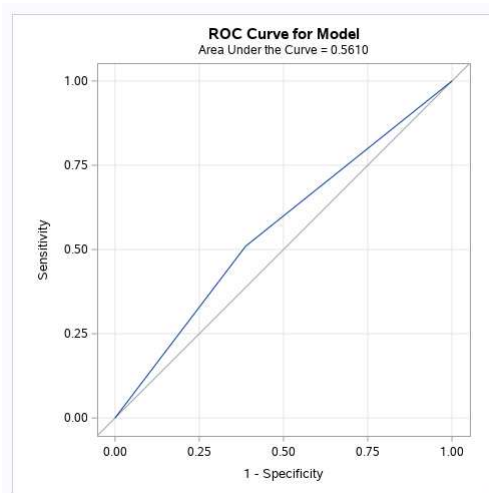
Slika 2.4: ROC krivulja za HDL (SAS ispis)



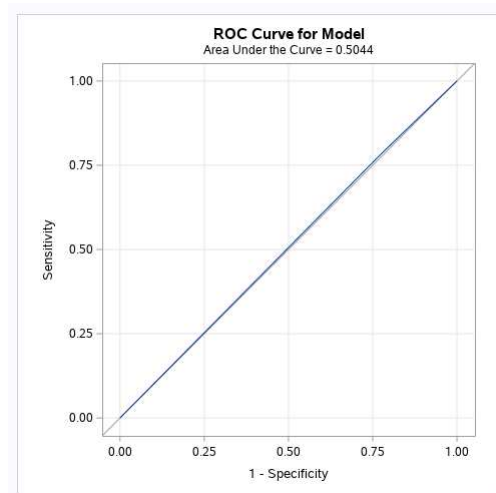
Slika 2.5: ROC krivulja za LDL (SAS ispis)



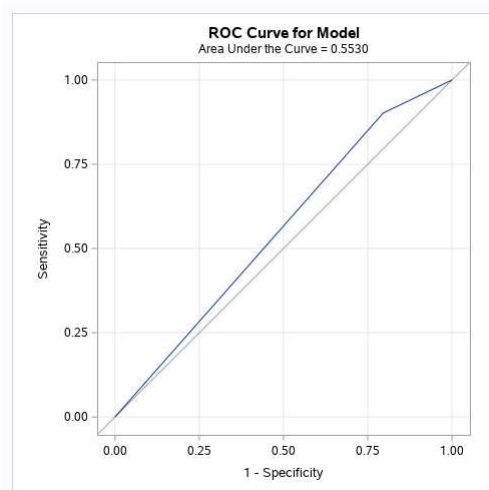
Slika 2.6: ROC krivulja za Trigliceridi (SAS ispis)



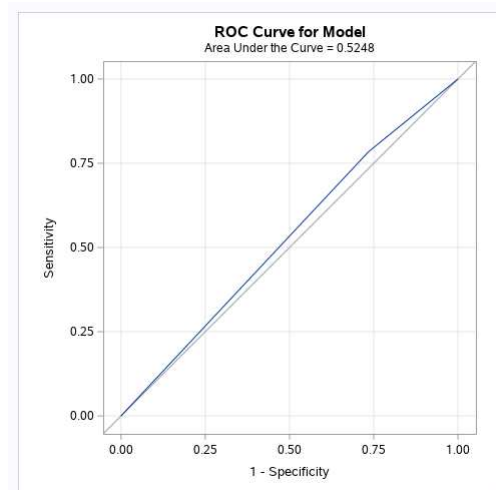
Slika 2.7: ROC krivulja za Spol (SAS ispis)



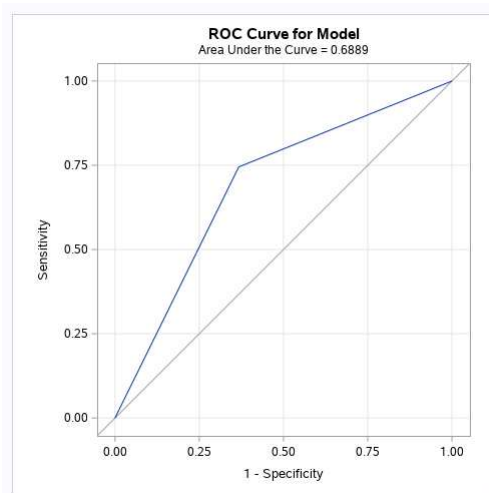
Slika 2.8: ROC krivulja za Pušenje (SAS ispis)



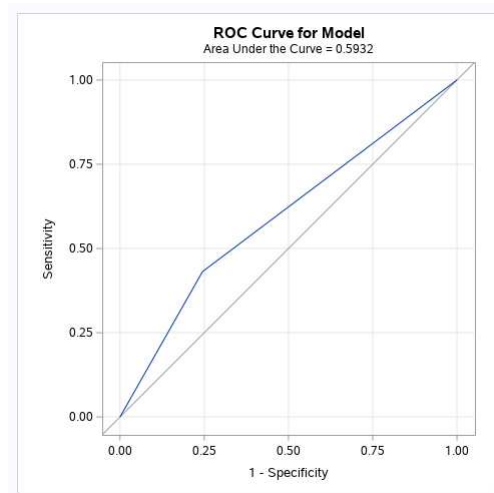
Slika 2.9: ROC krivulja za Alkohol (SAS ispis)



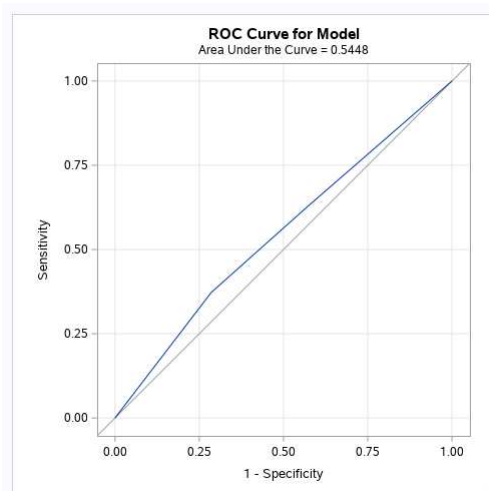
Slika 2.10: ROC krivulja za NSAR (SAS ispis)



