Prozni izveštaj za projekat

### Opis projekta i baze

Ovaj projekat je rađen sa ciljem otkrivanja faktora uspešnosti imunizacije protiv besnila. Podaci su prikupljeni u Pasterovom zavodu – Nacionalnoj referentnoj laboratoriji za besnilo. Podaci se nalaze u fajlu .xlsx, a anonimnost pacijenata je omogućena jer baza ne sadrži nikakve lične informacije. Podaci su prikupljani u periodu od 2017. do 2019. godine. Baza sadrži 17 nezavisnih varijabli relevantnih za istraživanje i 2 zavisne varijable. Nezavisne varijable su: godina, pol, godina rođenja, ustanova pošiljalac uzorka, trajanje vakcinacije, razlika završetka vakcinacije i početka ispitivanja, razlika početka vakcinacije i početka ispitivanja, razlika završetka vakcinacije i bustera, razlika prijema uzorka i uzorkovanja, razlika prijema uzorka i početka ispitivanja, razlika uzorkovanja i početka ispitivanja, ukupno dana od vađenja krvi, davanje seruma, broj jedinica, lokacija ozleda, broj ozleda i životinja koja je načinila ozledu. Zavisne varijable su: uspešnost imunizacije i rezultat analize.

Podaci su uz pomoć read\_excel() funkcije uneti u R i varijable su izabranje. Uz pomoc stringr i dplyr paketa podaci su sređivani kako bi se uklonili pogrešni unosi. Istaknute funkcije u ovom delu projekta jesu [grepl()](https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/grep) i [replace()](https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/replace).

Nakon uklanjanja neodgovarajućih unosa u varijable od interesa kreirana je funkcija pod nazivom remove\_outliers:

remove\_outliers <- function(x, na.rm = TRUE, ...) {

qnt <- quantile(x, probs=c(.25, .75), na.rm = na.rm, ...)

H <- 1.5 \* IQR(x, na.rm = na.rm)

y <- x

y[x < (qnt[1] - H)] <- NA

y[x > (qnt[2] + H)] <- NA

y

}

Funkcija je kreirana za svrhu uklanjanja outliera, jedan argument koji je potrebno uneti i jedan default argument. Argument koji se unosi jeste pojedinačna varijabla kojoj je potrebno ukloniti outliere a oni se definišu kao vrednosti koje se nalaze preko jedne i po vrednosti interkvartilnog raspone te varijable, sa obe strane distribucije. Argument na.rm = TRUE omogućava da se vrednosti računaju iako u varijabli ima nedostajućih podataka (na).

Nakon uklanjanja outliera sledi eksplorativna analiza podataka gde je korištena biblioteka psych, sa osvrtom na [describe()](https://www.rdocumentation.org/packages/psych/versions/2.2.3/topics/describe) funkciju koja je vrlo korisna za dobijanje deskriptivne statistike svake pojedinačne varijable ubacivanjem cele baze kao argumenta u funkciju.

Biblioteka ggplot2 se koristila za vizualizaciju podataka sa najistaknutijim funkcijama geom\_bar() i geom\_histogram() koje su služile za vizualizaciju pojedinačnih varijabli.

Shapiro-Wilk test([shapiro.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test)) se koristio da testira normalnost distribucija varijabli koje su kontinuirane.

Za inferencijalnu statistiku između zavisne i nezavsne varijable korišteni su χ2 test, t test(ili neparametrijska alternativa Wilcoxon, ukoliko uslovi za t test nisu zadovoljeni), ANOVA(ili neparametrijska alternativa Kruskal – Wallis ukoliko uslovi nisu zadovoljeni), Pirsonov produkt moment koeficijent korelacije. Funkcije primenjene za ove testove mogu se videti u Tabeli 1.

|  |  |
| --- | --- |
| Test | Funkcija |
| χ2 test | [chisq.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/chisq.test) |
| t test | [t.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/t.test) |
| Wilcoxon | [wilcox.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test) |
| ANOVA | [aov()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aov) |
| Kruskal – Wallis | [kruskal.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kruskal.test) |
| Pirsonov produkt moment koeficijent korelacije | [cor.test()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test) |

Tabela 1. Prikaz odgovarajućih funkcija za određeni test.

