Pregled statističnih modelov

Milena Kovač

24. oktober 2010

Delitev statističnih modelov

1. Linearni : nelinearni modeli
2. Eno- oz. večlastnostni modeli
3. Modeli s polnim in z nepolnim rangom
4. Modeli z ozirom na število vplivov
5. Hierarhični, križno klasificirani in kombinirani modeli
6. Sistematski, naključni in mešani modeli

2. Enolastnostni modeli

$$y_{ijk} = \mu + P_i + b\left(x_{ijk} - \bar{x}\right) + e_{ijk}$$

- Obdelujemo vsako odvisno spremenljivko y_{ijk} posebej
- Statistični modeli so lahko enaki ali različni
- Ne upoštevamo povezav med izbranimi lastnostmi
- Prednost
  - manjše število enačb
  - dobro, kadar lastnosti niso korelirani: $r_{tt'} = 0$
  - dobro, kadar ni manjkajočih meritev
- Slabost: kadar predpostavke ne držijo

2. Večlastnostni modeli

$$y_{1ijk} = \mu_1 + P_{1i} + b_1\left(x_{ijk} - \bar{x}\right) + e_{1ijk}$$

$$y_{2ijk} = \mu_2 + P_{2i} + b_{I2}\left(x_{ijk} - \bar{x}\right) + b_{II2}\left(x_{ijk} - \bar{x}\right)^2 + e_{2ijk}$$

$$cov\left(y_{1ijk},\, y_{2ijk}\right) = \sigma_{12}$$

- Obdelamo več koreliranih lastnosti hkrati: $y_{1ijk},\, y_{2ijk},\, \ldots\, y_{tijk}$
- Upoštevamo kovarianco $\sigma_{tt'}$ med lastnostmi
- Statistični modeli za posamezne lastnosti lahko različni
- Težave pri zelo koreliranih lastnostih: $r_{tt'} = 1$ ali $r_{tt'} = -1$
- Računsko zahtevnejša

3. Z ozirom na rang sistema enačb

- Modeli s polnim rangom
  - število parametrov = stopinj prostosti
  - sistem enačb ima samo eno rešitev
  - vsi parametri so ocenljivi
    - ⋆ modeli z enim sistematskim členom z razredi
    - ⋆ regresijski modeli

- Modeli z nepolnim rangom
  - število parametrov >stopinj prostosti
  - sistem enačb ima nič ali neskončno mnogo rešitev
  - vsi parametri so ocenljivi

3a. Modeli z enim sistematskim vplivom z razredom

$y_{ijk} = AB_{ij} + e_{ijk}$ $\qquad i = 1, 2, \ldots p;$ $\qquad j = 1, 2, \ldots q;$

$y_{ijkl} = AB_{ij} + a_{ijk} + e_{ijkl}$

- **Sistematski vpliv**: samo interakcija med vplivoma A_i in B_j

$y_{ijk} =$	AB_{ij}	$+e_{ijk}$	$y_{ijkl} =$	AB_{ij}	$+a_{ijk}$	$+e_{ijkl}$
Število parametrov	pq			pq	m	
Število stopinj prostosti	pq	n-pq		pq	m	n-pq

- Število neznanih parametrov za sistematske vplive = številu stopinj prostosti

- Del sistema enačb z **naključnimi vplivi** je polnega ranga
  - m =število živali s podatki, dodani tudi predniki iz porekla
  - izjema: genetsko identični osebki (enojajčni dvojčki, klonirani organizmi)
  - naključni vplivi ne zmanjšujejo stopinj prostosti za ostanek

3a. Regresijski modeli

- Enostavna (linearna) regresija: $y_i = \beta_0 + \beta x_i + e_i$
- Polinomska regresija: $y_i = \beta_0 + \beta_I x_i + \beta_{II} x_i^2 + \cdots + e_i$
- Kompleksne funkcije
  - laktacijske krivulje
- Multipla regresija: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$
  - več kot ena pojasnjevalna spremenljivka
  - ocena mesnatosti na liniji klanja

