

Logistička Regresija: Od Podataka do Predviđanja

Vodič za razumevanje, primenu i interpretaciju jednog od najmoćnijih alata u analitici.



Kritično Pitanje: Kako Predvideti Rizik?

U kardiohirurgiji, procena operativnog rizika je ključna. Pre svake intervencije, timovi moraju da odgovore na pitanje:

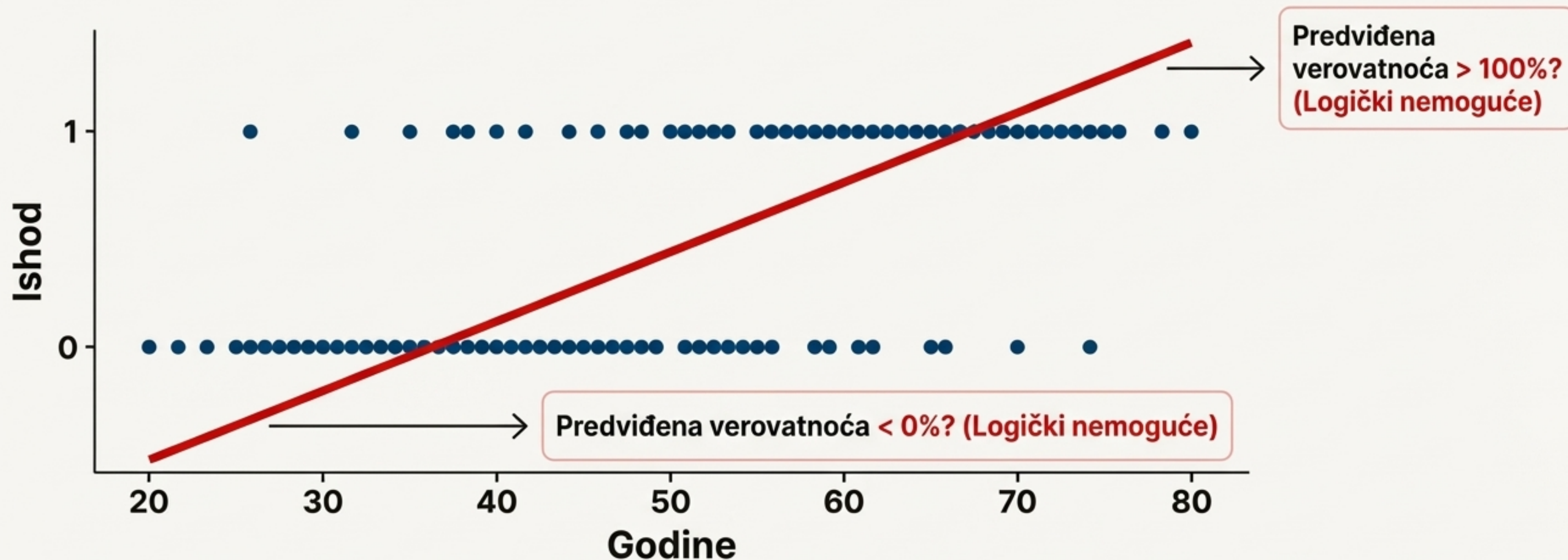
“Na osnovu faktora kao što su starost, prethodne operacije, funkcija leve komore i druge bolesti, kolika je **verovatnoća** da će pacijent preživeti 30 dana nakon operacije?”

Precizno predviđanje ishoda (smrt ili preživljavanje) omogućava bolju pripremu, informisanje pacijenata i objektivnu evaluaciju rezultata. Ovo je binarni problem: ishod je 0 (smrt) ili 1 (preživljavanje).



Zašto Jednostavna Linearna Regresija Nije Dovoljna?

Prvi instinkt bi bio da koristimo linearnu regresiju za predviđanje. Međutim, kada je ishod binaran (0 ili 1), ovaj pristup dovodi do apsurdnih rezultata.



Potreban nam je model koji prirodno ograničava predviđanja na interval između 0 i 1.