U dva naredna naslova mogu se videti analize rezultata sprovođenih testova sa njihovom interpretacijom kao i analize rizika i modela koji je odabran kao finalni model.

# Analiza rezultata

U narednom delu mogu se videti deskriptivna i inferencijalna statistika korištene za potrebe sprovođenja ovog projekta.

## Deskriptivna statistika:

### Kontinuirane varijable:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ime varijable | Aritmetička sredina |  | Geometrijska sredina | Medijana | Standardna devijacija | Raspon | Interval poverenja (96%) | Broj nedostajućih slučajeva |
| Trajanje vakcinacije | 15.62 |  | 15.33738 | 14 | 3.19 | 19 | 15.344 - 15.896 | 65 |
| Starost | 48.03 |  | 41.24463 | 50 | 21.69 | 88 | 46.263 - 49.797 | 1 |
| Razlika završetka vakcinacije i početka analize | 24.35 |  | 23.40073 | 23 | 7.01 | 40 | 23.739 - 24.961 | 74 |
| Razlika početka vakcinacije i početka analize | 40.95 |  | 39.89707 | 39 | 9.56 | 58 | 40.137 - 41.763 | 51 |
| Razlika prijema uzorka i uzorkovanja | 11.28 |  | 7.860798 | 8 | 9.66 | 39 | 10.483 - 12.077 | 15 |
| Razlika prijema uzorka i početka analize | 3.29 |  | 2.817094 | 4 | 1.87 | 11 | 3.137 - 3.443 | 8 |
| Razlika uzorkovanja i početka analize | 14.66 |  | 11.08176 | 12 | 9.99 | 43 | 13.833 - 15.487 | 19 |
| Analiza rezultata | 5.54 |  | 2.143151 | 2.4 | 10.60 | 94.40 | 4.676 - 6.403 | 1 |
| Broj jedinica | 1370.13 |  | 1293.273 | 1400 | 408.72 | 2034 | 1330.352-1409.907 | 174 |
| Broj ozleda | 1.70 |  | 1.443834 | 1 | 1.16 | 5 | 1.563 - 1.837 | 303 |

### Kategorijske varijable:

#### Pol:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Muški | 263 | 0.4534483 |
| Ženski | 317 | 0.5465517 |

#### Godina:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| 2017 | 112 | 0.1924399 |
| 2018 | 150 | 0.2577320 |
| 2019 | 320 | 0.5498282 |

#### Ustanova pošiljalac uzorka:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Beograd | 133 | 0.228522337 |
| Čačak | 6 | 0.010309278 |
| Ćuprija | 3 | 0.006872852 |
| Kikinda | 18 | 0.030927835 |
| Kosovska Mitrovica | 8 | 0.013745704 |
| Kragujevac | 7 | 0.012027491 |
| Kraljevo | 53 | 0.091065292 |
| Kruševac | 60 | 0.103092784 |
| Leskovac | 11 | 0.018900344 |
| Niš | 5 | 0.008591065 |
| Novi Sad | 48 | 0.082474227 |
| Pančevo | 34 | 0.058419244 |
| Pirot | 5 | 0.008591065 |
| Požarevac | 6 | 0.010309278 |
| Sarajevo | 14 | 0.024054983 |
| Smederevo | 43 | 0.073883162 |
| Smederevska Palanka | 4 | 0.006872852 |
| Sombor | 6 | 0.010309278 |
| Sremska Mitrovica | 15 | 0.025773196 |
| Subotica | 16 | 0.027491409 |
| Šabac | 7 | 0.012027491 |
| Užice | 3 | 0.005154639 |
| Valjevo | 26 | 0.044673540 |
| Vranje | 8 | 0.013745704 |
| Vršac | 16 | 0.027491409 |
| Zaječar | 15 | 0.025773196 |
| Zrenjanin | 11 | 0.018900344 |

#### Uspešno imunizvan:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Da | 491 | 0.8450947 |
| Ne | 90 | 0.1549053 |