3a. Model z linearno regresijo

$y_i =$	$\beta_0 +$	$\beta x_i +$	e_i
Število parametrov	1	1	
Število stopinj prostosti	1	1	n-2

- Model z linearno regresijo

  - β_0 - srednja vrednost (presečišče v $x_i = 0$)
  - β - regresijski (smerni) koeficient premice

- Število neznanih parametrov = številu stopinj prostosti

- Parametra β_0 in β sta ocenljiva in imata natanko eno rešitev

- Dodani naključni del modela ne spremeni zaključkov

3a. Polinomska regresija

$$y_i = \beta_0 + \beta_I x_i + \beta_{II} x_i^2 + \beta_{III} x_i^3 + e_i$$

$y_i =$	$\beta_0 +$	$\beta_I x_i +$	$\beta_{II} x_i^2 +$	$\beta_{III} x_i^3$	e_i
Štev. parametrov	1	1	1	1	
Štev. stopinj prostosti	1	1	1	1	n-4

- Polinom tretje stopnje s štirimi neznanimi parametri
- Vse je potrebno oceniti in so ocenljivi
- Stopnjo polinoma postavljamo pogosto na osnovi slike in strokovne razlage
- S stopnjami ne pretiravamo (do vključno pete potence)

3a. Stopnja polinoma

Model z visokimi potencami je neprimeren:

1. strokovna razlaga polinomov višje stopnje je problematična

2. v statististiki

   (a) velja zakon skromnosti,
   (b) morda polinom ni primerna funkcija

3. matematika garantira, da višje polinome lahko aproksimiramo z nižjimi polinomi:

$$X^n \cong \alpha + \left(\frac{\partial x^n}{\partial x}\right)_{x=\alpha}(x-\alpha) + \frac{1}{2}\left(\frac{\partial^2 x^n}{(\partial x)^2}\right)_{x=\alpha}(x-\alpha)^2 + \cdots =$$

$$\cong \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \ldots + \gamma_n x^k$$ Taylorjeve vrste

3a. Regresijski modeli - izjema

- polinom, kjer je ocena regresijskega koeficienta $\hat{b}_x = 0$
  - izpustimo člen z $\hat{b}_x = 0$
  - poiščemo ortogonalne polinome (členi regresije so neodvisni)

- visoka korelacija med neodvisnimi spremenljivkami
  - $cov\,(x_{1ijk}, x_{2ijk}) \approx 1$ ali $cov\,(x_{1ijk}, x_{2ijk}) \approx -1$
  - odvisni spremenljivki sta skoraj isto
  - vzamemo tisto, ki bolje pojasnjuje, ali
  - se bolje kombinira z ostalimi

3a. Multipla regresija

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$

$y_i =$	$\beta_0 +$	$\beta_1 x_{1i} +$	$\beta_2 x_{2i} +$	$\beta_3 x_{3i} +$	e_i
Štev. parametrov	1	1	1	1	
Štev. stopinj prostosti	1	1	1	1	n-4

- Tri neodvisne spremenljivke: x_{1i}, x_{2i}, x_{3i}

- Pri vseh treh pričakujemo linearno povezavo

- Vsi parametri (β_0, β_1, β_2, β_3) ocenljivi

3a. Multipla regresija s polinomi

$$y_i = \beta_0 + \beta_{I1} x_{1i} + \beta_{II1} x_{1i}^2 + \beta_{III1} x_{1i}^3 + \beta_{I2} x_{2i} + \beta_{II2} x_{2i}^2 + e_i$$

$y_i =$	Število parametrov	Število stopinj prostosti	
β_0	1	1	dve neodvisni spremenljivki
$+\beta_{I1} x_{1i}$	1	1	pri prvi x_{1i}
$+\beta_{II1} x_{1i}^2$	1	1	- polinom tretje stopnje
$+\beta_{III1} x_{1i}^3$	1	1	
$+\beta_{I2} x_{2i}$	1	1	pri drugi x_{2i}
$+\beta_{II2} x_{2i}^2$	1	1	- polinom druge stopnje
$+e_i$		n-4	

