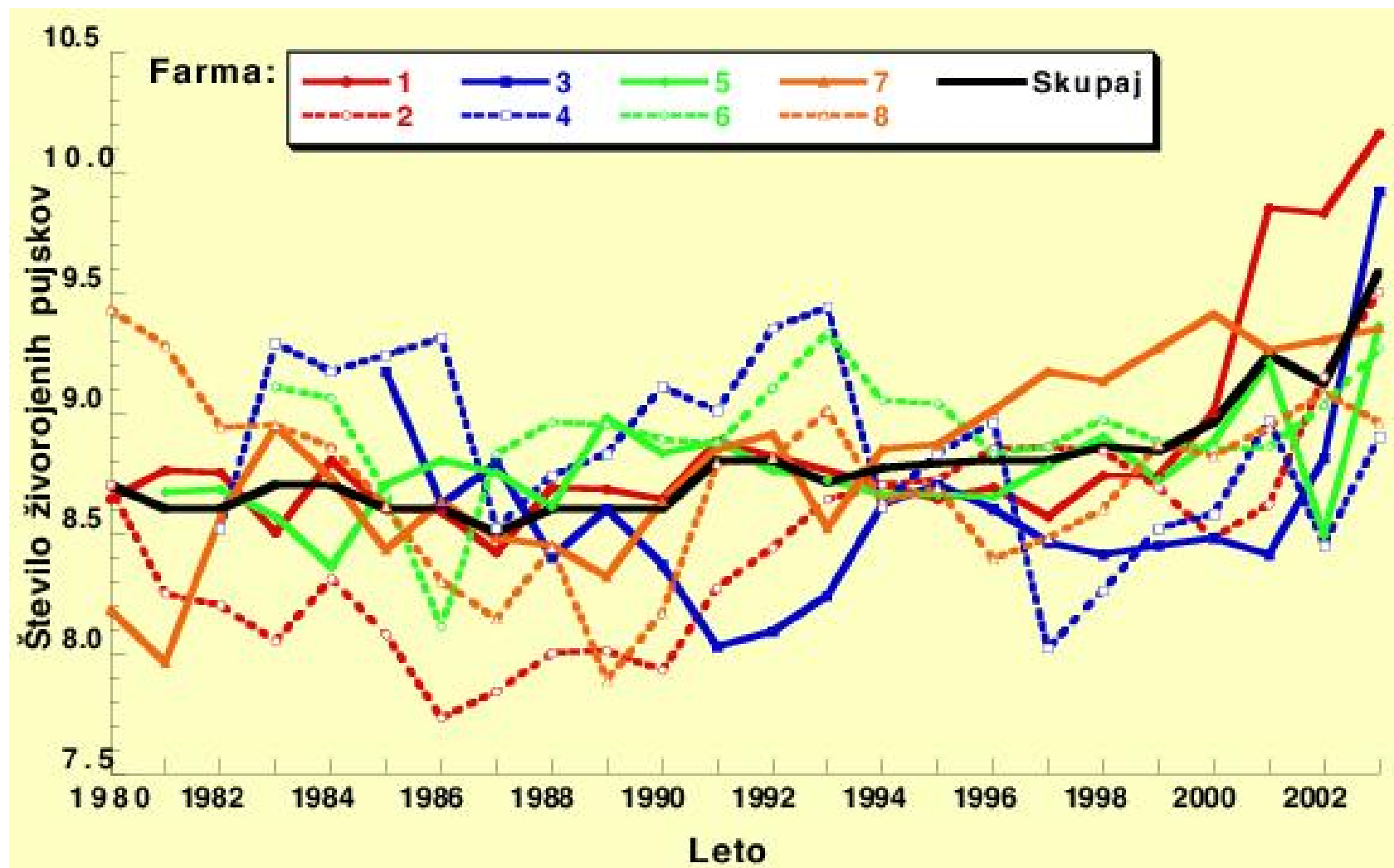


Statistični modeli - interakcija -

Milena Kovač

16. november 2012

Število živorojenih pujskov

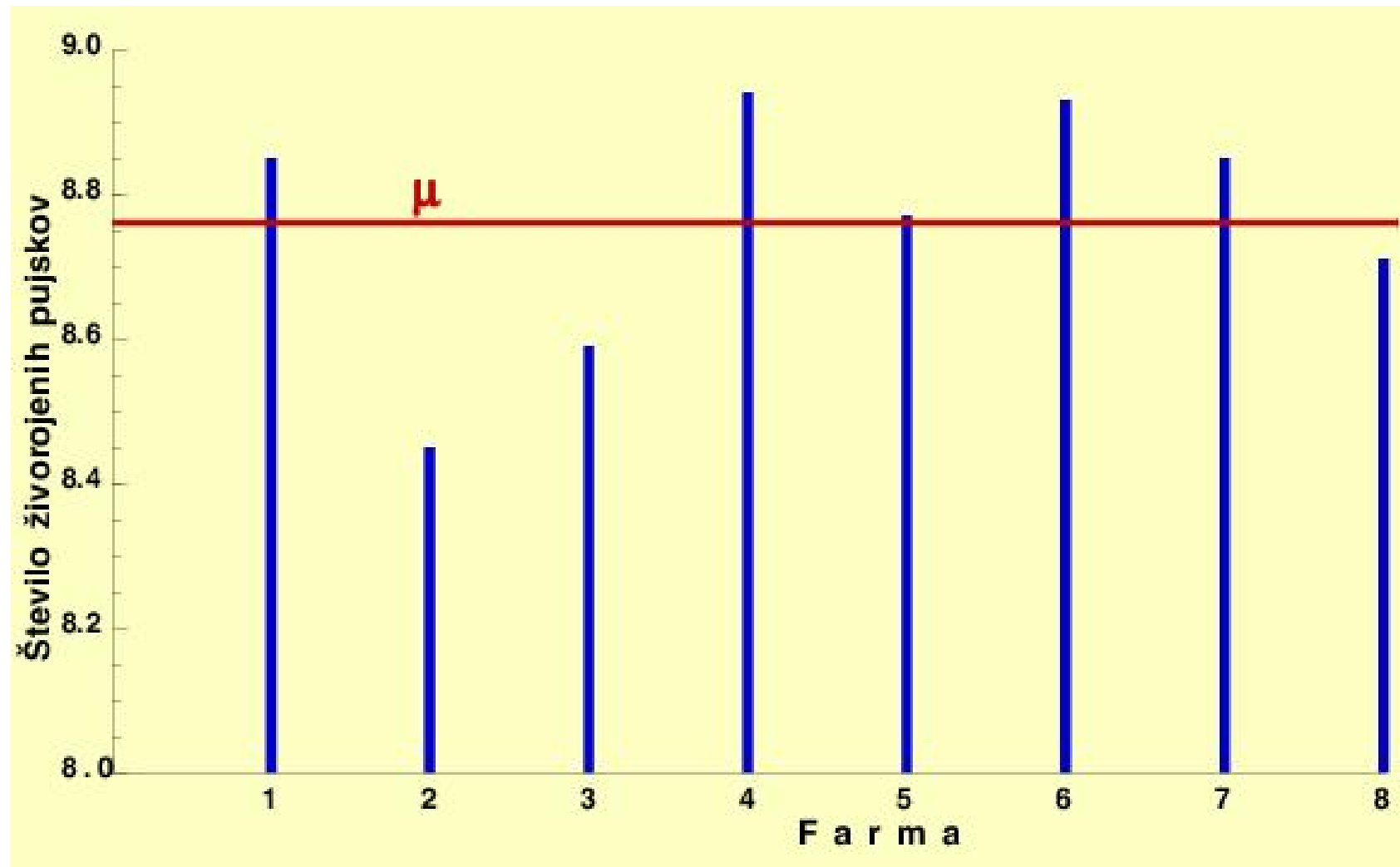


Sestavimo model!

- Vplivi: ■ leto, farma
- Odvisna spremenljivka: ■ število živorojenih pujskov
- Neodvisna spremenljivka: ■ leto, a ni primerne funkcije
- Označimo vplive: ■ leto (L), farma (F)

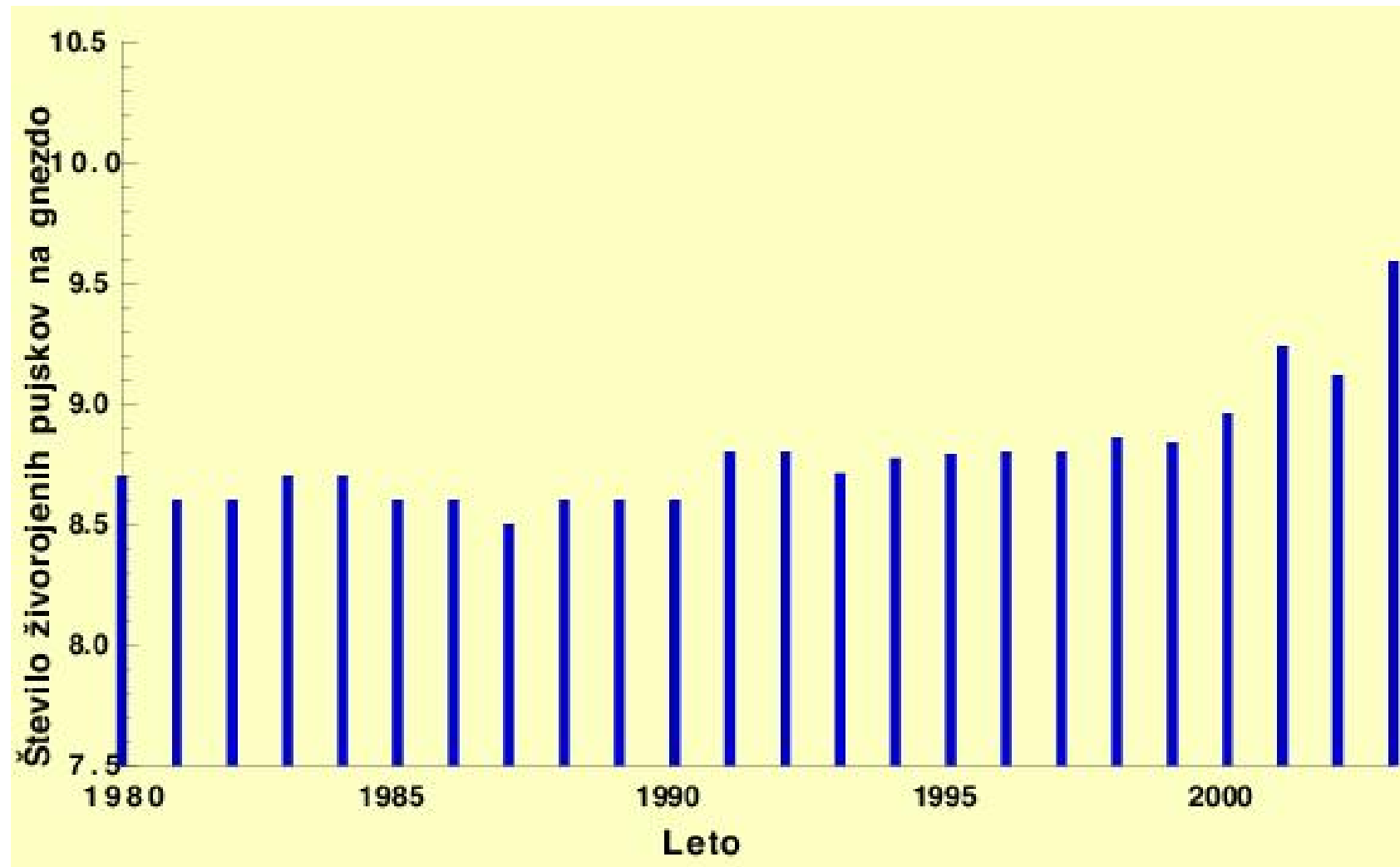
- Obrazložitev povezave: ■
Število živorojenih pujskov se z leti povečuje, a povečevanje ne moremo opisati s funkcijo zaradi nepričakovanih sprememb. Čeprav je ugodno, da se velikost gnezda povečuje, pa nas graf prepriča, da z regresijo ne bomo uspešni. Uporabili bomo kar vpliv z nivoji - za vsako leto poseben nivo. Med farmami obstajajo razlike, ki se z leti spreminjajo. Vzroki so različni. Trud, da bi bilo gnezdo povečano, je različno (ne)uspešen.