#### Dat serum:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Da | 429 | 0.8845361 |
| Ne | 56 | 0.1154639 |

#### Lokacija ozlede:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Ruke | 279 | 0.59615385 |
| Glava i vrat | 17 | 0.03632479 |
| Noge | 161 | 0.34401709 |
| Višestruke ozlede | 6 | 0.01282051 |
| Trup | 5 | 0.01068376 |

#### Animal:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| Divlja mačka | 1 | 0.002066116 |
| Divlja svinja | 1 | 0.002066116 |
| Kuna | 2 | 0.004132231 |
| Lisica | 10 | 0.020661157 |
| Mačka | 269 | 0.555785124 |
| Majmun | 13 | 0.026859504 |
| Miš | 1 | 0.002066116 |
| Pacov | 7 | 0.014462810 |
| Pas | 171 | 0.353305785 |
| Slepi miš | 8 | 0.016528926 |
| Srna | 1 | 0.002066116 |

#### Lot vakcine:

|  |  |  |
| --- | --- | --- |
|  | Frekvencija | Proporcija |
| H1931 | 1 | 0.002083333 |
| L1262-2 | 3 | 0.006250000 |
| M16661V | 96 | 0.200000000 |
| N16972V | 72 | 0.150000000 |
| N1J313V | 105 | 0.218750000 |
| P1C531 | 175 | 0.364583333 |
| P1D271 | 28 | 0.058333333 |
|  |  |  |

## Inferencijalna statistika:

### Analiza pojedinačnih nezavisnih varijabli sa varijablom uspešno imunizovan

#### Varijabla Pol:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Pol | |
|  |  | Muški | Ženski |
| Uspešno imunizovan | Da | 221 | 266 |
| Ne | 40 | 50 |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(1) = 0.027, p = 0.8699

Zaključak: Ne postoji značajna razlika na varijabli pol. Ne postoji povezanost između pola i uspešnosti imunizacije.

Varijabla Starost:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Starost.

Vrednosti rezultata iznose: W = 0.970, p < 0.01. Ovi rezultati ukazuju na to da varijabla Starost nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 14352, p < 0.01.

Zaključak: Postoji statistički značajna razlika, osobe koje nisu uspešno imunizovane su starije.



#### Varijabla Ustanova pošiljalac uzorka:

|  |  |  |
| --- | --- | --- |
|  | Uspešno imunizovan | |
|  | Da | Ne |
| Beograd | 122 | 11 |
| Čačak | 5 | 1 |
| Ćuprija | 4 | 0 |
| Kikinda | 17 | 1 |
| Kosovska Mitrovica | 8 | 0 |
| Kragujevac | 6 | 1 |
| Kraljevo | 45 | 8 |
| Kruševac | 49 | 11 |
| Leskovac | 9 | 2 |
| Niš | 5 | 0 |
| Novi Sad | 38 | 9 |
| Pančevo | 33 | 1 |
| Pirot | 4 | 1 |
| Požarevac | 6 | 1 |
| Sarajevo | 12 | 2 |
| Smederevo | 29 | 12 |
| Smederevska Palanka | 2 | 2 |
| Sombor | 6 | 0 |
| Sremska Mitrovica | 14 | 1 |
| Subotica | 10 | 6 |
| Šabac | 5 | 2 |
| Užice | 3 | 0 |
| Valjevo | 18 | 8 |
| Vranje | 6 | 2 |
| Vršac | 12 | 4 |
| Zaječar | 12 | 3 |
| Zrenjanin | 9 | 2 |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(26) = 41.416, p = 0.02817.

Kako postoji statistički značajna razlika post hoc test će biti primenjem kako bi se dalje istražile razlike između Ustanova pošiljalaca uzoraka.

