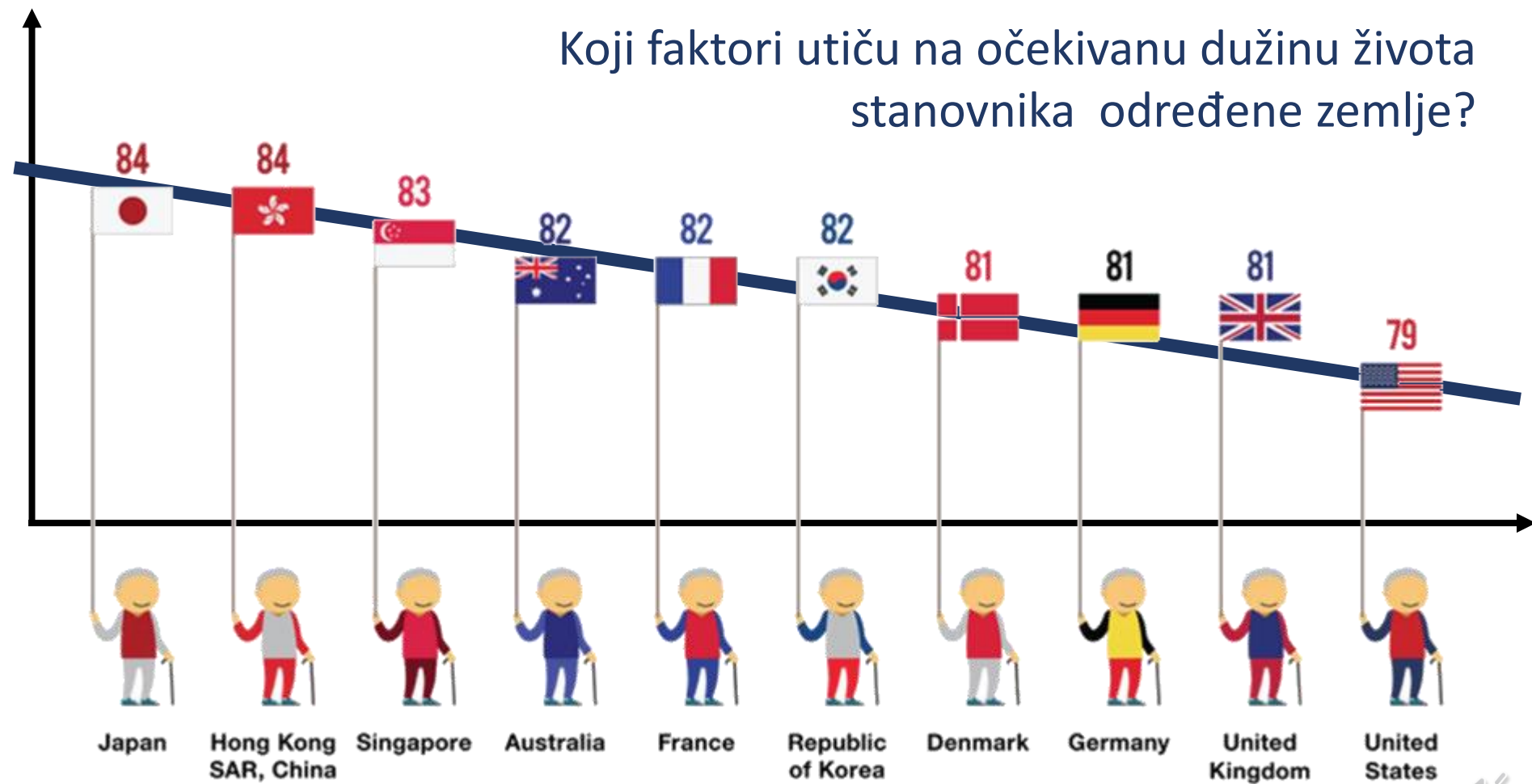


Jednostruka linearna regresija

Koji faktori utiču na očekivanu dužinu života stanovnika određene zemlje?





Jedna mogućnost:
pristup adekvatnim
sanitarijama



Podaci

% stanovništva sa pristupom adekvatnim sanitarijama

<http://apps.who.int/gho/data/node.main.46>

Prosečna dužina života

<http://apps.who.int/gho/data/node.main.688>

Country	SanitationFacilities	LifeExpectancyBirth
Australia	100.0	82.8
Bahamas	92.0	76.1
Belize	90.5	70.1
Chile	99.1	80.5
Greece	99.0	81.0
Guatemala	63.9	71.9
Guyana	83.7	66.2
Hungary	98.0	75.8
Latvia	87.8	74.6
Malta	100.0	81.7

Serbia?



Ne postoji zemlja sa SF = 96.4

SanitationFacilities = 96.4

LifeExpectancyBirth = ?

