Poglavje 1

STATISTIČNI MODELI - OSNOVNI POJMI

Pred nami so podatki, ki smo jih zbrali v poskusu, preizkusu ali v pogojih reje. Podatki so lahko številni, ali pa smo uspeli opraviti le nekaj meritev. Manjše število meritev je opravičljivo, kadar so meritve drage, ali pa so živali izpostavljene neugodnim pogojem. Če se le da, poskušamo opraviti zadostno število meritev. Potrebno število meritev lahko predvidimo pred začetkom preizkusa, v kolikor poznamo porazdelitev opazovanj in zanesljivost metode merjenja. Izogibamo se preizkusom z majhnim številom meritev. Le malo je še lastnosti, o katerih prav ničesar ne vemo in smo veseli, če najdemo povprečja in porazdelitve. Tako vedno poskrbimo že pri načrtovanju poskusa za hipoteze. Z njimi postavimo cilje in tako zagotovo vemo, kaj bomo s čim primerjali. Pri postavitvi hipotez gotovo ne bomo zadovoljni z najbolj preprosto, s katero dokažemo, če je srednja vrednost enaka 0 (ničelna hipoteza) ali pa se razlikuje od 0 (alternativna hipoteza). O načrtovanju poizkusov se bomo še pogovarjali, vendar pa se bomo najprej soočili s statističnim modelom in obdelavo podatkov.

1.1 Deterministični in stohastični modeli

V grobem ločimo dve vrsti modelov: deterministične in stohastične. Modele predstavimo z enačbami. ***Deterministični modeli*** (enačba 1.1) natančno določajo odvisno spremenljivko. Ko izberemo neodvisne spremenljivke (x_i), lahko odvisno spremenljivko (y_i) ***izračunamo***. Parametra β_0 in β_1 sta poznani konstanti. Deterministični model lahko ponazorimo z enačbo za izračun dnevnega prirasta (enačba 1.1) .

$$y_i = \beta_0 + \beta_1 * x_i; i = 1,\ 2,\ ...,\ n \qquad [1.1]$$

$$dnevni\ prirast = \frac{prirast}{obdobje} \qquad [1.2]$$

Stohastični modeli (enačba 1.3) vedno vsebujejo naključno (slučajno) spremenljivko - napako (e_i). Zaradi te napake moramo meritve ponavljati in parametre, ki opisujejo odvisno spremenljivko, lahko na koncu samo bolj ali manj zanesljivo ***ocenimo***.

$$y_i = \beta_0 + \beta_1 * x_i + e_i \qquad [1.3]$$

β_0 in β_1 sta v modelu neznanki - ***parametra*** porazdelitve. $\hat{\beta}_0$ in $\hat{\beta}_1$ pa njuni ***oceni***. $\hat{y}_i$ je ocena odvisne spremenljivke pri določeni vrednosti neodvisne spremenljivke x. To je tudi njena pričakovana vrednost meritve pri izbrani vrednosti x. V primeru iz enačbe 1.3 bomo dobili ocene parametrov pri $x = 0$, pogosto pa uporabimo tudi modele, ko izberemo, da je $x = \bar{x}$. Pričakovano vrednost ocenimo oz. izvrednotimo po enačbi (npr. v1.4).

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i + \hat{e}_i \qquad [1.4]$$

Ostanek lahko poimenujemo tudi odklon, napaka ali šum (*ang.* residual). Pri merah razpršenosti smo najprej začeli s povprečnim odklonom in ugotovili, da je njegova vrednost vedno 0. Tako lahko privzamemo, da je pričakovana vrednost za ostanek enaka $\hat{e}_i = 0$.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i \qquad [1.5]$$

POMNI! Parameter ni enak oceni, torej $\beta_0 \neq \hat{\beta}_0$. Prav tako $\beta_1 \neq \hat{\beta}_1$ in $y_i \neq \hat{y}_i$. Parameter predstavlja dejansko, konstantno vrednost, ki je ne moremo izmeriti niti izračunati. Lahko jo le ocenimo ali napovemo. V tem primeru dobimo oceno oziroma napoved, ki bolj ali manj natančno (merilo je ***standardni odklon***) ter bolj ali manj pristransko (merilo je ***pristranost*** - *ang.* bias) predstavlja omenjeni parameter.

Definirajmo še ***oceno napake***: $\hat{e}_i = y_i - \hat{y}_i$. To je razlika med opazovano vrednostjo y_i in njeno pričakovano vrednostjo $\hat{y}_i$.

Statistični model je abstrakcija realnosti in ne more nadomestiti kompletne slike. Z analizo podatkov želimo številne informacije “zgostiti" na predstavljivo raven. Človek obvlada le okrog 100 parametrov, podatkov pa je lahko zabeleženih na tisoče. Analiza nam torej nudi izvleček informacij. Izvlečki (parametri) naj bi zadovoljivo pojasnjevali podatke, vendar pa naj bi jih bilo le toliko, kot je nujno potrebno (***zakon skromnosti*** - *ang.* parsimony). Z odvečnimi parametri izgubljamo učinkovitost statističnega preizkusa in s tem zmanjšamo zanesljivost ter uporabnost analize oziroma zaključkov. Model lahko dostikrat preuredimo tako, da zmanjšamo število parametrov. Npr. razrede pri kvantitativnih sistematskih vplivih lahko pogosto nadomestimo z regresijo. Drugi primer so interakcije, ki odvzamejo sorazmerno veliko stopinj prostosti. Kadar niso pomembne, jih velja izpustiti iz modela.

1.2 Elementi modela

Opisi statističnih modelov so sestavljeni iz štirih delov - elementov:

1. ***Enačbe*** (*ang.* equations)
2. ***Pričakovane vrednosti*** (*ang.* excepted values)
3. ***Strukture varianc in kovarianc*** (*ang.* variance covariance structure, variance covariance matrices)
4. ***Predpostavk*** (*ang.* assumptions) in ***omejitev*** (*ang.* restrictions)

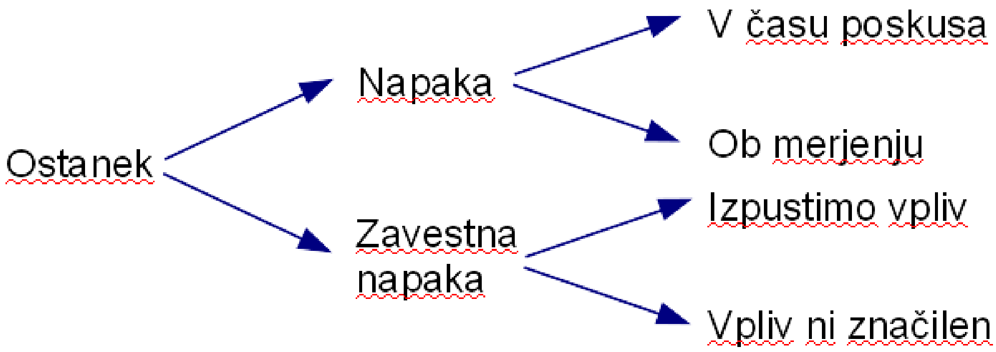
Najprej bomo porabili čas za opis enačbe statističnega modela, kasneje pa bomo obdelali tudi druge elemente modela. Pri običajnih poskusih, ki so načrtovani in temeljijo na naključnih vzorcih, so ostali trije elementi preprosti, z njimi pa se srečamo, kadar obdelujemo proizvodne podatke ali pa se ukvarjamo s selekcijo ali izločanjem živali.

Statistični model opisuje opazovanje, imenovano tudi meritev, podatek ali lastnost. Opazovanje ali meritev je funkcija sistematskih (*ang.* fixed) ter naključnih (*ang.* random) vplivov in ostanka. V statistiki opazovanje ali lastnost poimenujejo tudi odvisna spremenljivka. Tako odvisno spremenljivko pojasnjujejo posamezni vplivi, ki jim zato rečemo tudi pojasnjevalne oziroma neodvisne spremenljivke.

$$\begin{bmatrix} opazovanje, \\ meritev, \\ lastnost\ ali \\ odvisna\ spremenljivka \end{bmatrix} = f \begin{bmatrix} sistematski\ vplivi \\ in\ naključni\ vplivi \end{bmatrix} + ostanek \tag{1.6}$$

$$\begin{bmatrix} observation, \\ measurement, \\ trait, \\ dependent\ variable \end{bmatrix} = f \begin{bmatrix} fixed\ effects \\ random\ effects \end{bmatrix} + residual \tag{1.7}$$

Spoznali smo se že z lastnostmi in vplivi, tudi razdelitev na sistematske in naključne vplive nam ni neznana, zato se bomo ukvarjali s funkcijami vplivov in ostankom.



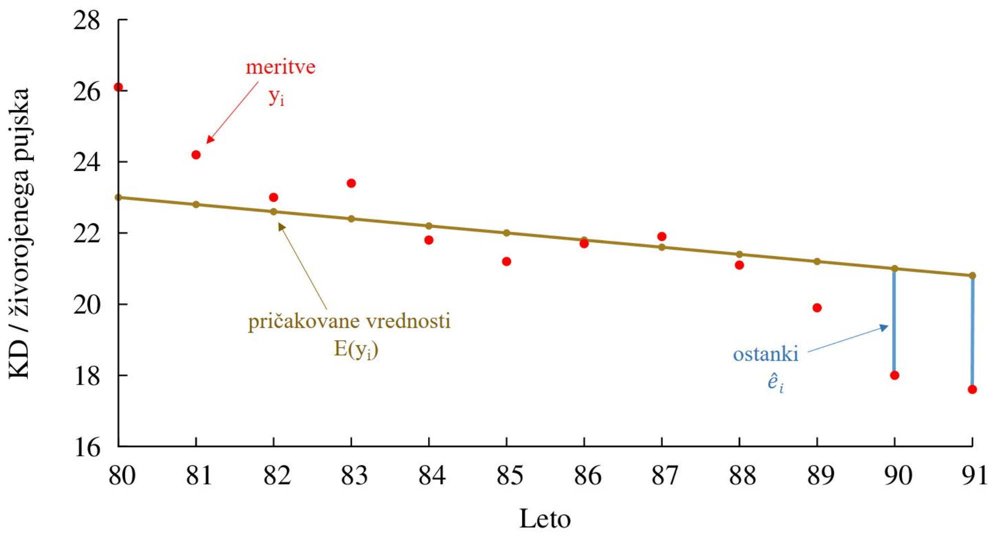
Slika 1.1: Ostanek

Ostanek (slika 1.1) je tudi naključna spremenljivka, a jo bomo zaradi posebnega značaja in pomena vedno obravnavali ločeno od ostalih naključnih spremenljivk. Ostanek se pojavlja kot naključna napaka pri izvedbi poskusa, pri meritvah ali pa je posledica bolj ali manj zavestnih napak. Zavestne napake zagrešimo, ko iz modela izpustimo manj pomembne sistematske vplive. Izpustimo lahko naključna spremenljivko ali pa izpustimo vpliv, ki ga ne poznamo, ker ga npr. v poskusu nismo beležili. Te napake niso nujno katastrofalne, v danem poskusu samo malo ponagajajo, na zaključke morda niti ne vplivajo. Skupaj pa se jih le nabere in je zato ostanek večji ostanek.

Že dober mizar večkrat premeri sobo in kote, ko se loti opremljanja sobe. Meritve ponavlja in se tako preverja, pa čeprav les, beton in kovina na spreminjajo veliko svoje oblike. V živinoreji pa delamo z biološkimi pojavi, ki so znatno bolj občutljivi na posamezne vplive, ker se biološki pojav (rast, plodnost) razvija dalj časa, nanj vpliva več dejavnikov in med seboj sodelujejo ali se ovirajo. Tudi same živali rade ponagajajo pri preizkusih. Tako je npr. bik v poskusu najprej pomagal pojesti krmo sosedoma, potem pa je njemu dodeljena krma ostala. Soseda sta uradno veliko pojedla in malo zrasla, medtem ko bik skoraj ni nič jedel in odlično rasel.

Da pojav dovolj dobro spoznamo, ne verjamemo eni sami meritvi, ampak jih ponovimo več hkrati pod istimi pogoji. Meritve so praviloma med seboj bolj različne kot mizarjevih. Da bi lahko presodili vplive dovolj zanesljivo, meritve torej večkrat ponovimo. Povprečje meritev opravljenih pod istimi pogoji, predstavlja ***primerjalno vrednost***. Primerjalna vrednost se lahko razlikuje, če smo vzorčili iz več podmnožic - če smo opazovali več skupin. Primerjalna vrednost se lahko spreminja od meritve do meritve, kadar imamo opravka s kvantitativnim vplivom. Izraz primerjalna vrednost uporabimo, ko hočemo poudariti uporabnost vrednosti. Meritev primerjamo s primerjalno vrednostjo in vemo, ali je dobra ali slaba. Kadar pa poudarimo vlogo v statistiki, pa primerjalni vrednosti rečemo pričakovana vrednost. Odstopanje meritev od primerjalne (pričakovane) vrednosti imenujemo ostanek.

PRIMER :
Na sliki 1.2 smo prikazali letna povprečja za lastnost število krmnih dni na živorojenega pujska, kar predstavlja lastno ceno pujska. Izbrali smo primer z malo točkami in model z linearno regresijo, da se bodo jasno pokazala odstopanja. Rdeče točke na sliki predstavljajo naše meritve (y_i), premica predstavlja pričakovane vrednosti ($\hat{y}_i$). Tudi točke na premici so pričakovane vrednosti pri določenem letu (spremenljivka na x-osi). Tako za vsako meritev lahko izračunamo svojo pričakovano vrednost. Za leto 1981 smo porabili v slovenskih rejah v povprečju kar 24 KD na živorojenega pujska, pričakovali pa bi nekaj manj kot 12 KD. V tem začetnem obdobju in tudi za zadnji dve leti premica ni najbolje prilegala podatkom.
Z modrimi črtami, ki povezujejo rdeče točke - meritve - s točkami na premici, smo označili ocene ostankov ($\hat{e}_i$). Ocenjene ostanke imamo pri vsaki meritvi, čeprav smo na sliki s črtami prikazali samo dve oceni pri zadnjih meritvah. Kadar sta vrednosti za meritev in pričakovano vrednost enaki, je ocena ostanka enaka 0, kot je to primer za leto 1986. Tudi v tem primeru imamo ostanek, ki je zelo majhen ali enak 0.



Slika 1.2: Meritve, pričakovane vrednosti in ostanki

PRIMER :
V poskusu so študirali vpliv različnih količin gnojila na rast paradižnikov (pregl. 1.1). Merili so velikost osem tednov po presajanju. Velikost paradižnikov je odvisna spremenljivka. Izmerjene vrednosti pa njene realizacije, dogodki. Učinek gnojila naj bi se pokazal v velikosti paradižnikov. Neodvisna spremenljivka bi lahko bila količina gnojila, vendar pa se bomo tokrat raje odločili za razrede: imenovali jih bomo skupine. Ker pričakujemo razlike med skupinami, bomo za vsako skupino izračunali pričakovano vrednost - povprečje. Izračunajmo še vse ostanke in pričakovano vrednost za ostanek!

Ker smo različno gnojili parcele, ne pričakujemo enakega pridelka, ampak predvidevamo razlike med skupinami. Tako v statistični model 1.8 vključimo poleg srednje vrednosti (μ) tudi vpliv skupine (S_i). Ker smo po skupinah imeli posajenih več rastlin, pri opazovanju dodamo indeksu skupine (i) tudi indeks rastline (j).

$$y_{ij} = \mu + S_i + e_{ij} \qquad [1.8]$$

Pričakovana vrednost v posameznih skupinah je različna ($\hat{y}_i$). Ostanek (e_{ij}) ima iste indekse kot opazovanje, saj pri vsaki meritvi lahko naredimo tudi napako pri merjenju. Ostankov ne poznamo, ocenjujemo jih lahko le kot odstopanja meritve od povprečja (pregl. 1.1). Vsota ostankov za vsako skupino posebej je 0.

V primeru, da bi razlik med skupinami ne pričakovali, potem v modelu (1.9) ni drugega kot srednja vrednost (μ). Opazovanja (y_i) imajo samo en indeks, ki označuje rastline v poskusu.

$$y_i = \mu + e_i \qquad [1.9]$$

Če predpostavimo, da vsi paradižnike enako rastejo ne glede na količino gnojila, izvrednotimo samo eno pričakovano vrednost ($\hat{y}$), ki jo predstavlja povprečje višine vseh paradižnikov (86.59 cm). Ocenjeni ostanki so v tem primeru veliko večji. Vsota ostankov po skupinah nima vrednosti 0, ampak sta vsoti v prvih dveh skupinah negativni, v zadnjih dveh pa pozitivni. Tako že na oko vidimo, da ima skupina

Tabela 1.1: Velikost paradižnikov (cm) po skupinah

Skupina 1			Skupina 2			Skupina 3			Skupina 4		
y_{1j}	$\hat{y}_1$	$\hat{e}_{1j}$	y_{2j}	$\hat{y}_2$	$\hat{e}_{2j}$	y_{3j}	$\hat{y}_3$	$\hat{e}_{3j}$	y_{4j}	$\hat{y}_4$	$\hat{e}_{4j}$
74	72	2	76	79	-3	87	90	-3	103	103	0
67	72	-5	80	79	1	91	90	1	99	103	-4
77	72	5	81	79	2	94	90	4	105	103	2
69	72	-3				88	90	-2	106	103	3
73	72	1							102	103	-1

Tabela 1.2: Velikost paradižnikov (cm) brez skupin

Skupina 1			Skupina 2			Skupina 3			Skupina 4		
y_i	$\hat{y}$	$\hat{e}_i$	y_i	$\hat{y}$	$\hat{e}_i$	y_i	$\hat{y}$	$\hat{e}_i$	y_i	$\hat{y}$	$\hat{e}_i$
74	86.59	-12.59	76	86.59	-10.59	87	86.59	0.41	103	86.59	16.41
67	86.59	-19.59	80	86.59	-6.59	91	86.59	4.41	99	86.59	12.41
77	86.59	-9.59	81	86.59	-5.59	94	86.59	7.41	105	86.59	18.41
69	86.59	-17.59				88	86.59	1.41	106	86.59	19.41
73	86.59	-13.59							102	86.59	15.41

(gnojenje) pomembno vlogo pri rasti paradižnikov. Skupna vsota ostankov je enaka 0, odstopanje je le zaradi zaokroževanja vrednosti na dve decimalki.

O statističnih modelih in metodah, kako izberemo, najprimernejšega se bomo še učili. Že sedaj pa lahko ugotovimo, da se statistični model 1.8 bolje prilega podatkom kot model 1.9. Biometrija ali statistika pa nam dajeta orodja, s katerimi dobimo odgovor, ali so razlike med skupinami dovolj velike, da je o tem vredno razpravljati. Različne vrednosti nikakor niso dovolj, da bi trdili, da so razlike dovolj velike.

1.3 Enačba modela

Datoteka s podatki o preizkusu mladic na rast in mesnatost vsebuje 11 zapisov (pregl. 1.3). Izmerjenih je bilo 11 mladic (živali), treh pasem v mesecih januar in februar. Mase pri merjenju so bile med 96 in 105 kg. Slanino so merili z dvema ponovitvama, dnevni prirast pa je izračunan iz podatkov o starosti in masi pri merjenju.

Pri postavljanju statističnega modela si v besedilu, preglednici ali sliki najprej označimo lastnosti in nato še vplive. Sistem označevanja lahko določite po svoje, v študijskem gradivu bomo pri prikazih lastnosti

Tabela 1.3: Podatki o preizkusu mladic na rast in zamaščenost

Žival	Pasma	Mesec	Masa (kg)	Dnevni prirast (g/dan)	Debelina slanine (mm)	
1	SL	JAN	102	540	13	13
2	SL	JAN	98	550	16	14
3	SL	FEB	105	550	16	16
4	SL	FEB	102	580	15	12
5	SVB	JAN	95	520	20	17
6	SVB	FEB	101	500	24	24
7	SVB	FEB	101	490	27	25
8	SML	JAN	97	560	26	27
9	SML	JAN	100	550	22	19
10	SML	FEB	97	600	23	25
11	SML	FEB	102	610	24	22

navedli z rdečim besedilom, vplive pa z zelenim. Predlagamo, da ločite tudi kvalitativne in kvantitativne vplive, npr. ene podčrtate.

V našem poskusu imamo tako dve odvisni spremenljivki - **dnevni prirast** in **debelino slanine**. **Pasma** in **mesec** sta sistematska vpliva z nivoji oz. kvalitativna vpliva. **Masa** je kvantitativni vpliv, pomemben pri debelini slanine, in ga vključimo v model z regresijo. Pričakujemo, da se prašičem, ki rastejo, povečuje tudi slanina. Debelino slanine lahko primerjamo le pri isti masi, kadar pa so odstopanja v masi na dogovorjenem intervalu, pa lahko za vsako žival opravimo korekcijo na skupno maso (npr. 100 kg) in potem primerjamo živali. Ker je dnevni prirast izračunan iz mase in trajanja pitanja, mase ne uporabljamo v modelu za dnevni prirast. Zadnji vpliv lahko predstavlja **žival**, kar predstavlja aditivni genetski vpliv oz. plemensko vrednost živali, ki se prenaša od staršev na potomce. Aditivni genetski vpliv je naključni, vpliv z nivoji, med katerimi poznamo sorodstvo. Nivoji (razredi) niso neodvisni, ker imajo sorodne živali iste gene, ki izvirajo od prednikov.

V modelu so lahko prisotne vse tri skupine vplivov, vendar to sploh ni nujno. Tako imamo modele samo s sistematskimi vplivi z nivoji, samo z regresijo, z ali brez naključnih vplivov ter vse možne kombinacije. V model vključimo tiste vplive, ki jih želimo proučiti, in tiste, ki jih v poskusu nismo mogli kontrolirati in bi nam lahko pri rezultatih nagajali. Na splošno velja, naj bo model preprost, čeprav tega ne smemo doseči za vsako ceno. Kasneje bomo omenili, katere kriterije upoštevamo pri dokončni izbiri statističnega modela.

1.4 Oznake v statističnem modelu

Enačba modela pri obdelavi poskusa opiše zbrane podatke y_{ijk} (leva stran enačbe) z nizom sistematskih (*ang.* fixed) in naključnih (*ang.* random) vplivov na desni. Oznaka y za lastnost oz. odvisno spremenljivko ni izbrana naključno, saj jo na grafikonih prikazujemo na y-osi. Indekse bomo pojasnili kasneje.

Napišimo enačbo za slanino iz preglednice 1.3 in vzemimo, da na debelino slanine vplivajo samo pasma, mesec, masa in žival.

Pasma je sistematski, kvalitativen in bo v modelu kot vpliv z razredi (*ang.* class) oz. nivoji. Nivoje štejemo jih z indeksom i. Tako ima vpliv pasme 3 nivoje in indeks i pa tri vrednosti: $i = 1$, ko je pasma SL, $i = 2$, ko je pasma SVB, in $i = 3$, ko je pasma SML. Vpliv označimo s veliko črko P, ki nas dobro spomni na naziv vpliva. Razredi pri pasmi so označeni z P_1, kar predstavlja še neznano vrednost za pasmo SL, P_2 za pasmo SVB in P_3 za pasmo SML. Kadar želimo v zapisu povedati, da nas zanima ena od pasem, pa jih ne želimo naštevati, to označimo z P_i, pri opisu pa navedemo, katere vrednosti ima indeks i z zapisom $i = 1, 2, 3$.

Mesec je sistematski, kvalitativen in bo v model vključen tudi kot vpliv z razredi. Razrede štejemo jih z indeksom j. Tako ima vpliv meseca 2 nivoja in indeks j pa dve vrednosti: $j = 1$ za mesec JAN in $j = 2$ za mesec FEB. Za vpliv meseca izberemo lahko oznako M, ki ima 2 nivoja in zato bomo iz podatkov poskusili oceniti dve neznanki: M_1 za mesec JAN in M_2za mesec FEB. Kadar želimo v zapisu povedati, da na podatek vpliva eden od mesecev, to ponazorimo z M_j.

Vpliva pasma in mesec imata vsak svoj indeks, ker imamo vsak mesec v preizkusu vse tri pasme; sta križno klasificirana. Indeks i označuje eno od treh pasem, indeks j pa enega od dveh mesecev.

Masa ob zakolu tudi vpliva na debelino hrbtne slanine. Če je prašič težji pri isti starosti, bo imel skoraj gotovo tudi več slanine. Povezava je načrtna, zato je vpliv sistematski, vendar pa je kvantitativen. Če je prašič težji za 5 kg, ima po vsej verjetnosti 5Vse kvantitativne vplive označimo z x (črko iz konca abecede). Izbor črke ni naključen, saj vpliv na grafikonih nanašamo na os x. Indekse pri neodvisni spremenljivki x bomo obrazložili kasneje, saj najprej določimo indekse vsem vplivom z razredi.

Žival je naključni in kvalitativen vpliv. Vsaka žival je razred zase in v podatkih jo predstavlja njena identifikacijska oznaka ali ime. V naših primerih bomo žival velikokrat označevali kar z zaporednimi številkami. Pri označevanju vplivov ne uporabljamo oznak s strešicami. Potem poskusimo najti sinonim ali pa se zatečemo k tujkam. Pri živali lahko uporabimo črko a (*ang. animal*). Ker vpliv živali predstavlja **a**ditivni genetski vpliv, lahko črko a opravičimo tudi s strokovnim izrazom, ki bi ga morali bolj pogosto uporabljati v praksi. Mala črka se uporablja za naključne kvalitativne vplive.
Sedaj pa določimo oznaki za žival še indekse. Posamezna žival je lahko samo ene (i-te) pasme in smo jo merili samo enkrat, torej le v enem (j-tem) mesecu. Tako rečemo, da je žival ugnezdena znotraj pasme in meseca, ter to označimo s tem, da dodamo indeksa i in j obeh nadrejenih vplivov. Vsak j -ti mesec pa smo preizkusili več živali pri vsaki i-ti pasmi, zato smo uporabili tudi dodatni indeks k, s katerim štejemo živali v skupini, ki jo določata izbrana pasma in mesec. Oznaka živali je tako a_{ijk}. Število živali v skupini smo označili z n_{ij}, kar dovoljuje, da je število živali v skupini različno. Do sedaj dobro poznamo indeksa i in j, za indeks k pa zapišemo, da zavzema vrednosti od 1 do n_{ij} v naslednji obliki $k = 1,\ 2,\ \cdots\ n_{ij}$.

Indeksov nismo podelili še meritvam, ostankom in neodvisni spremenljivki x. Začnimo z meritvami y in ostankom e, ki imata vedno iste indekse. Pri vsaki meritvi storimo napako in to tudi takrat, ko se slučajno nismo zmotili. Takrat je napaka enaka nič, a tega nikoli ne vemo, ker napake ne moremo izmeriti. Tako sta meritev in napaka pri tej meritvi vedno v paru. Pri debelini slanine smo opravili dve meritvi na žival. Da jih ločimo, potrebujemo dodaten indeks (l). Tako bomo prvo meritev na živali ijk označili z y_{ijk1} ($l = 1$) in drugo z y_{ijk2} ($l = 2$). Vsakokrat, kadar imamo več kot eno meritev, meritev in ostanek označimo z dodatnim indeksom. Tako bomo v našem primeru pri debelini slanine meritev navajali z y_{ijkl} in ostanek z e_{ijkl}, medtem ko pri dnevnem prirastu dodatnega indeksa l ne potrebujemo, meritev označimo z y_{ijk} in ostanek z e_{ijk}.

Pri neodvisni spremenljivki x postavimo indekse prav na koncu. Pogosto imajo neodvisne spremenljivke iste indekse kot odvisne spremenljivke (lastnosti) in ostanki. Lahko bi rekli, da nikoli nimajo dodatnega indeksa. Pri preizkusu prašiče enkrat stehtamo in debelino hrbtne slanine izmerimo s ponovitvami. Tako imamo dve meritvi (y_{ijk1} in y_{ijk2}) na žival ijk , žival pa je tehtana samo enkrat. Tako povsem zadostuje, da neodvisni spremenljivki x označimo z istimi indeksi, kot smo žival. V tem primeru ima neodvisna spremenljivka (x_{ijk}) en indeks manj kot opazovanje.

$$y_{ijkl} = \mu + P_i + M_j + b\left(x_{ijk} - \bar{x}\right) + a_{ijk} + e_{ijkl} \qquad [1.10]$$

kjer pomeni:

y_{ijkl}	- opazovanje (odvisna spremenljivka)	observation, trait
x_{ijk}	- masa ob zakolu (neodvisna spremenljivka)	independent variable
$\bar{x}$	- konstanta (povprečje neodvisne spremenljivke)	constant (average of x_{ijk})
μ	- srednja vrednost	mean
P_i	- sistematski vpliv pasme; i = 1, 2, 3	fixed effect
M_j	- sistematski vpliv meseca; j = 1, 2	
b	- regresijski koeficient	regression coefficient
a_{ijk}	- naključni vpliv živali; k = 1, 2, ... n_{ij}	random effect
e_{ijkl}	- ostanek, z modelom nepojasnjen; l= 1, 2	residual

Prisotne morajo biti informacije:

1. o številu opazovanj
2. o številu razredov (nivojev) pri vsakem vplivu: npr. i = 1, 2, ... n_i
3. opis, kako so bila opazovanja merjenja
4. kateri vplivi so sistematski in kateri naključni.