Vpliv farme



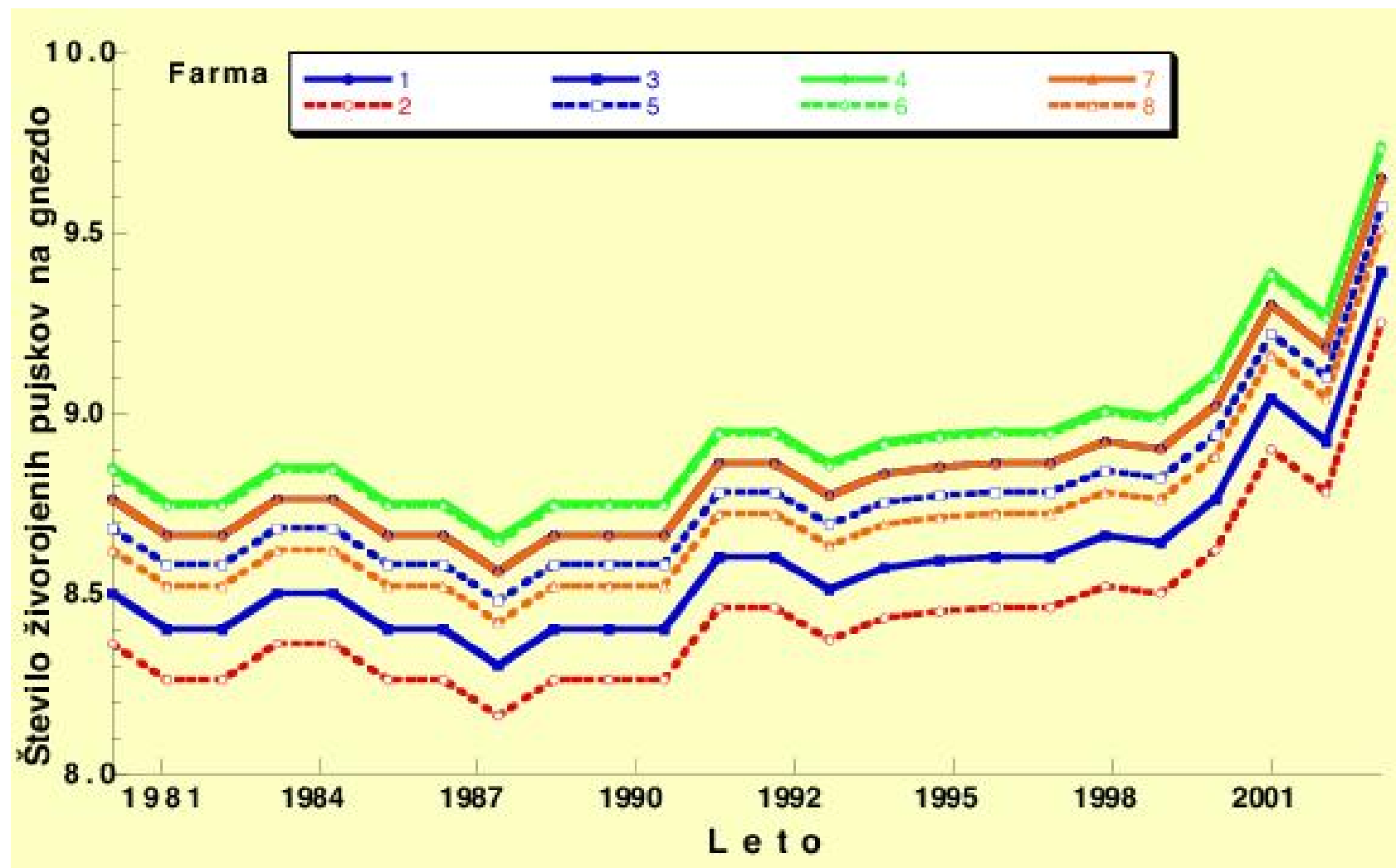
$$y_{ij} = \mu + F_i + e_{ij}$$

Vpliv leta



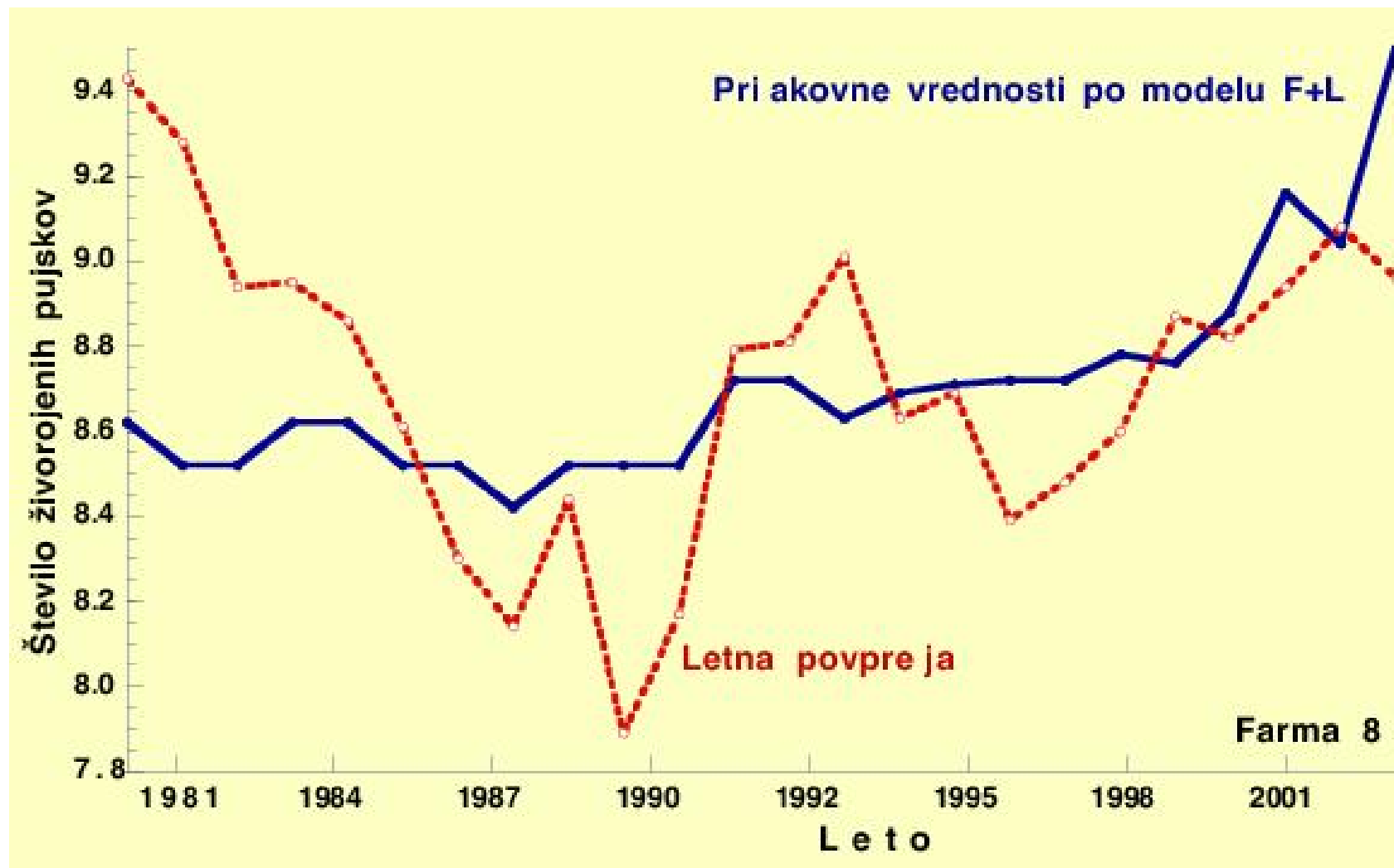
$$y_{ij} = \mu + L_i + e_{ij}$$

Vpliv farme in leta