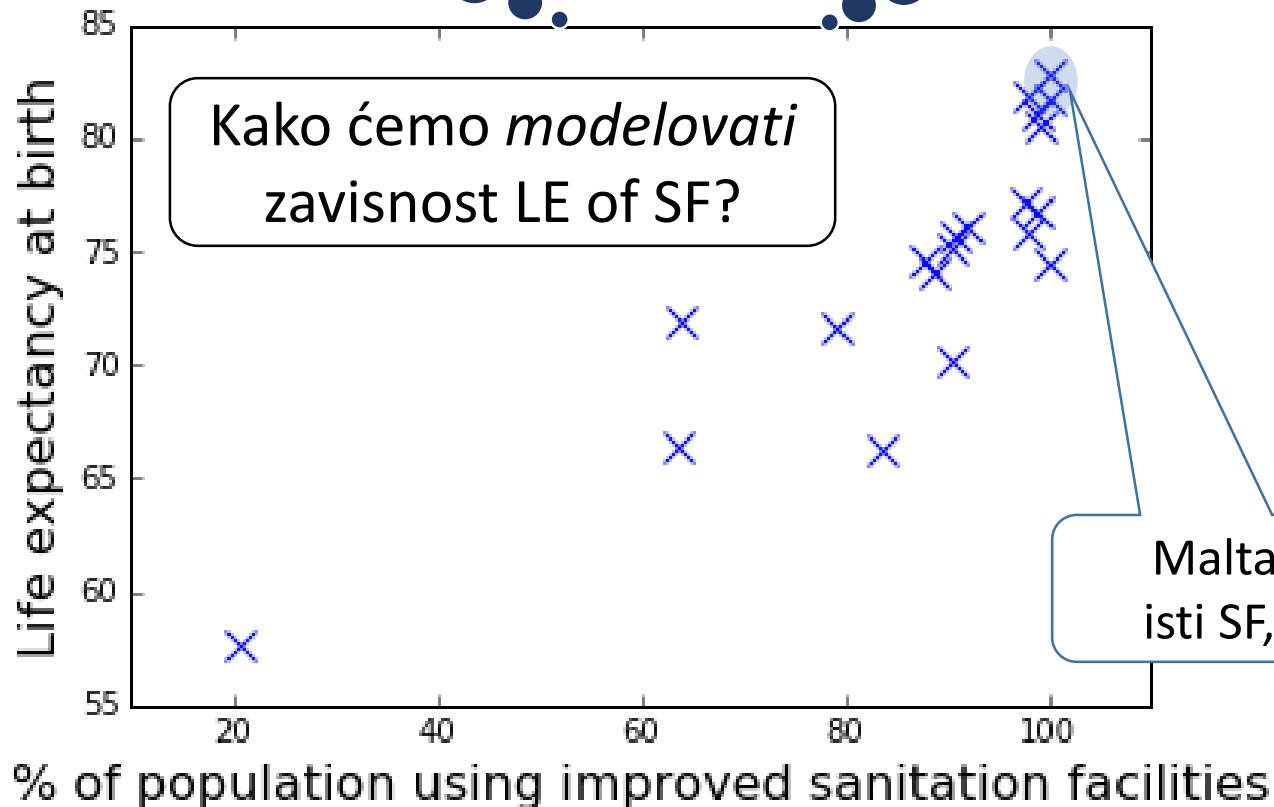
Da li da uzmemo prosek sličnih?