Nomenklatura:

- opazovanje: mala črka "y". Če imamo več različnih lastnosti, dodamo kot prvi indeks arabsko številko (npr. $y_{1...}$, $y_{2...}$, $y_{t...}$) ali črko (npr. $y_{S...}$, $y_{D...}$, $y_{t...}$), ki nas spominja na lastnost. Izjemoma uporabimo tudi črko z. Pike, ki sledijo indeksu za lastnost, predstavljajo ostale indekse, ki jih določimo v modelu.
- sistematski vplivi: **velike** ali grške črke. Izbrana črka naj spominja na ime vpliva.
- naključni vplivi: **male** črke; pogosto se uporablja "a" za aditivni genetski vpliv (plemenska vrednost), "u" na splošno
- konstante: pogosto male črke iz začetka abecede
- ostanek: mala črka "e"
- glavni vplivi: **ena** črka z **enim indeksom** (primer: vpliv pasme P_i)
- ugnezdeni vplivi: **ena** črka z **več indeksi** (primer: vpliv živali znotraj pasme a_{ij}). Ugnezdeni vpliv nosi tudi indekse nadrejenih vplivov in dobi tudi svojega. Kadar je vpliv ugnezden, je znotraj nadrejenega vpliva ("gnezda") več nivojev "podrejenega" vpliva ("pujskov"). Niti dva nadrejena nivoja, si ne delita istega podrejenega nivoja. Tako pujsek ne more biti hkrati rojen v dveh gnezdih. Nadrejeni in podrejeni vpliv sta hierarhično klasificirana.
- interakcije: **več** črk z **več** indeksi. Praviloma najprej navedemo nadrejene vplive, potem pa interakcije. Interakcijo poimenujemo tako, da najprej sestavimo črke - oznake vplivom, med katerimi pričakujemo interakcijo, nato pa še indekse. Praviloma se držimo abecednega vrstnega reda indeksov. Če nadrejene vplive izpustimo, lahko obdržimo tudi kombinacijo črk. Tako poudarimo, kako smo do vpliva prišli. Bralcem bo morda tudi bolj jasno, če bomo potem z linearnimi kombinacijami ocen za interakcije poskušali oceniti nadrejene interakcije ali nadrejene glavne vplive.
- Indekse podeljujemo praviloma po abecedi od črke i dalje. Do črk s strešicami praviloma ne pridemo, če pa se slučajno zgodi, pa jih preskočimo.

1.5 Postopek izgradnje modela

Postopek izgradnje modela bomo spremljali na primeru mladic iz preglednice1.3. Na nekaterih mestih bomo za ilustracijo dodali še kakšno predpostavko. Uporabili bomo tudi druge primere, zato bodite pozorni na spremno besedilo.

1.5.1 Glavni vplivi

1.5.1.1 Seznam glavnih vplivov

Sestavimo seznam vseh vplivov, ki bi jih lahko opazovali v modelu. Izberemo vplive, ki jih bomo v poskusu proučevali. To so tako imenovani glavni vplivi. Kadar poskus izvedemo v kontroliranih pogojih, je lista vplivov končana. Pri večini poskusov v živinoreji pa v praksi "nagajajo" še drugi vplivi, na katere nimamo vpliva ali pa bi jih bilo težko kontrolirati. Tako na prirejo vpliva npr. sezona, ker so različni klimatski pogoji, različna prehrana itd. Tako med glavne vplive uvrstimo tudi tiste, ki bi nas v poskusu lahko motili, a jih ne moremo kontrolirati.

Število vplivov pa mora biti v poskusu obvladljivo, zato ne pretiravamo. Pri načrtu poskusa je potrebno poznati možne vplive, seznam katerih pripravimo iz pregleda literature, lahko pa dodamo seveda tudi kakšnega, ki še ni nikogar zanimalo. Razčlenjenost poskusa mora biti taka, da so končni rezultati uporabni.

Označimo si vplive, ki bi jih kazalo v poskusu načrtno spremljati in ostale završemo. To izbiro naredimo, ko poskus načrtujemo! Pri izvedbi poskusa pa je dobro zbirati tudi dodatne informacije, ki bi morda lahko vplivali na poskus. Če se že ne lotimo sistematskega zbiranja dodatnih informacij, pa pripišimo zadeve vsaj takrat, ko se med poskusom kaj nepričakovanega zgodi. Tako lahko žival zboli, pri opazovanju obnašanja pride do nepričakovanega ropota, v poletnem času sta bila dva dneva izredno mrzla in vetrovna, kar bi lahko vplivalo na opazovano lastnost itd.

PRIMER: Dnevni prirast pri mladicah Vplivi, ki smo jih v poskusu beležili, so: pasma, mesec in žival. Sicer na prirast mladic lahko vpliva še veliko drugih vplivov, ki smo jih v poskusu držali konstantne ali pa jih nismo uspeli zabeležiti. Prirast bomo izvrednotili samo od rojstva do odbire, ne pa tudi na drugih intervalih rasti. Ker smo podatke zbrali v preizkusu mladic, bomo spremljali samo ženske živali, ne pa tudi kastratov ali merjascev. Ker nismo pričakovali, da bi rejec lahko vplival, ga nismo zapisali. V tem poskusu tudi nismo predvideli skupnega okolja v gnezdu, ki ima po dosedanjih izkušnjah pomemben vpliv. Kasneje bomo skupno okolje v gnezdu in skupno okolje čredi tudi vključevali, v prvem koraku pa se bomo zadovoljili s preprostejšim modelom. Nekateri vplivi pa so združeni: prehrana se nekoliko spreminja s časom, ker se spreminjajo sestavine. Vpliv prehrane je torej močno povezan s časom - sezono in vpliva nista ločljiva. Tako bomo skupen vpliv še naprej združevali in skupno imenovali sezonski vpliv. Pri prašičih smo za sezono uporabili mesec. Da bi poudarili, da smo sezone naredili po mesecih, smo vpliv poimenovali "mesec".

Posebno mesto namenjamo neodvisnim spremenljivkam. Za opis teh vplivov je največkrat primerna regresija, ki jo ponazorimo s funkcijo, kriterije za izbor funkcije bomo obravnavali kasneje v poglavju o regresijskih modelih (2.3.1). Vpliv imenujemo po neodvisni spremenljivki x, predstavljajo pa ga neznani parametri - regresijski koeficienti b ali β, s katerimi opišemo vpliv neodvisne (x_{ijk}) na odvisno spremenljivko (y_{ijkl}).

1.5.1.2 Označimo glavne vplive

enolično, kot smo se dogovorili: sistematske z veliko črko, naključne z malo črko. Kot glavni vpliv razumemo vpliv, katerega razredi so identificirani z enim samim indeksom. Indeksi so nam pri gradnji modela dobrodošli pripomoček, zato jih velja pripisati. V tem koraku niso nujni, vendar se bomo odločili in jih navajali, da se jih navadimo in razumemo njihov pomen.

PRIMER: Nadaljujmo s primerom dnevnega prirasta pri mladicah P_i - vpliv pasme; i = 1, 2, 3 (1=SL, 2=SVB, 3=SML)
M_j - vpliv meseca oziroma sezone; j = 1, 2 (1=JAN, 2=FEB)
a_{ijk} - vpliv živali , aditivni genetski vpliv oziroma plemenska vrednost živali; k = 1, 2, ... n_{ij}

Dogovorili smo se, da glavne vplive označimo z eno črko, praviloma začetnico naziva. Pri sistematskih vplivih uporabimo veliko črko, pri naključnih pa malo črko. V omenjenem poskusu smo izmerili samo 11 živali, zato bi lahko bili celo v dilemi, kako obravnavati žival. Živali je v poskusu malo, odločujoče pa je dejstvo, da imamo po živali eno samo meritev. V resnici je poskus premajhen, da bi dal kakorkoli uporabne rezultate. Uporabljamo ga samo za ponazoritev postopka, pri tem bomo razmišljali, kot bi bilo v poskusu več 10000 živali.

Vpliv pasme in meseca sta križno klasificirana, kar označimo tako, da navedemo različna indeksa. Živali vseh treh pasem smo merili v obeh mesecih. Žival je lahko samo ene pasme. Ker je merjena samo enkrat, je bila meritev lahko opravljena samo v januarju ali februarju, nikakor pa ne v dveh mesecih hkrati. To ponazorimo tako, da navedemo indeks pasme in indeks meseca. Vsak mesec smo pri vsaki pasmi zagotovo izmerili več živali, saj smo že v našem vzorcu to nakazali. Tako potrebujemo dodatni indeks k, ki bo preštel živali (pravzaprav meritve) i-te pasme izmerjene v j-tem mesecu. Število živali bo v primeru iz prakse različno.

Ne pozabite! Vplive moramo opisati. K opisu pa sodi tudi opis indeksa. Če imamo manjše število nivojev, te nivoje naštejemo, pri daljšem nizu napišemo dva, nadaljujemo s tremi pikicami in končamo

Tabela 1.4: Križno klasificirana vpliva pasme in meseca

	M_1	M_2	$\sum P$
P_1	n_{11}	n_{12}	$n_{1.}$
P_2	n_{21}	n_{22}	$n_{2.}$
P_3	n_{31}	n_{32}	$n_{3.}$
$\sum M$	$n_{.1}$	$n_{.2}$	$n_{..}$

Tabela 1.5: Vpliv meseca ugnezden znotraj pasme

	M_1	M_2	M_3	M_4	M_5	M_6	$\sum P$
P_1	n_{11}	n_{12}	0	0	0	0	$n_{1.}$
P_2	0	0	n_{23}	n_{24}	0	0	$n_{2.}$
P_3	0	0	0	0	n_{35}	n_{36}	$n_{3.}$
$\sum M$	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{.5}$	$n_{.6}$	$n_{..}$

z največjo vrednostjo . Ta vrednost je lahko eksplicitno navedena, če je vrednost znana. V primeru, da se končni vrednosti želimo izogniti, navedemo oznako. Običajno, kadar ni dvoumno je to n ali m, ker nas to spomni na število. Zlasti kadar n uporabimo v indeksu, pa iščemo druge rešitve: p in q sta tudi sorazmeroma pogosto uporabljena, ker jih običajno uporabljamo za število parametrov. Kadar rabimo, lahko v indeksu navedemo tudi vpliv kot npr. n_P za število pasem, n_a za število živali. V našem primeru smo uporabili n_{ij}, kar pomeni, da pričakujemo različno število živali i-te pasme v j-tem mesecu. Pazite, da so opisi nazorni in nedvoumni!

1.5.1.3 Določimo naravo vplivov

O naravi vplivov odloča struktura podatkov in jo najbolje ponazorimo s postavitvijo indeksov. Če to storimo že ob poimenovanju, pri tem koraku še enkrat preverimo, da je delo korektno opravljeno. Da pa na to ne bi pozabili, smo dali še poseben korak.

Tako ločimo križno klasificirane (pregl. 1.4) in ugnezdene vplive (pregl. 1.5). Križno klasificirani modeli so zaželeni, ker opisujejo splošna pravila in omogočajo prepoznavanje tudi posebnosti (specifičnih vplivov). Da sta dva (ali več) vpliva križno klasificirana (pregl. 1.4), moramo imeti meritve pri vseh možnih kombinacijah teh dveh (ali več) vplivov. Po drugi strani so ugnezdeni vplivi (pregl. 1.5) tisti, pri katerih skupina nivojev enega vpliva pripada drugemu vplivu in sicer samo enemu nivoju tega nadrejenega vpliva. Nivoji ugnezdenega vpliva so hierarhično razporejeni znotraj nadrejenega vpliva. Tako najprej navedemo nadrejeni in nato ugnezdeni vpliv. V našem primeru iz preglednice smo vsako pasmo preizkusili v dveh zaporednih, a različnih mesecih. Pri prvi pasmi smo preizkus opravili v prvih dveh mesecih (zimskih), pri drugi pasmi začetek pomladi, pri tretji pa konec pomladi in začetek poletja. Razlike med meseci so za večino proizvodnih lastnosti med prašiči tolikšne, da primerjava razlik med pasmami in meseci ni mogoča.

Prikazujemo pa še ponesrečen načrt poskusa (pregl. 1.6). Če bi hoteli uvrstiti ta načrt glede na naravo vplivov, bi bili vplivi križno klasificirani, vendar pa je obdelava takih podatkov zelo problematična zaradi praznih celic. Prazne celice so tisti meseci pri posamezni pasmi (oznaka PM_{ij}), ko nismo opravili nobene meritve. Načeloma bi zadoščala samo ena meritev, čeprav sedaj že dobro vemo, da bi bilo povprečje slabo ocenjeno. V tem primeru bi verjetno najbolje storili, če bi tri celice (PM_{13}, PM_{26} in PM_{31}) izločili ali pa jih obdelali v ločenih obdelavah s podatki iz tistih celic, kjer so primerjave mogoče. Pri poskusu smo opravili več dela in morda več analiz, kot bomo uspeli izluščiti iz poskusa.Čeprav si takih podatkov ne želimo, pa se nam včasih le zgodi, da moramo pristati na take podatke. Kadar rejec ne redi vseh pasem ali genotipov, ki bi jih radi opazovali, nimamo dosti izbire. Tudi pri nabiranju vzorcev v trgovinah ne moremo kupiti iste izdelke vseh prodajalcev v vseh trgovinah. V večjih trgovinah prodajajo izdelke številnih proizvajalcev, manjše pa se z izdelkom oskrbujejo pri enem proizvajalcu. Izberejo lahko celo

Tabela 1.6: Ponesrečen načrt poskusa

	M_1	M_2	M_3	M_4	M_5	M_6	$\sum P$
P_1	n_{11}	n_{12}	n_{13}	0	0	0	$n_{1.}$
P_2	0	0	n_{23}	n_{24}	0	n_{26}	$n_{2.}$
P_3	n_{31}	0	0	0	n_{35}	n_{36}	$n_{3.}$
$\sum M$	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{.5}$	$n_{.6}$	$n_{..}$

manjšega proizvajalca, ki zagotavlja posebno kakovost in ne sodeluje z večjimi trgovci.Tako je nemogoče priti do optimalnega načrta poskusa.

PRIMER: **Nadaljujmo primer za dnevni prirast pri mladicah** Križno klasificirane in ugnezdene vplive smo že določili. Tako sta pasma in mesec križno klasificirana vpliva, žival pa je ugnezdena znotraj pasme in tudi meseca, ker so bile mladice samo enkrat merjene.

PRIMER**: Vpliv leta in pasme pri molznicah** Ponazorimo še en križno klasificirani model (enačba 1.11). Tako npr. opažamo, da se količina mleka na kravo v standardni laktaciji iz leta v leto polagoma povečuje pri vseh pasmah. V modelu lahko imamo dva glavna vpliva: leto in pasmo, ki sta križno klasificirana. V vseh letih smo opazovali vse pasme krav in lahko primerjamo med seboj leta ali pasme. Tako dobimo splošne razlike med pasmami in med leti. Toda v sušnem letu, ko je bila pridelana krma slabše kakovosti in še ta v manjših količinah, pa je bila količina mleka manjša kot običajno. Izboljševanje iz leta v leto pa je splošni trend in bi ga lahko opisali s krivuljo. Kadar pa lahko pride do slabe letine, bomo raje uporabili vplive z nivoji, čeprav je tudi to splošni trend, saj je poslabšanje opazno pri vseh pasmah. Če pa je bil padec večji pri pasmi z večjo povprečno količino mleka, pa je to specifični vpliv pasme. V mislih smo imeli krave črno-bele pasme. Vedeti tudi moramo, da so se najbrž rejci črno-bele pasme drugače (specifično) obnašali: svojim kravam so, čeprav je bila krma draga, kupili vse, kar so potrebovale. Tako v resnici sploh niso občutile slabe letine na svoji koži in so obržale običajno količino mleka. Tudi to je specifični odziv pri črno-beli pasmi krav. V obeh primerih so reagirale črno-bele krave drugače, kot smo "na splošno pričakovali". V tem primeru imamo prisotno interakcijo (PL_{ij}). Interakcija ni razvidna iz samega poskusa. Kasneje bomo omenjali postopek, kako jih določimo in preverimo, če so pomembne.

$$y_{ijk} = \mu + P_i + L_j + PL_{ij} + e_{ijk} \qquad [1.11]$$

1.5.1.4 Določimo ugnezdene vplive

Ugnezdeni vplivi so vplivi, katerih skupina nivojev tega vpliva pripada enemu nivoju nadrejenega vpliva. Ugnezdeni vplivi niso v celoti identificirani, dokler niso opisani glavni vplivi (ali kombinacije vplivov), znotraj katerega so ugnezdeni. Tako ugnezdene vplive običajno navajamo v modelu za nadrejenimi vplivi. V seznamu jim dodamo oznako za "nadrejene" vplive (npr. DA: D je ugnezden znotraj A) ali pa pripišemo indekse (D_{ij}). Več o ugnezdenih vplivih in primere si lahko preberete v poglavju o hierarhičnih modelih (2.4.1).

1.5.1.5 Določitev možnih interakcije

Glavne vplive razvrstimo v stolpec (pregl. 1.7), pri regresiji navedemo regresijske koeficiente. Prvi vpliv izpustimo in za ostalimi vplivi napišemo dvopičje ter pripišemo prvi vpliv na desni strani dvopičja od vključno druge vrstice dalje. Interakcija je možna, če se črki v nazivu vpliva ali v indeksu na desni in levi strani ne ponovita. Poimenujemo jo tako, da sestavimo najprej črke sodelujočih vplivov po abecednem zaporedju indeksov. Indekse navedemo skupaj, za nizom črk. Kjer se na levi in desni strani

Tabela 1.7: Določanje možnih interakcij za dnevni prirast pri mladicah (korak 1)

Vpliv	Interakcija	Novi vpliv
P_i		
$M_j \quad : \quad P_i$	možna	PM_{ij}
$a_{ijk} \quad : \quad P_i$	ni možna	

Tabela 1.8: Določanje možnih interakcij za dnevni prirast pri mladicah (korak 2)

Vpliv	Interakcija	Novi vpliv
P_i		
$M_j \quad : \quad P_i$		
$a_{ijk} \quad : \quad P_i \quad : \quad M_j$	ni možna	
$PM_{ij} \quad : \quad M_j$	ni možna	

dvopičja pojavita bodisi ista črka za vpliv bodisi isti indeks, interakcija ni možna. Le pri regresijskih koeficientih je drugače. Vse regresijske koeficiente označujemo z isto črko b, zato upoštevamo predvsem indekse. Zadnja dva stolpca nista zaželena, ker pri celotnem postopku zmanjšata preglednost. Prvega lahko nadomestimo s tem, da tam, kjer je možna interakcija, pri desnem stolpcu vpliv označimo (npr. obkrožimo), drugače pa pustimo nespremenjeno ali prečrtamo. Zadnji pa tudi ni potreben, ker bomo nastalo interakcijo dopisali spodaj.

Možne interakcije dopišemo na konec liste (pregl. 1.8), se pomaknemo v tretjo vrstico, na koncu pripišemo podpičje in drugi vpliv in nato storimo isto pri vseh naslednjih vplivih, tudi pri tistih, ki so bili pravkar dodani. Lepo podpisujemo, da ne bi česa pozabili ali preskočili. Ko bomo določali možne interakcije, gledamo prvi stolpec na levi in zadnji stolpec na desni. Kriteriji in postopek so isti kot pri prvem koraku.

V drugem koraku ni bilo nobene nove možne interakcije, zato se lista možnih vplivov v preglednici 1.7 ni nič podaljšala. Postopek moramo nadaljevati z vplivom a_{ijk}. Ostala nam je samo ena kombinacija, ki tudi ni možna, ker na obeh straneh najdemo kar dva ista indeksa i in j.

V našem primeru je postopek končan. Praviloma bi nadaljevali z interakcijo PM_{ij}, vendar se lista pri tem vplivu konča. Torej nismo končali, ker smo se znašli na koncu osnovne liste glavnih vplivov. Če bi sledile dodatne interakcije, bi postopek nadaljevali.

1.5.1.6 Označitev sistematskih ali naključnih vplivov

Glavni vplivi so praviloma že določeni. Pri razvoju modela smo morda spoznali, da smo naredili napako. M To pač popravimo in postopek ponovimo. Interakcije prevzamejo naravo (značaj ali tip) vplivov. Tako so interakcije med samimi sistematskimi vplivi so sistematske. Če pa v interakciji nastopa vsaj eden naključni vpliv, pa bo interakcija naključna. Označimo ga lahko s kombinacijo velikih in malih črk (aT_{ijk}) ali samo z malimi črkami (at_{ijk}), da poudarimo naravo parametra. Tako kot za ostale naključne vplive, tudi za take interakcije potrebujemo znane variance ali pa jih računamo.

Tabela 1.9: Določanje možnih interakcij za dnevni prirast pri mladicah (korak 3)

Vpliv		nov vpliv
P_i		
$M_j \quad : \quad P_i$		
$a_{ijk} \quad : \quad P_i \quad : \quad M_j$		
$PM_{ij} \quad : \quad M_j \quad : \quad a_{ijk}$	ni možna	

PRIMER: **Nadaljujemo primer za dnevni prirast pri mladicah** Interakcija PM_{ij} je sistematska, ker sta oba sodelujoča vpliva sistematska.

7. Izločitev podvojenih vplivov Podvojeni vplivi se nam lahko pojavijo iz različnih vzrokov. Eden od vzrokov bi lahko bil pri ugnezdenih vplivih, kjer ima vsak ali samo kateri od nivojev pri nadrejenem vplivu le po en nivo ugnezdenega vpliva. Za primer omenimo situacijo, ko ima vsak rejec samo eno pasmo. Ko naredimo poskus samo na eni farmi ali na eni pasmi, tudi ne moremo vključiti vpliv farme oziroma pasme. Na ta način je podvojena srednja vrednost μ, ker vpliv z enim nivojem predstavlja populacijo.

Do odvečnih interakcij bi lahko prišlo, če se ne bi držali abecednega zaporedja indeksov. Drugačen vrstni red sicer ni napaka, je pa težje preveriti, če imamo podvojene vplive. Podvojen je vpliv, če se nahajajo iste črke v nazivu ali indeksih. Trojno interakcijo med vplivi A_j, B_i in C_k lahko poimenujemo ABC_{jik} ali BAC_{ijk}. Vplive bi lahko razvrstili tudi drugače, skupno je možnih šest poimenovanj. Vse te oznake pa predstavljajo isti vpliv: interakcijo pravilno poimenovano BAC_{ijk}. Če se držimo abecednega vrstnega reda indeksov, se bomo že pri določanju možnih interakcij izognili podvojitvam.

Izločimo tudi vplive brez zadostnih dodatnih informacij! Parametri pri sistematskih vplivih ne smejo imeti popolnoma iste indekse kot ostanek. To bi pomenilo, da smo za posamezni nivo pri takem sistematskem vplivu opravili samo eno meritev, kar pa je znatno premalo, da bi ga zadovoljivo ocenili. Podvojitev indeksov je možna pri naključnih vplivih. Za napoved je zadosti ena sama meritev za posamezni nivo (enoto, žival), dodatne informacije lahko dobimo še od koreliranih nivojev (sorodnih živali) ali koreliranih lastnosti, vir informacij pa so tudi znane variance in kovariance. Omenili smo pa že, da lahko naključni vpliv napovemo tudi, ko nimamo nobenih informacij. Se pač odločimo, da ne odstopa od pričakovane vrednosti (npr. plemenska vrednost je 0). Nekoliko bolj pozorni moramo biti na strukturo podatkov, ko ocenjujemo komponente (ko)varianc. Imeti moramo ustrezno količino informacij za posamezno komponento, vendar pa to presega okvir našega predmeta.

POMNI! Viri informacij so lahko različni. Najbolje je, da meritev izmerimo na živali sami, vendar pa to ni vedno mogoče. Če opravljamo preizkus na mlečnost pri govedu ali velikost gnezda pri prašičih, podatkov pri moških živalih ni mogoče izmeriti. Da pa bi napovedali plemensko vrednost pri moških živalih, uporabimo meritve na sorodnicah. To pa je že drugi vir informacij, ki ga povezujemo s poreklom. Tretji vir pa so korelirane lastnosti. Korelacija je lahko pozitivna ali negativna. Pri izbiri koreliranih lastnosti pa pazimo na to, da je korelacija med lastnostma visoka, pri merjeni lastnosti pa mora biti visoka tudi heritabiliteta (dednostni delež). Kot primer naj predstavimo meritve debeline podkožnega maščobnega tkiva pri prašičih. Če smo hudo natančni, nas maščobno podkožno tkivo sploh ne zanima, izboljšali bi radi samo mesnatost. Ker pa smo prepričani in vsi poskusi potrjujejo, da sta mesnatost in debelina podkožnega maščobnega tkiva tesno povezani in je merjenje podkožnega maščobnega tkiva priročnejše, pač merimo podkožno maščobno tkivo. Vir informacij so tudi znane variance in kovariance ali razmerja med njimi. Če vemo, da predstavljata naključni del modela le žival in ostanek in sta njuni varianci v razmerju 0.4 : 0.6, potem lahko "ostanek", ki ostane po odstranitvi sistematskih vplivov, razdelimo na dve komponenti. Brez razmerja pa delitev ne bi bila možna.

PRIMER: **Nadaljujmo primer za dnevni prirast pri mladicah** V tem primeru ni podvojenih vplivov, pri živali in ostanku pa imamo iste indekse, ker ima žival samo eno meritev. Da bi plemensko vrednost živali lahko ločili, potrebujemo še najmanj en vir informacij. Zadostuje že razmerje med aditivno genetsko varianco (σ_a^2) in varianco za ostanek (σ_e^2), zadovoljni pa smo lahko tudi s poreklom.

Pri regresiji moramo paziti, da modelov ne okrnimo. Linearno premico določata dva parametra: regresijski koeficient b in presečišče z ordinato, osjo y. Pri enostavni regresiji ni nevarnosti, medtem ko bi se pri ugnezdeni regresiji kaj hitro lahko pri statistični presoji črtali nadrejeni vpliv ali interakcijo. Vzemimo nekaj pravilnih kombinacij. V prvi enačbi (1.12) imamo regresijo b_j ugnezdno znotraj meseca M_j. Prisotnost vpliva M_j v modelu zagotavlja, da dobimo presečišče z osjo y za vsako premico posebej (slika 1.3).

$$\cdots + M_j + b_j \left(x_{ijklm} - 100\right) + \cdots \qquad [1.12]$$

Tudi v drugi enačbi (1.13) je regresija ugnezdena, tokrat znotraj interakcije PM_{ij}. Prav tako je prisotnost interakcije potrebna, dokler nismo prepričani, če lahko regresijo poenostavimo.

$$\cdots + PM_{ij} + b_{ij} \left(x_{ijklm} - 100\right) + \cdots \qquad [1.13]$$

V slednji enačbi (1.14) imamo kvadratno regresijo z dvema členoma, prav tako kot premica, tudi parabola potrebuje presečišče z osjo y. V našem primeru to zagotavlja prisotnost interakcije. Vsi trije členi imajo pri vseh treh členih iste indekse, vendar niso obravnavani kot podvojeni vplivi. Prvi člen zagotavlja presečišča z osjo y, drugi predstavlja naklon in tretji je povezan z ukrivljenostjo parabole.

$$\cdots + PM_{ij} + b_{Iij} \left(x_{ijklm} - 100\right) + b_{IIij} \left(x_{ijklm} - 100\right)^2 + \cdots \qquad [1.14]$$

Sedaj pa še nekaj nezaželenih kombinacij. V naslednjem modelu (1.15) smo črtali mesec kot glavni vpliv. Vse premice so vodene skozi eno točko na osi y. Premicam smo tako vsilili potek skozi eno točko (slika 1.4).