$$y_{ijk} = \mu + F_i + L_j + e_{ijk}$$

Primerjava model (modro) : podatki (rdeče)



(V tem primeru regresija za vpliv leta ni primerna!)

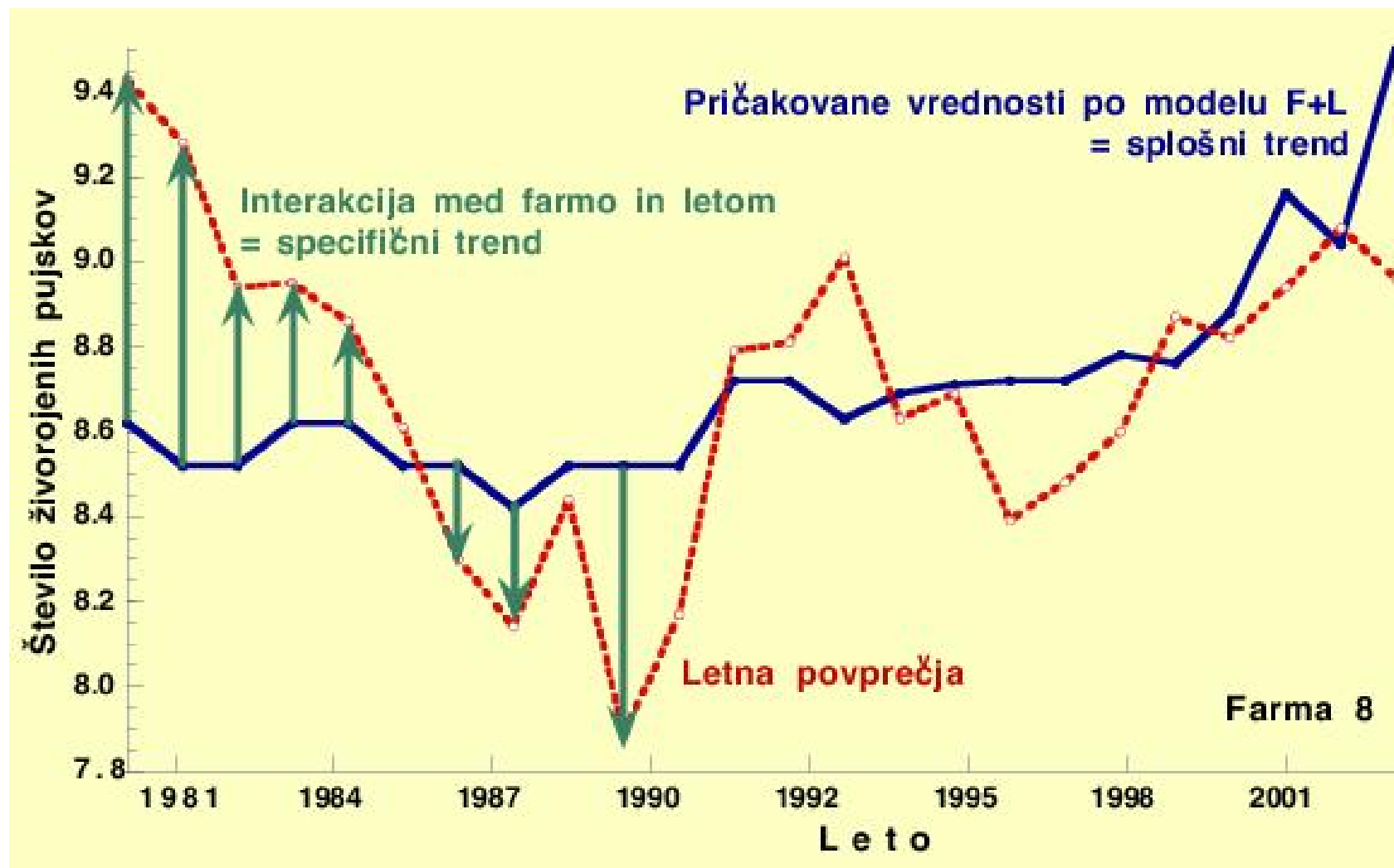
Dopolnimo model!