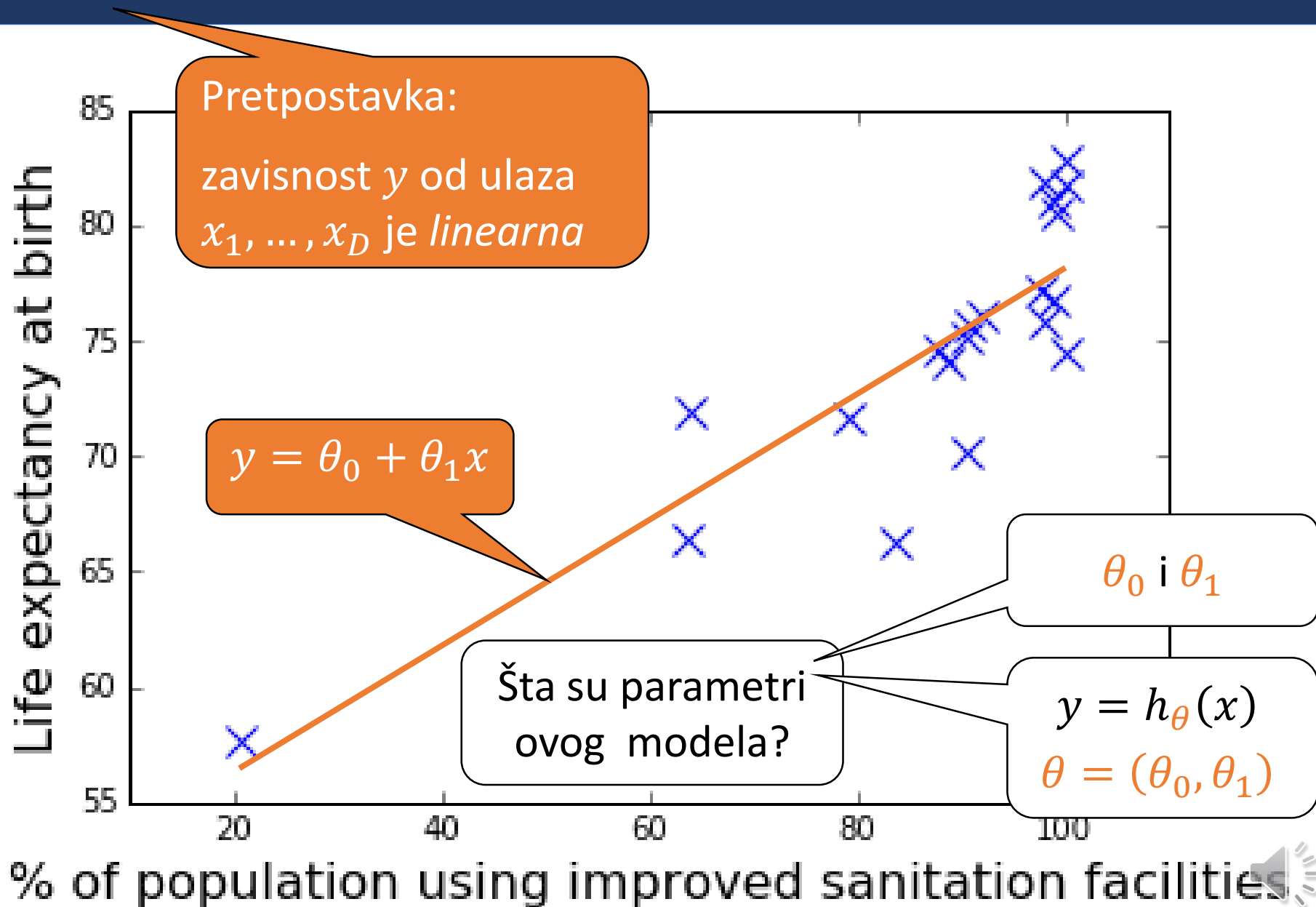
Podaci

Da li ovaj problem sadrži
sve suštinske komponente
za primenu ML-a?

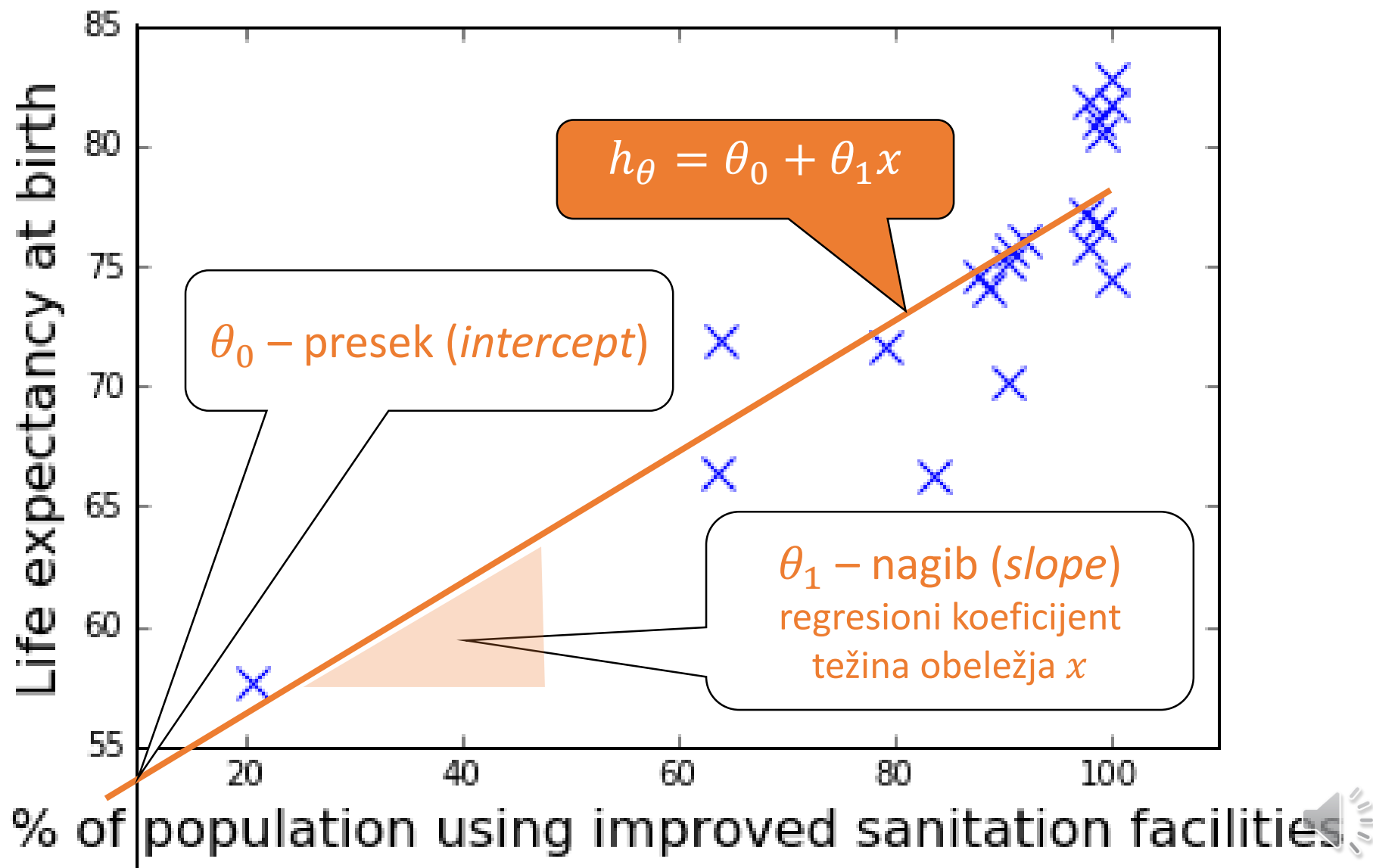
Nadgledano ili
nenadgledano učenje?