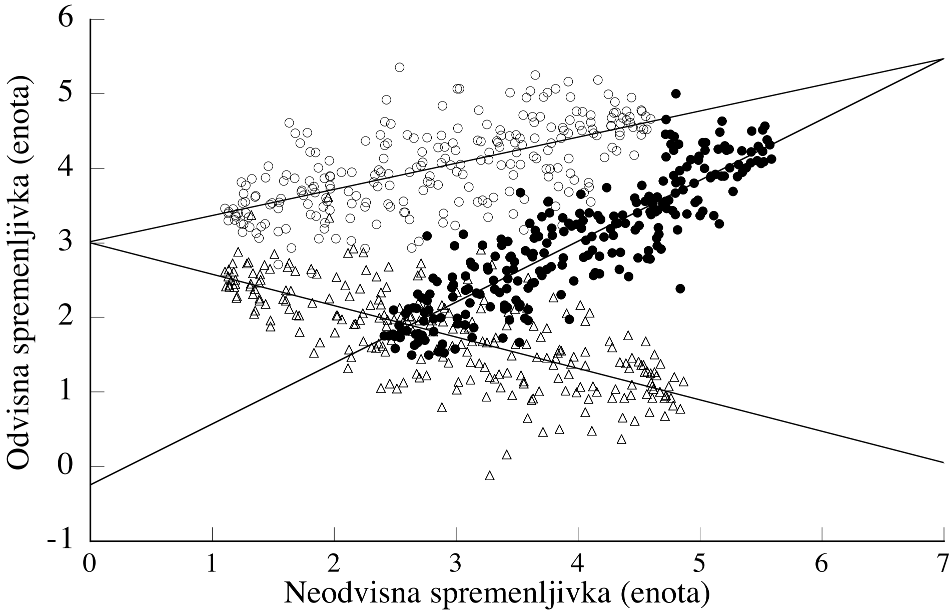
$$\cdots + \cancel{M} + b_j \left(x_{ijklm} - 100\right) + \cdots \qquad [1.15]$$

V naslednji enačbi (1.16) imamo regresijski koeficient za parabolo ugnezden samo znotraj meseca, linearni člen pa znotraj interakcije. Kot vemo, mora imeti vsaka parabola tri člene. Če mora biti ugnezdena znotraj interakcije, moramo to storiti tudi za kvadratni člen. Ko pa je zadostno ugnezdenje znotraj samo enega vpliva, pa to naredimo tudi pri linearnem členu. Tudi v tem primeru pa lahko v modelu ostane interakcija. Tudi tretji primer (1.17) je napačen, komentar pa je zelo podoben, kot v drugem napačnem primeru. Poskusite ga sami opisati.

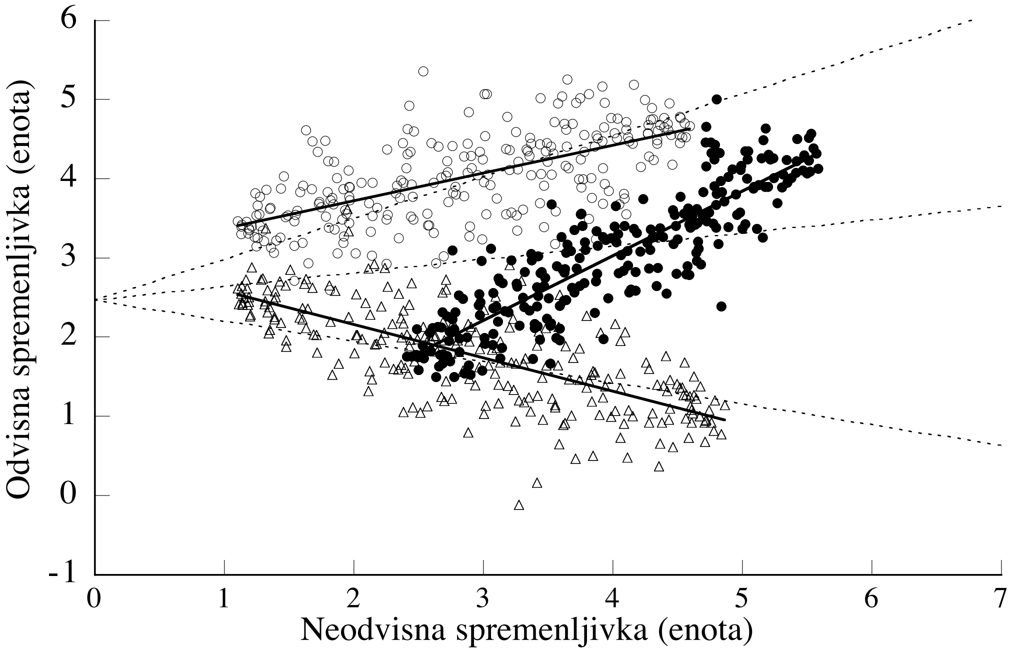
$$\cdots + PM_{ij} + b_{Iij} \left(x_{ijklm} - 100\right) + b_{IIj} \left(x_{ijklm} - 100\right)^2 + \cdots \qquad [1.16]$$

$$\cdots + M_j + b_{Iij} \left(x_{ijklm} - 100\right) + b_{IIj} \left(x_{ijklm} - 100\right)^2 + \cdots \qquad [1.17]$$

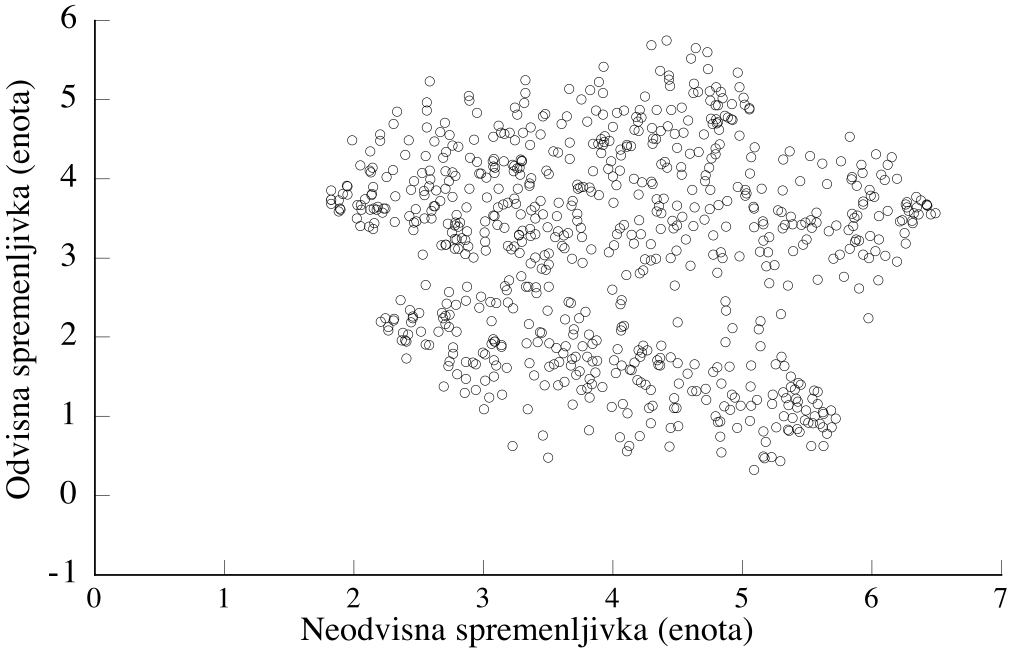
PRIMER: Oglejmo si primer še na sliki (1.3), kjer imamo tri nivoje nekega vpliva, npr. meseca M. Kot običajno je neodvisna spremenljivka oziroma vliv na osi x in odvisna spremenljivka na osi y. Podatke pri prvem nivoju smo prikazali z belimi krogci, drugega s črnimi in tretjega s trikotniki. Pri prvih dveh je regresijski koeficient pozitiven in pri zadnjem pa negativen. Očitno je, da moramo oceniti tri regresijske koeficiente. Če poiščemo presečišča regresijskih premic (polna črta) z osjo y, boste ugotovili, da so različna. Regresija je torej ugnezdena znotraj nekega glavnega vpliva, npr. meseca. Model bo vseboval člena iz enačbe 1.13.
Sedaj pa še poglejmo, kaj se zgodi, če iz modela črtamo "nadrejeni" vpliv meseca (1.4). Ponovno dobimo tri regresijske premice, ki izhajajo iz ene točke na osi y kot šop. Na naši sliki so prikazane črtkano. Kot vidimo iz slike, šop regresijskih premic ne opisuje dobro nobenega oblaka podatkov. Pri oblaku s črnimi krogci še posebej močno odstopa od prave vrednosti. Strokovno pravimo, da se premice ne prilegajo podatkom, zato podatkov ne opisujejo dobro.



Slika 1.3: Opazovanja pripadajo trem nivojem glavnega vpliva M v primeru 1.5.1.6



Slika 1.4: Premicam vsiljen potek skozi skupno presečišče na osi y v primeru 1.5.1.6



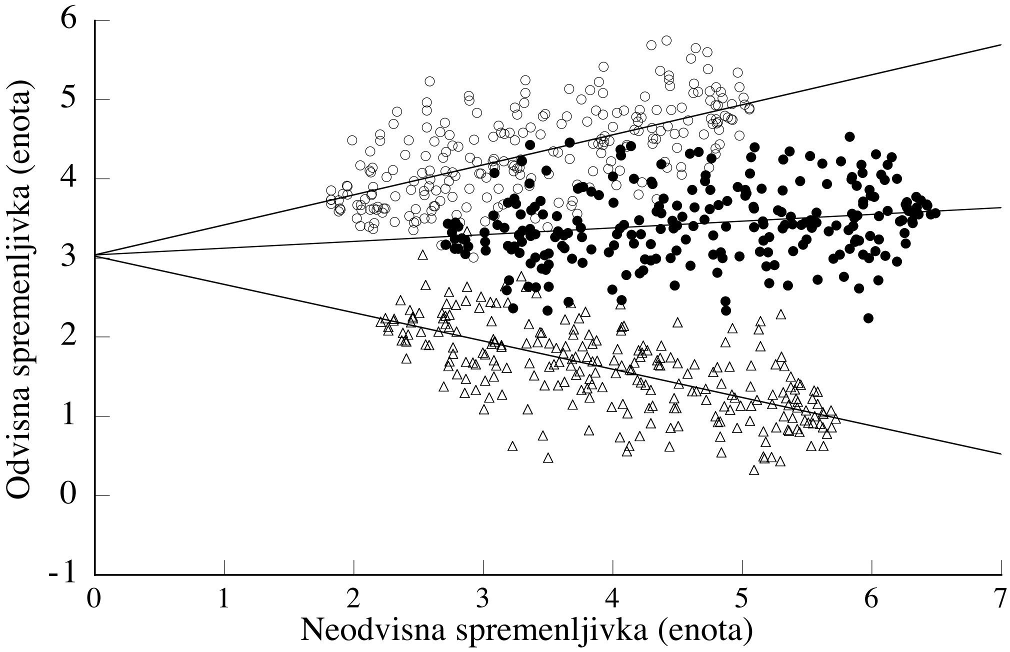
Slika 1.5: Vpliv neodvisne na odvisno spremenljivko v primeru 1.5.1.6

POMNI! Pri polinomih vedno ohranimo vse člene: presečišče z y-osjo, in vse člene do najvišje stopnje. Kadar je regresija ugnezdena, dobijo vsi regresijski koeficienti indekse, od "nadrejenega" vpliva.

PRIMER: Vzemimo podatke iz neke druge simulacije. Podatke smo nanesli na sliko 1.5. Ponovno imamo neodvisno spremenljivko x_i in odvisno spremenljivko y_i. Poskusite drug drugemu opisati, kako bi uredili model, da bi opisal kar najbolje situacijo na sliki!
V resnici so zopet predvideni trije snopi točk. Če ste dovolj pozorni, lahko te snopiče tudi prepoznate. Da pa bi jih bolj nazorno videli, smo jih ponazorili še s različnimi znaki na sliki 1.6. V tem primeru skupno presečišče z osjo y zadovoljivo odgovarja in lahko izjemoma po statistični obdelavi nadrejeni vpliv S izpustimo brez škode. Modeli 1.18 bi v tem primeru lahko bil poenostavljen, vendar tudi v takem primeru, če je le dovolj podatkov, obdržimo nadrejeni vpliv v modelu.

$$y_{ij} = \mu + S_i + b_i x_{ij} + e_{ij} \quad [1.18]$$

8. Napišemo osnovni in možni model. Model je potrebno jasno in v celoti opisati. Tu bomo opisali le prvi del modela - enačbo, vendar pa ne smemo pozabiti na ostale tri komponente, ki pa smo jih do sedaj le omenili. Katere so že?



Slika 1.6: Snopič regresijskih premic za primer 1.5.1.6

***PRIMER* :** **Nadaljujemo zgornji primer za dnevni prirast pri odbiri mladic.** Osnovni model 1.19 vsebuje samo glavne in ugnezdene vplive, ki smo jih napisali v levi stolpec v preglednici 1.7. V model za dnevni prirast (y_{ijk}) smo vključili srednjo vrednost populacije (μ), sistematska vpliva pasma (P_i) in mesec (M_j) ter naključni vpliv živali (a_{ijk}). Zadnji člen e_{ijk} predstavlja ostanek. V poskusu smo imeli tri pasme, dva meseca in 11 živali.

$$y_{ijk} = \mu + P_i + M_j + a_{ijk} + e_{ijk} \qquad [1.19]$$

Možni model 1.20 vključuje poleg vplivov iz osnovnega modela 1.19 še interakcijo med pasmo in mesecem (PM_{ij}).

$$y_{ijk} = \mu + P_i + M_j + PM_{ij} + a_{ijk} + e_{ijk} \qquad [1.20]$$

POMNI! Pri krajših modelih lahko opis modela navedemo v nekoliko zgoščeni obliki kar v besedilu, kot smo to naredili v zgoraj.

Do sem postopek opravimo prvič že pred pričetkom poskusa. Poleg tega preverimo število opazovanj, število (želenih) parametrov in število stopinj prostosti. Zaradi preglednosti bomo obdelali najprej vse korake pri določanju modelov, se nato vrnili k določanju omenjenih vrednosti in naredili še kakšen primer. Po opravljenem preizkusu vsekakor preverimo, če sta osnovni in možni model postavljena na začetku poskusa še vedno primerno izhodišče. Lahko bi sicer ugotovili, da so se pri izvedbi poskusa spreminjali pogoji, ki jih sicer nismo predvideli. Zaradi dobro vodenega dnevnika pa informacij nismo izgubili in jih lahko upoštevamo.

9. Statistično ovrednotenje opravimo s statistično analizo po izvedbi preizkusa. To delo boste lahko vadili na vajah, tu pa se bomo seznanili samo s pravili. Imamo deduktivni in induktivni postopek. Pri obeh metodah je kriterij verjetnost, da drži ničelna hipoteza. O hipotezah se bomo kasneje pogovarjali. Tule si bomo zapomnili le, da pri ničelni hipotezi predpostavimo, da ni razlik med nivoji pri izbranem

Tabela 1.10: Kriteriji za izločanje vplivov

P-vrednost	Storimo:	Izjema:
0.00 - 0.05	vedno obdržimo	—
0.05 - 0.40	skoraj gotovo obdržimo	zmanjkuje stopinj prostosti
0.40 - 0.60	skoraj gotovo črtamo	vsi bodo čudno gledali, če vpliva ne bo
0.60 - 1.00	skoraj gotovo črtamo	je cilj raziskave

vplivu. Vse ostale možnosti so pa predstavljene z alternativno hipotezo. Takrat, ko je bolj verjetna ničelna hipoteza in je P-vrednost bliže 1.00 (pregl. 1.10), vpliva praviloma ne upoštevamo. Le v izjemnih primerih, ko je to najpomembnejši rezultat raziskave in je vključen v naslov naloge, ga bomo obdržali. V nasprotnem primeru bi pač zgrešili naslov. Ko pa so p-vrednosti bliže 0.00, pa je vpliv potrebno obravnavati, zlasti pri prvih preizkusih. Zaradi pomanjkanja stopinj prostosti lahko črtamo tudi vplive, kjer sta ničelna in alternativna hipoteza enakovredni. V majhnih a dragih preizkusih, ko smo uspeli opraviti le nekaj analiz oziroma meritev, morda izpustimo tudi druge vplive, da le niso značilni.

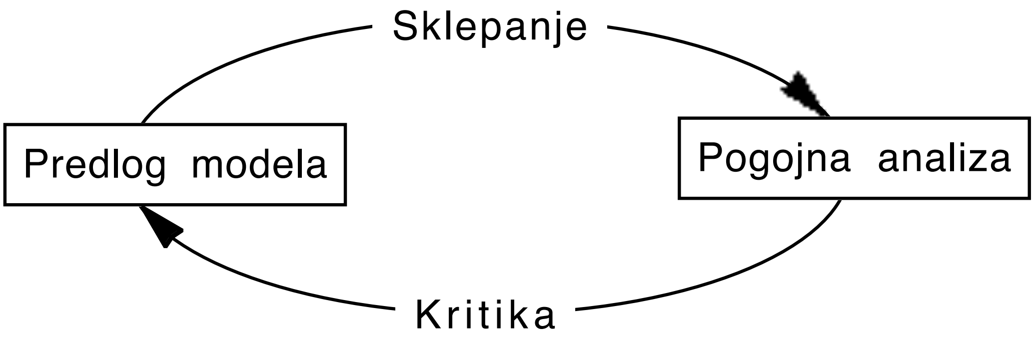
PRIMER: Vzemimo, da je p-vrednost 0.04. Verjetnost, da ničelna hipoteza drži, je zelo majhna, samo 4 %. Alternativna hipoteza je veliko bolj verjetna in sicer velja v 96 % primerov. Torej je vpliv najverjetneje precej pomemben in ga ne smemo zanemariti, črtati iz modela.

***Deduktivni postopek*:** Obdelavo lahko začnemo s teoretičnim, možnim modelom in izločimo vplive, ki nimajo pomembnega vpliva v našem poskusu (p-vrednost večja kot 0.60). Pri vrednostih med 0.60 in 0.40 se odločimo na osnovi strokovne presoje (število opazovanj in število parametrov, pričakovanja). Presodimo tudi, koliko slabši je model brez posameznih vplivov. Statistično značilne vplive vedno obdržimo v modelu in vplive s p-vrednostjo pod 0.40 pa praviloma tudi. Praviloma izločamo najprej interakcije z največ sodelujočimi vplivi, nadrejene črtamo, ko v modelu ni več podrejenih. Vplive izločamo postopoma in raje ponovimo izračun. Postopek ponavljamo, dokler se model spreminja ... Zelo tudi pazimo, da v modelih z regresijo, dovolimo presečišča z ordinato (y - osjo).

***Induktivni postopek*:** V primeru, ko imamo z možnim modelom težave, lahko začnemo z manjšim številom vplivov. Izberemo jih lahko tudi na osnovi osnovnih statističnih parametrov po vplivih, slikah, literaturi, itd. Model očistimo po zgornjem postopku in nato poizkušamo modele z dodanimi vplivi in spremljamo, koliko boljši je razširjeni model. Pri tem postopku se nam lahko zgodi, da katerega od pomembnih, značilnih vplivov ali kombinacije ne preizkusimo.

***Prilagajanje modela*:** Več bo v modelu parametrov, večji bo delež pojasnjene variance, bolj se bo model prilegal. Če izberemo toliko parametrov, kot imamo meritev, se bo model prilegal popolnoma. Ne bo pa uporaben. Vsekakor strmimo za enostavnejšim modelom (zakon skromnosti). To lahko zagotovimo samo s skrbno načrtovanim poskusom. Prav lahko pa se nam ob pridnem beleženju podatkov zgodi, da šele pri obdelavi odkrijemo vpliv, ki ga nismo načrtovali. Poruši se nam uravnotežena struktura podatkov, pojavijo se nivoji brez meritev, itd. Po drugi strani pa bi črtanje statistično (skoraj) značilnega sistematskega vpliva signifikantno povečalo varianco ostanka ter pripomoglo tudi k pristranski oceni. Ocene bi bile "onesnažene" z izpuščenim sistematskim vplivom. Črtanje naključnega vpliva praviloma poveča varianco ostanka. V primeru, ko so nivoji izpuščenega vpliva med seboj korelirani, pa so lahko ocene v reduciranem modelu tudi pristransko ocenjene.

10. Strokovna presoja Model, ki se pri statistični presoji izkaže za najbolj primernega, moramo še temeljito strokovno presoditi. Težko bi se izognili strokovni razlagi statistično značilnega vpliva, pa



Slika 1.7: Proces izgradnje modela

Tabela 1.11: Krmni preizkus

Vrsta obroka	Žival	Čas (T)	
(D)	(a)	1	2
1	1	y_{111}	y_{112}
	2	y_{121}	y_{122}
2	3	y_{231}	y_{232}
	4	y_{241}	

čeprav nas sploh ni zanimal. Moramo ga vsaj na kratko odpraviti, kar je možno, ko je rezultat povsem pričakovan. Malo več težav bodo povzročali vplivi, ki niso v skladu s pričakovanji. Potolažite se lahko z dejstvom, da so predvsem taki rezultati spodbujali znanstvenike k razmišljanju in odkrivanju novih spoznanj. Temeljito se lotite študija narave tako vplivov kot lastnosti v modelu. Če ste se vam je porodila nova ideja, ponovite postopek od začetka... Morda pa je zaplet mogoče pojasniti le z novim poskusom. Tudi ugotovitev, da nam ni šlo vse po sreči, ni katastrofa, če vem, da smo uporabili vse do sedaj pridobljeno znanje iz literature. Tudi informacija, zakaj je nekaj šlo po nepričakovani poti, je lahko dragocena.

11. Proces izgradnje modela Kot vidimo, izgradnja modela ni enkraten proces. Vračati se moramo nazaj. Morda je dovolj, da nekajkrat ponovimo statistično ovrednotenje, prav nič nenavadno pa ni, da bomo morda spremenili listo vplivov. Morda bomo iskali ponovne informacije, ki bi se znali skrivati v katerem od obstoječih informacijskih sistemov.

POMNI! Pri presoji modela ne smemo absolutno zagovarjati svoje teorije - hipoteze. Poskusimo se vživeti v vlogo kritika, osvetlimo tudi druge plati medalje!

VAJA 1:

Opazujemo štiri živali v dveh dneh. V preizkusu smo imeli tudi dva obroka, z vsakim smo krmili po dve živali. Merili smo lastnost "y" (npr. količino mleka).

a)Razvijte možni model, ko uporabite čas kot razred!

$$y_{ijk} = \mu + D_i + T_j + DT_{ij} + a_{ik} + aT_{ijk} + e_{ijk} \quad [1.21]$$

b)Razvijte možni model, ko uporabite čas kot neodvisno spremenljivko in jo vključite v model kot linearno regresijo!

$$y_{ijk} = \mu + D_i + \beta_i \left(x_{ijk} - \bar{x}\right) + a_{ij} + e_{ijk} \quad [1.22]$$

Tabela 1.12: Podatki o preizkusu mladic na rast in zamaščenost na različnih treh farmah

Žival	Pasma	Mesec	Farma	Masa (kg)	Debelina slanine (mm)		Dnevni prirast (g/dan)
1	SL	JAN	A	102	12	13	506
2	SL	JAN	B	98	15	14	550
3	SL	FEB	C	105	15	16	532
4	SL	FEB	A	102	14	12	577
5	SVB	JAN	B	95	19	17	512
6	SVB	FEB	B	101	23	24	499
7	SVB	FEB	C	101	26	24	466
8	SML	JAN	A	97	25	27	545
9	SML	JAN	C	100	21	19	549
10	SML	FEB	C	97	22	23	600
11	SML	FEB	B	102	23	15	610

Čeprav imate rezultat nakazan, prikažite celoten postopek in model tudi pravilno opišite!

VAJA 2:

Vzemimo preizkus mladic (**??**). Med glavne vplive bomo uvrstili pasmo (P_i), mesec (M_j), farmo (F_k), maso ob odbiri kot kvadratno regresijo in žival (a_{ijkl}).

a) Napišite osnovni model za debelino hrbtne slanine!

b) Razvijte možni model!

$$y_{ijklm} = \mu + P_i + M_j + F_k + b_I(x_{ijkl} - 100) + b_{II}(x_{ijkl} - 100)^2 + a_{ijkl} + e_{ijklm} \quad [1.23]$$

$$y_{ijklm} = \mu+P_i+M_j+F_k+PM_{ij}+PF_{ik}+MF_{jk}+PMF_{ijk}+b_{Iijk}(x_{ijkl}-100)+b_{IIijk}(x_{ijkl}-100)^2+a_{ijkl}+e_{ijklm} \quad [1.24]$$

V obeh modelih pomeni:

y_{ijklm} - opazovanje za debelino hrbtne slanine

μ - srednja vrednost populacije

P_i - vpliv pasme; i = 1, 2, 3

M_j - vpliv meseca; j = 1, 2

F_k - vpliv farme; k = 1, 2, 3, 4

PM_{ij}, PF_{ik}, MF_{jk} - dvojne interakcije med posameznimi vplivi

PMF_{ijk} - trojna interakcija med posameznimi vplivi

b_I, b_{Iijk} - regresijski koeficienti pri linearnih členih

b_{II}, b_{IIijk} - regresijski koeficienti pri kvadratnih členih

x_{ijkl} - masa živali ob odbiri

a_{ijkl} - vpliv živali, aditivni genetski vpliv, plemenska vrednost; l = 1,2, ..., n_{ijk}

e_{ijklm} - ostanek; m = 1, 2

n_{ijk} - število živali pasme i, v mesecu j in na farmi k

Tabela 1.13: Določitev možnih interakcij za debelino hrbtno slanino pri preizkusu mladic

Vplivi						
P_i						
M_j : P_i						
F_k : P_i	: M_j					
b_I : P_i	: M_j	: F_k				
b_{II} : P_i	: M_j	: F_k	: b_I			
a_{ijkl} : P_i	: M_j	: F_k	: b_I	: b_{II}		
PM_{ij}	: M_j	: F_k	: b_I	: b_{II}	: a_{ijkl}	
PF_{ik}	: M_j	: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Ii}	: M_j	: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{IIi}	: M_j	: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
MF_{jk}		: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Ij}		: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{IIj}		: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
PMF_{ijk}		: F_k	: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Ik}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{IIk}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Iik}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{IIik}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Ijk}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{IIjk}			: b_I	: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Iij}				: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Iijk}				: b_{II}	: a_{ijkl}	: PM_{ij}
b_{Iij}					: a_{ijkl}	: PM_{ij}
b_{Iijk}					: a_{ijkl}	: PM_{ij}

Tabela 1.14: Seznam parametrov v modelu

Vpliv	Lokacijski parametrov	Parametrov disperzije	Obrazložitev
Srednja vrednost populacije	μ	–	
		Sistematski vplivi	
Pasma	P_1, P_2, P_3	–	tri pasme
Mesec / sezona	M_1, M_2	–	dva meseca
		Naključni vplivi	
Žival	$\cdots a_{ijk} \cdots$	σ_a^2	plemenske vrednosti za živali z meritvami in sorodnike z (aditivno) genetsko varianco

1.5.2 Število parametrov, stopinje prostosti in rang sistema

Ponovimo osnovni model za dnevni prirast v preizkusu z mladicami. V poizkusu smo imeli skupno 11 meritev, vsaka žival je imela natanko eno meritev.

$$y_{ijk} = \mu + P_i + M_j + a_{ijk} + e_{ijk} \qquad [1.25]$$

1.5.2.1 Število parametrov

Parametri so vrednosti, ki jih želimo oceniti, da bi poznali posamezne nivoje. Tokrat nas zanimajo predvsem lokacijski parametri pri sistematskih in naključnih (pregl. 1.14) vplivih.

Poskusimo najprej našteti vse parametre pri sistematskih vplivih! Srednja vrednost populacije je ena sama, označena z μ. V preizkusu smo imeli tri pasme: slovenski landras (P_1), slovenski veliki beli prašič (P_2) in slovenski mesnati landras (P_3). Merili smo samo januarja (M_1) in februarja (M_2). V sistematskem delu imamo skupaj parametrov.

V našem modelu je naključni vpliv samo eden. To je žival, pravzaprav njena plemenska vrednost oz. aditivna genetska vrednost. V našem primeru imamo samo 11 živali, običajno pa plemensko vrednost napovedujemo za veliko število živali in to za živali z meritvami in njihove sorodnike brez meritev. Tako je praktično nemogoče prikazati napovedi za vse živali v preglednici ali sliki. Če bi imeli primer s 10000 živalmi, kar je malo tudi v majhnih populacijah, bi imeli tabelo z 10000 oznakami živali in njihovimi plemenskimi vrednostmi ali histogram z 10000 stolpci. Prav noben od obeh prikazov ni pregleden, zato lokacijske parametre (v tem primeru plemenske vrednosti za živali) prikazujemo s porazdelitvami. Pričakovana (oz. povprečna) vrednost pri naključnih vplivih ima vrednost 0. Za opis normalne porazdelitve potrebujemo še mero razpršenosti (oz. parameter disperzije). To je varianca za posamezni naključni vpliv, v našem primeru je to aditivna genetska varianca (σ_a^2). Lokacijske parametre pa navajamo praviloma le za podskupino nivojev pri naključnem vplivu, npr. izpišemo plemenske vrednosti bikov, katerih seme ponuja osemenjevalno središče ali pa plemenske vrednosti krav iz črede na posamezni kmetiji.