3a. Multipla regresija

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + e_i$$

$y_i =$	$\beta_0 +$	$\beta_I x_{1i} +$	$\beta_2 x_{2i} +$	$\beta_3 x_{1i} x_{2i}$	e_i
Štev. parametrov	1	1	1	1	
Štev. stopinj prostosti	1	1	1	1	n-4

- modela sta enaka

  - dve neodvisni spremenljivki x_{1i} in x_{2i}
  - linearni člen pri vsaki spremenljivki s regresijskima koeficientoma β_1 in β_2
  - produkt spremenljivk z regresijskim koeficientom β_3 ali β_{12}

3a. Ugnezdena regresija

$$y_{ij} = \beta_i + \beta_{1i}x_{1ij} + \beta_{2i}x_{2ij} + \beta_{12i}x_{1ij}x_{2ij} + e_{ij} \qquad i = 1,\ 2,\ 3,\ 4;$$
$$j = 1,\ 2,\ \ldots\ n_i$$

$y_{ij} =$	$\beta_i +$	$\beta_{Ii}x_{1ij} +$	$\beta_{2i}x_{2ij} +$	$\beta_{12i}x_{1ij}x_{2ij}$	e_{ij}
Štev. parametrov	4	4	4	4	
Štev. stopinj prostosti	4	4	4	4	$\sum n_i$-16

- dve neodvisni spremenljivki x_{1ij} in x_{2ij}
- regresija je ugnezdena znotraj sistematskega vpliva β_i
- v modelu je 12 regresijskih koeficientov
- sistem je polnega ranga, ima eno rešitev

3b. Ugnezdena regresija

$$y_{ij} = \mu + \beta_i + \beta_{1i}x_{1ij} + \beta_{2i}x_{2ij} + \beta_{12i}x_{1ij}x_{2ij} + e_{ij} \qquad i = 1, 2, 3, 4;$$
$$j = 1, 2, \ldots n_i$$

$y_{ij} =$	$\mu +$	$\beta_i +$	$\beta_{Ii}x_{1ij} +$	$\beta_{2i}x_{2ij} +$	$\beta_{12i}x_{1ij}x_{2ij}$	e_{ij}
Štev. parametrov	1	4	4	4	4	
Štev. stopinj prostosti	1	4-1	4	4	4	$\sum n_i$-1

- dve neodvisni spremenljivki x_{1ij} in x_{2ij}
- v modelu imamo tudi srednjo vrednost μ
- regresija je ugnezdena znotraj sistematskega vpliva β_i
- v modelu je 12 regresijskih koeficientov
- sistem **ni polnega ranga**, ima nič ali neskončno mnogo rešitev

3b. Modeli z nepolnim rangom

- **Več parametrov kot stopinj prostosti**

- Kadarkoli sta dva sistematska vpliva ali en sistematski vpliv in srednja vrednost

$$y_{ij} = \mu + A_i + e_{ij} \qquad i = 1, 2 \qquad j = 1, 2, \ldots n_i$$

- Modelu odgovarja naslednji sistem enačb

$$\begin{array}{rrrcr}
10\mu & +4A_1 & +6A_2 & = & 114 \\
4\mu & +4A_1 & +0 & = & 48 \\
6\mu & +0 & +6A_2 & = & 66
\end{array} \qquad \begin{array}{r} n =? \\ n_1 =? \\ n_2 =? \end{array}$$

- Sistem enačb ima linearno odvisne enačbe
  - Prva vrstica je vsota drugih dveh
  - Zadnja vrstica je razlika med drugo in prvo

- Poiskati moramo sistem enačb, kjer ni odvisnosti

3b. Reševanje sistema enačb z nepolnim rangom

$$A_1 = \frac{1}{4}\,(48 - 4\mu)$$

$$A_2 = \frac{1}{6}\,(66 - 6\mu)$$

$$10\mu = [114 - (48 - 4\mu) - (66 - 6\mu)] \Longrightarrow 0 = 0$$

- Zadnja enačba ni uporabna
- Rešitve dobimo, če si "izmislimo" oceno za μ
- **Restrikcija** = izbira vrednosti za "manjkajoče" parametre