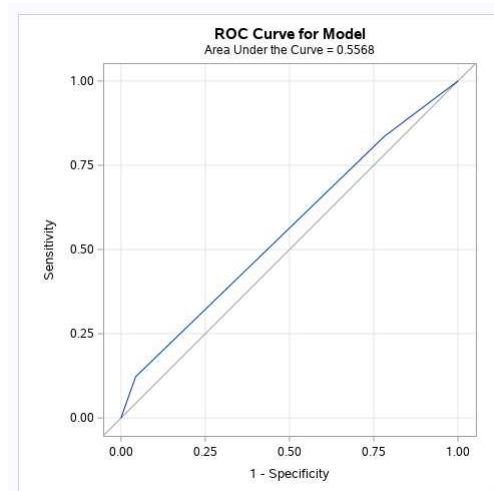
Slika 2.11: ROC krivulja za Tramadol (SAS ispis)



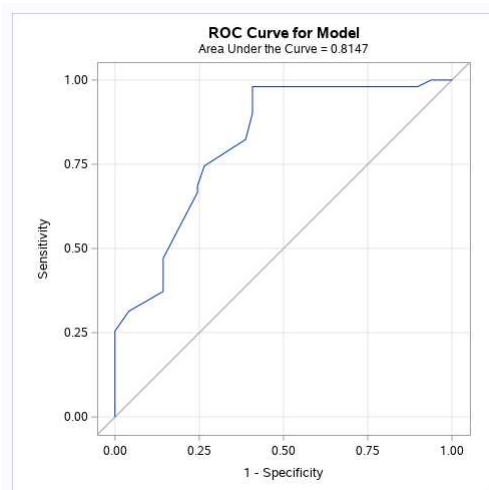
Slika 2.12: ROC krivulja za Antidepresivi (SAS ispis)



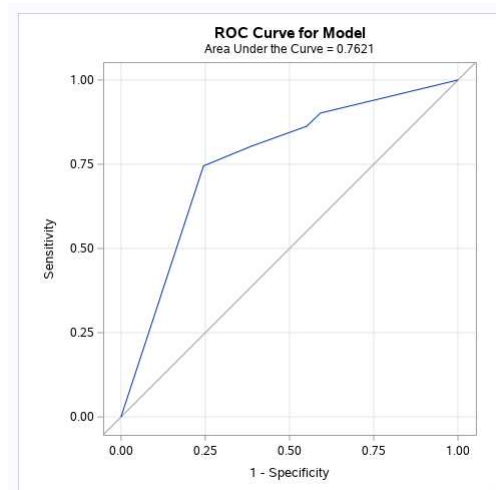
Slika 2.13: ROC krivulja za BMicat (SAS ispis)



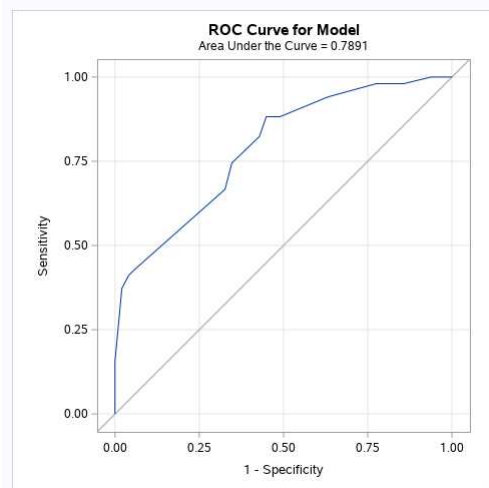
Slika 2.14: ROC krivulja za Polineuropatija_tip (SAS ispis)



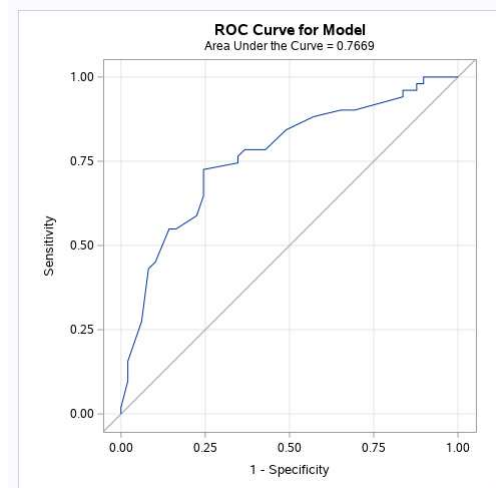
Slika 2.15: ROC krivulja za PF (SAS ispis)



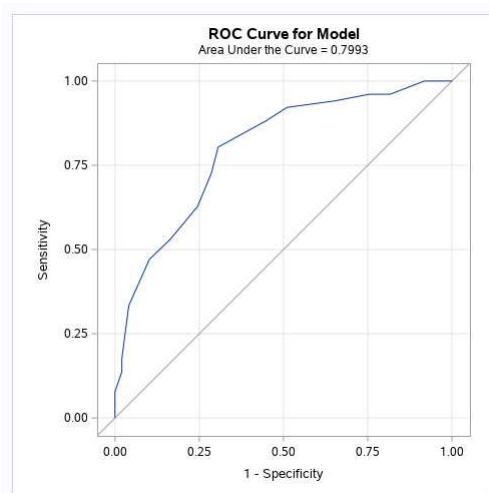
Slika 2.16: ROC krivulja za RP (SAS ispis)