Linearna regresija



Linearna regresija



Primena modela

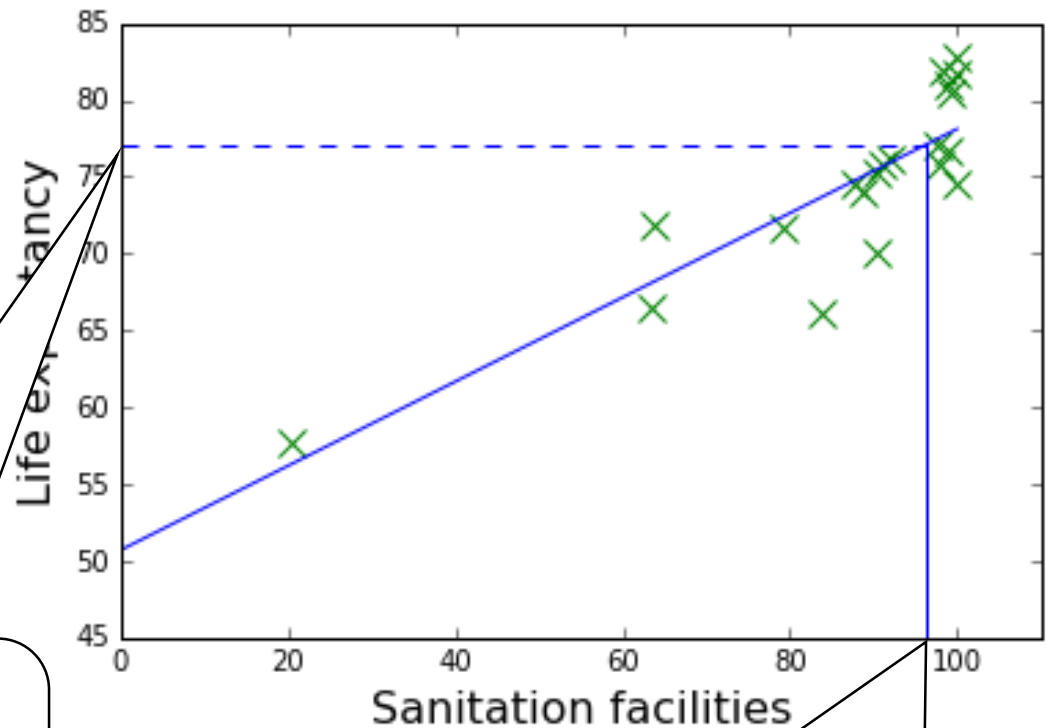
Od klase modela

$$h_{\theta}(x) = \theta_1(x) + \theta_0$$

odabrali smo model

$$h_{\hat{\theta}}(x) = 0.274 \cdot x + 50.7$$

koji najviše odgovara opservacijama



$$\hat{y}(96.4) = h_{\hat{\theta}}(96.4) = 0.274 \cdot 96.4 + 50.7 = 77.1$$

(tačna vrednost je 75.6)

Ako je u Srbiji procenat stanovništva sa pristupom odgovarajućim sanitarijama 96.4, koliki je očekivan životni vek stanovnika Srbije?