- Ponovno napišimo enačbo modela:

$$y_{ijk} = \mu + F_i + L_j + e_{ijk}$$

- Ali smo povezavo dovolj dobro opisali? ■
Število ŽP se tudi poslabša ali izboljša za nekaj let. Spremembe so neodvisne od preteklih let. Lahko je izbruhnila bolezen ... Na farmi 8 so letna povprečja precej odstopala od splošnega trenda, ki smo ga določili z vplivoma farma in leto.
- Dodatni vpliv: ■omogoča, da ima leto na vsaki farmi drugačen vpliv (FL)

Vpliv farme, leta in interakcija - farma 8



$$y_{ijk} = \mu + F_i + L_j + FL_{ij} + e_{ijk}$$

Kvalitativni vplivi

1. Farme bi lahko razvrstili na več načinov:
 - (a) po abecedi,
 - (b) po šifrah,
 - (c) na prvo mesto bi dali farmo, ki nas najbolj zanima.

2. Nivoji so ločeni in nepovezani, nimajo vrednosti

3. Za ponazoritev praviloma uporabljamo HISTOGRAME

4. Prirejani kvalitativni vplivi (leto) - izjema

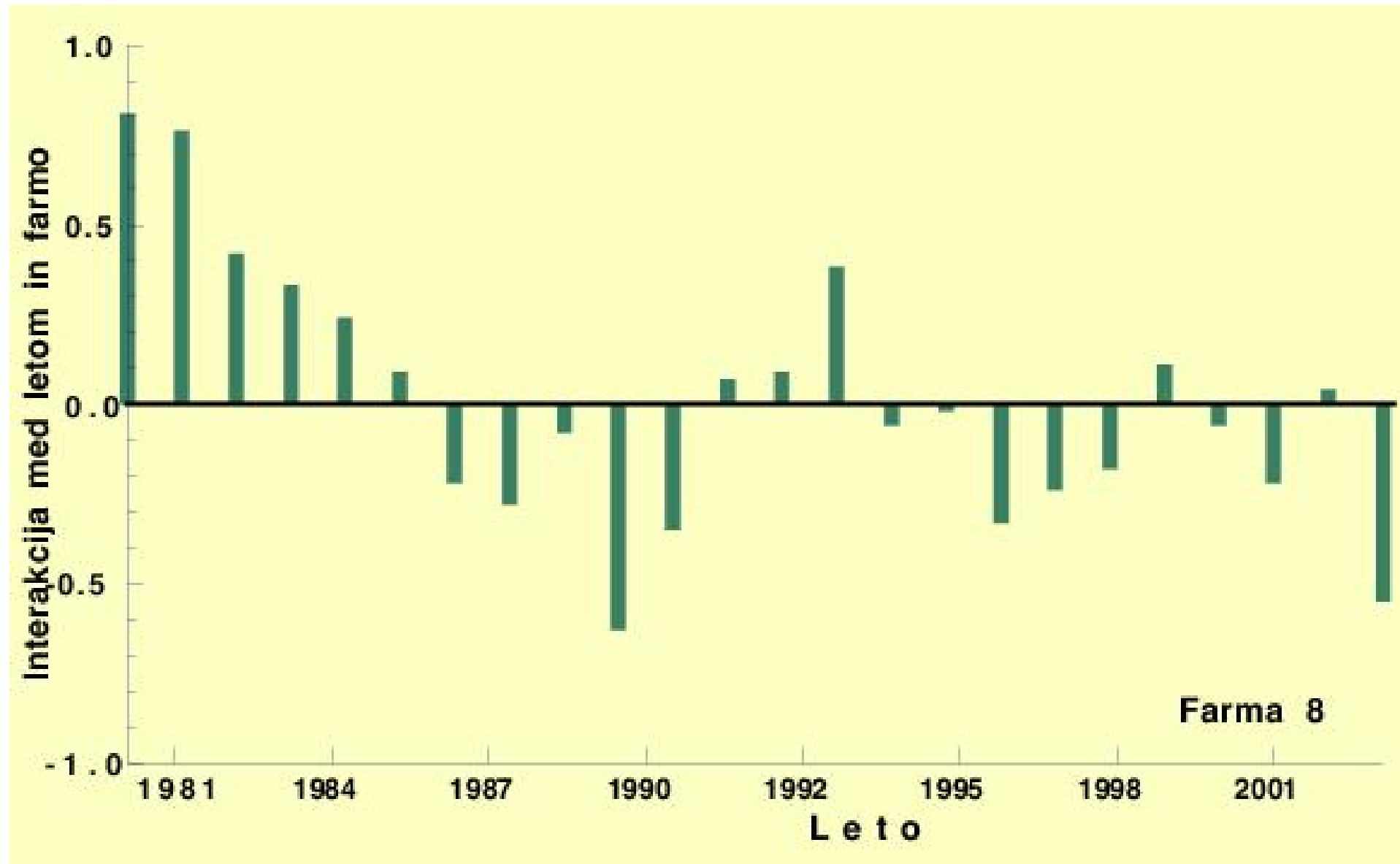
Model s kvalitativnimi vplivi

$$y_{ijk} = \mu + F_i + L_j + FL_{ij} + e_{ijk}$$

kjer pomeni:

- y_{ijk} - število živorojenih pujskov
 - μ - srednja vrednost
 - F_i - vpliv farme; $i = 1, 2, \dots, 9$
 - L_j - vpliv leta; $j = 1980, 1981, \dots, 2003$
 - FL_{ij} - interakcija med farmo in letom
- * e_{ijk} - ostanek

Interakcija med farmo in letom



Interakcija¹

Z vplivom leta dobimo **splošni** (skupaj za vse farme, črna črta na grafu), z interakcijo med farmo in letom pa **specifični** (na farmi, odstopanja posameznih črt od črne črte) vpliv leta

- bolezen izbruhne samo na eni farmi
- delavec je zaposlen samo na eni farmi
- klima se spreminja po letih in farmah
- hlevi so različni in jih adaptirajo ...
- spreminjajo postopke dela

¹v interakcijo se bomo še poglobili

Označitev interakcije

$$y_{ijk} = \mu + F_i + L_j + FL_{ij} + e_{ijk}$$

- oznaki vplivov F in L sestavimo v FL
- črke sestavimo po abecednem redu indeksov pri glavnih vplivih
- dodamo indekse na koncu oznake
- interakcijo med farmo in letom označimo FL_{ij}

Načrt poskusa pri kvalitativnih vplivih

Poskus poteka med 1980 do 2003 na več farmah prašičev.