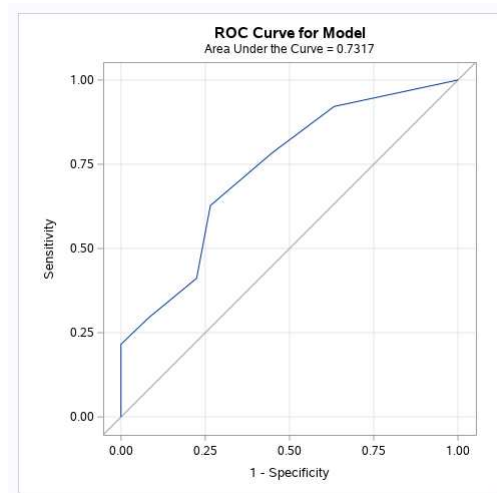
Slika 2.17: ROC krivulja za BP (SAS ispis)



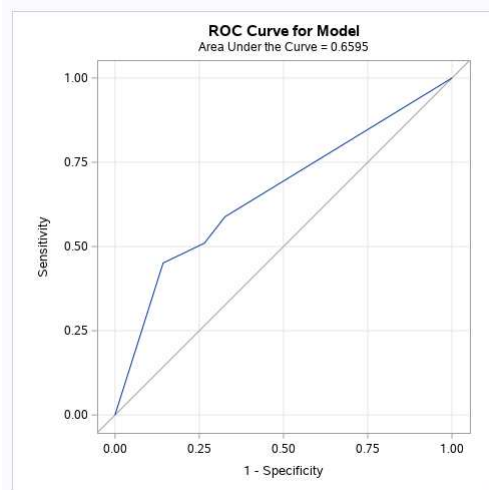
Slika 2.18: ROC krivulja za GH (SAS ispis)



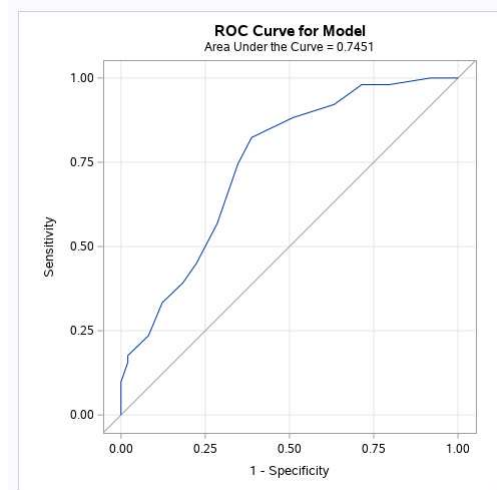
Slika 2.19: ROC krivulja za VT (SAS ispis)



Slika 2.20: ROC krivulja za SF (SAS ispis)



Slika 2.21: ROC krivulja za RE (SAS ispis)



Slika 2.22: ROC krivulja za MH (SAS ispis)

2.3 Rezultati multivarijatnih logističkih regresija

Prvi multivarijatni logistički model sadržava nezavisne varijable (bez faktorskih) koje su se pokazale statistički značajnim kod univarijatnih modela, dok drugi model sadržava i faktorske nezavisne varijable. Za razinu značajnosti, kao i u prethodnom, uzimamo $\alpha = 5\%$.

Tablica 2.16: Rezultati adekvatnosti multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i Antidepresivi

Df	-2LogL (Intercept only)	-2LogL (Intercept and Covariates)	Likelihood Ratio (χ^2)	p-vrijednost
2	138.589	123.140	15.4498	0.0004

Kriterij konvergencije je zadovoljen. Gledajući p-vrijednost u tablici 2.16, zaključujemo da je model statistički značajan.

Tablica 2.17: Rezultati procjene parametara multivarijatnog logističkog modela metodom maksimalne vjerodostojnosti gdje su nezavisne varijable Tramadol i Antidepresivi

Varijabla	Procjena intercepta	Procjena parametra	Standardna greška	Wald (χ^2)	p-vrijednost
Tramadol	-0.9320	1.5055	0.4371	10.8400	0.0010
Antidepresivi	-0.9320	0.4805	0.4362	0.6039	0.4371

Na osnovu p-vrijednosti u tablici 2.17, možemo zaključiti da samo varijabla Tramadol statistički značajno utječe na vjerojatnost da osoba pati od bolne polineuropatije.

Jednadžba multivarijatnog logističkog modela je (tablica 2.17)

$$\text{logit}(p) = -0.9320 + 1.5055 \cdot \text{Tramadol} + 0.4805 \cdot \text{Antidepresivi}.$$

Kao i kod univarijatne logističke regresije pokazat ćemo kako možemo izračunati p-vrijednost da osoba s određenim karakteristikama pati od bolne polineuropatije (koristimo formulu 2.1):

- vjerojatnost da osoba koja koristi lijek Tramadol, ali ne koristi antidepresive pati od bolne polineuropatije je

$$p(1, 0) = \frac{e^{-0.9320+0.4371 \cdot 1+0.4805 \cdot 0}}{1 + e^{-0.9320+0.4371 \cdot 1+0.4805 \cdot 0}} = 0.3787.$$

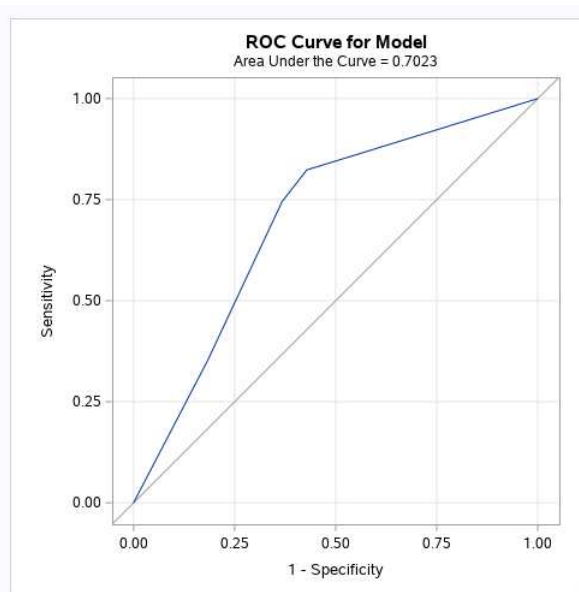
Tablica 2.18: Rezultati procjene omjera šansi multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i Antidepresivi

Varijabla	<i>Odds ratio</i> procjena	95% pouzdani interval
Tramadol	4.506	1.839 - 11.042
Antidepresivi	1.453	0.566 - 3.726

Iz 95% pouzdanih intervala (tablica 2.18) možemo zaključiti da je samo varijabla Tramadol statistički značajna varijabla jer njezin interval ne sadrži jedinicu.