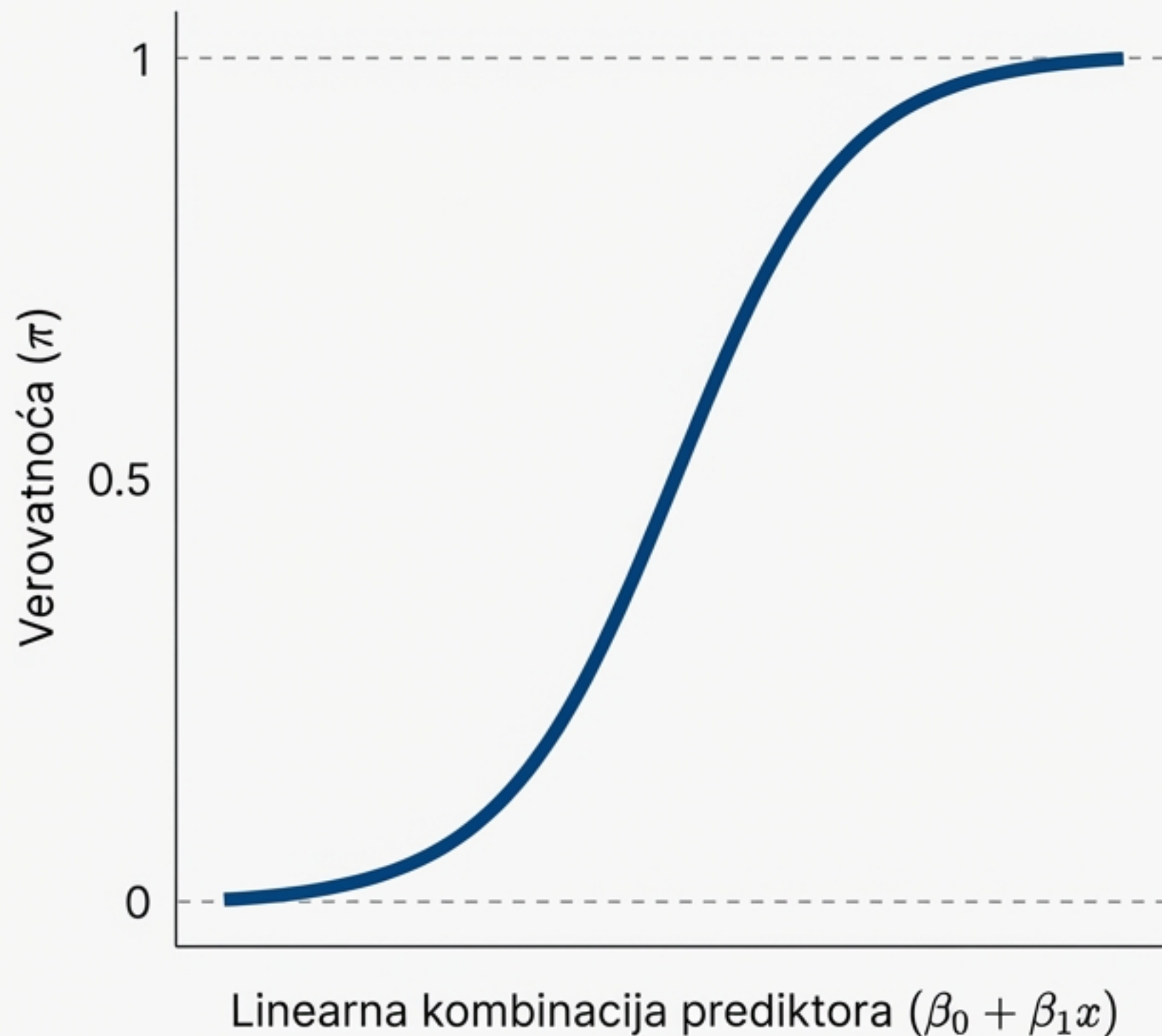
Rešenje: Logistička Funkcija (Sigmoidna Kriva)

Logistička regresija koristi S-krivu (logističku funkciju) koja preslikava bilo koju realnu vrednost u verovatnoću između 0 i 1. Bez obzira na vrednost ulaznih prediktora, izlaz je uvek smisljena verovatnoća.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

ili jednostavnije:

$$\pi(x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))}$$



Veza sa Linearnim Modelima: Logit Transformacija

Kako modelujemo nelinearnu S-krivu koristeći linearnu jednačinu? Kroz **Logit** transformaciju. Logit transformiše verovatnoću (koja je ograničena na $[0, 1]$) u vrednost koja može ići od $-\infty$ do $+\infty$.