Sedaj moramo samo prešteti število lokacijskih parametrov pri posameznem naključnem vplivu in jih vpisati v pregl. 1.15. Parametri so neznanke, ki jih želimo oceniti, in za vsako neznanko moramo v sistemu nastaviti po eno enačbo. Prav tako bi radi izvrednotili (= izračunali vrednost) plemenske vrednosti, torej so plemenske vrednosti za žival tudi neznanke in moramo nastaviti po eno enačbo za vsako žival. Sistem enačb ima v našem primeru 17 enačb. Število neznank v sistemu enačb poimenujemo razsežnost ali ***red sistema***. V našem primeru je red sistema 17.

1.5.2.2 Stopinje prostosti

Stopinje prostosti predstavljajo število ocenljivih parametrov. Ti parametri povedo vse o podatkih, kadar je izbrani model pravilen. Ostale ocene so le linearne kombinacije ocenljivih parametrov. Ko imamo npr. ocenjeno srednjo vrednost, je dovolj, da izvemo še oceni za dve pasmi, oceno tretje pa lahko dobimo s sklepanjem - lahko jo izračunamo iz omenjenih treh ocen.

Tabela 1.15: Določanje števila parametrov in stopinj prostosti v modelu 1.25

Vplivi	Število parametrov	Število stopinj prostosti	
Število meritev	11		
Sistematski del modela			
μ	1		1
P_i	3	3-1=	2
M_j	2	3-1=	1
Skupaj za model	6		4
Za ostanek		11-4=	7
Naključni vplivi			
a_{ijk}	11		11
Celoten sistem enačb			
Red	6+11=17		
Rang		4+11=	15

Za izhodišče vzamemo število meritev (pregl. 1.15). Vsaka meritev prinese eno stopinjo prostosti, torej imamo v našem preizkusu z dnevnim prirastom 11 stopinj prostosti. Dogovorimo se, da ste opravili preizkus vi in obdelali podatke. Torej so informacije vaše. Mi pa bi radi napisali članek in bi rada od vas kupili informacije, ki zbrane podatke zadostno opišejo. Postavili ste nam visoko ceno, torej bomo dobro premislili, katere ocene parametre bomo kupovali.

Največ o populaciji mi bo gotovo povedala kar srednja vrednost μ (pregl. 1.15) in je tudi najbolj zanesljiva ocena, saj smo za izračun porabili kar vse razpoložljive podatke. Ko plačamo ceno, je ta informacija naša (stopinje prostosti za model), vi pa imate eno manj (stopinje prostosti za ostanek).

Stopinje prostosti za križno klasificirane sistematske vplive. Potem poskusimo s kupčijo pri vplivu pasme. Prav gotovo moramo pridobiti informacijo za prvo pasmo (P_1). Tudi pri drugi pasmi nam ne pomaga nič drugega kot plačati novo informacijo (P_2). Pri zadnjem nivoju pa lahko z nekaj sklepanja sami pridemo do rezultata, saj velja enačba 1.26. Ker imam že tri informacije, lahko izračunam še četrto neznanko (P_3). Za vpliv pasme potrebujemo en parameter manj, kot smo jih želeli oceniti, zadnjega "ne potrebujemo" več.

$$\mu = \frac{P_1 + P_2 + P_3}{3} \qquad [1.26]$$

Če smo natančni, bomo uporabili nekaj znanj iz matematike. Enačba za srednjo vrednost μ in enačbe za vse nivoje pri vplivu pasme (P_1, P_2, P_3) so med seboj linearno odvisne: srednjo vrednost dobimo kot povprečje vseh parametrov za pasmo. Lahko pa enačbo razrešimo na katerikoli parameter pri vplivu pasme (enačba 1.27). Naredimo to npr. za zadnjo pasmo (P_3):

$$P_3 = 3\mu - P_1 - P_2 \qquad [1.27]$$

Tudi GLM v SASu pri ocenah za zadnji nivo izpiše vrednost 0. To ne pomeni, da je vpliv zadnje pasme enak 0, ampak pomeni, da vpliva pasme ne moremo oceniti. Ocenimo lahko le razlike med pasmami. Vse preostale pasme so torej primerjane s pasmo, ki se je znašla na zadnjem mestu.

Pri vplivu meseca lahko razmišljamo enako (enačba 1.28). Zadnjega, t.j. drugi mesec, nam ni treba kupiti, ostale pa moramo. Ker sta bila samo dva meseca (dva parametra), lahko pridobimo le eno novo informacijo. Vpliv meseca v tem primeru prispeva 1 stopinjo prostosti za model, število stopinj prostosti za model pa se za isto število zmanjša.

$$\mu = \frac{M_1 + M_2}{2} \qquad [1.28]$$

Tudi pri vplivu meseca, ki je križno klasificiran s pasmo, lahko zadnji parameter izračunamo tako, da od 2-kratne srednje vrednosti odštejemo vrednost za preostale mesece (enačba 1.29).

$$M_2 = 2\mu - M_1 \quad [1.29]$$

Zadnji nivo (M_2) ni ocenljiv in bo v GLM-u postavljen na vrednost nič, vrednosti za vse ostale mesece v podatkih pa bodo predstavljale odstopanje nivoja od zadnjega nivoja.

Stopinje prostosti za naključne vplive. Za naključne vplive uporabimo pri izračunu samo meritve, ampak tudi podobnost med njimi (npr. sorodstvo med živalmi pri genetskih vplivih, korelirane meritve) in strukturo varianc in kovarianc. Za vsak lokacijski parameter dobimo napoved. Ko je vrednost 0, je napoved za tisti nivo tudi 0. Napovedi ostalih nivojev so dejansko odstopanja od pričakovanih vrednosti (kar je tudi 0) in ne od nivojev, katerih napovedi so 0. Ko npr. pridemo do zadnje živali ne moremo uganiti njene plemenske vrednosti, ampak jo moramo dobiti v izračunu. Število stopinj prostosti za naključni vpliv je enako številu lokacijskih parametrov, nivojev pri naključnem vplivu. Število stopinj prostosti za naključne vplive potrebujemo za oceno komponent variance in kovarianc, česar pa pri tem predmetu še ne bomo delali.

Velja se zapomniti naslednje posebnosti pri naključnih vplivih:

1. Niz enačb za vsak naključen vpliv je polnega ranga.
2. Izjema je lahko pri genetskem vplivu pri potomcih iz enega oplojenega jajčeca (npr. enojajčnih dvojčkih) in klonih.
3. Pri izračunu stopinj prostosti za model in ostanek stopinj prostosti za naključne vplive ne upoštevamo.

Število stopinj prostosti za model in ostanek. Torej sta števili stopinj prostosti za model in za ostanek odvisna le od sistematskega dela modela (pregl. 1.15).

Število ***stopinj prostosti za model*** (1.30) je vsota vseh stopinj prostosti za sistematski del modela.

$$s.p.\, za\, model = \sum p_i \quad [1.30]$$

Število ***stopinj prostosti za ostanek*** (1.31) izračunamo tako, da od števila opazovanj odštejemo število stopinj prostosti za model.

$$s.p.\, za\, ostanek = število\, opazovanj - s.p.\, za\, model \quad [1.31]$$

VAJA 3:
Vzeli bomo podatke za dnevni prirast, a smo opravili meritve na 9563 prašičih na štirih farmah, štiriindvajset mesecev in treh pasmah. Vplivi so križno klasificirani. Prašiči so bili iz 873 gnezd, plemensko vrednost pa bi radi izračunali dodatno še za 1700 prednikov. Določimo število parametrov in stopinje prostosti po vplivih, za model in ostanek za naslednji model (enačba 1.32).

$$y_{ijklm} = \mu + P_i + M_j + F_k + PM_{ij} + PF_{ik} + MF_{jk} + PMF_{ik} + g_{ijkl} + a_{ijklm} + e_{ijklm} \quad [1.32]$$

Predno preverimo, kako dobimo število parametrov in stopinj prostosti za interakcije, preverimo še naključna vpliva. V modelu imamo naključni vpliv skupnega okolja v gnezdu (g_{ijkl}) in ker imamo 873 gnezd, ima vpliv tudi toliko nivojev, za katere želimo dobiti napovedi. Izmerili smo 9563 živali in dodali

Tabela 1.16: Določanje števila parametrov in stopinj prostosti v modelu 1.25

Vplivi	Število parametrov		Število stopinj prostosti	
Število meritev	9563			
Sistematski del modela				
μ		1		1
P_i		3	3-1=	2
M_j		24	24-1=	23
F_k		4	4-1=	3
PM_{ij}	3×24=	72	(3-1)×(24-1)=	46
PF_{ik}	3×4=	12	(3-1)×(4-1)=	6
MF_{jk}	24×4=	96	(24-1)×(4-1)=	69
PMF_{ijk}	3×24×4=	288	(3-1)×(24-1)×(4-1)=	138
Skupaj za model		500		288
Za ostanek			9563-288=	9275
Naključni vplivi				
g_{ijkl}		873		873
a_{ijklm}	9563+1700=	11263		11263
Celoten sistem enačb				
Red	500+873+11263=	12636		
Rang			288+873+11263=	12424

še 1700 sorodnikov, ki so lahko predniki, sovrstniki ali celo potomci. Skupaj imamo torej 11263 živali in za vsako od njih želimo napovedati plemensko vrednost.

POMNI! Plemenska vrednost poimenujemo parametre samo pri aditivnih genetskih vrednostih (pri vplivu živali). Pri drugih vplivih, ki opisujejo okolje, govorimo enostavno samo o vrednosti nivojev.

Število stopinj prostosti za interakcije v sistematskem delu modela. Ostale stopinje prostosti v modelu 1.32 smo že znali določiti, ostale pa so nam le interakcije.

Vzemimo interakcijo med pasmo (P_i) in farmo (F_k). Ta rezultate razmišljanja bomo zbrali v preglednici 1.17. Neznane parametre, ki jih moramo pridobiti iz podatkov, bomo navedli z oznako. Tiste, ki jih lahko izvrednotimo iz že pridobljenih vrednosti, pa bomo predstavili kar z enačajem (=) .

V prvem stolpcu preglednice 1.17 imamo naštete parametre za srednjo vrednost in dve pasmi, parameter za pasmo tri pa (P_3) smo nadomestili z =, vzroke pa smo obrazložili v poglavju 1.5.2.2. Isto velja za vpliv farme. Ker je vpliv farme križno klasificiran s pasmo, veljajo zanj ista pravila.

V spodnjem desnem delu preglednice imamo parametre za interakcijo med pasmo in farmo (PF_{ik}). Preštejmo parametre za interakcijo: štejemo tiste, ki so označeni in tiste, ki smo jih označili z enačajem. Ugotovimo lahko, da je nivojev pri interakciji 12, kar lahko dobimo tudi tako, da zmnožimo število nivojev pri vseh sodelujočih vplivih. Pri naši dvojni interakciji imamo tri pasme in štiri farme zato je nivojev pri interakciji med pasmo in farmo $3 * 4 = 12$.

Sedaj pa pojdimo v drugo vrstico preglednice, kjer so parametri za prvo pasmo (P_1) in vse interakcije te pasme in štirih farm. Pasmo imamo na štirih farmah, njen rezultat je povprečje pasme, dosežene na štirih farmah (enačba 1.33). Ker pa smo parameter P_1 že izračunali (je v modelu pred interakcijo), ni neznanka. Pridobiti moramo ocene za prve tri parametre PF_{1k}, četrtega pa ne moremo več oceniti, saj

Tabela 1.17: Ocenljivi parametri in število stopinj prostosti pri križno klasificiranih vplivih

μ	F_1	F_2	F_3	=
P_1	PF_{11}	PF_{12}	PF_{13}	=
P_2	PF_{21}	PF_{22}	PF_{23}	=
=	=	=	=	=

ga lahko izračunamo iz preostalih štirih parametrov. Enačba za parameter PF_{14} je linearna kombinacija enačb za pasmo ena in prve tri parametre za interakcijo.

$$P_1 = \frac{PF_{11} + PF_{12} + PF_{13} + PF_{14}}{4} \qquad [1.33]$$

Isto velja tudi za vse druge nivoje vpliva pasme (za vse vrstice), razen zadnjega.

Sedaj pa bi pogledali še po stolpcih. Poglejmo stolpec za prvo farmo (F_1). Prva dva parametra interakcije farme 1 s prvo (PF_{11}) in drugo (PF_{21}) pasmo smo morali oceniti. Zadnji parameter (PF_{31}) z interakcijo med prvo farmo in tretjo pasmo pa je lahko izračunana, po enačbi 1.34. Rezultate farme F_1 smo že pridobili za prve dve pasmi, ostane nam edino še neznana proizvodnost pasme 1 na farmi 4 (PF_{14}), saj rezultat za prvo farmo skupaj že imamo.

$$F_1 = \frac{PF_{11} + PF_{21} + PF_{31}}{3} \qquad [1.34]$$

Tako ravnamo še za drugo in tretjo farmo. Tudi pri četrti farmi (F_4) smo lahko dobili rezultate, prav tako tudi za prvi dve pasmi na četrti farmi, tako lahko napišemo dodatno enačbo, s katero bomo izračunali še parameter PF_{34}.

$$F_4 = \frac{PF_{14} + PF_{24} + PF_{34}}{3} \qquad [1.35]$$

Povzamemo lahko, da dobimo stopinje prostosti pri interakcijah tako, da pomnožimo stopinje prostosti pri sodelujočih glavnih vplivih. Število parametrov pri trojni interakciji (p_{PMF}) je enaka produktu (enačba 1.36) številu parametrov po posameznih vplivih, ki sestavljajo interakcijo.

$$p_{PMF} = p_P \times p_M \times p_F \qquad [1.36]$$

Število stopinj prostosti (enačba 1.37) pa dobimo tako, da pomnožimo stopinje prostosti za vključene vplive.

$$sp_{PMF} = (p_P - 1) \times (p_M - 1) \times (p_F - 1) \qquad [1.37]$$

Red in rang sistema. Red sistema predstavlja število (neznanih) parametrov v modelu. Za vsak nivo pri vsakem vplivu, tako sistematskem kot naključnem, imamo po eno neznanko. Za vsako od neznank bomo rabili po eno enačbo, zato je število neznank in število enačb enako.

Sistem enačb lahko vsebuje odvisne enačbe. Nekatere enačbe so linearne kombinacije drugih. V naših sistemih so linearno odvisne tiste enačbe, kjer je število stopinj prostosti različno od števila parametrov. Pri sistematskem delu modela je rang enak številu stopinj prostosti, pri naključnem pa številu parametrov. Za mešani model dobimo rang tako, da seštejemo obe vrednosti.

$$rang = število\ s.p.\ za\ model + število\ parametrov\ za\ naključni\ del\ modela \qquad [1.38]$$

Tabela 1.18: Seznam parametrov in število stopinj prostosti v modelu 1.32 za debelino slanine

Vplivi	Število parametrov		Število stopinj prostosti	
Število meritev	9563×3=28689			
Sistematski del modela				
μ		1		1
P_i		3	3-1=	2
M_j		24	24-1=	23
F_k		4	4-1=	3
PM_{ij}	3×24=	72	(3-1)×(24-1)=	46
PF_{ik}	3×4=	12	(3-1)×(4-1)=	6
MF_{jk}	24×4=	96	(24-1)×(4-1)=	69
PMF_{ijk}	3×24×4=	288	(3-1)×(24-1)×(4-1)=	138
b_{Iijk}	3×24×4=	288	3×24×4=	288
b_{IIijk}	3×24×4=	288	3×24×4=	288
Skupaj za model		1076		864
Za ostanek			9563-288=	9275
Naključni vplivi				
g_{ijkl}		873		873
a_{ijklm}	9563+1700=	11263		11263
Celoten sistem enačb				
Red	500+873+11263=	12636		
Rang			288+873+11263=	12424

Stopinje prostosti pri kvantitativnih vplivih. V model za debelino slanine, ki smo jo izmerili 3-krat na vsaki živali, bomo modelu 1.32 dodali še kvadratno regresijo ugnezdeno znotraj interakcije PMF_{ijk}. V modelu bomo tako imeli tri glavne vplive, tri dvojne in eno trojno interakcijo.

Ker imamo na vsaki od 9563 živali po tri meritve, imamo na voljo skupaj 28689 meritev (pregl. 1.18).

Število parametrov pri glavnih vplivih in interakcijah pri sistematskih vplivih je enako številu nivojev iz pregl. 1.16. Zaradi več meritev po živali, ni v poskusu prav nič gnezd kot pri dnevnem prirastu. Torej imamo živali iz 873 gnezd in toliko nivojev za skupno okolje v gnezdu. Prav tako smo merili debelino slanine na 9563 živalih in imamo 1700 sorodnikov. Ker sta v modelu dva naključna vpliva imamo tri komponente variance:

- varianco za skupno okolje v gnezdu (σ_g^2),
- genetsko varianco (σ_a^2) in
- varianco za ostanek (σ_e^2).

Pri ugnezdeni regresiji (1.36) imamo vedno par regresijskih koeficientov iz linearnega (b_{Iijk}) in kvadratnega (b_{IIijk}) člena, parameter PFM_{ijk} pa predstavlja presečišče z y-osjo. Ker imamo 288 parametrov PMF_{ijk}, je 288 tudi parametrov b_{Iijk} in b_{IIijk}. Če bi narisali sliko, bi imeli na sliki kar 288 parabol. Pri regresijah moramo oceniti vse parametre, torej je pri regresijskih koeficienti število stopinj prostosti enako število parametrov.

$$y_{ijklmn} = \mu + P_i + M_j + F_k + PM_{ij} + PF_{ik} + MF_{jk} + PMF_{ijk} +$$

$$+ b_{Iijk}(x_{ijklm} - 100) + b_{IIijk}(x_{ijklm} - 100)^2 + g_{ijkl} + a_{ijklm} + e_{ijklmn} \qquad [1.39]$$

Pravzaprav bi zadoščal na videz preprostejši model 1.40. Število parametrov za regresijo in za naključne vplive je isto kot pri modelu 1.18. Ker v sistematskem delu nimamo srednje vrednosti, glavnih vplivov in dvojnih interakcij, moramo oceniti vse parametre trojne interakcije. Trojna interakcija in 288 parametrov in 288 je tudi stopinj prostosti. Model 1.40 je tako polnega ranga. Iz teh 288 parametrov pa lahko z dodatnimi enačbami ocenimo parametre za dvojne interakcije, glavne vplive in srednjo vrednost. Za modela 1.18 in 1.40 rečemo, da sta ***ekvivalentna modela*** in dajeta iste (ekvivalentne) rezultate.

Pri spodnjem modelu sta red in rang sistema enačb enaka. Vsi dobimo iste rešitve, kar ni garantirano pri sistemih z nepolnim rangom, kot je to model 1.24. Model je polnega ranga in zato ima manj numeričnih problemov, je pa manj prikladen za interpretacijo.

$$y_{ijklmn} = PMF_{ijk} + b_{Iijk}(x_{ijklm} - 100) + b_{IIijk}(x_{ijklm} - 100)^2 + g_{ijkl} + a_{ijklm} + e_{ijklmn} \qquad [1.40]$$

1.5.3 Vaje iz razvoja modela

VAJA 4:
Pri miškah smo delali poskus s petimi različnimi krmami. Poskus smo izvajali v treh ponovitvah. V poskus smo vsakič vzeli 50 gnezd, kjer je bilo ob odstavitvi med 8 in 12 potomcev. V celoten poskus so bili vključeni trije očetje, ki so bili v vseh treh poskusih enakomerno zastopani. Odstavljene miške smo enakomerno porazdelili v skupine s krmo. Pri uvrščanju v skupine smo gledali tudi na to, da sta bila spola čim bolj enakomerno zastopana in da je bila povprečna odstavitvena masa med skupinami čim bolj enaka. Zanimala nas je rast mišk od odstavitve naprej. Živali smo tehtali vsak teden. Poskus je trajal mesec in pol.
Med glavne vplive bomo uvrstili pasmo (P_i), mesec (M_j), farmo (F_k), maso ob odbiri kot kvadratno regresijo in žival (a_{ijkl}).
a) Razvijte osnovni in možni model in ju napišite!
b) Določite število parametrov, stopinje prostosti, določite število podatkov, red in rang sistema pri obeh modelih!
c) Kritično presodite model!

VAJA 5:
Pri miškah smo delali poskus s petimi različnimi krmami. Poskus smo izvajali v treh ponovitvah. V poskus smo vsakič vzeli 50 gnezd, kjer je bilo ob odstavitvi med 8 in 12 potomcev. V vsakem poskusu so bili vključeni po trije različni očetje, ki so bili v poskusu enakomerno zastopani. Odstavljene miške smo enakomerno porazdelili v skupine s krmo. Pri uvrščanju v skupine smo gledali tudi na to, da sta bila spola čim bolj enakomerno zastopana in da je bila povprečna odstavitvena masa med skupinami čim bolj enaka. Zanimala nas je rast mišk od odstavitve do pubertete. Živali smo tehtali na začetku in koncu poskusa.
a) Razvijte osnovni in možni model in ju napišite!
b) Določite število parametrov, število stopinj prostosti za posamezne vplive, model in ostanek, določite število podatkov, red in rang sistema!
c) Kritično presodite model!

VAJA 6:
Naredili bomo poskus na pujskih. Iz vsakega gnezda bomo vzeli po 4 svinjke in 4 kastrate. Pri vsaki svinji bomo vzeli po 3 gnezda. Svinja je bila vedno parjena z istim merjascem, merjasec je bil parjen s 6 svinjami. V poskusu je bilo 5 merjascev. Pujske smo naključno razdelili enakomerno razdelili v 4 skupine, ki so dobivale različno vsebnost lizina v krmi in sicer 50 ng/kg, 60 ng/kg, 70 ng/kg in 80 ng/kg (vrednosti niso priporočila za prehrano prašičev). Stehtali jih bomo pri starosti 180 dni.
a) Razvijte osnovni in možni model in ju napišite!
b) Določite število parametrov, število stopinj prostosti za posamezne vplive, model in ostanek, določite število podatkov, red in rang sistema!
c) Kritično presodite model!

VAJA 7:
Naredili bomo poskus na pujskih. Iz vsakega gnezda bomo vzeli po 4 svinjke in 4 kastrate. Pri vsaki svinji bomo vzeli po 3 gnezda. Svinja je bila vedno parjena z istim merjascem, merjasec je bil parjen s 6 svinjami. V poskusu je bilo 5 merjascev. Pujske smo naključno razdelili enakomerno razdelili v 4 skupine, ki so dobivale krmo različnega izvora. Da bi izvedli preizkus nepristransko, smo krme označili s črkami A, B, C in D. Pujske bomo stehtali na začetku in koncu poskusa pri starosti 180 dni.
a) Razvijte osnovni in možni model in ju napišite!
b) Določite število parametrov, število stopinj prostosti za posamezne vplive, model in ostanek, določite število podatkov, red in rang sistema!
c) Kritično presodite model!

VAJA 8:
Naredili bomo poskus na pujskih. Iz vsakega gnezda bomo vzeli po 4 svinjke in 4 kastrate. Pri vsaki svinji bomo vzeli po 3 gnezda. Svinja je bila vedno parjena z istim merjascem, merjasec je bil parjen s 6 svinjami. V poskusu je bilo 10 merjascev. Po dva merjasca smo vsakokrat naključno razdelili v 5 skupin, ki so dobivale krmo različnega izvora. Da bi izvedli preizkus nepristransko, smo krme označili s črkami A, B, C, D in E. Pujske bomo stehtali na začetku in koncu poskusa pri starosti 180 dni.
a) Razvijte osnovni in možni model in ju napišite!
b) Določite število parametrov, število stopinj prostosti za posamezne vplive, model in ostanek, določite število podatkov, red in rang sistema!
c) Kritično presodite model!

REŠITEV1.5.3:

Kot rešitev navajamo osnovni 1.128 in možni 1.129 model, izpustili smo opis parametrov, odvisnih in neodvisnih spremenljivk ter indeksov. Vrstni red vplivov je lahko različen, potem bodo lahko tudi indeksi pri vplivih različni, kar ni nič napačnega, število indeksov V preglednici 1.21 povzemamo izračun števila parametrov in stopinj prostosti za možni model, za osnovnega preverite rezultate tako, da upoštevate samo tiste vplive, ki nastopajo v osnovnem, izhodiščnem modelu. Število podatkov ne moremo natančno predvideti, ker ni povsem jasno, koliko mladičev vzamemo iz gnezda. Vemo pa, da imamo 3 poskuse, pri vsakem imamo 50 gnezd, število mladičev v gnezdu pa variira med 8 in 10. Velikost gnezda smo označili z n_{jkl}. Skupno število (1.43) torej dobimo tako, da seštejemo mladiče iz vseh gnezd. Pričakujemo, da bo v 150-tih gnezdih najbrž v poskusu nekako 1350 mladičev, lahko pa skupno število variira med 1200 do 1500 mladičev. V primeru možnega modela bi radi ocenili 270 parametrov. To nekako pomeni, da podatkov ni prav veliko, zato zelo upamo, da se bomo lahko znebili nekaterih členov in poenostavili model.

$$y_{ijklmn} = \mu + K_i + P_j + O_k + S_l + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right) + b_{IIijkl}\left(x_{ijklmn} - \bar{x}\right)^2 + g_{jkm} + e_{ijklnm} \qquad [1.41]$$

Tabela 1.19: Izračun stopinj prostosti za model 1.129

Vpliv	Seznam parametrov	Štev. parametrov		Štev. stopinj prostosti	
Srednja vrednost	μ	1	1		1
Krma	K_1, K_2, K_3, K_4, K_5	5	5	5-1	4
Poskus	P_1, P_2, P_3	3	3	3-1	2
Oče	O_1, O_2, O_3	3	3	3-1	2
Spol	S_1, S_2	2	2	2-1	1
Interakcija med					
K in P	$KP_{11}, \cdots KP_{ij}, \cdots, KP_{53}$	5×3	15	(5-1) $\times$ (3-1)	8
K in O	$KO_{11}, \cdots KO_{ik}, \cdots, KO_{53}$	5×3	15	(5-1) $\times$ (3-1)	8
K in S	$KS_{11}, \cdots KS_{il}, \cdots, KS_{52}$	5×2	10	(5-1) $\times$ (2-1)	4
P in O	$PO_{11}, \cdots PO_{jk}, \cdots, PO_{33}$	3×3	9	(3-1) $\times$ (3-1)	4
P in S	$PS_{11}, \cdots PS_{jk}, \cdots, PS_{32}$	3×2	6	(3-1) $\times$ (2-1)	2
O in S	$OS_{11}, \cdots OS_{jk}, \cdots, OS_{32}$	3×2	6	(3-1) $\times$ (2-1)	2
K, P in O	$KPO_{111}, \cdots KPO_{ijk}, \cdots, KPO_{533}$	5× 3 × 3=	45		16
K, P in S	$KPS_{111}, \cdots KPS_{ijl}, \cdots, KPS_{532}$	5 × 3 × 2=	30		8
K, O in S	$KOS_{111}, \cdots KOS_{ijl}, \cdots, KOS_{532}$	5 × 3 × 2=	30		8
P, O in S	$POS_{111}, \cdots POS_{ijl}, \cdots, POS_{332}$	3 × 3 × 2=	18		4
K, P, O in S	$KPOS_{1111}, \cdots KPOS_{ijkl}, \cdots, KPOS_{5332}$	5×3×3×2=	90	(5-1)×(3-1)×(3-1)×(2-1)=	16
Regresijski koeficient za					
linearni člen	$b_{I1111}, \cdots, b_{Iijkl}, \cdots .. b_{I5332}$	5×3×3×2=	90		90
kvadratni člen	$b_{I1111}, \cdots, b_{Iijkl}, \cdots .. b_{I5332}$	5×3×3×2=	90		90
Število za model					
Gnezdo	... g_{kjm}...		150		150
Interakcija med					
K in g	... Kg_{ijkm}...	5×150=	750		750
S in g	... Sg_{jklm}...	2×150=	300		300

$$\begin{aligned} y_{ijklmn} = \; & \mu + K_i + P_j + O_k + S_l + KP_{ij} + KO_{ik} + KS + PO_{jk} + PS_{jl} + OS_{kl} + \\ & + KPO_{ijk} + KPS_{ijl} + KOS_{ikl} + POS_{jkl} + KPOS_{ijkl} + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right) + \\ & + b_{IIijkl}\left(x_{ijklmn} - \bar{x}\right)^2 + g_{jkm} + Kg_{ijkm} + Sg_{jklm} + e_{ijklnm} \end{aligned} \quad [1.42]$$

$$N = \sum_{j=1}^{3}\sum_{k=1}^{50} n_{jkm} \quad [1.43]$$

Pri posameznih nivojih pri naključnih vplivih je lahko malo opazovanj, ki se pri interakciji še porazdelijo. Tako se lahko zgodi, da bo lahko posamezni nivo tudi brez informacij. Ko bomo prešteli nivoje pri naključni interakciji, jih bo lahko kar nekaj manjkalo. Kadar želimo izračunati le lokacijske parametre, nas nivoji brez podatkov pri naključnih vplivih ne skrbijo in zanje lahko nastavimo enačbe.