Odds ratio procjena (tablica 2.18) kaže nam da osobe koje koriste lijek Tramadol imaju 4.506 puta (350.6%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste lijek Tramadol.

Još nam preostaje pogledati grafički prikaz ROC krivulje multivarijatnog logističkog modela. Vrijednost *c-statistike* je 0.702 pa ne možemo reći da model ima jako dobru prediktivnu snagu.



Slika 2.23: ROC krivulja multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i Antidepresivi (SAS ispis)

Sve ovo ćemo ponoviti, samo ćemo u model uz varijable Tramadol i Antidepresivi dodati i sve faktorske varijable.

Tablica 2.19: Rezultati adekvatnosti multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol, Antidepresivi i sve faktorske varijable

Df	-2LogL (Intercept only)	-2LogL (Intercept and Covariates)	Likelihood Ratio (χ^2)	p-vrijednost
10	138.589	83.704	54.8854	< 0.0001

Kriterij konvergencije je zadovoljen. Model je statistički značajan jer je p-vrijednost manja od zadane razine značajnosti (tablica 2.19).

Analizirajmo pospješuje li dodavanje faktora model.

$$H = -2\text{LogL}(\text{model bez faktora}) - -2\text{LogL}(\text{model s faktorima})) \sim \chi^2(8)$$

$$H = 123.140 - 83.704 = 39.436$$

$$p - \text{vrijednost} = 0.000004$$

Prema p-vrijednosti, zaključujemo da dodavanje faktora statistički značajno pospješuje model.

Tablica 2.20: Rezultati procjene parametara multivarijatnog logističkog modela metodom maksimalne vjerodostojnosti gdje su nezavisne varijable Tramadol, Antidepresivi i sve faktorske varijable

Varijabla	Procjena parametra	Standardna greška	Wald (χ^2)	p- vrijednost
Intercept	4.1727	1.3928	8.9751	0.0027
Tramadol	1.4279	0.6290	5.1544	0.0232
Antidepresivi	-0.5239	0.6363	0.6779	0.4103
PF	-0.0474	0.0169	7.8249	0.0052
RP	0.00846	0.0106	0.6376	0.4246
BP	-0.0318	0.0212	2.2509	0.1335
GH	0.00510	0.0236	0.0468	0.8288
VT	-0.0285	0.0237	1.4512	0.2283
SF	-0.00025	0.0158	0.0002	0.9875
RE	0.000629	0.00818	0.0059	0.9387
MH	-0.00151	0.0228	0.0044	0.9473

Zadnji stupac tablice 2.20 kaže nam da varijable Tramadol i PF statistički značajno utječu na vjerojatnost da osoba pati od bolne polineuropatije.

tablica 2.20 daje nam i jednadžbu multivarijatnog logističkog modela

$$\text{logit}(p) = 4.1727 + 1.4279 \cdot \text{Tramadol} - 0.5239 \cdot \text{Antidepresivi} - 0.0474 \cdot \text{PF} + 0.00846 \cdot \text{RP} - \\ 0.0318 \cdot \text{BP} + 0.00510 \cdot \text{GH} - 0.0285 \cdot \text{VT} - 0.00025 \cdot \text{SF} + 0.000629 \cdot \text{RE} - \\ 0.00151 \cdot \text{MH}.$$

Vjerojatnost da osoba koja koristi lijek Tramadol te ne koristi antidepresive s PF score-om 35, RP score-om 0, BP score-om 12, GH score-om 10, VT score-om 45, SF score-om 62.5, RE score-om 66.7 i MH score-om 60 pati od bolne polineuropatije je

$$p(1, 0, 35, 0, 12, 10, 45, 62.5, 66.7, 60) = \frac{e^x}{1 + e^x} = 0.90587,$$

gdje je

$$x = 4.1727 + 1.4279 \cdot 1 - 0.5239 \cdot 0 - 0.0474 \cdot 35 + 0.00846 \cdot 0 - 0.0318 \cdot 12 + \\ 0.00510 \cdot 10 - 0.0285 \cdot 45 - 0.00025 \cdot 62.5 + 0.000629 \cdot 66.7 - 0.00151 \cdot 60.$$

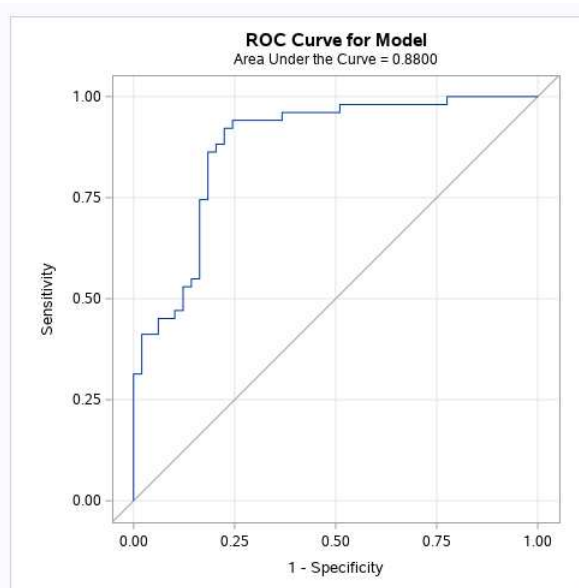
Tablica 2.21: Rezultati procjene omjera šansi multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol, Antidepresivi i sve faktorske varijable

Varijabla	Odds ratio procjena	95% pouzdani interval
Tramadol	4.170	1.216 - 14.306
Antidepresivi	0.592	0.170 - 2.061
PF	0.954	0.923 - 0.986
RP	1.008	0.988 - 1.030
BP	0.969	0.929 - 1.010
GH	1.005	0.960 - 1.053
VT	0.972	0.928 - 1.018
SF	1.000	0.969 - 1.031
RE	1.001	0.985 - 1.017
MH	0.998	0.955 - 1.044

Iz 95%-tnog intervala pouzdanosti (tablica 2.21) možemo zaključiti da su varijable Tramadol i PF statistički značajne varijable jer njihovi intervali ne sadrže jedinicu.