μ	0	10	20	$\cdots$	rešitve si niso
A_1	12	2	-8	$\cdots$	na videz prav nič
A_2	11	1	-9	$\cdots$	podobne, a
$A_1 - A_2$	1	1	1	$\cdots$	so ocenljive linearne
$\mu + A_1$	12	12	12	$\cdots$	kombinacije enake
$\mu + A_2$	11	11	11	$\cdots$	pri vseh rešitvah

3b. Reparametrizacija

- Preureditev enačb v sistem s polnim rangom

- Črtamo prvo vrstico in prvi stolpec

$$\begin{array}{llllll} 10\mu & +4A_1 & +6A_2 & = & 114 & \mu = 0 \\ 4\mu & +4A_1 & +0 & = & 48 & \Longrightarrow \\ 6\mu & +0 & +6A_2 & = & 66 & \end{array} \quad \begin{array}{lllll} / & / & / & & / \\ / & +4A_1 & +0 & = & 48 \\ / & +0 & +6A_2 & = & 66 \end{array}$$

- Lahko črtamo tudi zadnjo vrstico in zadnji stolpec

$$\begin{array}{llllll} 10\mu & +4A_1 & +6A_2 & = & 114 & \\ 4\mu & +4A_1 & +0 & = & 48 & \Longrightarrow \\ 6\mu & +0 & +6A_2 & = & 66 & A_2 = 0 \end{array} \quad \begin{array}{lllll} 10\mu & +4A_1 & / & & 114 \\ 4\mu & +4A_1 & / & = & 48 \\ / & / & / & & / \end{array}$$

3. Povzetek - rang sistema

- Modeli s polnim rangom - en niz rešitev

- Modeli z nepolnim rangom - mnogo nizov rešitev, ocenljive linearne funkcije iste

- Reparametrizacija - modele z nepolnim rangom pretvorimo v model s polnim rangom z novimi parametri

- Restrikcija - izberemo vrednosti za neocenljive parametre ali si izberemo dodatne enačbe

4. Z ozirom na strukturo podatkov

- Hierarhični modeli
  - samo ugnezdeni vplivi
  - vključeni indeksi "nadrejenih" vplivov
- Križno klasificirani modeli
  - samo križno klasificirani vplivi
  - indeksi različni
- Kombinirani modeli
  - vključujejo križno klasificirane in ugnezdene vplive