χ2 post hoc test sa Bonferroni korekcijom je primenjen. Rezultati su dati u sledećoj tabeli:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Uspešno imunizovan-Da | |  |
|  | Reziduali | p vrednost | korigovano p |
| Beograd | 2.637764 | 0.008345\* | 0.450655 |
| Čačak | -0.07629134 | 0.93918700 | 1 |
| Ćuprija | 0.8609981 | 0.3892390 | 1 |
| Kikinda | 1.188219 | 0.234747 | 1 |
| Kosovska Mitrovica | 1.221893 | 0.221748 | 1 |
| Kragujevac | 0.09244571 | 0.92634400 | 1 |
| Kraljevo | 0.09480077 | 0.92447300 | 1 |
| Kruševac | -0.6298369 | 0.5288010 | 1 |
| Leskovac | -0.2437824 | 0.8073990 | 1 |
| Niš | 0.9634633 | 0.3353150 | 1 |
| Novi Sad | -0.7115868 | 0.4767210 | 1 |
| Pančevo | 2.090512 | 0.036572 | 1 |
| Pirot | -0.2761928 | 0.7824000 | 1 |
| Požarevac | 1.056342 | 0.290812 | 1 |
| Sarajevo | 0.1315454 | 0.8953440 | 1 |
| Smederevo | -2.516122 | 0.011865\* | 0.640733 |
| Smederevska Palanka | -1.908546 | 0.056321 | 1 |
| Sombor | 1.056342 | 0.290812 | 1 |
| Sremska Mitrovica | 0.9614635 | 0.3363190 | 1 |
| Subotica | -2.458105 | 0.013967\* | 0.754231 |
| Šabac | -0.9570849 | 0.3385240 | 1 |
| Užice | 0.7449987 | 0.4562730 | 1 |
| Valjevo | -2.192446 | 0.028347\* | 1 |
| Vranje | -0.743318 | 0.457289 | 1 |
| Vršac | -1.058653 | 0.289758 | 1 |
| Zaječar | -0.4826023 | 0.6293780 | 1 |
| Zrenjanin | -0.2437824 | 0.8073990 | 1 |

Zaključak: Sveukupno, podaci sugerišu da postoji statistički značajna razlika kod uspešno i neuspešno imunizovanih pacijenata prema Ustanovi pošiljaocu uzorka, ali poređenje reizduala u post hoc testu ne može sa sigurnošću da pokaže između kojih grupa postoji značajnost. Iako Bonferonijeve prilagođene p vrednosti ne pokazuju značaj, 4 neprilagođene p vrednosti su bile značajne. Beograd sa pozitivnim rezidualima pokazuje da se javlja više uspešno imunizovanih pacijenata nego uspešnih, a Smederevo, Subotica i Valjevo sa negativnim rezidualima u grupi uspešno imunizovanih pacijenata ukazuju da se više neuspešno imunizovanih pacijenata javlja iz tih ustanova.

Odbijanje nulte hipoteze u početnom χ2 testu može biti posledica složenih kontrasta u proporcijama populacije.

#### Varijabla Trajanje vakcinacije:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Trajanje vakcinacije.

Vrednosti rezultata iznose: W = 0.058109, p < 0.01. Ovi rezultati ukazuju na to da varijabla Trajanje vakcinacije nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 18979, p = 0.1499.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Trajanje vakcinacije.



#### Varijabla Razlika prijema uzorka i uzorkovanja:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Razlika prijema uzorka i uzorkovanja.

Vrednosti rezultata iznose: W = 0.90527, p < 0.01. Ovi rezultati ukazuju na to da varijabla Razlika prijema uzorka i uzorkovanja nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 19648, p = 0.459.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Razlika prijema uzorka i uzorkovanja.



#### Varijabla Razlika prijema uzorka i početka analize:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Razlika prijema uzorka i početka analize

Vrednosti rezultata iznose: W = 0.91215, p < 0.01. Ovi rezultati ukazuju na to da varijabla Razlika prijema uzorka i početka analize nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 18888, p = 0.0749.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Razlika prijema uzorka i početka analize. P nivo na ovom testu je graničan.



#### Varijabla Razlika uzorkovanja i početka analize:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Razlika uzorkovanja i početka analize.

Vrednosti rezultata iznose: W = 0.92733, p < 0.01. Ovi rezultati ukazuju na to da varijabla Razlika uzorkovanja i početka analize nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 19004, p = 0.2775.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Razlika uzorkovanja i početka analize.



#### Varijabla Razlika završetka vakcinacije i početka analize:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Razlika završetka vakcinacije i početka analize.