Osobe koje koriste lijek Tramadol imaju 4.170 (317%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste lijek Tramadol. Povećanjem PF score-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 4.85% (tablica 2.21).

Vrijednost *c-statistike* je 0.880 pa možemo reći da model ima jako dobru prediktivnu snagu.



Slika 2.24: ROC krivulja multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol, Antidepresivi i sve faktorske varijable (SAS ispis)

Napravimo sada multivarijatni logistički model u kojem će nezavisne varijable biti samo Tramadol i PF. Zanima nas hoće li se možda prediktivna snaga modela povećati.

Tablica 2.22: Rezultati adekvatnosti multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i PF

Df	-2LogL (Intercept only)	-2LogL (Intercept and Covariates)	Likelihood Ratio (χ^2)	p-vrijednost
2	138.589	90.341	48.2489	< 0.0001

Kriterij konvergencije je zadovoljen. Iz p-vrijednosti u tablici 2.22, zaključujemo da je model statistički značajan.

Tablica 2.23: Rezultati procjene parametara multivarijatnog logističkog modela metodom maksimalne vjerodostojnosti gdje su nezavisne varijable Tramadol i PF

Varijabla	Procjena intercepta	Procjena parametra	Standardna greška	Wald (χ^2)	p-vrijednost
Tramadol	1.5107	1.7834	0.5528	10.4088	0.0013
PF	1.5107	-0.0540	0.0116	21.7945	< 0.0001

Ako pogledamo p-vrijednosti u tablici 2.23, možemo zaključiti da sve varijable statistički značajno utječu na vjerojatnost da osoba pati od bolne polineuropatije.

Jednadžba multivarijatnog logističkog modela je (tablica 2.23)

$$\text{logit}(p) = 1.5107 + 1.7834 \cdot \text{Tramadol} - 0.0540 \cdot \text{PF}.$$

Vjerojatnost da osoba koja ne koristi lijek Tramadol i s PF *score*-om 65 pati od bolne polineuropatije iznosi

$$p(0, 65) = \frac{e^{1.5107+1.7834 \cdot 0 - 0.0540 \cdot 65}}{1 + e^{1.5107+1.7834 \cdot 0 - 0.0540 \cdot 65}} = 0.119276$$

Tablica 2.24: Rezultati procjene omjera šansi multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i PF

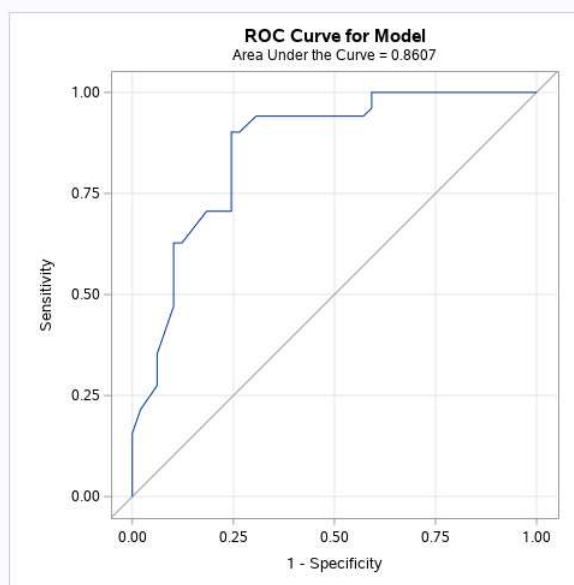
Varijabla	<i>Odds ratio</i> procjena	95% pouzdani interval
Tramadol	5.950	2.014 - 17.582
PF	0.947	0.926 - 0.969

Prema 95%-tnom intervalu pouzdanosti (tablica 2.24), možemo zaključiti da su sve varijable statistički značajne varijable jer njihovi intervali ne sadrže jedinicu.

Prema procjenama omjera šansi (tablica 2.24):

- Osobe koje koriste lijek Tramadol imaju 5.950 (495%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste lijek Tramadol.
- Povećanjem PF *score*-a za 1 smanjuje se omjer šanse da osoba pati od bolne polineuropatije za 5.55%.

Vrijednost *c-statistike* je 0.861 pa možemo reći da model ima jako dobru prediktivnu snagu.



Slika 2.25: ROC krivulja multivarijatnog logističkog modela gdje su nezavisne varijable Tramadol i PF (SAS ispis)

Usporedimo li *c-statistiku* s prethodnim modelom, vidimo da smo sada dobili nešto manju prediktivnu snagu modela. No, napravili smo model sa samo dvije varijable, dok smo u prethodnom imali njih deset.

2.4 Stepwise procedura

Postoji još nekoliko metoda selekcije varijabli za konstruiranje logističkog modela. Svaka metoda temelji se na tome da se za svaku nezavisnu varijablu provjerava kako ona utječe na model te se na temelju toga odlučuje hoće li se ona dodati, odnosno izbaciti iz modela.

Forward selekcija, tj. selekcija unaprijed, metoda je koja počinje s praznim modelom te dodaje nezavisne varijable jednu po jednu. U svakom koraku dodaju se ona varijabla koja najviše doprinosi adekvatnosti modela.