<i>Farma</i>	1980	1981	1982	\dots	2003	$\Sigma leta$
<i>A</i>	n_{11}	n_{12}	n_{13}	\dots	n_{1q}	$n_{1.}$
<i>B</i>	n_{21}	n_{22}	n_{23}	\dots	n_{2q}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots	\dots	\vdots	\vdots
<i>E</i>	n_{p1}	n_{p2}	n_{p3}	\dots	n_{pq}	$n_{p.}$
$\Sigma farme$	$n_{.1}$	$n_{.2}$	$n_{.3}$	\dots	$n_{.q}$	n

Načrt poskusa pri kvalitativnih vplivih

- vse celice naj bodo zapolnjene
- načrtujte dovolj opazovanj po celicah
- izenačena velikost skupin (celic)
- najmanj eno opazovanje na celico (premalo za resne rezultate!)
- delajmo skromne, enostavne poskuse ...
- podatki iz prakse (proizvodnje) - komplicirani modeli

Poskus z mladnicami

Farma	Žival	Pasma	Mesec	Masa	DP	DHS (mm)	
1	1	ŠL	Jan	102	540	13	13
1	2	ŠL	Jan	98	550	16	14
1	3	ŠL	Feb	105	550	16	16
1	4	ŠL	Feb	102	580	15	12
1	5	LW	Jan	95	520	20	17
1	6	LW	Feb	101	500	24	24
1	7	LW	Feb	101	490	27	25
1	8	NL	Jan	97	560	26	27
1	9	NL	Jan	100	550	22	19
1	10	NL	Feb	97	600	23	25
1	11	NL	Feb	102	610	24	22

Določimo vplive (mladice)

✓ Z mladkami smo opravili poskus na izbrani farmi s tremi pasmi v dveh mesecih. Živalim smo ugotovili starost in jim tehtali maso. Izračunati želimo plemensko vrednost (vpliv živali) za dnevni prirast (DP) in debelino hrbtnne slanine (DHS).

Vplivi: ■ pasma (P), mesec (M), masa, žival (a)

Neodvisna spremenljivka: ■ masa (x)

Lastnosti: ■ dnevni prirast, debelina hrbtnne slanine

Interakcija med pasmo in mesecem: PM

Sestavimo modela

- za dnevni prirast

$$y_{ijk} = \mu + P_i + M_j + PM_{ij} + a_{ijk} + e_{ijk}$$

- za debelino hrbtnne slanine

$$y_{ijkl} = \mu + P_i + M_j + PM_{ij} + b(x_{ijk} - 100) + a_{ijk} + e_{ijkl}$$

- vpliv živali smo označili z malo črko (naključni vpliv)
- za DHS smo vključili regresijo za maso
- pri DHS imamo dve meritvi na žival: več meritev kot živali

Stopinje prostosti

DP-o.	μ	P_i	M_j		a_{ijk}	model	ostanek	
param.	1	3	2		—	6	—	
s.p.	1	2	1		—	4	11-4=7	
DP-m.	μ	P_i	M_j	PM_{ij}		a_{ijk}	model	ostanek
param.	1	3	2			—	6	—
s.p.	1	2	1			—	4	11-4=7
DHS-o.	μ	P_i	M_j		b	a_{ijk}	model	ostanek
param.	1	3	2		1	—	7	—
s.p.	1	2	1		1	—	5	11-5=6
DHS-m.	μ	P_i	M_j	PM_{ij}	b_{ij}	a_{ijk}	model	ostanek
param.	1	3	2	6	6	—	18	
s.p.	1	2	1	2	6	—	12	11-12 = ???

Sistematski vplivi

1. imajo malo nivojev (osebkov, enot, razrede)
2. sorazmeroma veliko opazovanj po nivoju
3. proučujemo samo nivoje v poskusu
4. nivoji med seboj niso sorodni ali pa sorodstvo zanemarimo
5. ne posplošujemo rezultatov na druge nivoje
6. lahko načrtno ponovimo

Naključni vplivi

1. nivoji vzorčeni iz znane porazdelitve
2. imajo več nivojev (osebkov, enot, razrede)
3. nivoji so med seboj lahko sorodni, zato lahko
4. posplošimo rezultate na druge, sorodne nivoje
5. ga ne moremo ponoviti

Izjeme?

1. Vpliv živali

- (a) očetje so izbrani in imajo veliko potomcev,
- (b) matere imajo malo potomcev in so lahko celo naključno izbrane
- (c) potomci imajo lahko tudi samo eno meritev

2. Rejci (skupno okolje v čredu)

- (a) velike črede
- (b) samo male črede
- (c) reje so zelo različnih velikosti (farme s 15 do 5000 plemenskih svinj)