Primena modela

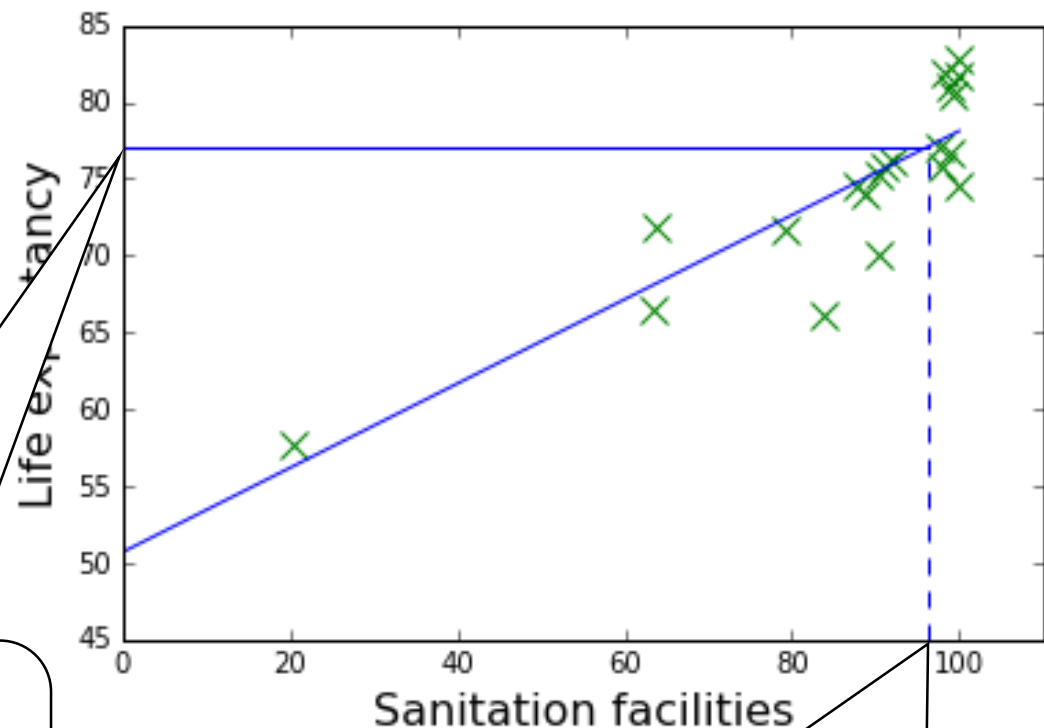
Od klase modela

$$h_{\theta}(x) = \theta_1(x) + \theta_0$$

odabrali smo model

$$h_{\hat{\theta}}(x) = 0.274 \cdot x + 50.7$$

koji najviše odgovara opservacijama



Koliki procenat stanovništva treba da ima pristup odgovarajućim sanitarijama da bi očekivani životni vek bio 77.1 godinu?

$$\hat{x} = \frac{77.1 - 50.7}{0.274} = 96.4$$



Interpretacija modela

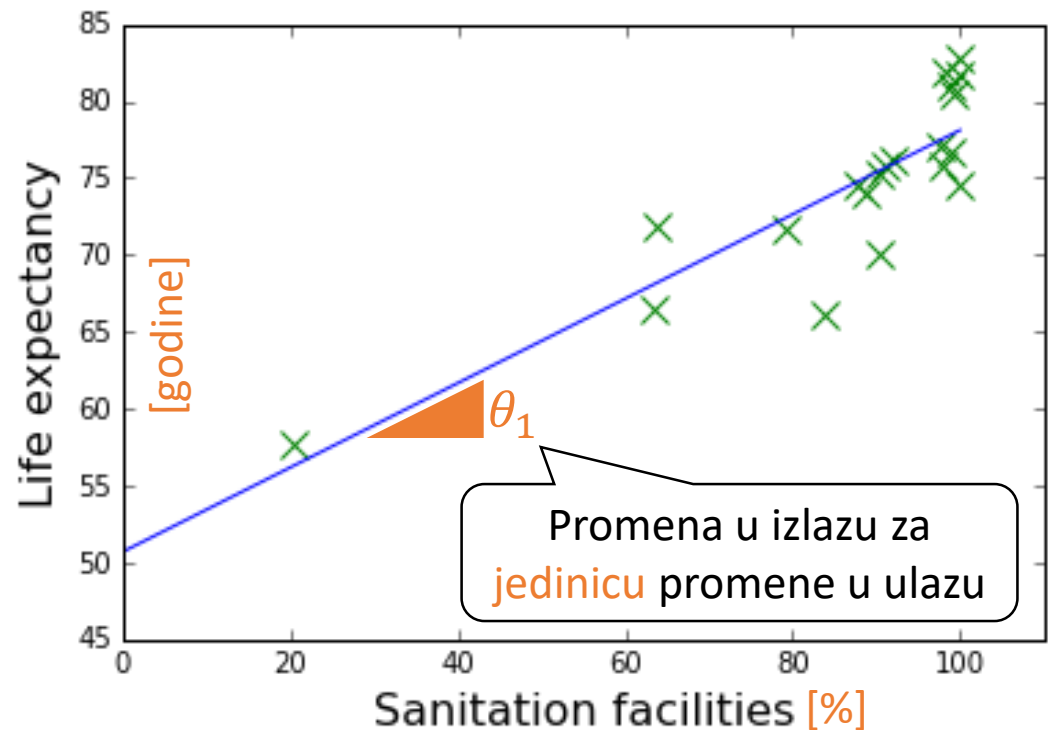
Od klase modela

$$h_{\theta}(x) = \theta_1(x) + \theta_0$$

odabrali smo model

$$h_{\hat{\theta}}(x) = 0.274 \cdot x + 50.7$$

koji najviše odgovara opservacijama



Ako 1% više stanovništva dobije pristup sanitarijama, životni vek poraste za 0.274 godine