Backward selekcija, tj. selekcija unatrag, metoda je koja počinje s punim modelom, odnosno modelom koji sadrži sve nezavisne varijable, te izbacuje nezavisnu varijablu jednu po jednu. U svakom koraku izbacuje se ona varijabla koja najmanje doprinosi adekvatnosti modela.

Stepwise selekcija kombinacija je *Forward* i *Backward* selekcije. Započinje kao *Forward* selekcija, ali svaka nezavisna varijabla koja uđe u model ne mora nužno i ostati u njemu. Nakon svakoga dodavanja nezavisne varijable, testira se može li se neka varijabla izbaciti

iz modela, a da pritom ne dođe do značajnog smanjenja adekvatnosti modela. Kada više nema varijabli koje se mogu dodati ili ako je model jednak nekom od modela iz prethodnih koraka, tada procedura završava. Upravo ćemo ovu selekciju sada i primijeniti.

Tablica 2.25: Rezultati adekvatnosti logističkog modela dobivenog *stepwise* selekcijom

Df	-2LogL (Intercept only)	-2LogL (Intercept and Covariates)	Likelihood Ratio (χ^2)	p-vrijednost
2	131.603	87.449	44.1546	< 0.0001

Kriterij konvergencije je zadovoljen. Na osnovu p-vrijednosti (tablica 2.25) možemo zaključiti da je model statistički značajan.

Tablica 2.26: Rezultati procjene parametara logističkog modela dobivenog *stepwise* selekcijom metodom maksimalne vjerodostojnosti

Varijabla	Procjena parametra	Standardna greška	Wald (χ^2)	p- vrijednost
Intercept	1.6292	0.6035	7.2872	0.0069
PF	-0.0532	0.0115	21.4963	< 0.0001
Tramadol	1.5922	0.5582	8.1353	0.0043

Prema rezultatima iz tablice 2.26, slijedi da sve varijable statistički značajno utječu na vjerojatnost da osoba pati od bolne polineuropatije.

Jednadžba pripadnog multivarijatnog logističkog modela dana je s (tablica 2.26)

$$\text{logit}(p) = 1.6292 + 1.5922 \cdot \text{Tramadol} - 0.0532 \cdot \text{PF}.$$

Tablica 2.27: Rezultati procjene omjera šansi logističkog modela dobivenog *stepwise* selekcijom

Varijabla	Odds ratio procjena	95% pouzdani interval
Tramadol	4.914	1.646 - 14.676
PF	0.948	0.927 - 0.970

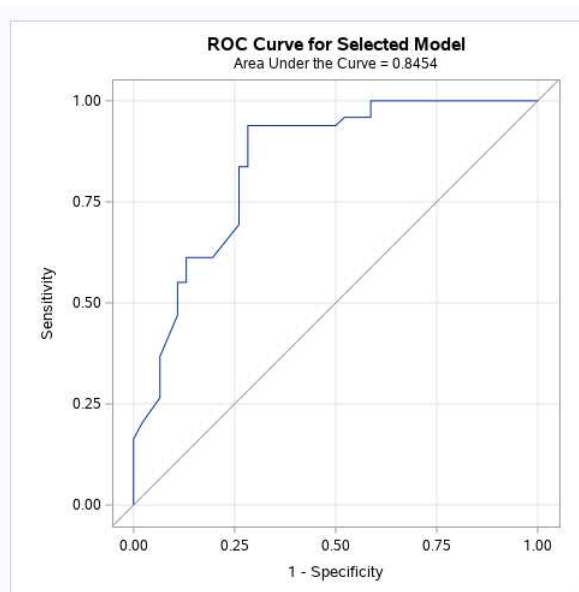
Gledajući 95% pouzdane intervale u tablici 2.27, možemo zaključiti da su sve varijable statistički značajne varijable jer njihovi intervali ne sadrže jedinicu.

Omjeri šanse za ovaj model nam ukazuju da osobe koje koriste lijek Tramadol imaju 4.914 (391.4%) veći omjer šanse da pate od bolne polineuropatije u odnosu na osobe koje ne koriste lijek Tramadol, a povećanjem PF *score*-a za 1 smanjuje se omjer šanse da osoba

pati od bolne polineuropatije za 5.46% (Talica 2.27).

Primijetimo da su omjeri šansi za model 2.24 i 2.27 malo drugačiji iako se u modelu nalaze iste varijable: Tramadol (5.95; 4.914) i PF *score* (0.947; 0.948). Također se i -2LogL modela razlikuju (tablica 2.22 i 2.25). Glavni razlog je što modeli nisu rađeni na istim bazama. Kada se koristila *stepwise* selekcija, u model je na početku ušlo svih 22 nezavisnih varijabli i, ako je nedostajao bilo koji podatak, cijela opservacija se brisala (N = 95). Dok su u model (tablica 2.22) ušle kao nezavisne varijable samo Tramadol i PF *score* (N = 100).

Vrijednost *c-statistike* je 0.845 pa možemo reći da model ima jako dobru prediktivnu snagu.



Slika 2.26: ROC krivulja logističkog modela dobivenog *stepwise* selekcijom (SAS ispis)