V modelu imamo tudi štiri komponente variance:

- varianco za skupno okolje v gnezdu (σ_g^2),
- varianco za skupno okolje pri interakciji med krmo in gnezdom (σ_{Kg}^2),
- varianco za skupno okolje pri interakciji med spolom in gnezdom (σ_{Sg}^2) in
- varianco za ostanek (σ_e^2).

REŠITEV 1.5.3:

V osnovni model 1.44 vključimo najprej spol (S_i) in krmo. Krme se razlikujejo zaradi različne vsebnosti lizina, vsebnosti so znane in jih bomo zato obravnavali kot kvantitativni vpliv, ki jo predstavlja neodvisna spremenljivka $\left(x_{ijklm}\right)$. V osnovni krmi primanjkuje lizina, ki je aminokislina neobhodno potrebna za rast živali in je telo ne more proizvesti samo. Presečišče z osjo y pri vsebnosti lizina 0 % tako nima pravega pomena in bomo rezultate težko presojali, saj ne vemo, kako prašiči rastejo brez lizina v krmi. Bolje je, da ordinatno os y prestavimo na eno od vsebnosti, ki jih krma vsebuje. V tem primeru je lahko najmanjša vsebnost (x_{min}) povsem primerna. Lizin je draga sestavina in bi bili kar zadovoljni, če ga lahko dodamo v manjših količinah. Torej se ima smisel osredotočiti na to točko. Rezultati niso zaradi tega prav nič potvorjeni, ne glede, kam bomo premaknili ordinato, morajo biti zaključki enaki. Tule bomo predpostavili kar linearno regresijo, imamo samo en regresijski koeficient (b), čeprav je potrebno to pri obdelavi preveriti. Na rezultate lahko vpliva tudi izbor merjascev $\left(M_j\right)$, ker niso naključni vzorec iz populacije moških prašičev. Da smo jih odbrali za pleme, so se morali dobro izkazati v preizkusu na lastnosti rasti. Podobno, čeprav manj intenzivno, so bile odbrane tudi svinje. Meritev po svinji je kar nekaj, saj ima v preizkusu tri gnezda z 8 pujski in lahko ocenimo oziroma odstranimo njen vpliv. Svinje $\left(F_{jk}\right)$ so ugnezdene znotraj merjasca. Po gnezdu imamo veliko meritev, le po 8 pujskov s po eno meritvijo. Pri navadnem krmnem poskusu bi ga nadvse radi izpustili, morda pa bi ga uspeli nadomestiti s čim drugim. Poskrbeli smo tudi za enakomerno zastopanost spolov in enakomerno porazdelitev po drugih vplivih. V takih uravnoteženih poskusih izpustitev gnezda kot vpliva ni napaka. V našem poskusu ga bomo torej z veseljem zavrgli, bomo ga pa najprej preverili. Ker je malo podatkov po gnezdu in kar nekaj gnezd v poskusu, bomo vpliv gnezda obravnavali kot naključni vpliv $\left(g_{jkl}\right)$. Gnezdo je ugnezdeno znotraj svinje in merjasca, a križno klasificirano s spolom.

$$y_{ijklm} = \mu + S_i + b\left(x_{ijklm} - x_{min}\right) + M_j + F_{jk} + g_{jkl} + e_{ijklm} \qquad [1.44]$$

Možni model 1.45 vključuje glavne vplive iz osnovnega modela 1.44. Pridobili smo tudi interakcije med spolom in očetom $\left(SM_{ij}\right)$ oziroma svinjo $\left(SF_{ijk}\right)$. Pri regresiji je možna ugnezditev znotraj interakcije SF_{ijk}, čeprav je povsem nepraktična. Če bi to res obstajalo, bi za vsako svinjo morali narediti krmni poskus in sicer v kombinaciji z vsako krmo. Za nameček bi jo morali svinjo krmiti z dvema krmama: ena bi bila bolj primerna za svinjke, druga za kastrate. Ko bi poskus za silo končali, bi bila svinja že ostarela in primerna za izločitev. Kaj bi nam rezultati sploh pomagali, če rezultati ne veljajo tudi za druge svinje?

Čemu torej tako kompleksen model na začetku? Z veseljem bomo vsak sestavljen vpliv in tudi kakšnega od glavnih črtali, ko bomo potrdili, da niso značilni. Model ima smisel le, če je obvladljiv, če ga lahko razložimo. Toda ne smemo se vnaprej odločiti, katere vplive bomo vključili v model. V postopku izgradnje modela preverimo, če je poskus potekal, kot smo si zastavili. Celo vpliva svinj in merjascev bi se radi iznebili. Reja prašičev bi bila prava nočna mora za gospodarja, če bi vsaki svinji moral pripraviti poseben obrok! Lahko pa bi se izkazalo, da je stara mati svoji svinji prinašala priboljške. Taka ljubljenka ali bolna žival bi v poskusu nagajala. V takšni analizi bi jo našli in odstranili iz analize.

V enem povsem resnem poskusu pitanja bikov se je zgodil nenavaden primer, ki dobro kaže na to, da se je potrebno lotiti analize sistematično. V poskusu je najlepše, že kar neverjetno, priraščal bik, ki je pojedel najmanj krme, celo tako malo, da bi moral beljakovine graditi iz dušika iz zraka. Raziskovalci smo napenjali možgane, a nismo našli problema. Klepet s oskrbnikom hleva pa je pokazal, da gre za sila navihanega bika. Najprej je pomagal pospraviti krmo v jaslih svojega desnega in levega soseda, nekaj njegove krme pa mu je celo ostalo v jaslih, saj kasneje pač ni bil več pripravljen deliti krme s kolegoma. Tudi pregled podatkov je pokazal, da je bila pripoved resnična.

Dobili smo še eno interakcijo in sicer med spolom in gnezdom $\left(Sg_{ijkl}\right)$. Vpliv je naključni: gnezdo smo razdelili na dve skupini in sicer po spolu. V eni skupinici je le polovica opazovanj v primerjavi z gnezdom, skupin pa je še enkrat toliko. Interakcija, ki ima vsaj en naključni vpliv, je naključna.

$$\begin{aligned} y_{ijklm} = \; & \mu + S_i + M_j + F_{jk} + SM_{ij} + SF_{ijk} + \\ & + b_{ijk}\left(x_{ijklm} - x_{min}\right) + g_{jkl} + Sg_{ijkl} + e_{ijklm} \end{aligned} \qquad [1.45]$$

Tabela 1.20: Ocenljivi parametri in stopinje prostosti pri ugnezdenih vplivih

	Merjasci				
Svinja	M_1	M_2	M_3	M_4	=
1	F_{11}	F_{21}	F_{31}	F_{41}	F_{51}
2	F_{12}	F_{22}	F_{32}	F_{42}	F_{52}
3	F_{13}	F_{23}	F_{33}	F_{43}	F_{53}
4	F_{14}	F_{24}	F_{34}	F_{44}	F_{54}
5	F_{15}	F_{25}	F_{35}	F_{45}	F_{25}
6	=	=	=	=	=

Tabela 1.21: Izračun stopinj prostosti za model 1.45

Vpliv	Seznam parametrov	Štev. parametrov		Štev. stopinj prostosti	
Srednja vrednost	μ	1	1		1
Spol	S_1, S_2	2	2	2-1	1
Merjasec	M_1, M_2, M_3, M_4, M_5	5	5	5-1	4
Svinja	$F_{11}, \cdots, F_{jk}, \cdots, F_{56}$	$5 \times 6=$	30	$5 \times$ (6-1) =	25
Interakcija med S in M	$SM_{11}, \cdots SM_{ij}, \cdots, SM_{25}$	$2 \times 5=$	10	(2-1)× (5-1)=	4
Interakcija med S in F	$SF_{111}, \cdots SF_{ijk}, \cdots, SF_{256}$	$2 \times 5 \times 6=$	60	(2-1)×5×(6-1)=	25
Regresijski koeficient za linearni člen	$b_{111}, \cdots, b_{ijk}, \cdots .. b_{256}$	$2 \times 5 \times 6=$	60	$2 \times 5 \times 6=$	60
Gnezdo	$\cdots g_{jkl} \cdots$		150		150
Interakcija med S in g	$\cdots Sg_{ijkl} \cdots$	$2 \times 150=$	300		300

Število opazovanj (1.24) lahko natančno izračunamo. Opravili smo 720 tehtanj.

$$N = 8\, pujskov\, x\, 3\, gnezda\, x\, 6\, svinj\, x\, 5\, merjascev\; = 720 \tag{1.46}$$

Najprej poskusimo ugotoviti število stopinj prosti pri ugnezdenih vplivih.

V modelu 1.45 imamo tri komponente variance:

- varianca za skupno okolje v gnezdu (σ_g^2),
- varianca za interakcijo med S in g (σ_{Sg}^2) in
- varianco za ostanek (σ_e^2).

1.6 Pričakovane vrednosti

Vzemimo model za debelino slanine (enačba 1.47), kjer smo vključili le vpliv pasme in mase kot sistematska vpliva, skupno okolje v gnezdu in aditivni genetski vpliv pa kot naključna.

$$y_{ijkl} = \mu + P_i + b\left(x_{ijk} - \bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl} \tag{1.47}$$

Pričakovano vrednost ali matematično upanje bomo označevali z veliko črko E, za njo pa v oklepaju navedemo spremenljivko, matematični izraz ali statistični model. Pričakovano vrednost opazovanja y_{ijk} bomo dobili tako, da opazovanje v izrazu nadomestimo z modelom [enačba 1.48].

$$E\left(y_{ijkl}\right) = E\left(\mu + P_i + b\left(x_{ijk} - \bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl}\right) = \tag{1.48}$$

Tabela 1.22: Prispevek sistematskih vplivov k pričakovani vrednosti

$E\left(y_{ijkl}\right) =$	od živali pričakujemo toliko,
$= E(\mu) +$	kot prispeva srednja vrednost
$+E(P_i) +$	+ kot prispeva pasma
$+\left(x_{ijk} - \bar{x}\right) E(b) +$	+ kot prispeva masa
$+\cdots$	

Operacije za izračun pričakovane vrednosti so podobne operacijam za vsoto, zato lahko desno stran razčlenimo (enačba 1.49).

$$= E(\mu) + E(P_i) + E\left(b\left(x_{ijkl} - \bar{x}\right)\right) + E\left(g_{ij}\right) + E\left(a_{ijk}\right) + E\left(e_{ijkl}\right) = \qquad [1.49]$$

Predno nadaljujemo, poskusimo opisati, kaj pričakovana vrednost v praksi pomeni (pregl. 1.22

V nadaljevanju poiščimo posamezne pričakovane vrednosti za posamezne člene iz enačbe 1.49.

1.6.1 Pričakovane vrednosti za konstanto

Pričakovana vrednost konstante (enačba 1.50 in 1.51) je konstanta.

$$E(c) = \frac{c + c + c + \cdots + c}{n} = \frac{nc}{n} = c \qquad [1.50]$$

Pričakovane vrednosti za konstante in sistematske spremenljivke so konstante. Konstante imajo določeno vrednost (npr. 1, 10, -2.4, 0), včasih paa jih označujemo s črkami na začetku abecede.

$$E(10) = \frac{10 + 10 + 10 + \cdots + 10}{n} = \frac{n \times 10}{n} = 10 \qquad [1.51]$$

1.6.2 Pričakovane vrednosti pri parametrih za sistematske vpliv

Pričakovana vrednost srednje vrednosti (enačba 1.52) je srednja vrednost. Ocena je nepristranska.

$$E(\mu) = \frac{\mu + \mu + \mu + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu \qquad [1.52]$$

Pričakovana vrednost za parameter (enačba 1.53), ki opisuje vpliv pasme 1, je parameter P_1. Dobimo jo tako, da za prispevke pasme pri meritvah pripadnikov te pasme izračunamo povprečje. Ocena je nepristranska.

$$E(P_1) = \frac{P_1 + P_1 + \cdots + P_1 + 0 + \cdots}{n_1} = \frac{n_1 P_1}{n_1} = P_1 \qquad [1.53]$$

Na isti način lahko dokažemo, da je pričakovana vrednost za parameter (enačba 1.54) za drugo pasmo (P_2) ali (enačba 1.55) katerokoli pasmo P_i ta-isti parameter.

$$E(P_2) = \frac{\cdots + 0 + P_2 + P_2 + \cdots + P_2 + 0 + \cdots}{n_2} = \frac{n_2 P_2}{n_2} = P_2 \qquad [1.54]$$

$$E(P_i) = \frac{P_i + P_i + \cdots + P_i}{n_i} = \frac{n_i P_i}{n_i} = P_i \qquad [1.55]$$

Kadar imamo neznanko pomnoženo s konstanto c, lahko konstanto izpostavimo (enačba 1.56) in tako dobimo, da je tudi končni rezultat produkt konstante z neznanko (parametrom).

$$E\left(3P_i\right) = 3E\left(P_i\right) = 3P_i \qquad [1.56]$$

Sistematske spremenljivke se obnašajo kot konstante, vendar pa imamo na voljo večjo zalogo vrednosti. Vsaka od teh vrednosti pa je konstanta. Kot primer si lahko ogledamo vpliv pasme. V slovenski populaciji prašičev imamo štiri tradicionalne pasme: slovenski landras, slovenski veliki beli prašič, pietren in slovenski mesnati landras. Za naključno izbranega prašiča pietren pričakujemo, da bo njegova proizvodnja enaka povprečni prireji pasme pietren. To velja, dokler nimamo dodatnih informacij, npr. meritve na prašiču ali sorodnikih. Vendar pa s temi dodatnimi informacijami opisujemo pogoje, v katerih redimo živali. Torej je nivo prireje vnaprej določen - sistematski (sistematski vplivi). Niso variabilni in med njimi ne obstaja podobnosti (korelacij, kovarianc). Če poznamo vrednosti nekaterih razredov sistematskih spremenljivk (pasem), ne moremo sklepati na lastnosti drugih.

Pri posameznih členih regresije imamo produkt regresijskega koeficienta (b) in neodvisne spremenljivke $\left(x_{ijkl} - \bar{x}\right)$. Pričakovana vrednost pomeni točko na krivulji, ki jo opisuje regresija, pri izbrani neodvisni spremenljivki x_{ijkl}. Da bi primerjali debelino slanine pri živalih, jo korigiramo na isto maso (izbrano vrednost z x_{ijkl}), ki je ista za vse živali. Prav zato jo lahko izpostavimo (enačb 1.57) in poiskati moramo pričakovano vrednost regresijskega koeficienta.

$$E\left(b\left(x_{ijkl} - \bar{x}\right)\right) = \left(x_{ijkl} - \bar{x}\right)E(b) \qquad [1.57]$$

Sedaj se lahko posvetimo pričakovane vrednosti za regresijski koeficient (enačba 1.58). V primeru, da je en sam regresijski koeficient, je enak za vse meritve. Pričakovano vrednost dobimo tako, da seštejemo regresijske koeficiente za vse meritve in delimo s številom meritev. Ponovno smo dobili parameter sam.

$$E\left(b\right) = \frac{b + b + b + \cdots + b}{n} = \frac{nb}{n} = b \qquad [1.58]$$

Kadar imamo regresijski koeficient ugnezden znotraj drugega vpliva (enačba 1.59), upoštevamo samo tiste meritve, ki pripadajo i-temu nivoju nadrejenega vpliva. Za vse meritve istega nivoja imamo isti regresijski koeficient (b_i), regresijske koeficiente pri drugih nivojih ne uporabimo pri izračunu pričakovane vrednosti za parameter $b_{i.}$.

$$E\left(b_i\right) = \frac{\cdots + 0 + b_i + b_i + b_i + \cdots + b_i + 0 + \cdots}{n_i} = \frac{n_i b_i}{n_i} = b_i \qquad [1.59]$$

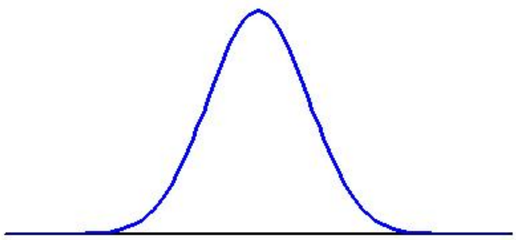
1.6.3 Pričakovana vrednost za ostanek

Omenili smo že, da so ostanki odstopanja od primerjane oz. povprečne vrednosti. Porazdelitev ostankov je pri številnih lastnostih, ki jih obravnavamo v živinoreji normalna s povprečjem 0 (slika 1.8

Odkloni so pozitivni in negativni, vsota ostankov (enačba 1.60) in s tem tudi povprečje (enačba 1.61) sta enaka 0.

$$\sum e_{ijkl} = 0 \qquad [1.60]$$

$$E\left(e_{ijkl}\right) = \frac{1}{n}\sum e_{ijkl} = \frac{1}{n}\times 0 = 0 \qquad [1.61]$$



Slika 1.8: Porazdelitev ostankov

1.6.4 Pričakovane vrednosti za parametre pri naključnih vplivih

Tudi lokacijski parametri pri naključnih vplivih so del odstopanj, ki je določen z razmerjem varianc. Dobili bomo negativne in pozitivne vrednosti in njihova vsota je praviloma 0. V enačbi 1.62 prikazujemo, da je vsota vrednosti parametrov za skupno okolje v gnezdu enaka 0.

$$\sum g_{ij} = 0 \qquad [1.62]$$

Tako nas ne čudi, da je tudi pričakovana vrednost za skupno okolje enaka 0.

$$E\left(g_{ij}\right) = \frac{1}{n_g}\sum g_{ij} = \frac{1}{n_g}\times 0=0 \qquad [1.63]$$

Tudi plemenska vrednost je del odstopanja, ki je določen z razmerjem med genetsko in fenotipsko varianco, poimenovanim tudi kot dednostni delež ali heritabiliteta. Vsota plemenskih vrednosti je zato 0 (enačba 1.64).

$$\sum a_{ijk} = 0 \qquad [1.64]$$

Tudi pričakovana vrednost za plemensko vrednost je pogosto 0 (enačba 1.65).

$$E\left(a_{ijk}\right) = \frac{1}{m}\sum a_{ijk} = \frac{1}{m}\times 0=0 \qquad [1.65]$$

Pri plemenski vrednosti imamo tudi izjemo, ko pričakovana vrednost ni enaka nič. Kadar dobimo živali iz različnih populacij, ki se genetsko razlikujejo, lahko izvor, časovno obdobje, namen in spol oblikujejo genetske skupine z različnimi pričakovanimi vrednostmi.

1.6.5 Uporabimo naučeno

Nadaljujmo z enačbo 1.48 in uporabimo naučeno. Osvežimo si prva dva koraka: pričakovano vrednost opazovanja poiščemo tako, da namesto opazovanja vstavimo desno stran modela in nato lahko izraz razčlenimo (enačba 1.66)

$$\begin{aligned} E\left(y_{ijkl}\right) &= E\left(\mu + P_i + b\left(x_{ijk} - \bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl}\right) = \\ &= E(\mu) + E(P_i) + E\left(b\left(x_{ijkl} - \bar{x}\right)\right) + E\left(g_{ij}\right) + E\left(a_{ijk}\right) + E\left(e_{ijkl}\right) = \end{aligned} \qquad [1.66]$$

Izpostavimo (enačba 1.67) še neodvisno spremenljivko, pri kateri bi radi izvrednotili pričakovano vrednost.

$$= E(\mu) + E(P_i) + \left(x_{ijkl} - \bar{x}\right)E(b) + E\left(g_{ij}\right) + E\left(a_{ijk}\right) + E\left(e_{ijkl}\right) = \qquad [1.67]$$

Sedaj pa v enačbi 1.68 vstavimo še rezultate, ki smo jih pravkar dobili za posamezne člene.

$$= \quad \mu + P_i + b\left(x_{ijkl} - \bar{x}\right) + 0 + 0 + 0 \qquad [1.68]$$

Za naključne spremenljivke smo privzeli, da je njihova pričakovana vrednost nič. Če pričakovana vrednost za naključni vpliv ni 0, lahko vedno razvijemo model, kjer bo ta pogoj izpolnjen. Tako tudi v živinoreji poznamo primere, ko je potrebno priznati, da pričakovana vrednost za naključni vpliv ni 0. Ko uvažamo genetski material (plemenske živali, seme plemenjakov) iz drugih rejskih programov, bi bilo nesmiselno in naivno trditi, da bomo kupili "enake" živali, kot jih imamo v svojih čredah. Kupili naj bi le boljše od naših, a časih kupimo tisto, kar v izvorni državi ne porabijo zase ali za boljše plačnike. Povprečna kakovost plemenskih živali iz različnih držav, glede na leto uvoza itd. je torej lahko drugačno, zato oblikujemo genetske skupine, v literaturi so poimenovane kot genetske skupine neidentificiranih staršev (*ang.* genetic groups of unknown parents, phanthom groups). V teh primerih pričakovana vrednost za plemenske vrednosti ni več 0. Na tej stopnji se bomo zadovoljili z informacijo, da obstajajo izjeme, kjer so pričakovane vrednosti za naključne vplive različne od 0, s takimi primeri pa zaenkrat še ne bomo ukvarjali.

Model je tako možno preurediti, da naša predpostavka drži. Pričakovana vrednost je tako in tako posledica nekega sistematskega vpliva. Namesto, da si otežujemo obdelavo, je primerneje ta sistematski vpliv vključiti v model. Naključne spremenljivke predstavljajo individualna odstopanja od pričakovane vrednosti, ki jo povzroča posamezni nivo. So variabilne in med njimi lahko nastopa podobnost (sorodne živali). Naključna spremenljivka je plemenska vrednost živali, ostanek, skupno okolje v gnezdu. Zaloge vrednosti so lahko končne ali neskončne množice, imajo pa (poznano) porazdelitev. Ko poznamo nekaj spremenljivk, lahko napovedujemo (*ang.* prediction) o lastnostih ostalih elementov množice (ostalih živali iz populacije). Vsekakor bodo napovedi boljše, če imamo veliko podatkov iz populacije. Za napovedovanje plemenskih vrednosti radi uporabimo meritve iz prireje npr. mleka, rasti itd.

1.6.6 Pričakovana vrednost pri pogoju

Pričakovana vrednost za katerokoli žival v populaciji je torej enaka pričakovani vrednosti populacije. Če tako želimo kupiti plemensko svinjo pasme slovenski landras in iščemo samo po oglasnikih, o kakovosti plemenske mladice nič ($E\left(a_{ijk}\right) = 0$) ne vemo. Vemo le, da lahko pri njej pričakujemo toliko pujskov (enačba 1.69), kot jih pričakujemo pri prašičih (μ) izbrane pasme (P_i).

$$E\left(\mu + P_i + g_{ij} + a_{ijk} + e_{ijkl}\right) = \mu + P_i \qquad [1.69]$$

Kako pa lahko bolje izberemo plemensko žival? Če je čreda, iz katere prihaja potencialna mladica $PUJSA\,PEPA$, v kontroli prireje, potem lahko za mladico zahtevamo njeno plemensko vrednost. Ko smo žival izbrali, pa se njena pričakovana vrednost spremeni. Če ima dobro plemensko vrednost, pričakujemo boljšo prirejo kot pri katerikoli svinji te pasme. Če bi imela slabo plemensko vrednost, pa bi imela slabšo prirejo kot ostale svinje iste pasme.

Sedaj pa to preverimo še z enačbami. Da nas zanima pričakovana vrednost, ko smo izbrali mladico z imenom $PUJSA\,PEPA$, zapišemo tako kot v enačbi 1.70

$$E\left(y_{ijkl} | a_{ijk} = \text{"}PUJSA\,PEPA\text{"}\right) = \qquad [1.70]$$

ali krajše v enačbi 1.71, kjer smo poudarili, da bomo izbirali živali. Uporabili smo nov znak - pokončno črto (|), kar pomeni "ob pogoju".

$$E\left(y_{ijkl} | a_{ijk}\right) = \qquad [1.71]$$

Nadaljujmo s krajšo verzijo. Opazovanje y_{ijkl} nadomestimo z desno stranjo modela (enačba 1.72)

$$= E\left(\mu + P_i + g_{ij} + a_{ijk} + e_{ijkl} | a_{ijk}\right) = \mu + P_i + a_{ijk} \quad [1.72]$$

in razstavimo (enačba 1.73).

$$= E\left(\mu | a_{ijk}\right) + E\left(P_i | a_{ijk}\right) + E\left(g_{ij} | a_{ijk}\right) + E\left(a_{ijk} | a_{ijk}\right) + E\left(e_{ijkl} | a_{ijk}\right) = \quad [1.73]$$

Pričakovana vrednost parametrov v sistematskem delu modela se zaradi izbire naključnega vpliva ne spremeni. Tako je

$$E\left(\mu | a_{ijk}\right) = \mu \quad [1.74]$$

in

$$E\left(P_i | a_{ijk}\right) = P_i \quad [1.75]$$

.

Tudi pričakovana vrednost za druge, neizbrane naključne vplive se ne spremeni (enačba 1.76), saj so vplivi neodvisni. Žival je res del skupine živali v gnezdu, a naključni vpliv živali je aditivna genetska vrednost živali, gnezdo, v katerem je bila ta žival rojena, pa je okolje v tem gnezdu (npr. temperatura v gnezdu, prepih ...)

$$E\left(g_{ij} | a_{ijk}\right) = E\left(g_{ij}\right) = 0 \quad [1.76]$$

Pričakovana vrednost za ostanek (enačba 1.77 se tudi ne spremeni, ko živali izberemo.

$$E\left(e_{ijkl} | a_{ijk}\right) = E\left(e_{ijkl}\right) = 0 \quad [1.77]$$

Ostal nam je samo člen (enačba 1.78), kjer iščemo pričakovano vrednost za plemensko vrednost, ko izberemo žival. Pričakujemo torej, da pri izbrani živali na njeno prirejo vpliva njena plemenska vrednost.

$$E\left(a_{ijk} | a_{ijk}\right) = a_{ijk} \quad [1.78]$$

Sedaj lahko ta spoznanja vnesemo v enačbo 1.73. Dobili smo pričakovano vrednost opazovanja y_{ijkl}ob pogoju, da smo izbrali žival.

$$= \mu + P_i + 0 + a_{ijk} + 0 \quad [1.79]$$

Prašiče smo vzeli kot primer le, ker smo hoteli pokazati, da se pri drugem naključnem vplivu za skupno okolje v gnezdu ni nič spremenilo.

POMNI! Ko izberemo nivo naključnega vpliva, pričakovana vrednost za izbrani nivo ni več le 0, ampak je vrednost tega nivoja (a_{ijk}). Ko izberemo nivo naključnega vpliva, se ta naključni vpliv "obnaša" kot sistematski.