Korak 1: Šansa (Odds)

$$\text{Šansa} = \pi(x) / (1 - \pi(x))$$

(Verovatnoća da se događaj desi podeljena sa verovatnoćom da se ne desi)

Korak 2: Logit (Log-Odds)

$$\text{logit}[\pi(x)] = \ln(\text{Šansa}) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right)$$

Ključni Uvid

Sada možemo modelovati logit kao linearnu funkciju prediktora:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 x$$

Logistička regresija je **Generalizovani Linearni Model (GLM)** koji koristi Logit kao 'funkciju poveznice' (link function) da poveže binarni ishod sa linearnim prediktorima.

Kamen Temeljac Interpretacije: Odnos Šansi (Odds Ratio)

Dok koeficijent β u linearnoj regresiji predstavlja promenu u ishodu, u logističkoj regresiji on predstavlja promenu u **log-šansama**. Da bismo ovo preveli u intuitivnu meru, koristimo **Odnos Šansi (OR)**.

Odnos šansi je $\exp(\beta)$. On nam govori koliko puta se menja **šansa** za ishod ($Y=1$) za svako povećanje prediktora `x` za jednu jedinicu.

OR > 1



Povećanje `x` za jednu jedinicu povećava šansu da se ishod desi.

OR < 1



Povećanje `x` za jednu jedinicu smanjuje šansu da se ishod desi.

OR = 1



Prediktor `x` nema uticaj na šansu ishoda.

Naglasak: `exp( $\beta$ )` je multiplikativni efekat na **šansu**, ne na verovatnoću.

Primer iz Prakse: Uticaj Pušenja na Koronarnu Srčanu Bolest (KSB)

Scenario: Analiziramo podatke da bismo utvrdili da li je pušenje (kodirano kao 1=pušač, 0=nepušač) faktor rizika za KSB.

Rezultati iz SPSS-a (Jednostavna Logistička Regresija)

Inter Medium	B	S.E.	Wald	df	Sig.	Exp(B)
PUSENJE	1.986	.591	11.274	1	.001	7.286
Constant	-1.099	.385	8.147	1	.004	.333

Korak-po-Korak Interpretacija:

Korak 1: Identifikuj Koeficijent

Regresioni koeficijent za pušenje je $B = 1.986$.

Korak 2: Izračunaj Odnos Šansi (OR)

$OR = \exp(B) = \exp(1.986) = 7.286$

Korak 3: Jasna Interpretacija

> Pušači imaju **7.28 puta veću šansu** da obole od KSB u odnosu na nepušače.