Vrednosti rezultata iznose: W = 0. 94096, p < 0.01. Ovi rezultati ukazuju na to da varijabla Razlika završetka vakcinacije i početka analize nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 13614, p = 0. 1318.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Razlika završetka vakcinacije i početka analize.



#### Varijabla Razlika početka vakcinacije i početka analize:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije varijable Razlika početka vakcinacije i početka analize.

Vrednosti rezultata iznose: W = 0. 93998, p < 0.01. Ovi rezultati ukazuju na to da varijabla Razlika početka vakcinacije i početka analize nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 18403, p = 0. 8891.

Zaključak: Ne postoji statistički značajna razlika za grupe uspešno i neuspešno imunizovanih na varijabli Razlika početka vakcinacije i početka analize.



#### Varijabla Dat serum:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Dat serum | |
|  |  | Da | Ne |
| Uspešno imunizovan | Da | 359 | 50 |
| Ne | 70 | 6 |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(1) = 1. 1767, p = 0. 278.

Zaključak: Ne postoji značajna razlika na varijabli Dat serum.



#### Varijabla Broj jedinica:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije Broj jedinica.

Vrednosti rezultata iznose: W = 0. 98207, p < 0.01. Ovi rezultati ukazuju na to da varijabla Broj jedinica nije normalno distribuirana.

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 9812.5, p = 0. 03491.

Zaključak: Između uspešno imunizovane i neuspešno imunizovane grupe postoji statistički značajna razlika na nivou od 0.05, neuspešno imunizovana grupa ima viši Broj jedinica.



#### Varijabla Lokacija ozlede:

|  |  |  |
| --- | --- | --- |
|  | Uspešno imunizovan |  |
|  | Da | Ne |
| Ruke | 241 | 40 |
| Glava i vrat | 1 | 3 |
| Noge | 134 | 27 |
| Višestruke ozlede | 5 | 1 |
| Trup | 4 | 1 |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(4) = 0.67984, p = 0.9538.

Zaključak: Ne postoji značajna razlika na varijabli Lokacija ozleda.

#### Varijabla Životinja:

|  |  |  |
| --- | --- | --- |
|  | Uspešno imunizovan | |
|  | Da | Ne |
| Divlja mačka | 1 | 0 |
| Divlja svinja | 1 | 0 |
| Kuna | 2 | 0 |
| Lisica | 8 | 2 |
| Mačka | 222 | 58 |
| Majmun | 11 | 2 |
| Miš | 0 | 1 |
| Pacov | 5 | 2 |
| Pas | 150 | 21 |
| Slepi miš | 8 | 0 |
| Srna | 2 | 0 |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(10) = 11.42, p = 0.3257.

Zaključak: Ne postoji značajna razlika na varijabli Životinja.

#### Varijabla Lot vaccine:

|  |  |  |
| --- | --- | --- |
|  | Uspešno imunizovan | |
|  | Da | Ne |
| H1931 | 1 | 0 |
| L1262-2 | 5 | 1 |
| M16661V | 99 | 15 |
| N16972V | 74 | 6 |
| N1J313V | 96 | 17 |
| P1C531 | 146 | 31 |
| P1D271 | 21 | 8 |
|  |  |  |

Koristeći neparametrijsku analizu χ2 test dobijaju se sledeći rezultati: χ2(6) = 9.2988, p = 0.1575.

Zaključak: Ne postoji značajna razlika na varijabli Lot vakcine.

#### Varijabla Potencija vakcine:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije Potencije vakcine. Rezultati su sledeći: W = 0.76129, p < 0.01. Rezultati sugerišu da distribucija varijable Potencija vakcine značajno odstupa od normalne distribucije.

Wilcoxon signed-rank test je primenjen kako bi se testirala razlika između grupe uspešno i grupe neuspešno imunizovanih pacijenata. Rezultati su sledeći: W = 11965, p = 0.9068.

Zaključak: Rezultati sugerišu da između grupe uspešno i neuspešno imunizovanih pacijenata ne postoji razlika na varijabli Potencija vakcine.



### Analize pojedinačnih nezavisnih varijabli sa zavisnom varijablom Analiza rezultata:

Shapiro-Wilks test koristio se za proveru normalnosti distribucije Rezultat analize.