1.6.7 Vaje za pričakovano vrednost

VAJA 9: Preverite, da velja naslednje

a) $E\left(y_{ijkl}|\, g_{ij}\right) = \mu + P_i + g_{ij}$

b)$E\left(y_{ijkl}|\, g_{ij}, a_{ijk}\right) = \mu + P_i + g_{ij} + a_{ijk}$

VAJA 10: Vzemimo naslednji model (enačba 1.80). Poiščite naslednje pričakovane vrednosti!

$$\begin{aligned} y_{ijklmn} = \quad & \mu \;+\; S_i \;+\; M_j \;+ SM_{ij} \;+\; b_{ijk}\left(x_{ijklm} - x_{min}\right) \;+ \\ & + h_k + \; g_{jkl} \;+\; Sg_{ijkl} \;+ Sh_{ik} + a_{ijklm} + \; e_{ijklmn} \end{aligned} \qquad [1.80]$$

a) $E(y_{ijklmn}) =$

b) $E(\,7M_j \;+ {}^1\!/_3 SM_{ij} \;+\; b_{ijk}\left(x_{ijklm} - x_{min}\right) \;+ h_k +\; 4g_{jkl} \;+\; Sg_{ijkl} \;+ Sh_{ik} + 3a_{ijklm}) =$

c) $E(y_{ijklmn}|\, g_{jkl}) =$

d) $E(y_{ijklmn}|\, h_k) =$

e) $E(y_{ijklmn}|\, Sh_{ik}) =$

f) $E(y_{ijklmn}|\, Sg_{ijkl}) =$

g) $E(y_{ijklmn}|\, a_{ijklm}) =$

h) $E(y_{ijklmn}|\, g_{jkl},\; a_{ijklm}) =$

i) $E(\,M_j \;+ {}^1\!/_3 SM_{ij} \;+\; b_{ijk}\left(x_{ijklm} - x_{min}\right) \;+ h_k +\; 4g_{jkl} \;+\; Sg_{ijkl} \;+ Sh_{ik} + 3a_{ijklm}|\, a_{ijklm}) =$

1.7 Struktura varianc in kovarianc

Tretji element modela se nanaša na strukturo varianc in kovarianc. Vedno izhajamo iz fenotipske variance in fenotipskih kovarianc.

$$var\left(y_{ijkl}\right) = \qquad [1.81]$$

Zanima nas torej varianca opazovanj in podobnost (kovarianca) med njimi. Najenostavneje se problema lotimo na ta način, da nanizamo opazovanja v vrstico in stolpec (slika 1.9) in jih potem po parih primerjamo.

Na sliki 1.9 dovoljujemo, da imajo opazovanja različno varianco (različno napako meritev) in tudi kovariance so lahko različne. Če poznamo variance in kovariance, potem lahko nastavimo sisteme enačb in jih rešimo. Tako dobimo ocene za sistematske parametre in napovedi za naključne lokacijske parametre.

Opazovanje nadomestimo z desno stranjo modela. Operacije so podobne kvadriranju veččlenika.

$$var\left(y_{ijkl}\right) = var\left(\mu + P_i + b\left(x_{ijk} - \bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl}\right) = \qquad [1.82]$$

Enačbo 1.82 razčlenimo.

$$\begin{gathered} = \quad var(\mu) + var(P_i) + var\left(b\left(x_{ijk} - \bar{x}\right)\right) + var\left(g_{ij}\right) + var\left(a_{ijk}\right) + var\left(e_{ijkl}\right) + \\ +2cov(\mu, P_i) + 2cov\left(\mu, b\left(x_{ijk} - \bar{x}\right)\right) + 2cov\left(\mu, g_{ij}\right) + 2cov\left(\mu, a_{ijk}\right) + 2cov\left(\mu, e_{ijkl}\right) + \\ +2cov\left(P_i, b\left(x_{ijk} - \bar{x}\right)\right) + 2cov\left(P_i, g_{ij}\right) + 2cov\left(P_i, a_{ijk}\right) + 2cov\left(P_i, e_{ijkl}\right) + \\ +2cov\left(b\left(x_{ijk} - \bar{x}\right), g_{ij}\right) + 2cov\left(b\left(x_{ijk} - \bar{x}\right), a_{ijk}\right) + 2cov\left(b\left(x_{ijk} - \bar{x}\right), e_{ijkl}\right) + \\ +2cov\left(g_{ij}, a_{ijk}\right) + 2cov\left(g_{ij}, e_{ijkl}\right) + \\ +2cov\left(a_{ij}, e_{ijk}\right) \end{gathered} \qquad [1.83]$$

	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n
y_1	σ_1^2	σ_{12}	$\cdots$	σ_{1i}	$\cdots$	σ_{1n}
y_2		σ_2^2	$\cdots$	σ_{2i}	$\cdots$	σ_{2n}
$\vdots$				$\vdots$		$\vdots$
y_i				σ_i^2	$\cdots$	σ_{in}
$\vdots$		Simetrična		$\vdots$	$\cdots$	$\vdots$
y_n					$\cdots$	σ_n^2

Slika 1.9: Matrika kovarianc med opazovanji

1.7.1 Varianca za parametre iz sistematskega dela modela

Za začetek poiščimo varianco srednje vrednosti (enačba 1.84. Ker še nismo spretni, je najbolje, da uporabimo kar definicijo za varianco: varianca je povprečni kvadratni odklon (od pričakovane vrednosti).

$$var\left(\mu\right) = \frac{\sum_i \sum_j \left(\mu - E\left(\mu\right)\right)^2}{s.p.} = \qquad [1.84]$$

Ko vstavimo v enačbo 1.85, da je pričakovana vrednost za srednjo vrednost kar srednja vrednost, vidimo, da v števcu seštevamo same ničle. Vsota kvadratnih odklonov je nič, število stopinj prostosti ($d.f.$) je več kot 0, zato je rezultat **0**.

$$= \frac{\sum_i \sum_j \left(\mu - \mu\right)^2}{s.p.} = \frac{0}{s.p.} = 0 \qquad [1.85]$$

Izpeljimo varianco še za sistematske vplive z razredi (enačba 1.86) tako, da ponovno uporabimo definicijo za varianco.

$$var\left(P_i\right) = \frac{\sum_i \sum_j \left(P_i - E\left(P_i\right)\right)^2}{s.p.} = \qquad [1.86]$$

Ponovno ugotavljamo (enačba 1.87), da je razlika v števcu enaka 0, zato sta tudi vsota kvadratnih odklonov in povprečni kvadratni odklon enaka 0.

$$= \frac{\sum_i \sum_j \left(P_i - P_i\right)^2}{s.p.} = \frac{0}{s.p.} = 0 \qquad [1.87]$$

S povsem istim postopkom (enačbi 1.88 in 1.89) lahko dokažemo, da je varianca prispevka kvantitativnega sistematskega vpliva tudi enaka 0.

$$var\left(b\left(x_{ij}-\bar{x}\right)\right)=\frac{\sum_j^n\left(b\left(x_{ij}-\bar{x}\right)-E\left(b\left(x_{ij}-\bar{x}\right)\right)\right)^2}{s.p.}= \qquad [1.88]$$

$$=\frac{\sum_j^{n_i}\left(b\left(x_{ij}-\bar{x}\right)-b\left(x_{ij}-\bar{x}\right)\right)^2}{s.p.}=\frac{0}{s.p.}=0 \qquad [1.89]$$

Stopinj prostosti nam niti ni bilo potrebno določati. Če je model primeren za niz podatkov, so stopinje prostosti zagotovo večje od 0.

Za sistematske vplive si velja zapomniti postopek, s katerim smo prišli do rezultata in seveda je praktično vedeti, da je varianca parametrov iz sistematskega dela modela enaka 0.

POMNI! Pri sistematskih vplivih izračunamo posamezne lokacijske parametre. Ker jih je malo, je nesmiselno govoriti o porazdelitvi.

$$var\left(sistematski\right)=0 \qquad [1.90]$$

$$var\left(cx\right)=cov\left(cx,\ cx\right)=c^2var\left(x\right)=c^2\sigma_x^2 \qquad [1.91]$$

$$cov\left(c_1x_1,\ c_2x_2\right)=c_1c_2cov\left(x_1,x_2\right)=c_1c_2\sigma_{x1x2}=c_1c_2\sigma_{x12} \qquad [1.92]$$

To si je lahko predstavljati pri kvalitativnih lastnostih (npr. pasme). Ne moremo jih razvrstiti po velikosti. Ko poznamo eno, ne vemo nič o drugi. Pri vsaki meritvi je nivo poznan in določen. Tako ni nič naključnega (varianca=0).

1.7.2 Varianca za ostanek

Čeprav formulo za varianco za ostanek že poznamo, ponovimo postopek (enačba 1.93), da znanje utrdimo.

$$var\left(e_{ijkl}\right)=\frac{\sum\left(e_{ijkl}-E\left(e_{ijkl}\right)\right)^2}{n-s.p._M}= \qquad [1.93]$$

Ker je pričakovana vrednost za ostanek enaka 0, kvadriramo ostanke, jih seštejemo in delimo s stopinjami prostosti za ostanek, kar je razlika med številom opazovanj (n) in številom stopinj prostosti za model ($s.p._M$).

$$=\frac{\sum\left(e_{ijkl}-0\right)^2}{n-s.p._M}=\frac{\sum e_{ijkl}^2}{n-s.p._M}=\sigma_e^2 \qquad [1.94]$$

1.7.3 Varianca za naključne vplive

Pri naključnih vplivih bomo postopali enako kot pri ostanku. Poiščemo odklone nivojev za skupno okolje v gnezdu od pričakovanih vrednosti (enačba 1.95).

$$var\left(g_{ij}\right) = \frac{\sum_i^p \sum_j^{m_i} \left(g_{ij} - E\left(g_{ij}\right)\right)^2}{n_g} = \qquad [1.95]$$

Zapis enačbe 1.95 je pravilen, a ker imamo pri naključnih vplivih običajno več indeksov, bomo v nadaljevanju v skalarni obliki poenostavili zapis vsote (enačba 1.96). Ugotovili smo že, da je pričakovana vrednost za skupno okolje v gnezdu (enačba 1.63) enaka 0. Odklon je torej enak vrednosti posameznega nivoja, vrednosti kvadriramo, seštejemo in delimo s številom gnezd (n_g). Tako dobimo varianco za skupno okolje v gnezdu, ki jo označimo s σ_g^2 .

$$= \frac{\sum \left(g_{ij} - 0\right)^2}{n_g} = \frac{\sum g_{ij}^2}{n_g} = \sigma_g^2 \qquad [1.96]$$

Podobno lahko sklepamo za vpliv živali oz. plemensko vrednost (enačba 1.97). Pričakovana vrednost za plemenske vrednosti je 0, zato je odklon enak kar plemenski vrednosti.

$$var\left(a_{ijk}\right) = \frac{\sum \left(a_{ijk} - E\left(a_{ijk}\right)\right)^2}{n_a} = \qquad [1.97]$$

Plemenske vrednosti kvadriramo, seštejemo in delimo s številom plemenskih vrednosti (n_a). V enačbi 1.98 smo dobili genetsko varianco (σ_a^2).

$$= \frac{\sum \left(a_{ijk} - 0\right)^2}{n_a} = \frac{\sum a_{ijk}^2}{n_a} = \sigma_a^2 \qquad [1.98]$$

POMNI! Pri naključnih vplivih je varianca večja od nič. Če je (skoraj) nič, tega naključnega vpliva ni smiselno imeti v modelu.

$$var\,(naključni) > 0$$

$$var\left(g_{ij}\right) = \sigma_g^2$$

$$var\left(a_{ijk}\right) = \sigma_a^2$$

1.7.4 Kovarianca med sistematskima vplivoma

Poiščimo kovarianco med srednjo vrednostjo in parametrom za izbran sistematski vpliv (npr. pasmo, enačba 1.99). Kovarianca je po definiciji povprečni produkt odklonov dveh spremenljivk. Odklone dobimo, da od parametra odštejemo njegovo pričakovano vrednost.

$$cov\left(\mu, P_i\right) = \frac{\sum \left(\mu - E\left(\mu\right)\right)\left(P_i - E\left(P_i\right)\right)}{s.p.} = \qquad [1.99]$$

Vstavimo pričakovane vrednosti (enačba 1.100) in poračunajmo! V obeh oklepajih dobimo vrednost 0, produkti so tudi 0 in, ko ničle seštejemo, je vsota tudi 0. Število stopinj prostosti mora biti večje od 0, da lahko števec delimo in za rezultat dobimo vrednost 0.

$$\frac{\sum \left(\mu - \mu\right)\left(P_i - P_i\right)}{s.p.} = \frac{\sum 0 \cdot 0}{s.p.} = 0 \qquad [1.100]$$

POMNI! Med sistematskimi vplivi ni podobnosti.

$$cov\left(sistematski, sistematski\right) = 0$$

1.7.5 Kovarianca med sistematskim in naključnim vplivom

Enačbo 1.101 za kovarianco nastavimo po definiciji!

$$cov\left(P_i, a_{ij}\right) = \frac{\sum \left(P_i - E\left(P_i\right)\right)\left(a_{ij} - E\left(a_{ij}\right)\right)}{s.p.} = \qquad [1.101]$$

Vstavimo pričakovane vrednosti (enačba 1.102) in poračunajmo! Ker je pričakovana vrednost parametrov iz sistematskega dela modela parameter sam, je razlika 0. Ne glede na to, koliko znaša drugi odklon, so vsi posamezni produkti 0 in prav tako tudi vsota. Število stopinj prostosti je večje od 0, zato je tudi končni rezultat 0.

$$= \frac{\sum \left(P_i - P_i\right)\left(a_{ij} - 0\right)}{s.p.} = \frac{\sum 0 \cdot a_{ij}}{s.p.} = 0 \qquad [1.102]$$

POMNI! Med sistematskimi in naključnimi vplivi ni podobnosti.

$$cov\left(sistematski, naključni\right) = 0$$

Kovarianca med sistematskima vplivoma in med sistematskim ter naključnim vplivom je tudi enaka vrednosti 0 zaradi značaja sistematskih vplivov (obnašajo se kot konstante).

1.7.6 Kovarianca med naključnim vplivom in ostankom

Nastavimo enačbo po definiciji!

$$cov\left(a_{ij},\, e_{ijk}\right) = \frac{\sum_i \sum_j \left(a_{ij} - E\left(a_{ij}\right)\right)\left(e_{ij} - E\left(e_{ij}\right)\right)}{n - sp_M} =$$

- Vstavimo pričakovane vrednosti in poračunajmo

$$= \frac{\sum_i \sum_j \left(a_{ij} - 0\right)\left(e_{ij} - 0\right)}{n - sp_M} =$$

$$= \frac{\sum_i \sum_j a_{ij}e_{ij}}{n - sp_M} = ? = 0$$

- je teoretično celo možna, v praksi nezaželena
- poskus izvedemo pošteno, nepristransko, potem je enaka 0

$$\boxed{cov\left(naključni, ostanek\right) = 0}$$

1.7.7 Kovarianca med naključnimi vplivi

$$\boxed{cov\left(naključni, naključni\right)\begin{cases} = 0 \\ \neq 0 \end{cases}}$$

- kovarianca = 0 ⇒vpliva sta neodvisna
- kovarianca ≠ 0 ⇒vpliva sta odvisna
  - direktni, maternalni in paternalni aditivni genetski vplivi
    oče in mati sta sorodna z živaljo
- Pogosto torej predpostavimo, da je 0

$$= var\left(a_{ij}\right) + var\left(e_{ijk}\right) + 2cov\left(a_{ij}, e_{ijk}\right) \qquad [1.103]$$

1.7.8 Uporabimo naučeno

Hoteli smo izpeljati strukturo fenotipske variance (enačba 1.82), po razčlenitvi v enačbi 1.83 pa smo si ogledali posamezne člene. Prepišimo enačbo 1.83, da bomo lažje sledili posameznim členom. Nad posamezne člene smo z rdečo barvo napisali rezultate, ki smo jih pred tem dobili.

$$\begin{aligned} var\left(y_{ijkl}\right) = \; & \overset{=0}{var(\mu)} + \overset{=0}{var(P_i)} + \overset{=0}{var\left(b\left(x_{ijk}-\bar{x}\right)\right)} + \overset{=\sigma_g^2}{var\left(g_{ij}\right)} + \overset{=\sigma_a^2}{var\left(a_{ijk}\right)} + \overset{=\sigma_e^2}{var\left(e_{ijkl}\right)} + \\ & +2\overset{=0}{cov(\mu, P_i)} + 2\overset{=0}{cov\left(\mu, b\left(x_{ijk}-\bar{x}\right)\right)} + 2\overset{=0}{cov\left(\mu, g_{ij}\right)} + 2\overset{=0}{cov\left(\mu, a_{ijk}\right)} + 2\overset{=0}{cov\left(\mu, e_{ijkl}\right)} + \\ & +2\overset{=0}{cov\left(P_i, b\left(x_{ijk}-\bar{x}\right)\right)} + 2\overset{=0}{cov\left(P_i, g_{ij}\right)} + 2\overset{=0}{cov\left(P_i, a_{ijk}\right)} + 2\overset{=0}{cov\left(P_i, e_{ijkl}\right)} + \\ & +2\overset{=0}{cov\left(b\left(x_{ijk}-\bar{x}\right), g_{ij}\right)} + 2\overset{=0}{cov\left(b\left(x_{ijk}-\bar{x}\right), a_{ijk}\right)} + 2\overset{=0}{cov\left(b\left(x_{ijk}-\bar{x}\right), e_{ijkl}\right)} + \\ & +2\overset{=0}{cov\left(g_{ij}, a_{ijk}\right)} + 2\overset{=0}{cov\left(g_{ij}, e_{ijkl}\right)} + 2\overset{=0}{cov\left(a_{ij}, e_{ijk}\right)} \end{aligned} \quad [1.104]$$

Dobili smo, da je fenotipska varianca ($var\left(y_{ijkl}\right)$) enaka vsoti varianc za skupno okolje v gnezdu (σ_g^2), genetski varianci (σ_a^2) in varianci za ostanek (σ_e^2). Če so opazovanja opravljena v isti populaciji in na isti način so praviloma identično porazdeljene. Vse tri komponente variance in njihova vsota - fenotipska varianca - tako veljajo za vse meritve. Rečemo tudi, da so variance homogene.

$$= \sigma_g^2 + \sigma_a^2 + \sigma_e^2 \quad [1.105]$$

V tem primeru so bile vse kovariance enake 0, vendar se na to ne moremo zanesti. Pri vsaki kovarianci posebej moramo razmisliti, če je res enaka 0. Predno si ogledamo nekatere primere kovarianc, ki so različne od nič, poskusimo postopek skrajšati.

1.7.9 Krajši postopek izpeljave strukture varianc in kovarianc

Vzemimo isti primer. Fenotipska varianca ($var\left(y_{ijkl}\right)$) je enaka varianci odstopanja od pričakovane vrednosti (enačba 1.69).

$$var\left(y_{ijkl}\right) = var\left(y_{ijkl} - E\left(y_{ijkl}\right)\right) = \quad [1.106]$$

Zamenjajmo opazovanje z desno stranjo modela in pričakovano vrednostjo opazovanja.

$$= var\left(\mu + P_i + b\left(x_{ijk}-\bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl} - \mu - P_i - b\left(x_{ijk}-\bar{x}\right)\right) = \quad [1.107]$$

Ko izraz poračunamo, ugotovimo, da iščemo varianco samo za naključni del modela (enačba 1.108).

$$= var\left(g_{ij} + a_{ijk} + e_{ijkl}\right) = \quad [1.108]$$

Ta izraz bomo sedaj zelo hitro razčlenili (enačba 1.109.

$$= var\left(g_{ij}\right) + var\left(a_{ijk}\right) + var\left(e_{ijkl}\right) + 2cov\left(g_{ij}, a_{ijk}\right) + 2cov\left(g_{ij}, e_{ijkl}\right) + 2cov\left(a_{ijk}, e_{ijkl}\right) = \quad [1.109]$$

Sedaj pa vstavimo še parametre za varianco, kovariance pa seveda tudi pri krajšem postopku enake 0. Dobili smo isti rezultat kot v enačbi 1.105.

$$= \sigma_g^2 + \sigma_a^2 + \sigma_e^2 + 2\times 0 + 2\times 0 + 2\times 0 \quad [1.110]$$

Daljši in krajši postopek dasta iste rezultate. Daljši postopek smo ubrali zato, da smo se naučili nekatera pravila, ki jih lahko uporabimo pri razmišljanju, kar se nam postopek zatakne. Pri svojem delu lahko vedno uporabimo krajši postopek, nikakor pa ju ne smemo pomešati. Zelo pomembno je, da natančno pišete oznake, indekse, postavljate oklepaje. Enačbe morajo držati!

Na ta način lahko preverimo strukturo fenotipske variance za vsa opazovanja, ki smo jih napisali na diagonali na sliki 1.9.

1.7.10 Fenotipska kovarianca med opazovanji

Sedaj pa preverimo nekatere fenotipske kovariance (enačba 1.111), kjer kovarianca obstaja. Vzemimo dve opazovanji. Drugo opazovanje $y_{i'j'k'l'}$ ima indekse z drugimi vrednostmi.

$$cov\left(y_{ijkl}, y_{i'j'k'l'}\right) = \qquad [1.111]$$

Vzemimo krajši postopek, pri katerem tudi pri kovarianci lahko od opazovanj najprej odštejemo njihove pričakovane vrednosti (enačba 1.112).

$$= cov\left(y_{ijkl} - E\left(y_{ijkl}\right), y_{i'j'k'l'} - E\left(y_{i'j'k'l'}\right)\right) = \qquad [1.112]$$

Pričakovano vrednost opazovanja predstavlja sistematski del enačbe modela in, ko jo odštejemo od enačbe modela, nam ostaneta naključna dela (enačba 1.113

$$= cov\left(g_{ij} + a_{ijk} + e_{ijkl}, g_{i'j'} + a_{i'j'k'} + e_{i'j'k'l'}\right) = \qquad [1.113]$$

Sedaj izraz 1.114 razčlenimo.

$$\begin{aligned} &= cov\left(g_{ij}, g_{i'j'}\right) + cov\left(g_{ij}, a_{i'j'k'}\right) + cov\left(g_{ij}, e_{i'j'k'l'}\right) + \\ &+ cov\left(a_{ijk}, g_{i'j'}\right) + cov\left(a_{ijk}, a_{i'j'k'}\right) + cov\left(a_{ijk}, e_{i'j'k'l'}\right) + \\ &+ cov\left(e_{ijkl}, g_{i'j'}\right) + cov\left(e_{ijkl}, a_{i'j'k'}\right) + cov\left(e_{ijkl}, e_{i'j'k'l'}\right) = \end{aligned} \qquad [1.114]$$

V izrazu 1.114 smo kovariance med parametri dveh različnih naključnih vplivov obarvani rdeče. Vse rdeče obarvane kovariance so enake 0, kar smo se že naučili. Sedaj pa si oglejmo še člene, ki so obarvani zeleno. Začnimo s skupnim okoljem v gnezdu. Če sta meritvi vzeti na živalih iz različnih gnezd (druga vrstica,kjer velja $i \neq i' \vee j \neq j'$), okolji nista podobni in je kovarianca med njima enaka 0. Če pa sta meritvi iz istega gnezda (prva vrstica, kjer velja $i = i' \wedge j = j'$), pa sta meritvi podobni. Črtice pri indeksih za drugo gnezdo so nepotrebne. Ko i' in j' nadomestimo z i -jem in j-jem, dobimo kovarianco med okoljem znotraj istega gnezda, kar pa je enako varianci za skupno okolje v gnezdu (σ^2_g;).

$$cov\left(g_{ij}, g_{i'j'}\right) = \begin{cases} cov(g_{ij}, g_{ij}) = \sigma^2_g; & i = i' \wedge j = j' \\ 0; & i \neq i' \vee j \neq j' \end{cases} \qquad [1.115]$$

Poglejmo si sedaj tudi genetsko kovarianco (izraz 1.116. Kadar sta živali nesorodni (tretja vrstica), nimata nič skupnega in je podobnost med njima enaka 0. Kadar gre za isto žival, imamo samo njeno plemensko vrednost in je kovarianca enaka genetski varianci (zgornja vrstica). Pri genetski kovarianci pa imamo še tretjo možnost (srednja vrstica), ko sta živali med seboj sorodni. Sorodstvo v tem primeru opišemo z deležem skupnih genov ($a_{kk'}$). Tako npr. imata starš in njegov potomec 1/2 skupnih genov, prav tako pravi sestri/brata, pol-sestri/pol-brata pa imata 1/4 skupnih genov, kar velja tudi za par stari starš - vnuk, stric/teta in nečak/nečakinja. Delež skupnih genov za ožje sorodnike, brez parjenja v sorodu lahko skoraj uganemo, izračun za zahtevnejše primere pa boste obravnavali pri kvantitativni genetiki.

$$cov\left(a_{ijk}, a_{i'j'k'}\right) = \begin{cases} cov(a_{ijk}, a_{ijk}) = \sigma^2_a; & ista\,žival \\ a_{kk'}\sigma^2_a; & sorodnika \\ 0; & nesorodni\,živali \end{cases} \qquad [1.116]$$

1.7.11 Fenotipska varianca ob pogoju

Še vedno ostanimo pri modelu 1.47. Za fenotipsko varianc0 ($var\left(y_{ijkl}\right)$) smo prikazali strukturo v enačbi 1.69). Videli smo tudi, da se spremeni pričakovana vrednost, če pri naključnem vplivu izberemo nivo ($E\left(y_{ijkl}|a_{ijk} = \text{"}PUJSA\ PEPA\text{"}\right)$). Izbrali smo torej svinjo "$PUJSA\ PEPA$" in dobili pričakovano vrednost. Sedaj pa bomo preverili, če se spremeni tudi fenotipska varianca (izraz 1.106).

$$var\left(y_{ijkl}|a_{ijk} = \text{"}PUJSA\ PEPA\text{"}\right) = \qquad [1.117]$$

Zapis lahko skrajšamo za izbrano ime, kot smo to storili tudi pri pričakovani vrednosti. Zanima nas torej fenotipska varianca ob pogoju, da smo izbrali žival a_{ijk} (izraz 1.118). Izberemo lahko tudi nekaj živali: morda nas zanimajo plemenjaki iz osemenjevalnega centra ali plemenice iz naše črede.

$$= var\left(y_{ijkl}|a_{ijk}\right) = \qquad [1.118]$$

Pojdimo po krajši poti. Fenotipsko vrednost y_{ijkl} nadomestimo z odstopanjem fenotipske vrednosti od pričakovane vrednosti (izraz 1.119). Izvrednotiti želimo (fenotipsko) varianco ob pogoju, da izberemo žival, zato bomo upoštevali tudi pričakovano vrednost ob istem pogoju (enačba 1.79).

$$= var\left(y_{ijkl} - E\left(y_{ijkl}|a_{ijk}\right)|a_{ijk}\right) = \qquad [1.119]$$

Zamenjajmo opazovanje z desno stranjo modela in pričakovano vrednostjo opazovanja (enačba 1.120).

$$= var\left(\mu + P_i + b\left(x_{ijk} - \bar{x}\right) + g_{ij} + a_{ijk} + e_{ijkl} - \mu - P_i - b\left(x_{ijk} - \bar{x}\right) + a_{ijk}|a_{ijk}\right) = \qquad [1.120]$$

Ko izraz poračunamo, ugotovimo, da iščemo varianco samo za naključni del modela brez aditivnega genetskega vpliva (enačba 1.121).

$$= var\left(g_{ij} + e_{ijkl}|a_{ijk}\right) = \qquad [1.121]$$

Ta izraz bomo sedaj zelo hitro razčlenili (enačba 1.122).

$$= var\left(g_{ij}|a_{ijk}\right) + var\left(e_{ijkl}|a_{ijk}\right) + 2cov\left(g_{ij}, e_{ijkl}|a_{ijk}\right) = \qquad [1.122]$$

Izbira nivojev naključnega vpliva pomeni, da se izbrani nivo obnaša kot sistematski vpliv. Ko imamo izbrano žival, pri njenih meritvah ne pričakujemo odstopanj zaradi plemenske vrednosti, saj je v primeru izbrane živali genotip določen. Na ostale variance pa izbor živali ne vpliva. Tako velja, da varianca za naključne vplive, ki jih nismo izbrali, ostaja enaka kot v izhodiščnem modelu (enačba 1.123).

$$var\left(g_{ij}|a_{ijk}\right) = var\left(g_{ij}\right) = \sigma_g^2 \qquad [1.123]$$

Tudi varianca za ostanek se ne spremeni (enačba 1.124), če izberemo naključni vpliv ali ne.

$$var\left(e_{ijkl}|a_{ijk}\right) = var\left(e_{ijkl}\right) = \sigma_e^2 \qquad [1.124]$$

Preostale kovariance, ki ne vključujejo izbranega naključnega vpliva, ostanejo enake kot

$$cov\left(g_{ij}, e_{ijkl}|a_{ijk}\right) = cov\left(g_{ij}, e_{ijkl}\right) = 0 \qquad [1.125]$$

Sedaj pa vstavimo še parametre za varianci, kovarianca med skupnim okoljem v gnezdu in napako ostaja enaka 0. Rezultat vidimo v izrazu 1.126

$$= \sigma_g^2 + \sigma_e^2 + 2\times 0 \qquad [1.126]$$

Daljši in krajši postopek dasta iste rezultate. Daljši postopek smo ubrali zato, da smo se naučili nekatera pravila, ki jih lahko uporabimo pri razmišljanju, kar se nam postopek zatakne. Pri svojem delu lahko vedno uporabimo krajši postopek, nikakor pa ju ne smemo pomešati. Zelo pomembno je, da natančno pišete oznake, indekse, postavljate oklepaje. Enačbe morajo držati!