4a. Hierarhični modeli

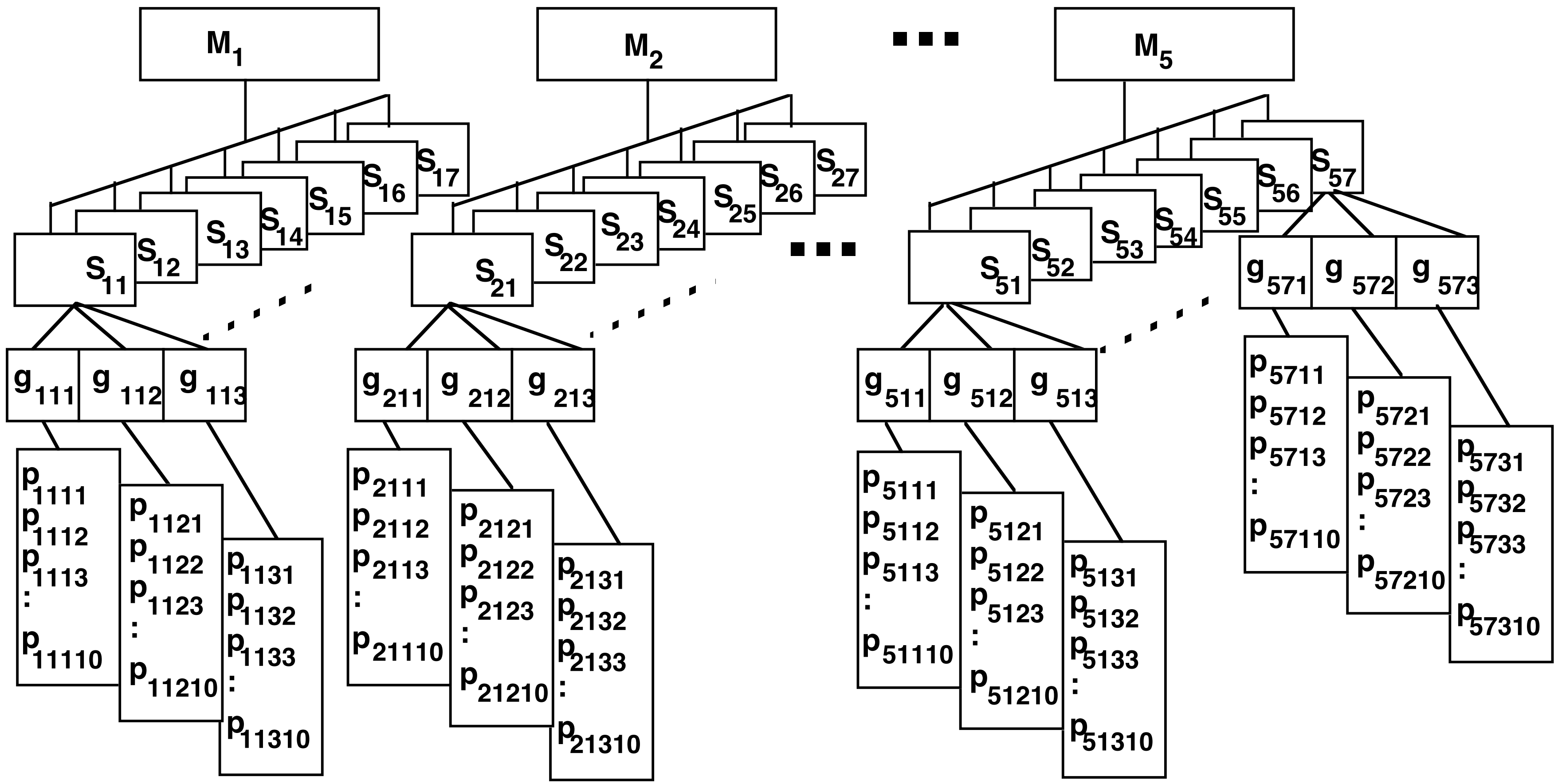
Vpliv	Pasma 1	Pasma 2	Pasma 3
Krma 1	n_{11}		
Krma 2	n_{21}		
Krma 3		n_{32}	
Krma 4		n_{42}	
Krma 5			n_{53}
Krma 6			n_{63}

$$y_{ijk} = \mu + P_i + K_{ij} + e_{ijk} \quad i = 1, 2, 3; \;\; j = 1, 2, \ldots 6; \;\; k = 1, 2, \ldots n_{ij}$$

- Krma (vpliv z več nivoji) je ugnezdena znotraj pasme (vpliva z manj nivoji)
- Nobene krme ne pokladamo dvema pasmama
- Vsaki pasmi pokladamo "specifične" krme

- Iz poskusa nikakor ne moremo razbrati, kako ena od pasem izkorišča krme, pokladane drugima pasmama

4a. Grafični prikaz hierarhičnega modela



$$y_{ijklm} = \mu + M_i + S_{ij} + g_{ijk} + p_{ijkl} + e_{ijklm}$$

$$i = 1, 2, \ldots 5 \qquad j = 1, 2, \ldots 7 \qquad k = 1, 2, 3 \qquad l = 1, 2, \ldots 10$$

- Merjasec je parjen vedno s sedmimi samicami, vsaka samica ima v poskusu tri gnezda, iz vsakega gnezda so v poskus vključili 10 pujskov