Interpretacija modela

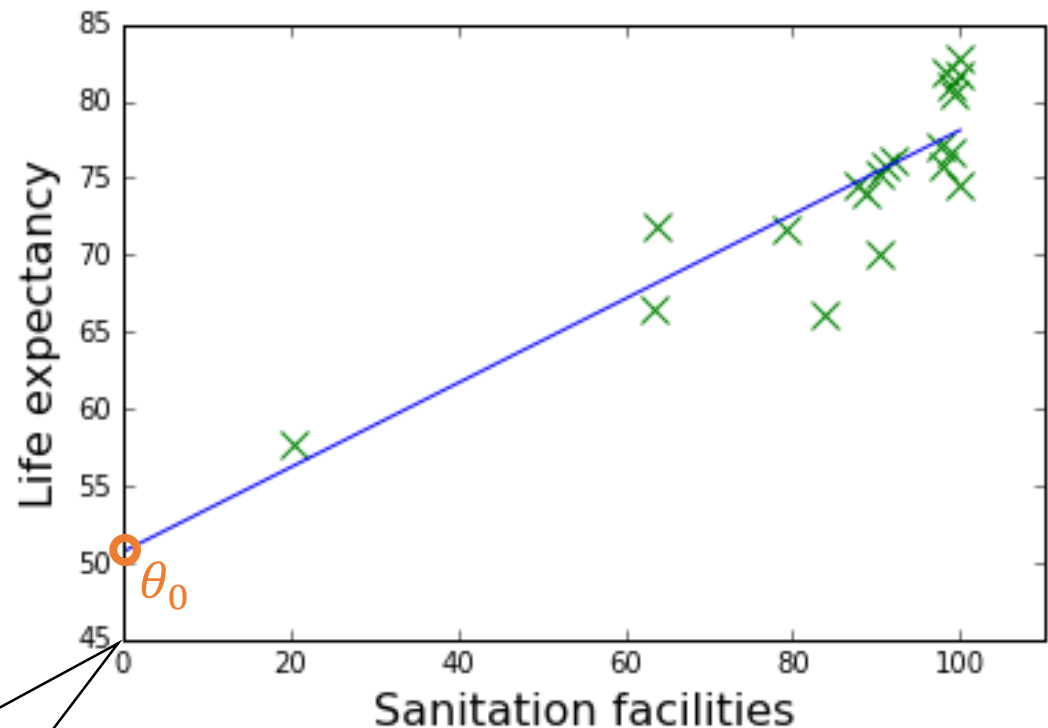
Od klase modela

$$h_{\theta}(x) = \theta_1(x) + \theta_0$$

odabrali smo model

$$h_{\hat{\theta}}(x) = 0.274 \cdot x + 50.7$$

koji najviše odgovara opservacijama



Koliki je životni vek kada niko nema pristup sanitarijama?

U opštem slučaju θ_0 nije interpretabilno

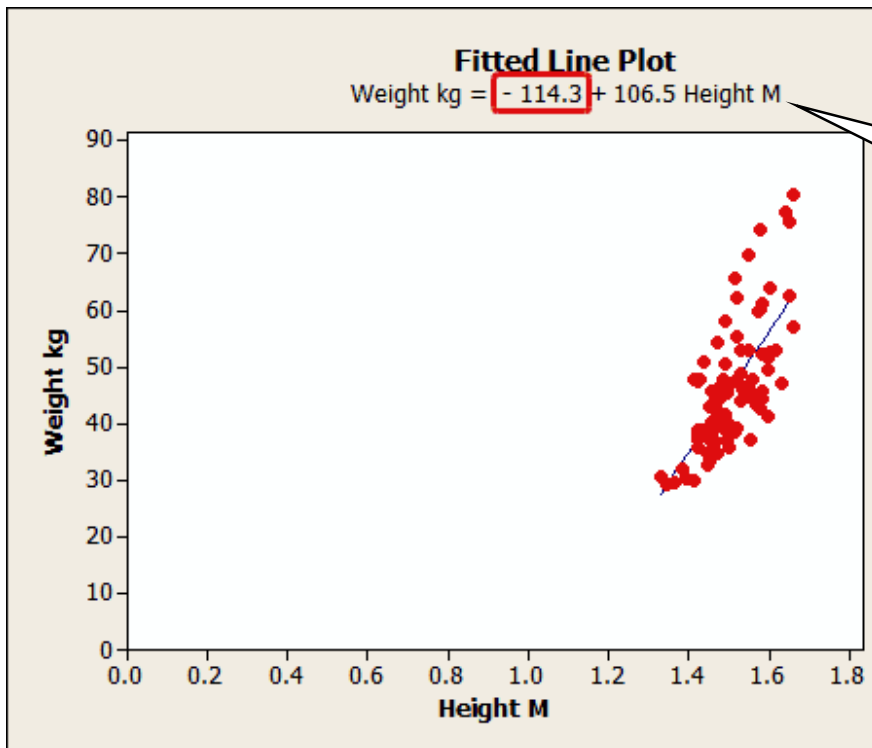


Interpretacija modela

Najčešće θ_0 nije interpretabilno

Matematički θ_0 je očekivana vrednost y kada su x postavljeni na 0

Praktično Da li je primer gde su svi $x = 0$ realno moguć?