Na ta način lahko preverimo strukturo fenotipske variance za vsa opazovanja, ki smo jih napisali na diagonali na sliki 1.9.

1.7.12 Vaje za strukturo varianc in kovarianc

VAJA 11: Preverite, da velja naslednje

a) $var\left(y_{ijkl}|\, g_{ij}\right) = \sigma_a^2 + \sigma_e^2 + 2\times 0$

b)$var\left(y_{ijkl}|\, g_{ij}, a_{ijk}\right) = \sigma_e^2$

c)$cov\left(y_{ijkl}, y_{ijkl'}\right) = \sigma_g^2 + \sigma_a^2 + \sigma_{ell'}$

d)$cov\left(y_{ijkl}, y_{ijk'l'}\right) = \sigma_g^2 + \frac{1}{2}\sigma_a^2 + \sigma_{ell'}$

e)$cov\left(y_{ijkl}, y_{ijkl'}|a_{ijk}\right) = \sigma_g^2 + \sigma_{ell'}$

f)$cov\left(y_{ijkl}, y_{ijkl'}|g_{ij}\right) = \sigma_a^2 + \sigma_{ell'}$

VAJA 12: Vzemimo naslednji model (enačba 1.127). Poiščite strukturo za naslednje variance in kovariance!

$$\begin{aligned} y_{ijklmn} = \quad & \mu + S_i + M_j + SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + \\ & + h_k + g_{jkl} + Sg_{ijkl} + Sh_{ik} + a_{ijklm} + e_{ijklmn} \end{aligned} \qquad [1.127]$$

a) $var(y_{ijklmn}) =$

b) $cov(y_{ijklmn}, y_{ijklmn'}) =$

c) $cov(y_{ijklmn}, y_{ijklm'n'}) =$; kjer sta starša ista

d) $cov(y_{ijklmn}, y_{ijkl'm'n'}) =$; kjer je mati druga, oče pa isti

e) $cov(y_{ijklmn}, y_{ijkl'm'n'}) =$; kjer sta živali nesorodni

f) $cov(y_{ijklmn}, y_{ijklm'n'}) =$; kjer sta živali nesorodni

g) $var(y_{ijklmn}|\, a_{ijklm}) =$

h) $var(y_{ijklmn}|\, g_{jkl}) =$

i) $var(y_{ijklmn}|\, h_k) =$

j) $var(y_{ijklmn}|\, Sh_{ik}) =$

k) $var(y_{ijklmn}|\, Sg_{ijkl}) =$

l) $var(y_{ijklmn}|\, g_{jkl},\, a_{ijklm}) =$

m) $cov(y_{ijklmn}, y_{ijklmn'}|\, a_{ijklm}) =$

n) $cov(y_{ijklmn}, y_{ijklm'n'}|\, a_{ijklm}) =$; kjer sta starša ista

o) $cov(y_{ijklmn}, y_{ijkl'm'n'}|\, a_{ijklm}) =$; kjer je mati druga, oče pa isti

p) $cov(y_{ijklmn}, y_{ijkl'm'n'} | a_{ijklm})$ =; kjer sta živali nesorodni

q) $cov(y_{ijklmn}, y_{ijklm'n'} | a_{ijklm})$ =; kjer sta živali nesorodni

r) $cov(y_{ijklmn}, y_{ijklm'n'} | g_{jkl})$ =; kjer sta živali nesorodni

s) $cov(y_{ijklmn}, y_{ijklm'n'} | Sg_{ijkl})$ =; kjer sta živali nesorodni

t) $cov(y_{ijklmn}, y_{ijkl'm'n'} | Sh_{ik})$ =; kjer je mati druga, oče pa isti

u) $cov(y_{ijklmn}, y_{ijkl'm'n'} | h_k)$ =; kjer je mati druga, oče pa brata

VAJA 13: Za isti model poiščite strukturo za naslednje variance in kovariance! Konstante so izbrane za lažjo preveritev izpeljave!

a) $var(7M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm}) =$

b) $cov(7M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm},$

$$5M_j + 1/2SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 3g_{jkl} + Sg_{ijkl} + Sh_{ik} + 1/4a_{ijklm}) =$$

c) $cov(7M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl'} + Sg_{ijkl'} + Sh_{ik} + 3a_{ijkl'm'},$

$5M_j + 1/2SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 3g_{jkl} + Sg_{ijkl} + Sh_{ik} + 1/4a_{ijklm})$ =; če je mati ista, očeta pa različna

d) $var(M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} | a_{ijklm}) =$

e) $cov(M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm},$

$$5M_j + 1/2SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 3g_{jkl} + Sg_{ijkl} + Sh_{ik} + 1/4a_{ijklm} | a_{ijklm}) =$$

f) $cov(M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl'} + Sg_{ijkl'} + Sh_{ik} + 3a_{ijkl'm'},$

$5M_j + 1/2SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 3g_{jkl} + Sg_{ijkl} + Sh_{ik} + 1/4a_{ijklm})$ =; če je mati ista, očeta pa različna

g) $var(S_i + 7M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn}) =$

h) $var(S_i + M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn} | a_{ijklm}) =$

i) $cov(S_i + M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn},$

$$S + M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn} | a_{ijklm}) =$$

j) $cov(S_i + M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn},$

$$S_i + M_j + 1/3SM_{ij} + b_{ij}\left(x_{ijklm} - x_{min}\right) + h_k + 4g_{jkl} + Sg_{ijkl} + Sh_{ik} + 3a_{ijklm} + 4/7e_{ijklmn} | Sg_{ijk}) =$$

1.8 Predpostavke in restrikcije

Nekaj predpostavk smo omenili in upoštevali že pri izpeljavi strukture varianc in kovarianc. Teh predpostavk nam ni potrebno ponavljati. Ko model spremenimo, se lahko spremenijo tudi predpostavke. Poglejmo si najprej nekaj običajnih predpostavk:

- Pogosto predpostavimo porazdelitve naključnih spremenljivk in ostankov. Najraje imamo naravno porazdelitev, pa čeprav je porazdelitev morda koničasta, nagnjena ali celo diskretna. Tako npr. število živorojenih pujskov v gnezdu nima normalne porazdelitve, pač pa Poissonovo, saj ne more imeti svinja v enem gnezdu 13,46 pujska. Obdelava podatkov je ob predpostavki, da so naključne spremenljivke normalno porazdeljene, preprostejša, kot če bi hoteli delati s pravo porazdelitvijo, rezultati pa so pri številu živorojenih pujskih v gnezdu enako dobri.

- Predpostavimo tudi, da so ostanki (napake meritev) so med seboj nekorelirani, če niso merjeni na isti enoti (živali). Pri izvajanju poskusov in meritvah moramo skrbeti, da so napake, ki jih ne moremo preprečiti med seboj neodvisne. Nekatere merilne naprave moramo med meritvami umerjati, tehtnico moramo počistiti, če žival na njej blati, živalim moramo preprečiti krajo krme ...
- Za živali v poskusu predpostavimo, da so sorodne, kot je zbrano v podatkih o poreklu, ostale niso sorodne. Pri izhodiščni (osnovni) populaciji vemo, da so med njimi najbrž tudi sorodniki, a podatkov o tem nimamo. Tudi sicer se lahko na osnovi poskusov odločimo, da bomo nekatere oddaljene sorodnike izbrisali, ker k obdelavi ne pripomorejo bistveno, zahteva pa se veliko več računalniških zmogljivosti. Tako se tem sorodnikom lahko odpovemo, a moramo predpostaviti, da so njihovi potomci, ki ostanejo v obdelavi in imajo izbrisane prednike, nesorodni.
- Dobri rejci se izogibajo parjenju v (ožjem) sorodstvu. Potomci takega parjenja so inbridirani. Tako kot pri prejšnji alineji upoštevamo v obdelavi samo tisti inbriding, ki ga je moč prepoznati izračunati) iz pripravljenih podatkov o poreklu. Za ostale živali predpostavljamo, da niso inbridirane ...
- Skupina živali iz skupnega okolja si delijo to okolje in so si zaradi tega bolj podobne, kar opiše varianca za to skupno okolje. Okolja pri različnih nivojih pa si niso podobna, kovarianca med njimi je enaka 0. Poglejmo dva primera
  - Mladiči iz istega gnezda imajo skupno okolje. Če se je pokvaril grelec, je zeblo vse pujske, če je imela svinja malo mleka, so pujski bili lačni. Težave enega gnezda se pojavijo naključno. Tako ni veliko verjetno, da bo grelec v sosednjih gnezdih deloval bolje ali pa pregorel, ker se je v opazovanem gnezdu pokvaril grelec. Isto lahko trdimo za mlečnost svinje. Raznolikost med gnezdi opisuje varianca za skupno okolje v gnezdu, kovarianca med skupnima okoljema v gnezdu pa je 0.
  - Če pa je pujske zeblo, ker je zmanjkalo elektrike in so vsi grelci prenehali greti, pa je bil vzrok v skupnem okolju v čredi. Prav tako se npr. razlikuje kakovost krme, ker je bila (ne)ugodna rastna doba ali pa težave pri spravilu. V tehnologiji reje, sami oskrbi, kakovosti krme, higiene v hlevu se kmetije razlikujejo. Morda imamo celo reje, ki posnemajo nekatere druge kmetije, sledijo nasvete istega svetovalca itd., vendar teh povezav ne poznamo in je težko oceniti in upoštevati. Zato pogosto predpostavimo, da so okolja v čredah neodvisna, kovarianca med njimi je enaka 0, varianca za skupno okolje v čredi pa opisuje razpršenost nivojev.
- Podobnost med fenotipskimi vrednostmi povzroča lahko še permanentno (trajno) okolje pri več ponovitvah na isti enoti. Tako lastnosti mlečnosti posamezne krave ne določajo samo genetski vplivi, ampak tudi okolje v času vzreje telice in okolje v času priprave krave na naslednjo telitev. Vzreja živali v mladosti trajno vpliva na njeno prirejo. Če je bila vzreja odlična, pričakujemo, da bodo lastnosti boljše v njeni celi življenjski dobi. Če pa bo vzreja pomanjkljiva, bo krava imela manjšo prirejo. Razpršenost permanentnih okolij opisuje varianca, različna permanentna okolja pa so neodvisna. Kovarianca med njimi je 0.

Kadar izpustimo naključni vpliv iz modela, bo "vpliv" pristal največkrat v ostanku, kar naj bi povzročilo kovarianco med ostanki pri meritvah iz skupnega okolja. Kljub temu, da so dejansko ostanki korelirani, predpostavimo, da niso. Če tega nočemo, moramo v model vrniti skupno okolje.

Pogosto predpostavimo, da obstajajo kovariance le med nivoji, razredi naključnih spremenljivk. Če pa imamo v modelu več aditivnih genetskih vplivov (neposredni, maternalni, paternalni), pa si te predpostavke ne smemo privoščiti! Če je med pomembnimi maternalnimi lastnostmi tudi mlečnost matere, neposredni genetski vpliv na rast potomca ni neodvisen. Hranila se v telesu potrebujejo tako za nalaganje mišičnine kot prirejo mleka. Metabolizem je skupen in

si konkurirata ali pomagata. Torej nista neodvisni.

V primeru pa, da želimo oceniti iz podatkov tudi variance (napake meritev) in kovariance, pa bi pri taki strukturi imeli več enačb kot parametrov. Da lahko ocenimo varianco, rabimo ponovitve - poizkuse pod istimi pogoji. To pa pomeni, da je struktura varianc bolj urejena.

Pri poskusu predpostavimo, katere kovariance so lahko nič, ostale pa moramo bodisi poznati bodisi zagotoviti primerno strukturo podatkov za analizo.

Predpostavke o porazdelitvi, ki niso opisane z zgornjimi parametri.

ID - identično porazdeljene spremenljivke (vse smo izmerili z istim ostankom)

IID- identično in neodvisno (*ang.* independent - ostanki med nivoji niso korelirani) porazdeljene spremenljivke

IIDN - IID normalno porazdeljene spremenljivke

1. Opiši konstantne pogoje (standardizirano okolje), npr. vse merjene živali so istega spola in starosti. Drugi vplivi morajo biti v modelu.
2. Opiši časovne spremenljivke (sezona)
3. Pomni, da so neomenjeni vplivi obravnavni od bralcev kot konstantni ali pa vzbujajo sume o zasnovi poskusa.

Lahko je tudi več točk - odvisno od modela. Model ni popoln, če niso opisani prav vsi deli modela. Le tako je možno poznati vse omejitve pri modelu ali poskusu. Da bi bil model boljši, je pogosto potrebno poskus znova zastaviti, kar pa bi morda bilo predrago in dolgotrajno, če bi ga hoteli spremeniti (dopolniti) ali ponoviti. Četudi nam v poskusu ne gre vse po načrtih, pa se lahko iz njega kaj naučimo in to sporočimo tudi drugim. Tako lahko še vedno oddamo diplomo, prispevek za revijo. Edini predpogoj je, da smo se pri opravljenem poizkusu trudili in nismo bili malomarni, ali prikrojili podatke. Torej model bi morali napisati pred postavitvijo poskusa. S tem bi morda uspeli izboljšati plan poskusa, hkrati pa bi si zagotovili, da bi pravilno postavili in testirali hipoteze.

1.9 Vaje iz pisanja in razvoja statističnih modelov

V poskusu (pregl. 1.23) smo v dvajsetih čredah dve leti proučevali vpliv dopolnilne krme za breje in doječe ovce na rojstno in odstavitveno maso jagnjet ter dnevni prirast. Pri tem smo za vsako jagnje zabeležili: rejca, sezono rojstva, dopolnilno krmo ovce v času brejosti, število jagnjet v gnezdu, spol, zaporedno jagnjitev kot jagnjitev mladice (M) ali stare ovce (S), očeta, mater ter starost ob odstavitvi (dni). Jagnjeta smo stehtali ob rojstvu ter ob odstavitvi (v kg, na 100 g natančno).

Naloga 14:

Upoštevajte vpliv rejca in žival kot naključna vpliva!

a) Naštejte odvisne spremenljivke!

b) Naštejte glavne vplive za rojstno maso, jih označite in opišite v skladu z dogovorom!

c) Napišite osnovni model za rojstno maso jagnjet!

d) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc v osnovnem modelu za rojstno maso!

e) Napišite osnovni model za odstavitveno maso jagnjet! Opišite vse oznake v modelu!

f) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc v osnovnem modelu za odstavitveno maso!

g) Razvijte možni model za odstavitveno maso! Napišite vse elemente možnega modela!

Tabela 1.23: Podatki iz preizkusa dopolnilne krme pri doječih ovcah

Rejec	Sezona jagnjitve	Ovca mati	Zap. jag.	Dop. krma	Oče	Štev. živor.	Jag-nje	Spol jag.	Rojstna masa (kg)	Odstavitev star.	Odstavitev masa	Odstavitev DP (g/dan)
MK	2018-03	14	M	1	12	1	1	m	3.2	120	25.2	183
MK	2017-03	15	S	1	13	2	2	m	3.5	115	30.6	236
MK	2017-03	15	S	1	13	2	3	z	3.4	115	28.7	220
MK	2017-04	16	S	2	13	1	4	z	3.6	125	34.3	270
MK	2018-03	17	M	2	12	2	5	z	3.9	130	34.8	238
MK	2018-04	17	M	2	12	2	6	m	2.8	118	24.3	182
AN	2018-03	18	S	2	13	1	7	z	3.7	122	29.9	215
AN	2018-04	19	S	3	12	3	8	m	3.4	115	26.4	200
AN	2018-04	19	S	3	12	3	9	z	3.9	114	27.3	205
AN	2018-04	19	S	3	12	3	10	z	3.5	114	28.7	221
AN	2017-03	20	M	3	13	1	11	m	3.6	132	27.8	183

h) Napišite vse elemente osnovnega modela za dnevni prirast! Opišite vse oznake v modelu!

i) Razvijte možni model za dnevni prirast!

j) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc v možnem modelu za dnevni prirast!

k) V vseh modelih določite število parametrov in stopinj prostosti za posamezne vplive, model in ostanek!

Naloga 15:

Upoštevajte vpliv rejca kot sistematski vpliv in žival kot naključni vpliv!

a) Naštejte odvisne spremenljivke!

b) Naštejte glavne vplive za rojstno maso, jih označite in opišite v skladu z dogovorom!

c) Napišite osnovni model za rojstno maso jagnjet!

d) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc za rojstno maso jagnjet!

e) Naštejte glavne vplive za odstavitveno maso jagnjet, jih označite in opišite v skladu z dogovorom!

f) Napišite vse elemente v osnovnem modelu za odstavitveno maso jagnjet!

g) Razvijte možni model za odstavitveno maso jagnjet!

h) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc v možnem modelu za odstavitveno maso!

i) Ponovite vajo še za dnevni prirast!

j) V vseh modelih določite število parametrov in stopinj prostosti za posamezne vplive, model in ostanek!

Naloga 16:

V model vključite samo sistematske vplive!

a) Naštejte odvisne spremenljivke!

b) Naštejte glavne vplive za rojstno maso, jih označite in opišite v skladu z dogovorom!

c) Napišite osnovni model za rojstno maso jagnjet!

d) Poiščite pričakovano vrednost in izpeljite strukturo varianc in kovarianc za maso!

e) Naštejte glavne vplive za odstavitveno maso, jih označite in opišite v skladu z dogovorom!

f) Napišite vse elemente osnovnega modela za odstavitveno maso jagnjet!

g) Razvijte možni model za odstavitveno maso jagnjet in zanj napišite vse elemente!

h) Ponovite vajo še za dnevni prirast!

i) V vseh modelih določite število parametrov in stopinj prostosti za posamezne vplive, model in ostanek!

Naloga 17:

Tudi v tej nalogi bomo uporabili podatke iz pregl. 1.23 in prav tako nas je zanimala masa ob odstavitvi. Za obdelavo smo izbrali naslednji model:

$$y_{ijkl} = \mu + K_i + S_j + KS_{ij} + b_{Ij}\left(x_{ijkl} - 120\right) + b_{IIj}\left(x_{ijkl} - 120\right)^2 + b_{IIIj}\left(x_{Mijkl} - 120\right)^3 + h_k + a_{ijkl} + e_{ijkl}$$

Oznake so v skladu z dogovori. Privzemite, da je model pravilen!

a) Opišite vplive z nivoji, kot je to potrebno pri navajanju modela!

b) Katera spremenljivka je neodvisna? Navedite oznako in ime!

c) Katero funkcijo smo uporabili za opis povezave med odvisno in neodvisno spremenljivko?

d) Koliko krivulj pričakujemo na osnovi modela? Utemeljite odgovor!

e) Ali je konstanta 120 v modelu primerno izbrana? Utemeljite odgovor!

f) Izpišite in preštejte parametre za sistematske vplive!

g) Koliko meritev ima vsaka žival? Obrazložite!

h) Koliko nivojev ima vpliv živali?

i) Izpeljite pričakovano vrednost in strukturo fenotipske variance!

j)Napišite vse elemente osnovnega modela!

k)Preverite, če je možna interakcija med dopolnilno krmo in čredo!

l)Izpeljite možni model in napišite vse elemente tega modela!

Naloga 18:

Tudi v tej nalogi bomo uporabili podatke iz pregl. 1.23 in prav tako nas je zanimala masa ob odstavitvi. Za obdelavo smo izbrali naslednji model:

$$y_{ijklm} = \mu + K_i + S_j + KS_{ij} + b_S\left(x_{Sijklm} - 120\right) + b_{Ii}\left(x_{Rijklm} - \bar{x}_R\right) + b_{IIj}\left(x_{Rijkl} - \bar{x}_R\right)^2 + h_k + g_{kl} + a_{ijklm} + e_{ijklm}$$

Oznake so v skladu z dogovori. Privzemite, da je model pravilen!

a) Opišite vplive z nivoji, kot je to potrebno pri navajanju modela!

b) Katera spremenljivka je neodvisna? Navedite oznako in ime!

c) Katero funkcijo smo uporabili za opis povezave med odvisno in neodvisno spremenljivko?

Naloga 19:

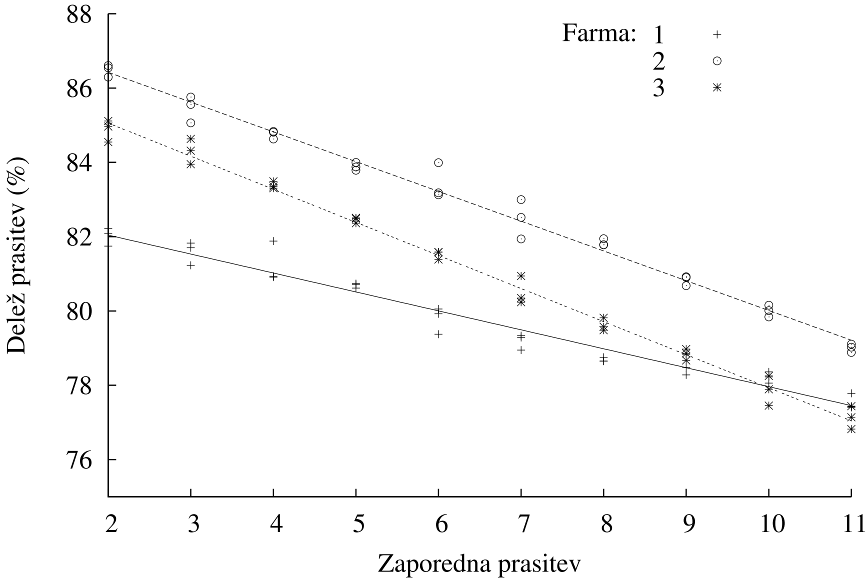
Kaj pomenijo oznake?

- K_i	________________	- σ^2	________________
- $\hat{K}_i$	________________	- h_{ik}	________________
- y_{ijk}	________________	- $\sigma_{\bar{x}}$	________________
- $\hat{y}_{ijk}$	________________	- μ	________________

Naloga 20:

Rezultati preizkusa so prikazani na sliki 1.10.

a) Katera je odvisna spremenljivka?



Slika 1.10: Vpliv zaporedne prasitve na delež prasitev

b) Kaj je opazovana lastnost?

c) Izpišite vplive, ki so nakazani na sliki! Opišite jih in označite v skladu z dogovorom!

d) Napišite enačbo modela, ki ga lahko razberete iz slike!

e) Izpeljite pričakovano vrednost in strukturo variance!

f) Interpretirajte rezultate!

Naloga 21:

Rezultati preizkusa so prikazani na sliki 1.11.

a) Katera je odvisna spremenljivka?

b) Kaj je opazovana lastnost?

c) Izpišite vplive, ki so nakazani na sliki! Opišite jih in označite v skladu z dogovorom!

d) Napišite enačbo modela, ki ga lahko razberete iz slike!

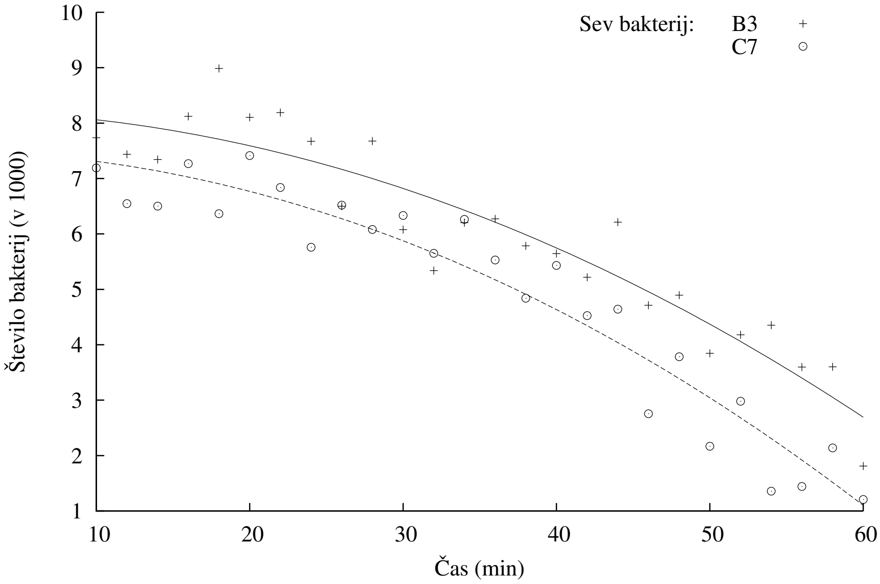
e) Izpeljite pričakovano vrednost in strukturo variance!

f) Interpretirajte rezultate!

Naloga 22:

V tabeli 1.24 je načrt krmnega poskusa pri prašičih v pitanju. Želeli smo preizkusiti štiri krmne mešanice z različno vsebnostjo beljakovin. Spremljali smo dnevne priraste, dnevno porabo krme od 30 do 100 kg in debelino hrbtne slanine. V preizkus so bile vključene tri pasme. V tabeli je navedeno število živali posameznih pasem, krmljenih s posameznimi krmami. Tako smo npr. 28 živali pri pasmi 22 krmili s krmo, ki je vsebovala 18 % beljakovin. Preizkus smo uspeli izvesti pri enem rejcu. V poskusu je bilo predvideno, da iz vsakega gnezda (prašiči imajo istega očeta in isto mater) vključimo po štiri pujske, praviloma po dve svinjki in dva kastrata, vendar nam je ob začetku poskusa nekaj pujskov manjkalo. Pujski iz istega gnezda niso bili razporejeni v skupino z isto krmo. Lastnosti smo merili enako natančno.

Utemeljite svojo trditev! Da boste uspešno odgovorili na vprašanja, predlagam, da si vplive izpišete, opišete in označite!



Slika 1.11: Vpliv trajanja učinkovanja dezinfekcijskega sredstva na število pri dveh sevih bakterij

Tabela 1.24: Načrt preizkusa in število opazovanj v posamezni skupini

	Število	Število	Število	Krma - vsebnost beljakovin (%)			
Pasma	očetov	mater	gnezd	14	16	18	20
11	8	32	32	32	26	16	6
22	7	28	28	16	8	28	24
33	8	32	32	16	32	20	26

a) Koliko je v poskusu vseh merjenih živali - pitancev?

b) Katere spremenljivke so odvisne?

c) Katere vplive obravnavamo kot kvantitativne vplive?

d) Katere vplive vključimo kot vplive z razredi?

e) Katere vplive opišemo z regresijo?

f) Kateri vplivi so sistematski?

g) Kateri vplivi so naključni?

h) Je načrt poskusa dobro zasnovan? Utemeljite!

i) Kolikim živalim bi lahko izvrednotili plemensko vrednost?

Naloga 23:

Iz prejšnje naloge (preglednica 1.24) vzemite debelino hrbtne slanine kot lastnost! Vplive smo označili po dogovorjenem sistemu. Pri odgovorih naj bodo razvidni postopki ali podane utemeljitve! Izhajajte iz naslednjega modela:

$$y_{ijklm} = \mu + P_i + S_j + b(x_{ijklm} - 14) + g_{ik} + a_{ijkl} + e_{ijklm}$$

a) Opišite oznake v enačbi modela!

b) Ali imamo v modelu naključne vplive? Naštejte jih!

c) V modelu določite število parametrov in stopinj prostosti po posameznih vplivih, za model in ostanek! Določite rang in red sistema! Lahko upoštevate samo sistematski del modela.

d) Ali ima žival eno ali več meritev za debelino hrbtne slanine? Utemeljite!

e) Kako so porazdeljeni ostanki?

f) Kakšno vlogo ima konstanta 14 v modelu?

Naloga 24:

Za izhodišče vzemite model za število živorojenih pujskov v gnezdu (preglednica **??**), ki vključuje naslednje vplive: mesec, genotip, skupina in starost ob pripustu. Starost ob pripustu v model vključite kot kvadratno regresijo. Napišite model tako, da opravite primerjavo števila živorojenih pujskov pri mladicah, ki so ob pripustu stare 220 dni.

a) Izpišite vplive, jih opišite in označite v skladu z dogovorom!

b) Napišite osnovni model!

c) Za osnovni model določite število parametrov in stopinj prostosti za posamezne vplive, model ter ostanek! Napišite rang in red sistema enačb!

Naloga 25:

Napišite teoretični model za dnevni prirast iz preglednice **??** in pojasnite odločitve.

Naloga 26:

Napišite teoretični model za debelino hrbtne slanine iz preglednice **??** in pojasnite odločitve.