Vrednosti rezultata iznose: W = 0.47664, p < 0.01. Ovi rezultati ukazuju na to da varijabla Rezultat analize nije normalno distribuirana. Kako je pretpostavka o normalnosti distribucije narušena primenjivaće se neparametrijski testovi kada je varijabla Rezultat analize zavisna.

#### Varijabla Godina:

Za inferencijalnu statistiku koristi se Kruskal-Wallis test. Rezultati su sledeći: χ2(2) 2.8367, p = 0.2421.

Ne postoji statistički značajna razlika u Rezultatu analize na varijabli Godina.



#### Varijabla Pol:

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 42650, p = 0.5753.

Zaključak: Ne postoji statistički značajna razlika između pola na varijabli Rezultat analize.

Varijabla starost:

Koeficijent korelacije između varijable Starost i Rezultat analize iznosi r(578) = -0.177, p < 0.01.

Zaključak: Između varijabli Starost i Rezultat analize postoji statistički značajna slaba negativna korelacija. Porast Starost prati pad vrednosti na varijabli Rezultat analize.



#### Varijabla Ustanova pošiljalac uzorka:

Za inferencijalnu statistiku koristi se Kruskal-Wallis test. Rezultati su sledeći: : χ2(26) = 36.659, p = 0.0802.

Zaključak: Ne postoji statistički značajna razlika na varijabli Rezultat analize u odnosu na različite Ustanove.



#### Varijabla Trajanje vakcinacije:

Koeficijent korelacije između varijable Trajanje vakcinacije i Rezultat analize iznosi r(515) = 0.065, p = 0.1394.

Zaključak: Između varijable Trajanje vakcinacije i Rezultata analize ne postoji linearna zavisnost.



#### Varijabla Razlika prijema uzorka i uzorkovanja:

Koeficijent korelacije između varijable Razlika prijema uzorka i uzorkovanja i Rezultat analize iznosi r(565) = -0.029, p = 0.4932.

Zaključak: Između varijabli Razlika prijema uzorka i uzorkovanja i Rezultat analize ne postoji korelacija.



#### Varijabla Razlika prijema uzorka i početka analize:

Koeficijent korelacije između varijable Razlika prijema uzorka i početka analize i Rezultat analize iznosi r(572) = -0.007, p = 0.8549.

Zaključak: Između varijabli Razlika prijema uzorka i početka analize i Rezultat analize ne postoji korelacija.



#### Varijabla Razlika završetka vakcinacije i početka analize:

Koeficijent korelacije između varijable Razlika završetka vakcinacije i početka analize i Rezultat analize iznosi r(506) = -0.039, p = 0.369.

Zaključak: Između varijabli Razlika završetka vakcinacije i početka analize i Rezultat analize ne postoji korelacija.



#### Varijabla Razlika početka vakcinacije i početka analize:

Koeficijent korelacije između varijable Razlika završetka vakcinacije i početka analize i Rezultat analize iznosi r(529) = -0.016, p = 0.6978.

Zaključak: Između varijabli Razlika završetka vakcinacije i početka analize i Rezultat analize ne postoji korelacija.



#### Varijabla Broj jedinica:

Koeficijent korelacije između varijable Ukupno dana od vađenja krvi i Rezultat analize iznosi r(406) = -0.029, p = 0.5563.

Zaključak: Između varijabli Broj jedinica i Rezultat analize ne postoji korelacija.

#### Varijabla Dat serum:

Za inferencijalnu statistiku koristi se Wilcoxon test sume rangova. Rezultati su sledeći: W = 6650.5, p = < 0.01.

Zaključak: Postoji statistički značajna razlika između između grupe koja je primila i koja nije primila serum, značajno više rezultate na varijabli Rezultat analize ostvaruje grupa koja nije primila serum.



#### Varijabla Lokacija ozlede:

Za inferencijalnu statistiku koristi se Kruskal-Wallis test. Rezultati su sledeći: χ2(4) = 3.3816, p = 0.4961.

Zaključak: Ne postoji statistički značajna razlika na varijabli Lokacija ozleda i Rezultat analize.



#### Varijabla Broj ozleda:

Koeficijent korelacije između varijable Broj ozleda i Rezultat analize iznosi r(277) = -0.057, p = 0. 3402.