2.5 Zaključak

Kako bismo procijenili vjerojatnost da osoba razvije bolnu polineuropatiju koristili smo se trima metodama. Prva metoda koju smo koristili je univarijatna logistička regresija. Od 22 nezavisne varijable, na razini značajnosti $\alpha = 5\%$, statistički značajne su ispale Tramadol, Antidepresivi te sve faktorske varijable. Bilo je i za očekivati da će sve faktorske varijable ispasti statistički značajne jer su one zbirni rezultati 8 dimenzija SF-36 upitnika koji ispituje kvalitetu života. Razlog zašto neke ostale varijable nisu ispale statistički značajne možda leži u tome što nismo koristili originalnu bazu već samo dio nje. Sljedeća metoda koju smo koristili je multivarijatna logistička regresija. U prvi model smo uključili sve one varijable, osim faktorskih, koje su u univarijatnoj logističkoj regresiji ispale statistički značajne. Kao rezultat smo dobili da je samo varijabla Tramadol statistički značajna te je prediktivna snaga modela 70.2%. Zatim smo u drugi model uključili i sve faktorske varijable. Tada smo, uz Tramadol, kao statistički značajnu varijablu dobili još i varijablu PF (fizičko funkcioniranje) te je prediktivna snaga modela 88%. Također smo analizom zaključili da dodavanje faktorskih varijabli značajno pospješuje model. S obzirom da su nam u drugom modelu statistički značajne ispale samo varijable Tramadol i PF, zanimalo nas je hoćemo li dobiti prediktivno jači model ako ga napravimo samo s te dvije varijable. Dobili smo model čija je prediktivna snaga 86.1%. Ako nema značajne razlike u prediktivnoj snazi modela, onda biramo onaj s manje varijabli. Dakle, odabrali bismo model u kojem su nezavisne varijable samo Tramadol i PF. Zadnji model dobili smo *stepwise* selekcijom. Statistički značajne varijable ispale su PF i Tramadol te je prediktivna snaga modela 84.5%. Vrlo je bliska prethodnoj, ali ipak različita. Također se razlikuje i -2LogLikelihood modela. Zašto dolazi do razlika kada oba modela sadrže iste dvije nezavisne varijable? *Stepwise* selekcija uzima u obzir sve nezavisne varijable, odnosno radi se selekcija nad cijelom bazom, dok multivarijatna logistička regresija gleda samo one varijable koje joj zadamo. Kada se gleda cijela baza, tada moramo uzeti u obzir da sigurno ima i nekih *missing*-a, odnosno razlikuje se broj podataka. Zato do razlike u rezultatima dolazi onda kada se ne gledaju iste baze. Uočimo da je u svim modelima varijabla Tramadol ispala statistički značajna. To je sukladno očekivanju jer je Tramadol lijek za ublažavanje bolova te ga zasigurno koristi skupina pacijenata koja pati od bolne polineuropatije.

Poglavlje 3

Dodatak

3.1 SAS kod

```
title"Deskriptiva - nezavisne kontinuirane";
proc means data = DIPL.nezavisne;
var Dob Dijabetes_god HbA1c HDL LDL Trigliceridi;
run;

title"Deskriptiva - nezavisne kategorijske";
proc freq data = DIPL.nezavisne;
table Spol Pušenje Alkohol NSAR Tramadol Antidepresivi BMIcat Polineuropatija_tip;
run;

title"Deskriptiva - nezavisne faktorske";
proc means data = DIPL.faktori;
var PF RP BP GH VT SF RE MH;
run;

title"Deskriptiva - zavisna";
proc freq data = DIPL.nezavisne;
table Polineuropatija;
run;

title" Univarijatna logistička regresija - Dob";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Dob /outroc = outroc;
run;

title" Univarijatna logistička regresija - Dijabetes_god";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Dijabetes_god /outroc = outroc;
```

```
run;
title" Univarijatna logistička regresija - HbA1c";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = HbA1c /outroc = outroc;
run;
title" Univarijatna logistička regresija - HDL";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = HDL /outroc = outroc;
run;
title" Univarijatna logistička regresija - LDL";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = LDL /outroc = outroc;
run;
title" Univarijatna logistička regresija - Trigliceridi";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Trigliceridi /outroc = outroc;
run;
title" Univarijatna logistička regresija - Spol";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Spol /outroc = outroc;
run;
title" Univarijatna logistička regresija - Pušenje";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Pušenje /outroc = outroc;
run;
title" Univarijatna logistička regresija - Alkohol";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Alkohol /outroc = outroc;
run;
title" Univarijatna logistička regresija - NSAR";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = NSAR /outroc = outroc;
run;
title" Univarijatna logistička regresija - Tramadol";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Tramadol /outroc = outroc;
```

```
run;
title" Univarijatna logistička regresija - Antidepresivi";
proc logistic data = DIPL.nezavisne descending;
model Polineuropatija = Antidepresivi /outroc = outroc;
run;

title"Modeliranje BMIcat pomoću dummy varijabli BMIcat 2 baseline";
data DIPL.dummy; set DIPL.nezavisne;
if BMIcat = 1 then do; BMI25 = 1; BMI30 = 0; end;
if BMIcat = 2 then do; BMI25 = 0; BMI30 = 0; end;
if BMIcat = 3 then do; BMI25 = 0; BMI30 = 1; end;
run;

title" Univarijatna logistička regresija - BMIcat";
proc logistic data = DIPL.dummy descending;
model Polineuropatija = BMI25 BMI30 /outroc = outroc;
run;

title"Modeliranje Polineuropatija_tip pomoću dummy varijabli tip 2 baseline";
data DIPL.dummy; set DIPL.dummy;
if Polineuropatija_tip = 1 then do; tip1 = 1; tip3 = 0; end;
if Polineuropatija_tip = 2 then do; tip1 = 0; tip3 = 0; end;
if Polineuropatija_tip = 3 then do; tip1 = 0; tip3 = 1; end;
run;

title" Univarijatna logistička regresija - Polineuropatija_tip";
proc logistic data = DIPL.dummy descending;
model Polineuropatija = tip1 tip3 /outroc = outroc;
run;

title" Univarijatna logistička regresija - PF";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = PF /outroc = outroc;
run;

title" Univarijatna logistička regresija - RP";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = RP /outroc = outroc;
run;

title" Univarijatna logistička regresija - BP";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = BP /outroc = outroc;
```



```
run;
title" Univarijatna logistička regresija - GH";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = GH /outroc = outroc;
run;
title" Univarijatna logistička regresija - VT";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = VT /outroc = outroc;
run;
title" Univarijatna logistička regresija - SF";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = SF /outroc = outroc;
run;
title" Univarijatna logistička regresija - RE";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = RE /outroc = outroc;
run;
title" Univarijatna logistička regresija - MH";
proc logistic data = DIPL.faktori descending;
model Polineuropatija = MH /outroc = outroc;
run;
title"Multivarijatna logistička regresija - Tramadol i Antidepresivi";
proc logistic data = DIPL.baza descending;
model Polineuropatija = Tramadol Antidepresivi /outroc = outroc;
run;
title"Multivarijatna logistička regresija - Tramadol, Antidepresivi i svi faktori";
proc logistic data = DIPL.baza descending;
model Polineuropatija = Tramadol Antidepresivi PF RP BP GH VT SF RE MH /outroc =
outroc;
run;
title"Multivarijatna logistička regresija - Tramadol i PF";
proc logistic data = DIPL.baza descending;
model Polineuropatija = Tramadol PF /outroc = outroc;
run;
title"Logistička regresija - stepwise";
proc logistic data = DIPL.baza descending;
```

```
model Polineuropatija = Dob Dijabetes_god HbA1c HDL LDL Trigliceridi Spol Pušenje  
Alkohol NSAR Tramadol Antidepresivi BMI25 BMI30 tip1 tip3 PF RP BP GH VT SF RE  
MH /selection = stepwise outroc = outroc;  
run;
```