- Koliko pujskov je v poskusu? 1050

4b. Križno klasificirani modeli

Vpliv	Pasma 1	Pasma 2	Pasma 3
Krma 1	n_{11}	n_{21}	n_{31}
Krma 2	n_{12}	n_{22}	n_{32}
Krma 3	n_{13}	n_{23}	n_{33}
Krma 4	n_{14}	n_{24}	n_{34}
Krma 5	n_{15}	n_{25}	n_{35}
Krma 6	n_{16}	n_{26}	n_{36}

$$y_{ijk} = \mu + P_i + K_j + e_{ijk} \quad i = 1, 2, 3 \quad j = 1,\ 2,\ \ldots\ 6 \quad k = 1,\ 2,\ \ldots n_{ij}$$

- Pasmam smo krmili vse razpoložljive krme

- Model je lahko enostaven, lahko pa vključuje tudi interakcijo

$$y_{ijk} = \mu + P_i + K_j + PK_{ij} + e_{ijk} i = 1, 2, 3 j = 1,\ 2,\ \ldots\ 6 k = 1,\ 2,\ \ldots n_{ij}$$

4b. Manjkajoče celice pri KK modelu

Vpliv	Pasma 1	Pasma 2	Pasma 3
Krma 1	n_{11}	n_{12}	n_{13}
Krma 2	n_{21}	0	n_{23}
Krma 3	0	n_{32}	n_{33}
Krma 4	n_{41}	n_{42}	n_{43}
Krma 5	n_{51}	n_{52}	0
Krma 6	n_{61}	n_{62}	0

- Modela sta enaka kot v prejšnem primeru
- Vpliva sta križno klasificirana, čeprav niso vse celice zasedene
- Paziti moramo pri postavljanju ocenljivih hipotez
- Presojo značilnosti vplivov delamo z vsoto kvadratov tipa IV

4b. Število ocenljivih parametrov

Farma	Pasma B_1	Pasma B_2	Pasma B_3
A_1	Y_{111}, Y_{112}	Y_{121}, Y_{122}	Y_{131}
A_2	Y_{211}, Y_{212}	Y_{221}, Y_{222}	

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{ijk} \quad i = 1, 2 \quad j = 1, 2, 3 \quad k = 1, \ldots n_{ij}$$

- parametri: μ, A_1, A_2, B_1, B_2, B_3, AB_{11}, AB_{12}, AB_{13}, AB_{21}, AB_{22}
- celica AB_{23} nima podatkov, zato je ne moremo oceniti
- meritev 9, zahtevano 12 parametrov - **prerazkošni model**
- ocenimo lahko samo 5 parametrov (število zasedenih celic)
- povprečje najmanjših celic je vedno ocenljivo, zato sta možna tudi modela

$$y_{ij} = \mu_i + e_{ij} \qquad i = 1, 2, \ldots 5 \qquad j = 1, \ldots n_i$$

$$y_{ijk} = AB_{ij} + e_{ijk} \qquad i = 1, 2 \qquad j = 1, 2, 3 \qquad k = 1, \ldots n_{ij}$$

4c. Kombinirani modeli

- V živinoreji pogosti
- Križno klasificirani in ugnezdeni vplivi
- Preveritev statističnega modela in izpisa
  - statistični paketi izluščijo strukturo iz podatkov sami
  - strukturo moramo poznati zaradi interpretacije
  - stopinje prostosti in pojavljanje vrednosti 0 pri ocenah

5. Z ozirom na naravo parametrov

- Modeli s sistematskimi vplivi (fixed model)

$$y_{ijk} = \mu + P_i + b\,(x_{ijk} - \bar{x}) + e_{ijk}$$

- Modeli z naključnimi vplivi (random model)

$$y_{ijk} = \mu + u_i + v_j + (uv)_{ij} + e_{ijk}$$

- Mešani modeli (mixed model)

$$y_{ijk} = \mu + P_i + b\,(x_{ijk} - \bar{x}) + a_{ij} + e_{ijk}$$

6. Z ozirom na število vplivov

Model		Število parametrov (p)	
$y_{ij} = \mu + \alpha_i + e_{ij}$ enorazsežni	$i = 1, 2, \ldots p_\alpha$ $j = 1, 2, \ldots n_i$	$\underline{1 + p_\alpha}$	
$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$ dvorazsežni brez interakcije	$i = 1, 2, \ldots p_\alpha$ $j = 1, 2, \ldots p_\beta$ $k = 1, 2, \ldots n_{ij}$	$1 + p_\alpha + p_\beta$	
$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$ dvorazsežni z interakcijo	$i = 1, 2, \ldots p_\alpha$ $j = 1, 2, \ldots p_\beta$ $k = 1, 2, \ldots n_{ij}$	$1 + p_\alpha + p_\beta + p_{\alpha\beta}$	(