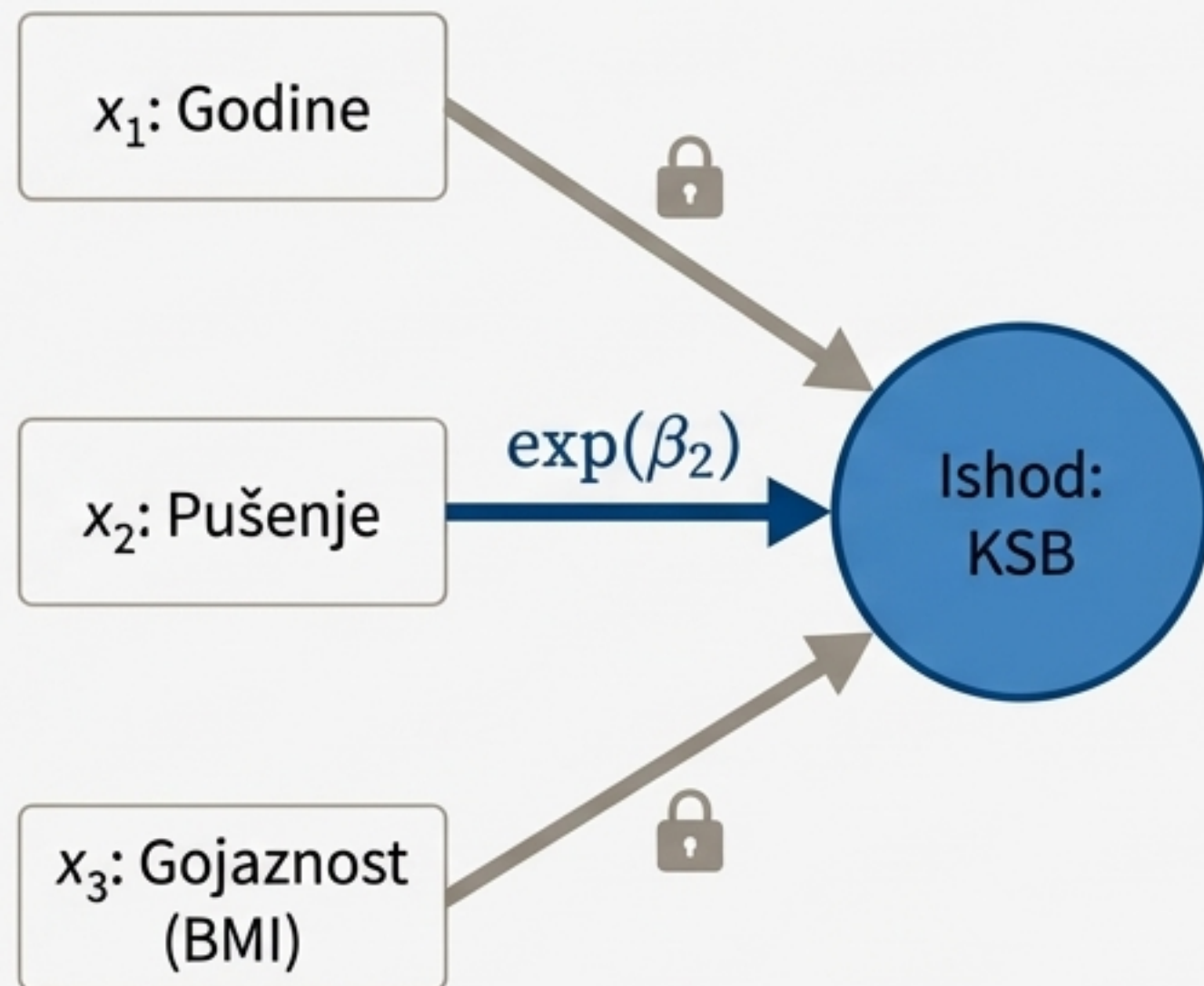
Od Jednog ka Više Faktora: Višestruka Logistička Regresija

U stvarnosti, ishodi retko zavise od samo jednog faktora. Višestruka logistička regresija nam omogućava da modelujemo uticaj više prediktora istovremeno.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

$\exp(\beta_i)$ sada predstavlja odnos šansi za promenu x_i za jednu jedinicu, **dok se vrednosti svih ostalih prediktora ($x_1, \dots, x_{i-1}, x_{i+1}, \dots$) drže konstantnim.**

Ovo nam omogućava da izmerimo jedinstveni uticaj pušenja na KSB, čak i kada uzmemo u obzir uticaj starosti i gojaznosti.



Čitanje Rezultata: Model KSB sa Više Prediktora

Ispod je prikazan izlaz iz SPSS-a za model koji predviđa KSB na osnovu starosti, pušenja i gojaznosti (BMI).

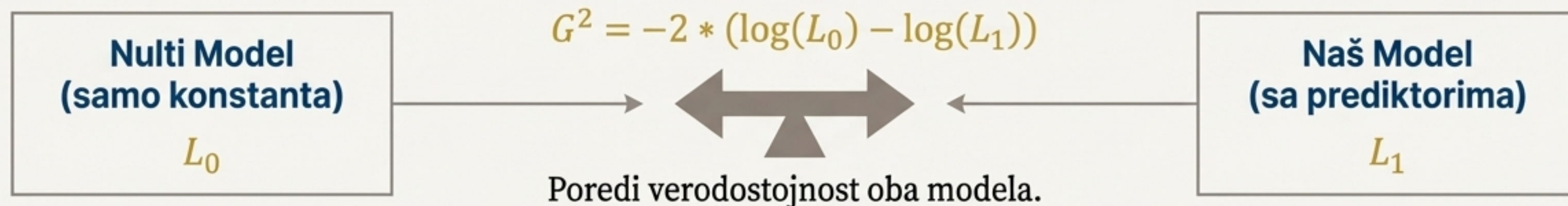
	B	S.E.	Wald	df	Sig.	Exp(B)
GODINE	.106	.031	11.467	1	.001	1.112
PUSENJE	2.615	.819	10.197	1	.001	13.666
BMI	1.271	.749	2.883	1	.090	3.565
Constant	-7.587	2.055	13.631	1	.000	.001