Bibliografija

- [1] *Dijabetička neuropatija*, <https://www.svkatarina.hr/endokrinologija-dijabetes-i-bolesti-metabolizma/dijabeticka-neuropatija>.
- [2] P. D. Allison, *Logistic Regression Using the SAS System: Theory and Application*, SAS Institute Inc, Cary, NC, USA, 1999.
- [3] B. Dawson i R. G. Trapp, *Basic & Cinical Biostatistics (4th Edition)*, Mc Graw-Hill, 2004.
- [4] V. Dermanovic Dobrota, P. Hrabac, D. Skegro, R. Smiljanic, S. Dobrota, I. Prkacin, N. Brkljacic, K. Peros, M. Tomic, V. Lukinovic-Skudar i V. Basic Kes, *The impact of neuropathic pain and other comorbidities on the quality of life in patients with diabetes*, (2014), <https://pubmed.ncbi.nlm.nih.gov/25468384/>.
- [5] M. Hercigonja-Szekeres, *Vrijednost tekstualnih podataka u opservacijskim epidemiološkim istraživanjima*, 2010., <https://core.ac.uk/download/pdf/11699865.pdf>.
- [6] D. W. Hosmer i S. Lemeshow, *Applied Logistic Regression (Second Edition)*, John Wiley and Sons, Inc., New York, USA, 2000.
- [7] M. Huzak, *Vjerojatnost i matematička statistika*, PMF-MO, nastavni materijali, 2006.
- [8] A. Jazbec, *Odabrane statističke metode u biomedicini*, PMF-MO, nastavni materijali, 2021.
- [9] T. Kralj i P. Lazić, *Statistički praktikum 1*, PMF-MO, nastavni materijali, 2020.
- [10] P. Lazić i V. Wagner, *Statistički praktikum 2*, PMF-MO, nastavni materijali, 2021.
- [11] V. Mužić Radović, *Kako doskočiti dijabetičkoj polineuropatiji*, 2011., <https://www.vasezdravlje.com/bolesti-i-stanja/kako-doskociti-dijabetickoj-polineuropatiji>.

Sažetak

U ovom radu najprije smo objasnili što je regresijska analiza i koji je njezin glavni cilj te pokazali koje su glavne razlike između linearne i logističke regresijske analize. Zatim smo definirali logistički model, objasnili metodu maksimalne vjerodostojnosti koju koristimo za procjenu parametara, pokazali na koji način pronalazimo najbolji model te kako interpretiramo parametre modela. Također smo definirali i ROC krivulju te objasnili njezino značenje. Sve to smo zatim primijenili i na praktičnom primjeru. Glavni cilj nam je bio pronaći koje varijable i koliko statistički značajno utječu na to da dijabetička polineuropatija bude bolna u odnosu na bezbolnu. Koristili smo bazu koja sadrži 22 varijable (zdrastveni podaci) o 100 pacijenata. Analizu smo proveli primjenom univarijatne i multivarijatne logističke regresije te *stepwise* selekcijom. Zaključak je da su statistički najznačajnije varijable Tramadol (lijek) i PF (fizičko funkcioniranje).

Summary

In this thesis it was first explained what is regression analysis and what is its main goal, it was also shown what are the main differences between linear and logistic regression analysis. Then the logistic model was defined, maximum likelihood method was explained which is used for parameters estimation, it was shown in which way the best model is found and how model parameters are interpreted. Also, ROC curve was defined and its meaning was explained. All of that was applied on practical example. Main goal was to find which variables and statistically to which extent they influence on diabetic polyneuropathy to be painful or painless. Database with 22 variables (health data) for 100 patient records was used. Analysis was carried out by applying univariate and multivariate logistic regression and the stepwise selection. It was concluded that the most statistically significant variables are Tramadol (medicament) and PH (physical functioning).

Životopis

Rođena sam 8. kolovoza 1996. godine u Dubrovniku. Osnovnu Školu Ante Kovačić te prirodoslovno-matematički smjer Gimnazije Lucijana Vranjanina završila sam u Zagrebu. Nakon završetka srednje škole, 2015. godine upisala sam preddiplomski studij Matematika na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Po završetku preddiplomskog studija, 2020. godine upisala sam diplomski studij Matematička statistika na istom fakultetu. Ljetni semestar akademske godine 2021./2022. provela sam na Erasmus+ studentskoj razmjeni na Faculty of Mathematics and Computer Science u Wrocławu. Kroz osnovnu i srednju školu te do pred kraj preddiplomskog studija aktivno sam se bavila odbojkom. Članica sam odbojkaške sekcije Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu te sam nekoliko godina bila i voditeljica sekcije. Tijekom studiranja obavljala sam različite studentske poslove.