Zaključak: Između varijabli Broj ozleda i Rezultat analize ne postoji korelacija.



#### Varijabla Životinja:

Za inferencijalnu statistiku koristi se Kruskal-Wallis test. Rezultati su sledeći: χ2(10) = 13.889, p = 0.1781.

Zaključak: Ne postoji statistički značajna razlika na varijabli Životinja i Rezultat analize.



#### Varijabla Lot vakcine:

Za inferencijalnu statistiku koristi se Kruskal-Wallis test. Rezultati su sledeći: χ2(6) = 7.4475, p = 0.281.

Zaključak: Ne postoji statistički značajna razlika na varijabli Lot vakcine i Rezultat analize



#### Varijabla Potencija vakcine:

Koeficijent korelacije između varijable Broj ozleda i Rezultat analize iznosi r(407) = -0.04, p = 0.4192.

Zaključak: Između varijabli Broj ozleda i Rezultat analize ne postoji korelacija.



# Analiza rizika za varijablu Uspešno imunizovan

Za sprovođenje analize rizika korištena je binarna logistička regresija sa varijablom Uspešno imunizovan kao zavisnom.

U prvom modelu kao prediktori uzete su sve varijable koje su na prethodnim testovima inferencijalne statistike imale značajne rezultate sa varijablom Uspešnost imunizacije i/ili Rezultat analize, a to su sledeće: Starost, Ustanova pošiljalac uzorka i Dat serum. U prvom modelu kao značajni su se pokazali prediktori Starost i Ustanova pošiljalac uzorka.

Drugi model se sastojao od Uspešnosti imunizacije kao zavisne varijable i dva značajna prediktora iz prethodnog modela: Starost i Ustanova pošiljalac uzorka.

Treći model sastojao se od Uspešnosti imunizacije kao zavisne varijable i Starosti kao prediktora.

Četvrti model sastojao se od Uspešnosti imunizacije kao zavisne varijable i Ustanova pošiljalac uzorka kao prediktora.

Svi modeli su trenirani na 80% nasumično izabranih pacijenata iz baze, a metrika modela je proveravana na preostalih 20%.

Za potrebe sprovođenja logističke regresije korištena je funkcija [glm()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm), sa argumentom family = binomial(link = “logit”) kako bi se na generalizovanom linearnom modelu sprovodila logistička regresija. Biblioteka cater koristila se zbog funkcije [createDataPartition()](https://www.rdocumentation.org/packages/caret/versions/6.0-91/topics/createDataPartition) kako bi se postojeća baza podataka razdvojila na set za treniranje i set za testiranje modela.

Varijablu uspešnosti imunizacije bilo je potrebno iz karakter varijable promeniti u numeričku I to se vršilo pomoću funkcije [replace()](https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/replace).

Funkcija compute.eval.metrics() kreirana je kako bi se računala metrika modela.

compute.eval.metrics <- function(cm) {

TP <- cm[2,2]

TN <- cm[1,1]

FP <- cm[1,2]

FN <- cm[2,1]

acc <- (TP + TN) / (TP + TN + FP + FN)

prec <- TP / (TP + FP)

recall <- TP / (TP + FN)

f1 <- 2 \* (prec \* recall) / (prec + recall)

c(accuracy = acc, precision = prec, recall = recall, f1 = f1)

}

Matrice konfuzije za drugi treći i četvrti model mogu se videti u narednim tabelama:

Drugi model:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predviđene vrednosti | |
|  |  | Ne | Da |
| Prave vrednosti | Ne | 8 | 19 |
| Da | 15 | 112 |

Treći model:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predviđene vrednosti | |
|  |  | Ne | Da |
| Prave vrednosti | Ne | 0 | 27 |
| Da | 0 | 127 |

Četrvti model:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predviđene vrednosti | |
|  |  | Ne | Da |
| Prave vrednosti | Ne | 0 | 27 |
| Da | 0 | 127 |

Kako treći i četvrti model uopšte ne predviđaju neuspešnu imunizovanost kod pacijenata, za finalni model odabran je drugi model koji sadrži Ustanovu pošiljalac uzorka i Starost kao prediktore zavisne varijable Uspešno imunizovan.