- **GODINE:** Za svaku dodatnu godinu starosti, šansa za KSB se **povećava** za **11.2%** (**OR=1.112**), držeći pušenje i gojaznost konstantnim. Efekat je **statistički značajan** ($p < .05$).
- **PUSENJE:** Pušači imaju **13.7 puta** veću šansu za KSB od nepušača, držeći starost i gojaznost konstantnim (**OR=13.666**). Efekat je **statistički značajan** ($p < .05$).
- **BMI (Gojaznost):** Gojazne osobe imaju **3.6 puta** veću šansu za KSB, ali ovaj efekat **nije statistički značajan** na nivou od 5% ($p = .090$).

Da li je Naš Model Uopšte Koristan? Test Odnosa Verodostojnosti

Pitanje: Kako znamo da li naš model sa prediktorima (starost, pušenje, BMI) predviđa ishod bolje od prostog pogađanja na osnovu ukupne proporcije?

Rešenje: **Test Odnosa Verodostojnosti** (Likelihood Ratio Test ili Deviance Test).



Analogija: Ovo je ekvivalent **F-testu** u linearnoj regresiji. Testira da li je model u celini statistički značajan.

Princip i Statistika:

- **Princip:** Upoređuje 'verodostojnost' (likelihood) našeg modela (L_1) sa verodostojnošću 'nultog' modela koji sadrži samo konstantu (L_0).
- **Test Statistika:** $G^2 = -2 * (\log(L_0) - \log(L_1))$. Ova statistika prati χ^2 (Hi-kvadrat) raspodelu.

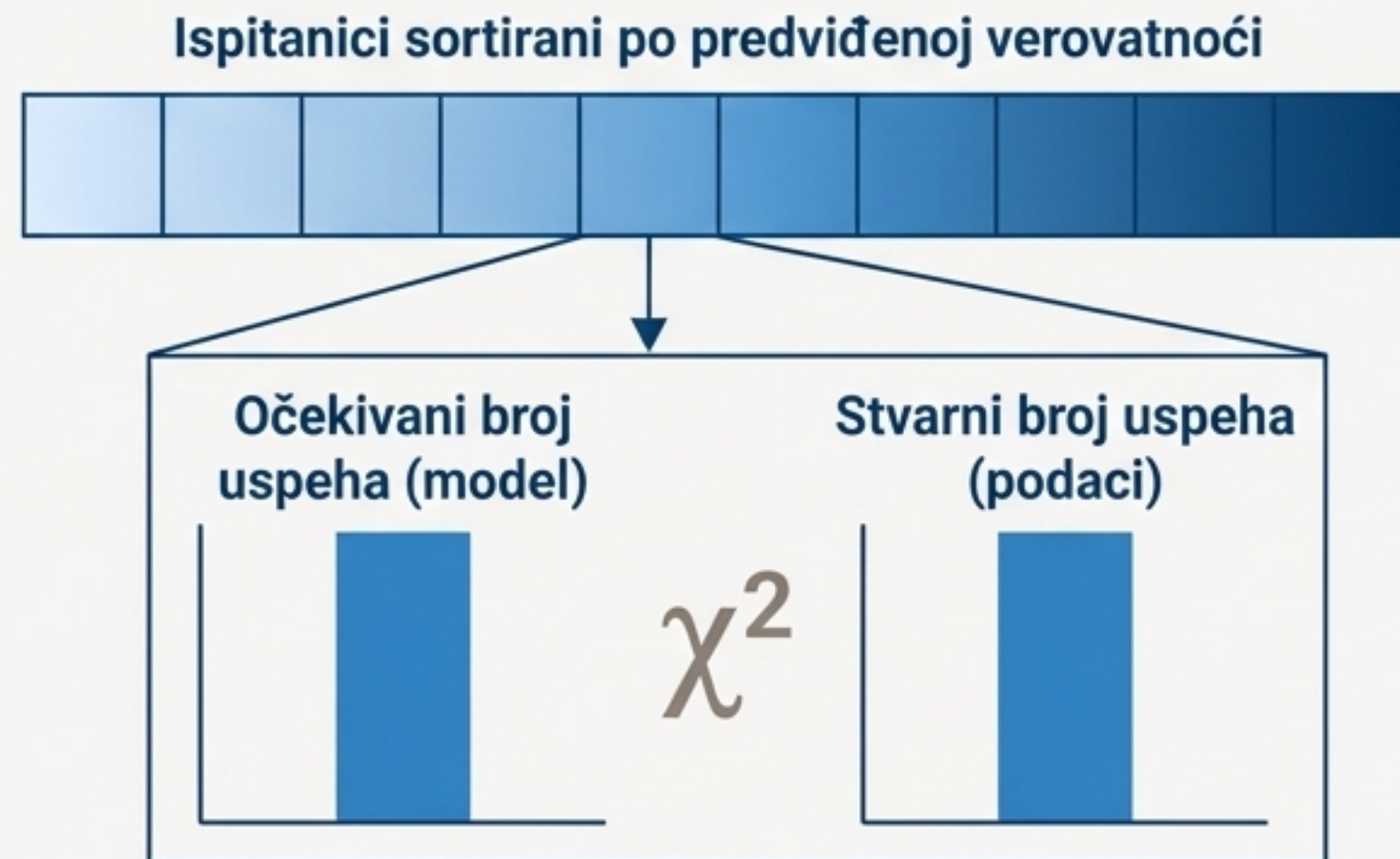
Interpretacija: Mala p-vrednost (< 0.05) znači da naš model sa prediktorima **značajno bolje** objašnjava ishod od modela bez njih.