Čovek koji ima visinu 0 je težak – 114.3 kg

Ali, da li je realno moguće da se pojavi čovek sa visinom 0?



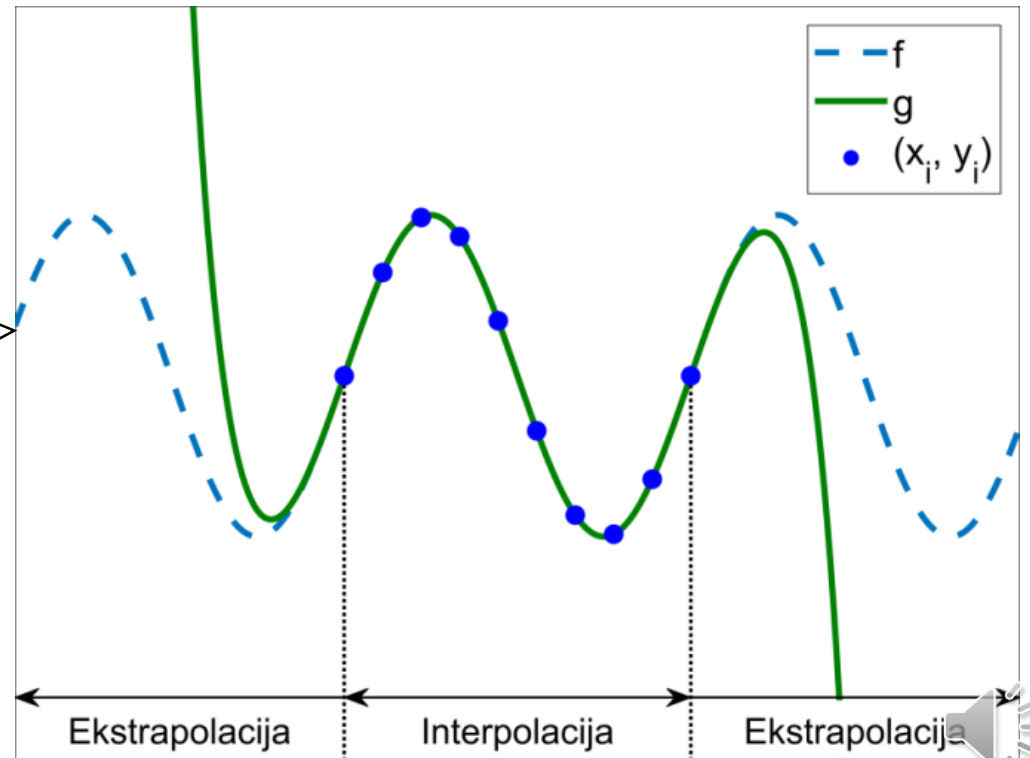
Interpretacija modela



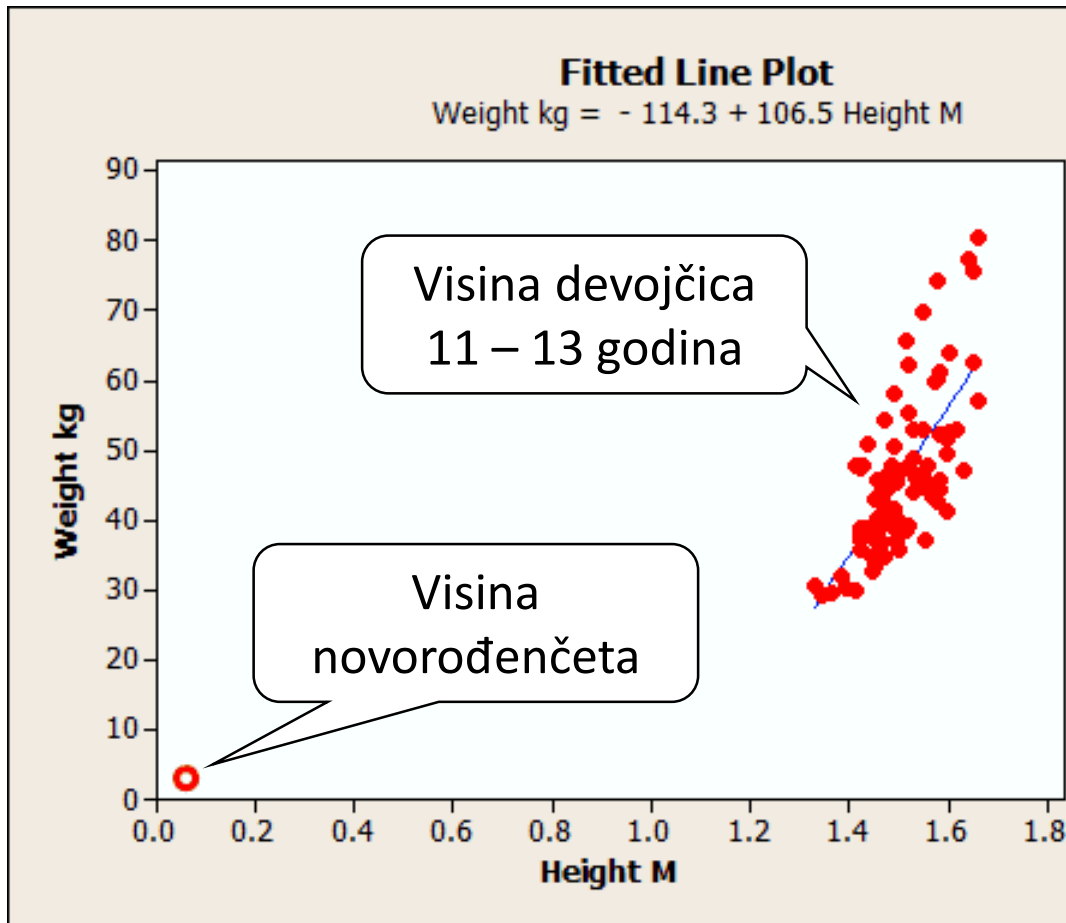
Ako je realno postaviti sve x na 0, da li je onda θ_0 uvek interpretabilno?