Metrika modela može se videti u slećoj tabeli:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Tačnost | Preciznost | Odziv | F1 mera |
| Model 2 | 0.7792208 | 0.8549618 | 0.8818898 | 0.8682171 |

Vrednosti Model likelihood ratio testa iznosi χ2(1) = 30.71, p < 0.01, iz čega se može zaključiti da je Model 2 značajno drugačiji od modela koji ima samo intercept, odnosno, ne sadrži ni jedan prediktor. McFadden-ov pseudo R² vrednost za ovaj model iznosi 0.140841.

Confusion matrix for the model:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predicted values | |
|  |  | No | Yes |
| Real values | No | 8 | 19 |
| Yes | 15 | 112 |

Koeficijenti modela:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | β | Količnik šansi  (odds ratio) | Standardna  greška | p nivo |
| Intercept | -3.410827 |  | 0.369949 | < 0.01 |
| Age | 0.031576 | 0.969 | 0.006102 | < 0.01 |
| Ustanova pošiljalac uzorka |  |  |  |  |
| Čačak | - 1.1960 | 0.302 | 1.2223 | 0.3278 |
| Ćuprija | 6.3888 | 1101683.294 | 91.9609 | 0.9446 |
| Kikinda | 0.3200 | 1.377 | 1.1219 | 0.7755 |
| Kosovska Mitrovica | 6.2673 | 1026394.156 | 30.5924 | 0.8377 |
| Kragujevac | -0.7614 | 0.467 | 1.1807 | 0.5190 |
| Kraljevo | -0.7991 | 0.450 | 0.6213 | 0.1983 |
| Kruševac | -0.9446 | 0.389 | 0.6202 | 0.1277 |
| Leskovac | -1.8249 | 0.161 | 0.9189 | 0.0470 |
| Niš | 5.8214 | 624703.433 | 91.9610 | 0.9495 |
| Novi Sad | -1.5844 | 0.205 | 0.5955 | 0.0078 |
| Pančevo | 0.4913 | 1.635 | 1.1102 | 0.6581 |
| Pirot | -1.7857 | 0.168 | 1.3264 | 0.1782 |
| Požarevac | 0.3904 | 1.478 | 1.1367 | 0.7312 |
| Sarajevo | -0.5316 | 0.588 | 1.1541 | 0.6451 |
| Smederevo | -1.8626 | 0.155 | 0.5774 | 0.0013 |
| Smederevska Palanka | -2.988 | 0.050 | 1.3320 | 0.0248 |
| Sombor | -0.8551 | 0.425 | 1.1726 | 0.4659 |
| Sremska Mitrovica | 0.0360 | 1.037 | 1.1212 | 0.9744 |
| Subotica | -2.4255 | 0.088 | 0.7786 | 0.0018 |
| Šabac | -2.0983 | 0.123 | 1.2710 | 0.0988 |
| Užice | 6.2464 | 989507.713 | 62.4024 | 0.9203 |
| Valjevo | -1.9296 | 0.145 | 0.6595 | 0.0034 |
| Vranje | -1.9475 | 0.143 | 1.3105 | 0.1373 |
| Vršac | -1.7978 | 0.166 | 0.7444 | 0.0157 |
| Zaječar | -0.3246 | 0.723 | 1.1481 | 0.7774 |
| Zrenjanin | -0.7189 | 0.487 | 1.1642 | 0.5369 |

Zaključak: Logistička regresija je sprovedena kako bise analizirali faktori rizika neuspešne imunizacije. Starost pacijenta predstavlja jedan od dva faktora rizika pronađenih u ovoj studiji. Pronađeno je da je šansa da imunizacija bude uspešna opada za 3.1% za svaku godinu starosti pacijenta, kada se intercept i Ustanova pošiljalac uzorka drže konstantnim. Takođe je pronađeno da su ustanove iz narednih mesta: Leskovac, Novi Sad, Smederevo, Smederevska Palanka, Subotica, Valjevo i Vršac značajni u modelu. Pacijenti iz ovih ustanova imaju veću verovatnoću za uspešnu imunizaciju.