- Razdelitev ni pomembna pri uporabi statističnih formul

- Omenjene statistične modele bi lahko obdelali tudi ročno

Kompleksni modeli

- Preproste modele srečamo
  - v skrbno načrtovanih poskusih s kontroliranimi pogoji
  - v učnem procesu, za ročno računanje
  - prikazovanje formul v skalarni algebri
- Za podatke iz prireje so modeli praviloma bolj sestavljeni
  - obvezna uporaba statističnih paketov
  - matrična algebra
- Obravnavali smo **linearne modele**

Pogojno linearni modeli

- izhodiščni model ni linearen

- obstaja **transformacija**, da postane model linearen

Inverzni model

$$y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i)^{-1}$$

- Je model linearen?

$$\frac{\partial y_i}{\partial \beta_0} = \frac{\partial \left[(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i)^{-1} \right]}{\partial \beta_0}$$

- Model ni linearen, ker prvi parcialni odvodi vsebijejo parametre

- Transformirajmo odvisno spremenljivko (inverzna vrednost y_i^{-1})

$$y_i^* = \frac{1}{y_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

$$\frac{\partial y_i^*}{\partial \beta_0} = 1 \qquad \frac{\partial y_i^*}{\partial \beta_1} = x_{1i} \qquad \frac{\partial y_i^*}{\partial \beta_2} = x_{2i}$$

- Prvi parcialni odvodi brez parametrov

- Po transformaciji je model linearen

Logit-model z napako v eksponentu

$$y_i = \frac{1}{1 + exp\left\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i\right\}}$$

- e_i predstavlja napako v eksponentu
- Model je nelinearen
- Poskusimo poiskati transformacijo

$$\frac{1}{y_i} - 1 = e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i}$$

$$ln \underbrace{\left(\frac{1}{y_i} - 1\right)}_{logit} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

- Transformacija uspela, model pogojno linearen

Logit model z ostankom izven eksponenta

Če je napaka izven eksponenta, pri transformaciji dobimo njen logaritem.

$$y_i = \frac{1}{1 + exp\left\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right\} * e_i}$$

$$ln \quad \underbrace{\left(\frac{1 - y_i}{y_i}\right)}_{logit} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \qquad \underbrace{ln\, e_i}_{novi\, ostanek}$$

Primer: dnevni prirast rastočih živali

$\bar{x}$	300	500	900
σ	30	50	100

Log-transformacije

$$y_i = \alpha x_i^{\gamma} * e_i$$

Log-transformirani model (**??**) je v našem primeru linearen, saj v prvih odvodih ni več parametrov. Logaritem spremenljivke x_i je le transfomacija neodvisne spremenljivke, ki smo jo ob izvedbi poskusa izmerili. Oceniti želimo parametra α in γ.

$$ln\left(y_i\right) = \alpha + \gamma ln\left(x_i\right) + e_i$$

$$\frac{\partial\left(ln(y_i)\right)}{\partial\alpha} = 1$$

$$\frac{\partial\left(ln(y_i)\right)}{\partial\gamma} = ln\left(x_i\right)$$

Za izbor transformacije poznamo metode, s katerimi lahko izberemo najbolj primerno, vendar se na to področje ne bomo podajali. Logaritemska transformacija je najmočnejša in o njej razmišljamo, kadar so vrednosti posameznih opazovanj zelo različne (najmanj desetkratne). Pri večjih vrednostih je tudi večja napaka oz. ostanek. Pomagamo si lahko tudi z grafom.