Ne uvek, prisetite se interpolacije i ekstrapolacije

Ne treba da koristite regresioni model za predikciju tačaka van opsega trening podataka



Interpretacija modela

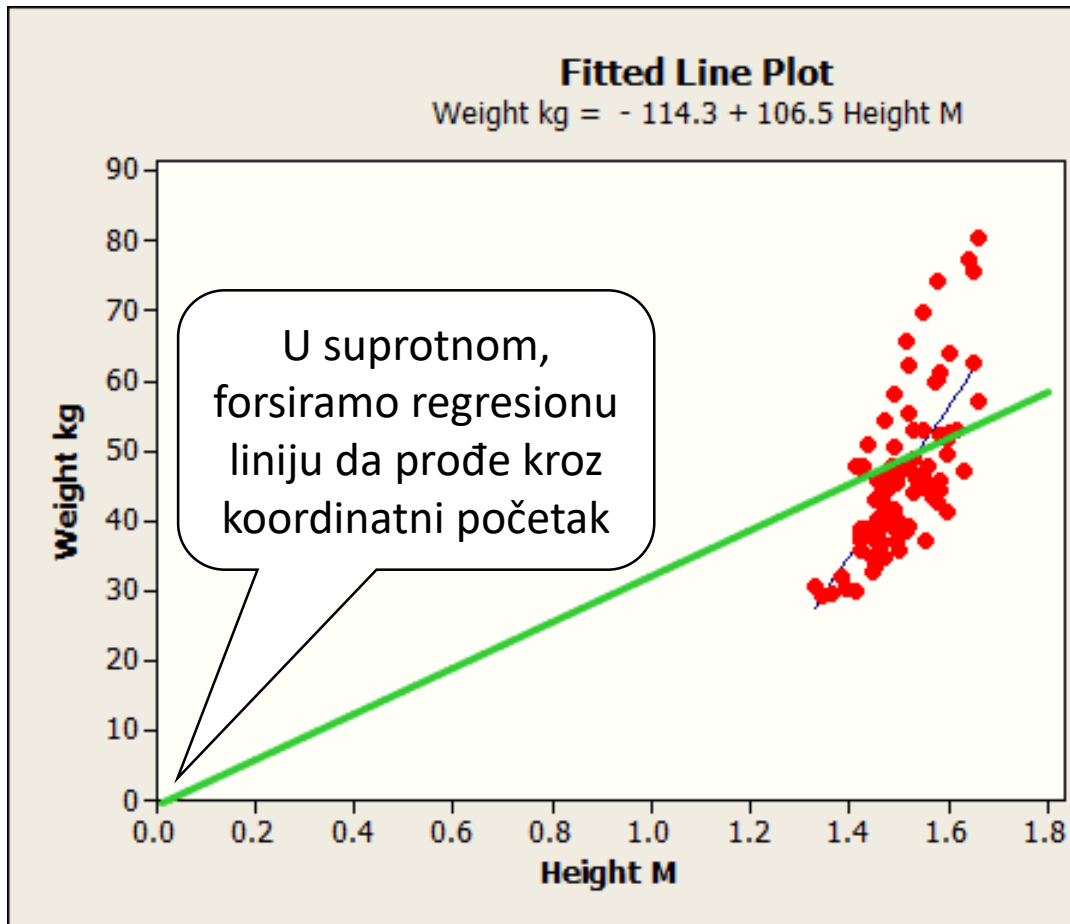


Ako nemamo trening podatke
u okolini 0-vrednosti, ne vredi
da interpretiramo θ_0

Čak i ako imamo, θ_0 ne mora
biti interpretabilno



Interpretacija modela



Ipak je od krucijalnog značaja da θ_0 uvrstimo u model



Rezime

- Upoznali smo se sa modelom linearne regresije
- Videli smo kako da ga interpretiramo
 - Konstantu θ_0 generalno ne vredi interpretirati
 - Ali je korisna za model
 - Vrednost regresionog modela jeste razumevanje kako se y menja sa promenom x