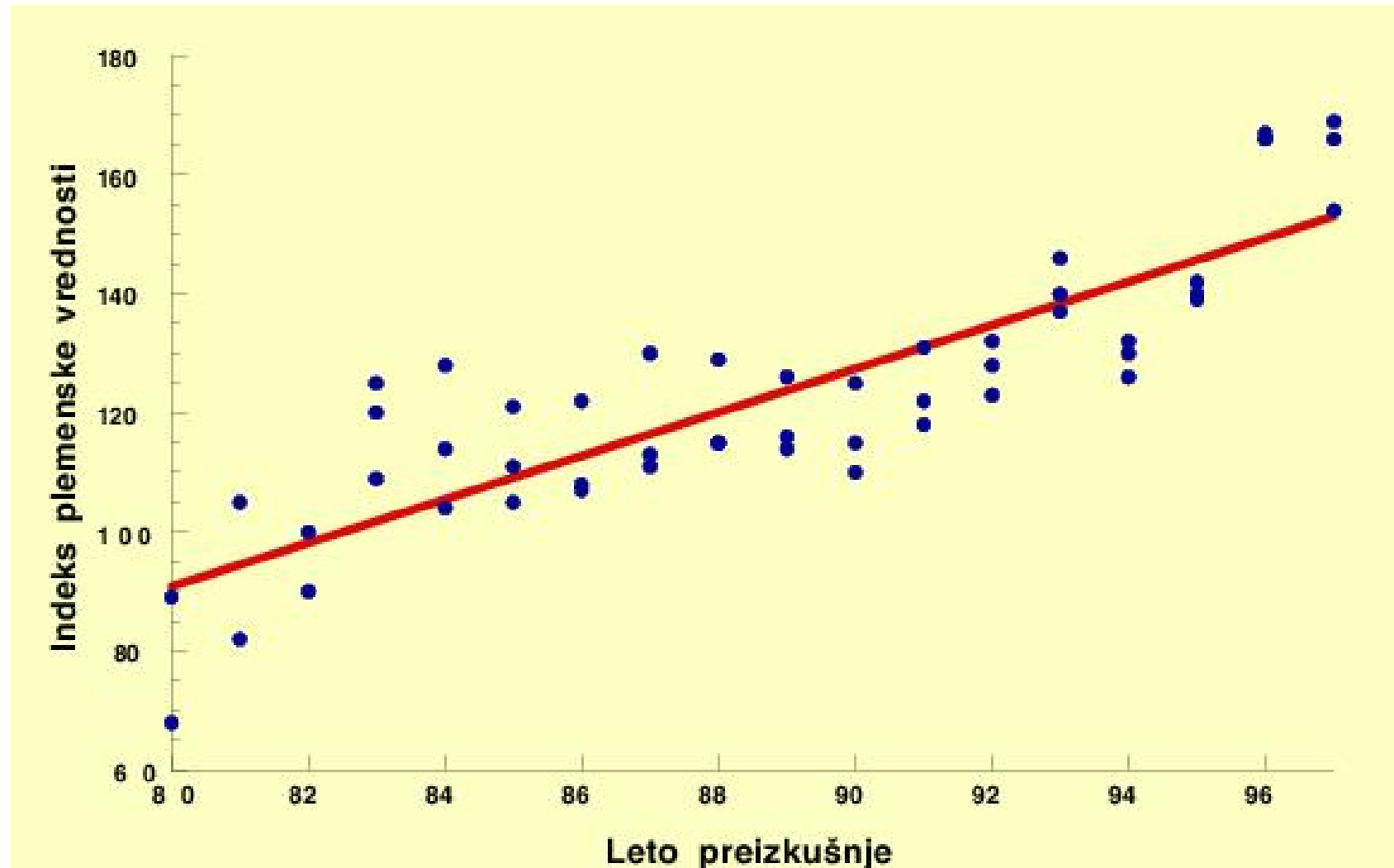
Dilema? Če imamo veliko nivojev, vzamemo naključni vpliv.

Indeksi pri neodvisni spremenljivki (pojasnilo)

- x ima več indeksov kot b , kateremu pripada
- nosi indekse vseh "nadrejenih" vplivov
- mora biti jasno, pri kateremu b -ju pripada (ima indeks pri b -ju)
- mora biti jasno, s katerimi opazovanji (y) je v parih
- indeks za stopnjo polinoma ni dodan neodvisni spremenljivki
- več neodvisnih spremenljivk \Rightarrow
 - dodatni indeks na začetku (npr. $x_{1...}$ in $x_{2...}$).
 - dodatne indekse dobijo tudi b -ji.