(Pogojno) nelinearni modeli

- Prvi parcialni odvodi vsebujejo parametre
- Ni primerne transformacije
- Proizvodne funkcije
  - rast
  - v reprodukcijskem ciklusu
  - v življenjski dobi ...

Rastne krivulje

$$y_i = \sum a_i \left(1 + tanh \left(b_i \left(t - d_1\right)\right)\right)$$

$$y_i = \underbrace{A\left(1 - Be^{-kt_i}\right)^{-1}}_{\eta(\boldsymbol{\beta})} + e_i$$

kjer pomeni:

y_i - opazovanje

t_i - čas, starost $\qquad i = 1, 2, ..., n$

B - masa ob rojstvu

A - odrasla velikost

k - parameter, ki je povezan z ukrivljenostjo

e_i - ostanek

e - ... konstanta

Parametri v modelu: $(A,\, B,\, k) = \boldsymbol{\beta}$

$$\frac{\partial y_i}{\partial A} = \left(1 - Be^{-kt_i}\right)^{-1}$$

$$\frac{\partial y_i}{\partial B} = A\frac{e^{-kt_i}}{\left(1 - Be^{-kt_i}\right)^2}$$

$$\frac{\partial y_i}{\partial k} = AB\frac{e^{-kt_i}}{\left(1 - Be^{-kt_i}\right)^2}$$

Model je nelinearen, ker v prvih odvodih ostajajo parametri. Model ima nekatere parametre, ki so bolj linearni od drugih. Parameter A se pojavlja v odvodih redkeje in v bolj enostavnih izrazih kot drugi parametri. Tako je parameter A lažje oceniti. V odvodih pa sta ostala tudi druga dva parametra B in k.

Nelinearne modele rešujemo iterativno tudi v primeru, če uporabimo metodo najmanjših kvadratov ali metodo največje zanesljivosti.

Pseudo-linearni model (linearna aproksimacija)

$$y_i = \eta\left(\beta_1, \beta_2, \beta_3, ...\right) + e_i$$

Pri teh modelih nelinearno enačbo nadomestimo z linearnim modelom. Model ni čisto pravi, vendar na opazovanem intervalu ne bomo dobili pomembnih odstopanj. Nelinearni krivulji se bomo približali s polinomom ali kakšno drugo funkcijo. Morda bomo pred tem preoblikovali, transformirali neodvisno spremenljivko. Pri tem pa je pomembno, da je model, s katerim poskušamo opisati pravo funkcijo, linearen.

Laktacijske krivulje

Obstaja tudi možnost sestavljanja različnih funkcij na različnih intervalih. Pri rastni krivulji bi na začetku rasti uporabili polinom druge stopnje s pozitivnim regresijskim koeficientom pri kvadratnem členu. V času največje rasti zadostuje linearna regresija. Ko pa se živali približujejo odrasli velikosti, pa hitrost rasti pojenja. Na tem intervalu je ponovno primernejši polinom druge stopnje, regresijski koeficient pri kvadratnem členu pa bo negativen.

Ekvivalentni modeli

- Različice modela, ki enakovredno opisujejo podatke
  - preoblikovanje zaradi interpretacije ali
  - numerične stabilnosti
- Pogoji:
  - isto število ocenljivih parametrov
  - nastopajo isti vplivi
  - črtamo lahko "nadrejene" vplive
- Reparametrizacija

Reparametrizacija - primer

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{ijk}$$

$$y_{ijk} = \mu + A_i + AB_{ij} + e_{ijk}$$

$$y_{ijk} = \mu + A_i + B_{ij} + e_{ijk}$$

$$y_{ijk} = \mu + AB_{ij} + e_{ijk}$$

$$y_{ijk} = AB_{ij} + e_{ijk}$$

$$y_{ijk} = \mu + A_i + B_j + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = \mu + A_i + AB_{ij} + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = \mu + A_i + B_{ij} + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = \mu + AB_{ij} + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = AB_{ij} + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$

$$y_{ijk} = \mu + A_i + B_j + b_i\left(x_{ijk} - \overline{x}\right) + e_{ijk}$$