Koliko se Dobro Model Uklapa u Podatke?

Pitanje: Znamo da je model značajan, ali da li se dobro kalibriše? Da li su predviđene verovatnoće bliske stvarnim proporcijama u podacima?

Alat za Procenu: Hosmer-Lemeshow test.

- ✓ **Visoka p-vrednost (> 0.05):** Dobar znak! Znači da nema statistički značajne razlike između predviđenih i stvarnih vrednosti. Model se dobro uklapa.
- ✗ **Niska p-vrednost (< 0.05):** Loš znak. Ukazuje na loše uklapanje modela.



Dodatna Provera: Hosmer-Lemeshow test. Dijagnostika reziduala se takođe koristi za identifikaciju uticajnih tačaka (outliers) koje mogu narušiti model.

Ključni Uvidi: Logistička Regresija u 4 Tačke

0 → 1

1. Svrha

Koristi se za predviđanje **binarnog ishoda** (npr. da/ne, uspeh/neuspeh, prisutan/odsutan) na osnovu jednog ili više prediktora.



2. Mehanizam

Modelira **verovatnoću** ishoda koristeći nelinearnu **logističku (S) funkciju**. Veza sa linearnim modelom se uspostavlja preko **Logit** transformacije ($\log(\text{šansa})$).

$\exp(\beta)$

3. Interpretacija

Glavni alat za tumačenje je **Odnos Šansi (Odds Ratio)**, koji se dobija kao $\exp(\beta)$. On pokazuje multiplikativni efekat prediktora na šansu ishoda.



4. Primena

Nezaobilazan alat u medicini (stratifikacija rizika), finansijama (predviđanje neplaćanja kredita), marketingu (predviđanje kupovine) i društvenim naukama.

Iza Binarnog: Šta Ako Imamo Više od Dve Kategorije?

Principi logističke regresije se mogu proširiti i na zavisne promenljive sa više od dva ishoda.

Nominalna Logistička Regresija



Kada se koristi:

Za zavisnu promenljivu sa 3+ kategorije koje **nemaju prirodan redosled**.

Primer:

Izbor internet pretraživača (Google, Bing, Yahoo!), izbor boje proizvoda (crvena, plava, crna).

Kako radi: Modelira verovatnoću pripadanja svakoj kategoriji u odnosu na jednu referentnu kategoriju.

Ordinalna Logistička Regresija



Kada se koristi:

Za zavisnu promenljivu sa 3+ kategorije koje **imaju prirodan redosled**.

Primer:

Ocena (5-10), stepen zadovoljstva (nezadovoljan, neutralan, zadovoljan), zdravstveno stanje (dobro, stabilno, kritično).

Kako radi: Modelira kumulativne verovatnoće (verovatnoću da ishod bude u određenoj kategoriji ili ispod nje).



Logistička Regresija: Pretvaranje složenosti u jasne, delotvorne uvide.

Ovaj materijal je zasnovan na analizi akademskih radova i praktičnih primera kako bi pružio sveobuhvatan pregled metode.