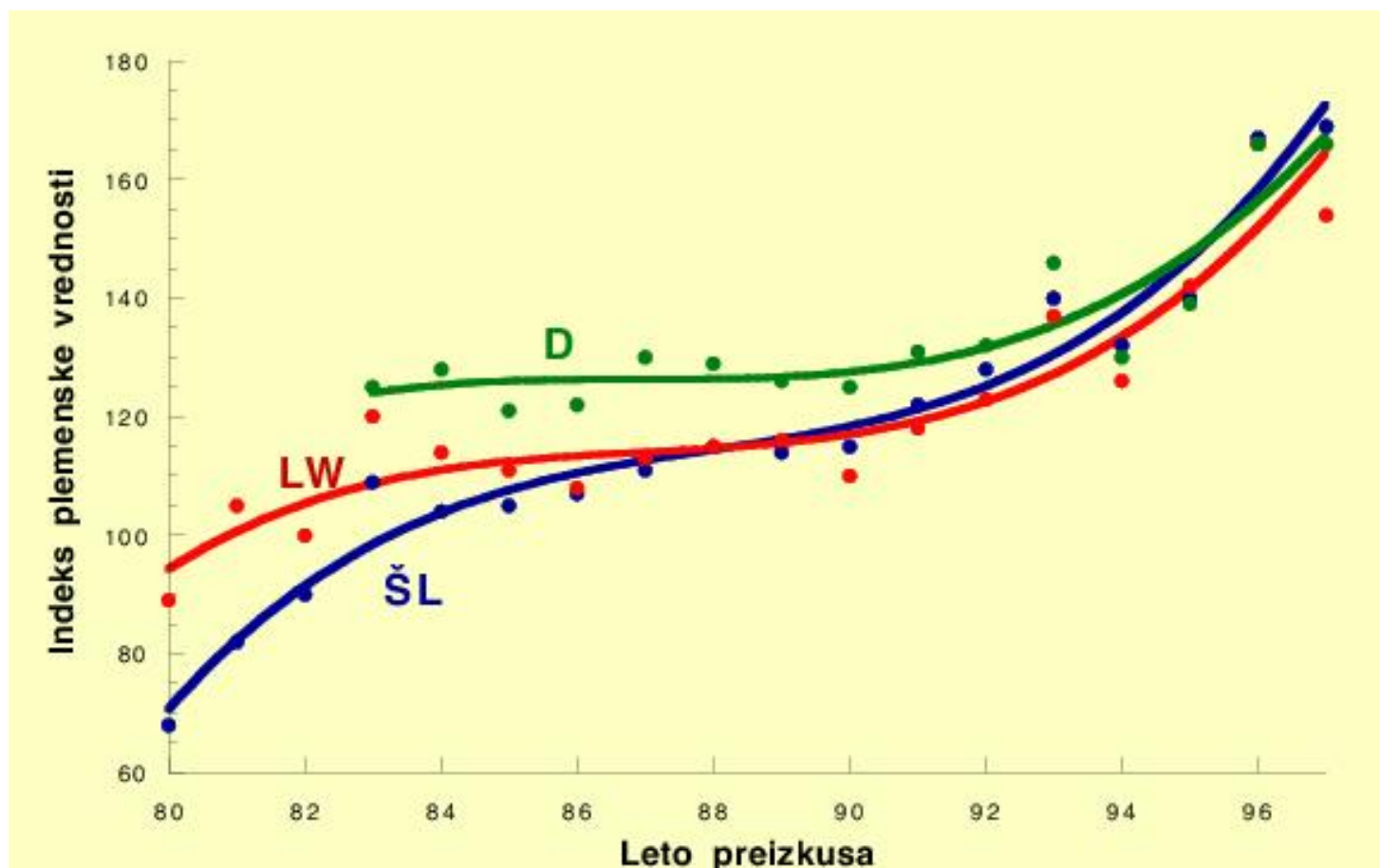
- z indeksi ne pretiravamo

Primer 1



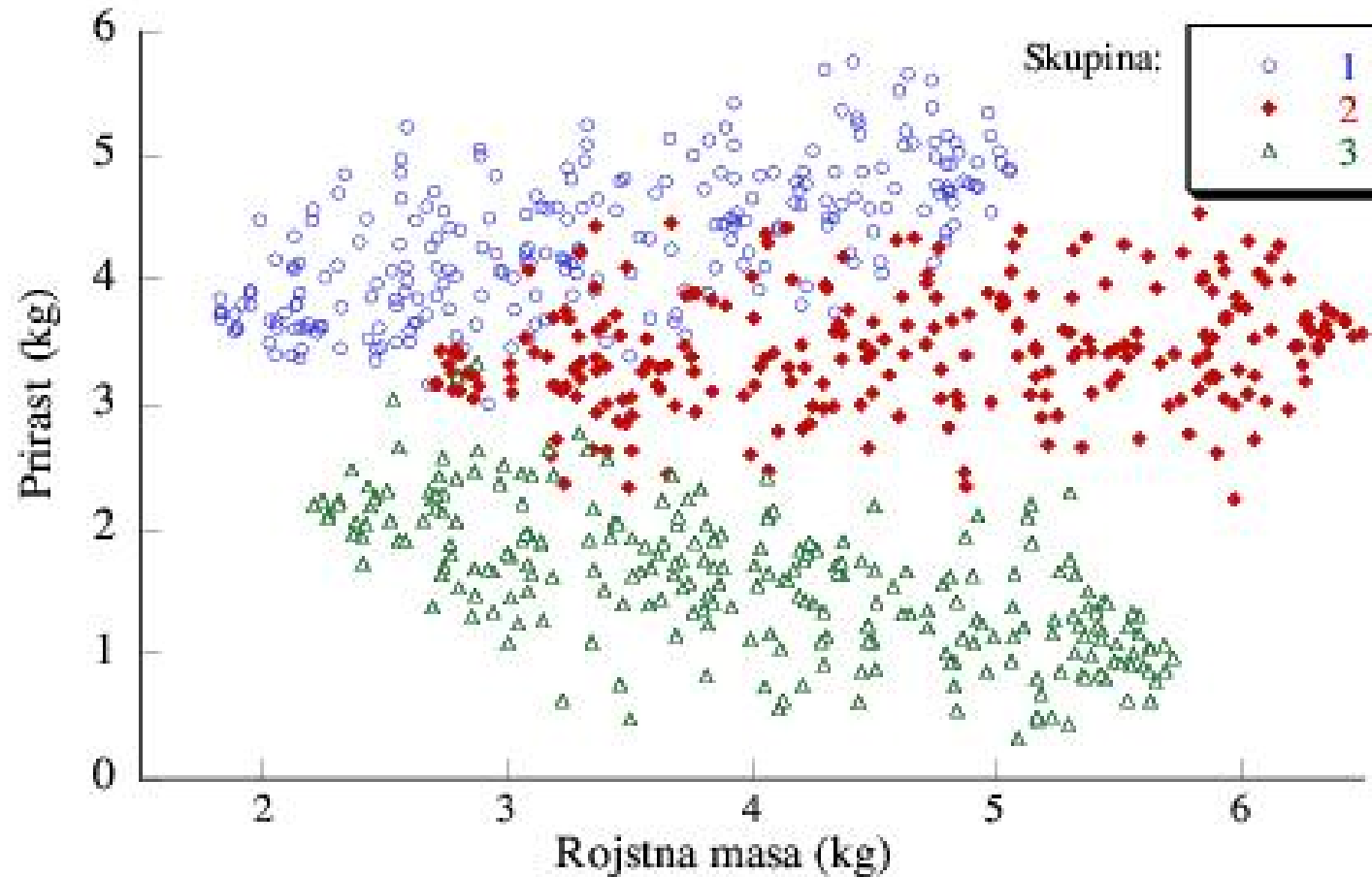
$$y_i = \mu + b(x_i - 80) + e_i$$

Primer 2



$$y_{ij} = \mu + P_i + b_{IIi}(x_{ij} - 80) + b_{IIIi}(x_{ij} - 80)^2 + b_{IIIi}(x_{ij} - 80)^3 + e_{ij}$$

Primer 3



$$y_{ij} = \mu + b_i(x_{ij} - 1) + e_{ij}$$