Naloga 27:

Opišite oznake v modelu, izvrednotite pričakovane vrednosti in ugotovite strukturo variance za model

$$y_{ijklmn} = \mu + K_i + P_j + O_k + S_l + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right) + b_{IIijkl}\left(x_{ijklmn} - \bar{x}\right)^2 + g_{jkm} + e_{ijklnm} \qquad [1.128]$$

Izvrednotite pričakovane vrednosti, varianco in kovarianco za naslednje izraze (O vsebini, pomenu in smislu modela ali posameznih členov ne razmišljajte! O linearnih kombinacijah in njihovih interpretacijah se bomo ukvarjali v nadaljevanju.)

a) $\left(y_{ijklmn}\right) =$

b) $\left(y_{ijklmn}|g_{jkm}\right) =$

c) $\left(S_l + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right)\right) =$

d) $\left(S_l + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right)|g_{jkm}\right) =$

e) $\left(\mu + 3P_j + \frac{1}{2}O_k + \frac{1}{4}g_{jkm} + e_{ijklnm}\right) =$

f) $\left(\mu + 3P_j + \frac{1}{2}O_k + \frac{1}{4}g_{jkm} + e_{ijklnm}|g_{jkm}\right) =$

g) $\left(\frac{1}{4}g_{jkm} + e_{ijklnm}\right) =$

h) $\left(\frac{1}{4}g_{jkm} + e_{ijklnm}|g_{jkm}\right) =$

Naloga 28:

Izvrednotite pričakovane vrednosti in ugotovite strukturo variance za model!

$$\begin{aligned} y_{ijklmn} = \; & \mu + K_i + P_j + O_k + S_l + KP_{ij} + KO_{ik} + KS + PO_{jk} + PS_{jl} + OS_{kl} + \\ & + KPO_{ijk} + KPS_{ijl} + KOS_{ikl} + POS_{jkl} + KPOS_{ijkl} + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right) + \\ & + b_{IIijkl}\left(x_{ijklmn} - \bar{x}\right)^2 + g_{jkm} + Kg_{ijkm} + Sg_{jklm} + e_{ijklnm} \end{aligned} \qquad [1.129]$$

Izvrednotite pričakovane vrednosti, varianco in kovarianco za naslednje izraze! O vsebini, pomenu in smislu modela ali posameznih členov ne razmišljajte! O linearnih kombinacijah in njihovih interpretacijah se bomo ukvarjali v nadaljevanju.

a) $\left(y_{ijklmn}\right) =$

b) $\left(y_{ijklmn}|g_{jkm}\right) =$

c) $\left(y_{ijklmn}|Kg_{ijkm}\right) =$

d) $\left(KPS_{ijl} + KOS_{ikl} + POS_{jkl} + KPOS_{ijkl} + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right)\right) =$

e) $\left(KPS_{ijl} + KOS_{ikl} + POS_{jkl} + KPOS_{ijkl} + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right)|Kg_{ijkm}\right) =$

f) $\left(\mu + K_i + P_j + O_k + g_{jkm} + Kg_{ijkm} + Sg_{jklm} + e_{ijklnm}|Kg_{ijkm}\right) =$

g) $\left(KPS_{ijl} + KOS_{ikl} + POS_{jkl} + KPOS_{ijkl} + b_{Iijkl}\left(x_{ijklmn} - \bar{x}\right)|g_{jkm}, Kg_{ijkm}\right) =$

h) $\left(\mu + K_i + P_j + O_k + g_{jkm} + Kg_{ijkm} + Sg_{jklm} + e_{ijklnm}|g_{jkm}, Kg_{ijkm}\right) =$

Poglavje 2

KLASIFIKACIJA MODELOV

2.1 Linearni, pogojno linearni in nelinearni modeli

Modele razdelimo na linearne, pogojno linearne in (pogojno) nelinearne modele. Razdelitev opravimo na osnovi prvih parcialnih odvodov odvisne spremenljivke z ozirom na vse parametre. Če se v nobenem parcialnem odvodu ne pojavi parameter, je model linearen. Nekateri modeli po naravi niso linearni, najdemo pa tako transformacijo, da transformiran model zadosti pogojem. V nekaterih primerih pa se lahko poslužimo tudi aproksimacije zahtevne nelinearne funkcije. To velja še posebej, če opazujemo spremenljivke le na manjšem intervalu.

Linearni modeli so zaželjeni, ker so računsko manj zahtevni in enostavnejši za interpretacijo. Pogosto pa pojava ne moremo poenostaviti tako preprosto. Rast organizmov od rojstva do odrasle velikosti in razgradnja hrane od zaužitja do neprebavljivih ostankov, koagulacija mleka, laktacijske krivulje so pogosto proučevane lastnosti v živinoreji, ki jih moremo v celoti opisati z linearnimi modeli. Le ti zadostujejo samo na zadostno majhnih intervalih.

PRIMER:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk} \qquad [2.1]$$

kjer pomeni:

y_{ijk}	- opazovanje
μ	- srednja vrednost
α_i	- sistematski vpliv α; i = 1, 2, ... p
β_j	- sistematski vpliv β; j = 1, 2, ... q
$\alpha\beta_{ij}$	- interakcija med vplivoma α in β
e_{ijk}	- ostanek; k = 1, 2, ... n_{ij}

V modelu 2.1 smo za sistematske vplive uporabili male grške črke, kar bomo v skalarni obliki bolj redko uporabljali. Uporabimo jih predvsem, kadar iz matrične oblike izpeljujemo skalarno obliko. Zanima nas, če je model linearen ali nelinearen. Poiskati moramo vse parcialne odvode z ozirom na vse neznane parametre.
Pri odvajanju se moramo navaditi, da so neznanke parametri iz modela. Namesto opazovanja y_{ijk} vnesemo desno stran modela 2.1, ki pojasnjuje variabilnost v modelu. Izraz lahko razstavimo na člene. Vplivi so neodvisni, zato so odvodi členov, ki jih odvajamo na drug parameter kot ga vsebuje v izrazu, nadomestimo z vrednostjo nič. Tudi ostanki so neodvisni od parametrov in člene z ostanki nadomestimo z vrednostjo odvoda, ki je enak 0. Le členi, ki vključujejo parametre, na katere odvajamo, dobijo od 0 različno.
Najprej odvajajmo na parameter μ (enačba 2.2). V prvem členu dobimo vrednost 1, pri ostalih pa 0, ker so parametri neodvisni od μ.

$$\frac{\partial y_{ijk}}{\partial \mu} = \frac{\partial\left(\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}\right)}{\partial \mu} = \frac{\partial \mu}{\partial \mu} + \frac{\partial \alpha_i}{\partial \mu} + \frac{\partial \beta_j}{\partial \mu} + \frac{\partial \alpha\beta_{ij}}{\partial \mu} + \frac{\partial e_{ijk}}{\partial \mu} = 1 + 0 + 0 + 0 + 0 \quad [2.2]$$

Nadaljujmo z vplivom α. V tej skupini imamo p med seboj neodvisnih parametrov. Za vsak parameter moramo poiskati parcialni odvod. Enačbe bodo podobne enačbi 2.3, kjer bi menjali le indekse i' in sicer od 1, 2 do p. Tam, kjer sta i in i' enaka, dobimo vrednost različno od 0. V vsaki enačbi je le en člen tak, ostali pa vsebujejo druge parametre.

$$\frac{\partial y_{ijk}}{\partial \alpha_{i'}} = \frac{\partial \mu}{\partial \alpha_{i'}} + \frac{\partial \alpha_i}{\partial \alpha_{i'}} + \frac{\partial \beta_j}{\partial \alpha_{i'}} + \frac{\partial \alpha\beta_{ij}}{\partial \alpha_{i'}} + \frac{\partial e_{ijk}}{\partial \alpha_{i'}} = 0 + \left\{ \begin{array}{ll} 1; & i = i' \\ 0; & i \neq i' \end{array} \right\} + 0 + 0 + 0 \qquad [2.3]$$

Pazite, da vas ne zavede interakcija. Oznaka za interakcijo vsebuje sicer oznaki glavnih vplivov, vendar pa ne predstavlja produkta! Je le oznaka, ki bi jo lahko zamenjali s povsem drugo oznako. Torej interakcije ne moremo izraziti kot funkcijo glavnih vplivov, je torej od njiju (ali njih) neodvisna. Vrednost odvodov, kjer nastopa interakcija in eden od glavnih vplivov, je 0.
Tudi odvoda na parametre za vpliv β (2.4) in interakcijo $\alpha\beta$ (2.5) dobimo po istem zgledu.

$$\frac{\partial y_{ijk}}{\partial \beta_{j'}} = \frac{\partial \mu}{\partial \beta_{j'}} + \frac{\partial \alpha_i}{\partial \beta_{j'}} + \frac{\partial \beta_j}{\partial \beta_{j'}} + \frac{\partial \alpha\beta_{ij}}{\partial \beta_{j'}} + \frac{\partial e_{ijk}}{\partial \beta_{j'}} = 0 + 0 + \left\{ \begin{array}{ll} 1; & j = j' \\ 0; & j \neq j' \end{array} \right\} + 0 + 0 \qquad [2.4]$$

$$\frac{\partial y_{ijk}}{\partial \alpha\beta_{i'j'}} = \frac{\partial \mu}{\partial \alpha\beta_{i'j'}} + \frac{\partial \alpha_i}{\partial \alpha\beta_{i'j'}} + \frac{\partial \beta_j}{\partial \alpha\beta_{i'j'}} + \frac{\partial \alpha\beta_{ij}}{\partial \alpha\beta_{i'j'}} + \frac{\partial e_{ijk}}{\partial \alpha\beta_{i'j'}} = 0+0+0+\left\{ \begin{array}{ll} 1; & i = i' \wedge j = j' \\ 0; & i \neq i' \vee j \neq j' \end{array} \right\} + 0 [2.5]$$

Vsi prvi odvodi so brez parametrov, torej je model 2.1 linearen.

PRIMER:

Vzemimo model iz vaje 1.5.3. Prepišimo model in ga razširimo še za kvadratni člen pri regresiji.

$$y_{ijklm} = \mu + S_i + b_1\left(x_{ijklm} - x_{min}\right) + b_2\left(x_{ijklm} - x_{min}\right)^2 + M_j + F_{jk} + g_{jkl} + e_{ijklm} \qquad [2.6]$$

kjer pomeni:

y_{ijklm} - opazovanja

μ - srednja vrednost

S_i - vpliv spola; i = 1, 2

b_1 - regresijski koeficient za linearni člen

b_2 - regresijski člen za kvadratni člen

x_{ijklm} - neodvisna spremenljivka za vsebnost lizina v krmi

x_{min} - minimalna količina lizina v krmi

M_j - vpliv merjasca; j = 1, 2, 3, 4, 5

F_{jk} - vpliv svinje; k = 1, 2, ... 6

g_{jkl} - vpliv skupnega okolja v gnezdu; l = 1, 2, 3

e_{ijklm} - ostanek; m = 1, 2, ... 4

Sedaj pa poskusimo odvajati. Prva neznanka je μ. Prvi parcialni odvod prikazujemo v enačbi 2.7. Samo pri prvem členu je odvod različen od 0.

$$\frac{\partial y_{ijk}}{\partial \mu} = \frac{\partial \mu}{\partial \mu} + \frac{\partial S_i}{\partial \mu} + \frac{\partial b_1\left(x_{ijklm} - x_{min}\right)}{\partial \mu} + \frac{\partial b_2\left(x_{ijklm} - x_{min}\right)^2}{\partial \mu} + \frac{\partial M_j}{\partial \mu} + \frac{\partial F_{jk}}{\partial \mu} + \frac{\partial g_{jkl}}{\partial \mu} + \frac{\partial e_{ijk}}{\partial \mu} = 1 + 0 + \cdots [2.7]$$

$$\frac{\partial y_{ijk}}{\partial S_{i'}} = \frac{\partial \mu}{\partial S_{i'}} + \frac{\partial S_i}{\partial S_{i'}} + \frac{\partial b_1\left(x_{ijklm} - x_{min}\right)}{\partial S_{i'}} + \frac{\partial b_2\left(x_{ijklm} - x_{min}\right)^2}{\partial S_{i'}} + \frac{\partial M_j}{\partial S_{i'}} + \frac{\partial F_{jk}}{\partial S_{i'}} + \frac{\partial g_{jkl}}{\partial S_{i'}} + \frac{\partial e_{ijk}}{\partial S_{i'}} = [2.8]$$

$$= 0 + \left\{ \begin{array}{l} 1;\ i = i' \\ 0;\ i \neq i' \end{array} \right\} + 0 + 0 + 0 + 0 + 0 + 0 \qquad [2.9]$$

$$\frac{\partial y_{ijk}}{\partial b_1} = \frac{\partial \mu}{\partial b_1} + \frac{\partial S_i}{\partial b_1} + \frac{\partial b_1\left(x_{ijklm} - x_{min}\right)}{\partial b_1} + \frac{\partial b_2\left(x_{ijklm} - x_{min}\right)^2}{\partial b_1} + \frac{\partial M_j}{\partial b_1} + \frac{\partial F_{jk}}{\partial b_1} + \frac{\partial g_{jkl}}{\partial b_1} + \frac{\partial e_{ijk}}{\partial b_1} = [2.10]$$

$$= 0 + 0 + \left(x_{ijklm} - x_{min}\right) + 0 + 0 + 0 + 0 + 0 \qquad [2.11]$$

$$\frac{\partial y_{ijk}}{\partial b_2} = \frac{\partial \mu}{\partial b_2} + \frac{\partial S_i}{\partial b_2} + \frac{\partial b_1\left(x_{ijklm} - x_{min}\right)}{\partial b_2} + \frac{\partial b_2\left(x_{ijklm} - x_{min}\right)^2}{\partial b_2} + \frac{\partial M_j}{\partial b_2} + \frac{\partial F_{jk}}{\partial b_2} + \frac{\partial g_{jkl}}{\partial b_2} + \frac{\partial e_{ijk}}{\partial b_2} = [2.12]$$

$$= 0 + 0 + 0 + \left(x_{ijklm} - x_{min}\right)^2 + 0 + 0 + 0 + 0 \qquad [2.13]$$

$$\frac{\partial y_{ijk}}{\partial M_{j'}} = \frac{\partial \mu}{\partial M_{j'}} + \frac{\partial S_i}{\partial M_{j'}} + \frac{\partial b_1\left(x_{ijklm} - x_{min}\right)}{\partial M_{j'}} + \frac{\partial b_2\left(x_{ijklm} - x_{min}\right)^2}{\partial M_{j'}} + \frac{\partial M_j}{\partial M_{j'}} + \frac{\partial F_{jk}}{\partial M_{j'}} + \frac{\partial g_{jkl}}{\partial M_{j'}} + \frac{\partial e_{ijk}}{\partial M_{j'}} = [2.14]$$

$$= 0 + 0 + 0 + 0 + \left\{ \begin{array}{l} 1;\ j = j' \\ 0;\ j \neq j' \end{array} \right\} + 0 + 0 + 0 \qquad [2.15]$$

$$\frac{\partial y_{ijk}}{\partial F_{j'k'}} = \frac{\partial \mu}{\partial F_{j'k'}} + \frac{\partial S_i}{\partial F_{j'k'}} + \frac{\partial b_1 \left(x_{ijklm} - x_{min}\right)}{\partial F_{j'k'}} + \frac{\partial b_2 \left(x_{ijklm} - x_{min}\right)^2}{\partial F_{j'k'}} + \frac{\partial M_j}{\partial F_{j'k'}} + \frac{\partial F_{jk}}{\partial F_{j'k'}} + \frac{\partial g_{jkl}}{\partial F_{j'k'}} + \frac{\partial e_{ijk}}{\partial F_{j'k'}} = \quad [2.16]$$

$$= 0 + 0 + 0 + 0 + 0 + \left\{ \begin{array}{l} 1;\ j = j' \wedge k = k' \\ 0;\ j \neq j' \vee k \neq k' \end{array} \right\} + 0 + 0 \quad [2.17]$$

$$\frac{\partial y_{ijk}}{\partial g_{j'k'l'}} = \frac{\partial \mu}{\partial g_{j'k'l'}} + \frac{\partial S_i}{\partial g_{j'k'l'}} + \frac{\partial b_1 \left(x_{ijklm} - x_{min}\right)}{\partial g_{j'k'l'}} + \frac{\partial b_2 \left(x_{ijklm} - x_{min}\right)^2}{\partial g_{j'k'l'}} + \frac{\partial M_j}{\partial g_{j'k'l'}} + \frac{\partial F_{jk}}{\partial g_{j'k'l'}} + \frac{\partial g_{jkl}}{\partial g_{j'k'l'}} + \frac{\partial e_{ijk}}{\partial g_{j'k'l'}} = \quad [2.18]$$

$$= 0 + 0 + 0 + 0 + 0 + 0 + \left\{ \begin{array}{l} 1;\ j = j' \wedge k = k' \wedge l = l' \\ 0;\ j \neq j' \vee k \neq k' \vee l \neq l' \end{array} \right\} + 0 \quad [2.19]$$

V nobenem od parcialnih odvodov ni nobenega parametra. Model je linearen.

V nadaljevanju poskusimo nekoliko spremenjene modele. Modeli so izpeljani iz modela 2.6. Spremembe so nerejene namenoma za ilustracijo pogojno linearnih modelov. Za analizo odbire pri mladicah, za katerega smo razvili izvorni model, niso priporočljivi.

$$y_{ijklm} = \mu + S_i + b_1 \left(x_{ijklm} - x_{min}\right) + b_2^2 \left(x_{ijklm} - x_{min}\right)^2 + M_j + F_{jk} + g_{jkl} + e_{ijklm} \quad [2.20]$$

$$y_{ijklm} = \mu + S_i + b_1 sin \left(x_{ijklm} - x_{min}\right) + M_j + F_{jk} + g_{jkl} + e_{ijklm} \quad [2.21]$$

$$y_{ijklm} = \mu + S_i + sin\left(b_1\right) \left(x_{ijklm} - x_{min}\right) + M_j + F_{jk} + g_{jkl} + e_{ijklm} \quad [2.22]$$

Vaje:

Naloga 29:

Določi, ali je model linearni, pogojno-linearni oziroma nelinearni model! Določite tudi, ali so modeli sistematski, naključni ali mešani!

$$y_{ijk} = \mu + A_i + B_{ij} + e_{ijk} \quad [2.23]$$

$$y_{ijk} = \mu + K_i + S_j + KS_{ij} + e_{ijk} \quad [2.24]$$

$$y_{ijk} = \mu + F_i + B_j + b_i \left(x_{ijk} - 100\right) + e_{ijk} \quad [2.25]$$

$$y_{ijk} = \mu + F_i + B_j + b_{|i} \left(x_{ijk} - 100\right) + b_{||i} \left(x_{ijk} - 100\right)^2 + e_{ijk} \quad [2.26]$$

$$y_{ijk} = \mu + F_i + B_j + b_i\, sin \left(x_{ijk}\right) + e_{ijk} \quad [2.27]$$

$$y_{ijk} = \mu + F_i + B_j + sin^b \left(x_{ijk}\right) + e_{ijk} \quad [2.28]$$

$$y_{ijk} = \mu + F_i + B_j + b_i\, exp \left(c \left(x_{ijk} - 100\right)\right) + e_{ijk} \quad [2.29]$$

Naloga 30:

Pri naslednjih modelih določite in utemeljite, ali so linearni! Določite tudi, ali so modeli sistematski, naključni ali mešani!

$$y_{ijk} = \mu + L_i + r_j + Lr_{ij} + e_{ijk} \quad [2.30]$$

$$y_{ijk} = \mu + L_i + R_j + LR_{ij} + b_{ij} * x_{ijk} + e_{ijk} \quad [2.31]$$

$$y_{ij} = \mu + L_i + b_i * \exp\left(x_{ij} - \overline{x}\right) + a_{ij} + e_{ij} \quad [2.32]$$

$$y_{ijk} = \mu + L_i + R_j + LR_{ij} + b_i * \sqrt{x_{ij}} + e_{ijk} \quad [2.33]$$

2.2 Z ozirom na število lastnosti

2.2.1 Enolastnostni modeli

$$y_{ijk} = \mu + P_i + b\left(x_{ijk} - \bar{x}\right) + e_{ijk} \quad [2.34]$$

Opazujemo in merimo samo eno odvisno spremenljivko. Lastnosti merjenih v poskusu je lahko tudi več, vendar jih obravnavamo kot nekorelirane, nepovezane. Dobimo povsem iste izračune, če jih analiziramo ločeno ali pa skupaj, istočasno.

2.2.2 Večlastnostni modeli

Opazujemo več lastnosti hkrati. Lastnosti so med seboj korelirane (podobne ali nasprotne). Iz ene lastnosti sklepamo na drugo lastnost in ta dodatni vir informacij tudi koristimo. Lastnosti so lahko povsem različno porazdeljene (različne porazdelitvene funkcije), lahko se razlikujejo v modelih, lahko so merjene v različnih okoljih. Podatki lahko celo manjkajo, ki lahko izpadejo naključno (npr. ko odstranimo osamalce) ali pa načrtno, sistematsko (selekcija).

Pri večlastnostnih modelih moramo opozoriti na lastnosti s popolno ali skoraj popolno podobnostjo. Korelacija med njima je 1 ali blizu 1 oziroma -1 ali blizu 1.

2.3 Z ozirom na rang sistema enačb

2.3.1 Modeli s polnim rangom

Vsi parametri so ocenljivi. Imajo eno samo rešitev.

<u>Regresijski modeli</u>

Regresijski modeli (enostavna, polinomska regresija, multipla regresija - regresija z več kot eno pojasnjevalno spremenljivko). Z regresijo proučujemo odnose med merjenimi spremenljivkami. Običajno so to modeli s polnim rangom. V primerih, ko je eden od regresijskih koeficientov enak nič (visoka korelacija med neodvisnimi spremenljivkami, nepotreben element v polinomu), je lahko model tudi z nepolnim rangom. Toda v takih primerih poenostavimo - izpustimo parameter, ki je enak nič in ponovimo analizo.

$opazovanalastnost = f(pojasnjevalnespremenljivke)$

$[response = f(predictors)]$

Linearna regresija je poseben razred povezanosti, opisana z ravno črto - regresijsko premico [2.35], kadar je v modelu ena neodvisna spremenljivka oziroma z ravninami pri regresijah z več neodvisnimi spremenljivkami [2.36]. Modeli tega tipa se lahko uporabljajo v več namenov, zelo razširjeni pa sta opisovanje odnosov med spremenljivkami in napovedovanje vrednosti. Odvisno spremenljivko lahko pojmujemo kot funkcijo lastnosti, določenih (konstantnih) pri poskusu: npr., debelina slanine se povečuje s težo.

PRIMER: Linearna regresija

$$y_i = \beta_0 + \beta x_i + e_i \qquad [2.35]$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i \qquad [2.36]$$

Pri ***polinomski regresiji*** za povezavo neodvisne in odvisne spremenljivke uporabljamo polinom druge ali višje stopnje [2.37]. Polinom druge stopnje ponazorimo s parabolo, pri več neodvisnih spremenljivkah, če je pri vseh spremenljivkah primerna kvadratna regresija, pa paraboloid. Toda v splošnem so povezave različne [2.38]. Povezavo s prvo neodvisno spremenljivko dobro opisuje parabola, za drugo je zadostna linearna regresija, pri tretji pa moramo prirediti polinom tretje stopnje. V primeru z več spremenljivkami lahko nastopajo tudi produkti [2.39].

PRIMER: Polinomska regresija

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \qquad [2.37]$$

$$y_i = \beta_0 + \beta_{11} x_{1i} + \beta_{12} x_{1i2} + \beta_{13} x_{1i3} + \beta_{21} x_{2i} + \beta_{22} x_{2i2} + e_i \qquad [2.38]$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i \qquad [2.39]$$

Model z visokimi potencami je neprimeren:

1. razlaga (strokovna)
2. statistično (zakon skromnosti)
3. matematično

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

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

Torej x^n se lahko izrazi s polinomom nižje stopnje. Polinome nižjih stopenj (npr. x^4) je težje aproksimirti, vendar ni nobenega problema izraziti x^{15}.

2.3.2 Modeli z nepolnim rangom

Sistemi z nepolnim rangom imajo več parametrov kot linearno neodvisnih enačb. Zato nimajo nobene rešitve ali pa jih imajo neskončno število. Ocenljivi parametri so dejansko le funkcije odvečnih parametrov. Šele, ko se omejimo na določene vrednosti za neznane parametre, lahko enolično določimo ostale ("ocenljive"). To odločitev imenujemo **restrikcija**.

PRIMER: Sistem dveh enačb s tremi neznankami (parametri x, y in z). Sistem lahko delno rešimo. Tako lahko dva parametra (spremenljivki) (npr. x in y) izrazimo kot funkcijo tretjega (z v našem primeru). $(x, y) = f(z)$

Tretja spremenljivka lahko zavzame katerokoli vrednost znotraj prostora parametrov (slanina 0-30). Ko poznamo njeno vrednost, lahko izračunamo ostale. Sistem enačb ima neskončno število rešitev. Ne moremo torej pričakovati ene same rešitve. Rezultate bomo preverjali tako, da bomo ocenili povprečje razreda.

PRIMER: Če npr. določimo z-ju vrednost nič, dobita parametra x in y vrednosti 22 in -6. Omejili smo se na eno rešitev.

$x + 2y + z = 10$	$x = 10 - 2y - z$	$= 22 - 5z$	$x = f_x(z)$
$x+3y-z = 4$	$y = -6 + 2z$		$y = f_y(z)$

Isti rezultat bi dobili, če bi že na začetku za parameter z uporabili "izbrano vrednost". Sistem bi se poenostavil na sistem dveh enačb (dve polni celici) z dvema neznankama. Dobili bi torej reducirani sistem s polnim rangom in tako z eno samo rešitvijo, če le-ta obstaja.

Rang sistema in linearno odvisne enačbe ni vedno enostavno določiti. Odvisen je tako od modela, kar je enostavno določiti, kot strukture podatkov, pri čemer pa so odvisne enačbe lahko zelo zakrite. Linearno odvisne enačbe, ki izvirajo iz modela, lahko vedno odpravimo z **reparametrizacijo**. Težje delo pa imamo z enačbami, ki so linearno odvisne zaradi strukture podatkov. Pomaga le temeljita analiza strukture podatkov. Vedno lahko določimo število ocenljivih parametrov, saj lahko ocenimo vedno samo srednje vrednosti posameznih najnižje ugnezdenih zasedenih celic.

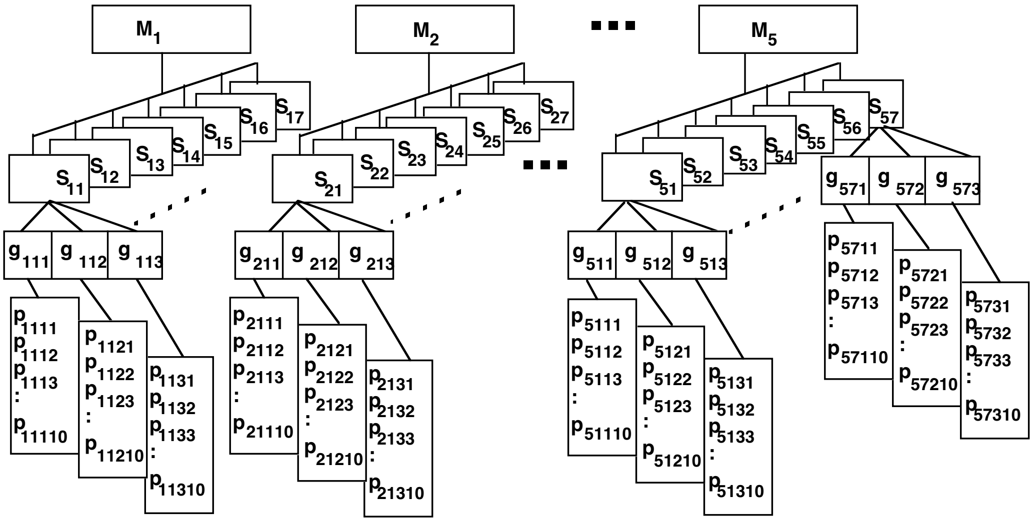
2.4 Z ozirom na strukturo podatkov

2.4.1 Hierarhični modeli

V hierarhičnem modelu imamo samo ugnezdene vplive.

Vzemimo pujske (a) iz enega gnezda! Vsi pujski pripadajo gnezdu (g), gnezdo pa samo eni svinji, svoji materi (S). Ne morejo imeti dveh mater hkrati. Svinja je praviloma za vsako gnezdo, zlasti pa v poskusnih razmerah, parjena samo z enim merjascem, medtem ko pri osemenjevanju lahko z enim merjascem pripustimo več svinj hkrati. Svinje so torej ugnezdene znotraj merjasca (M). Poskusimo določiti. Začnemo z vplivom, ki je najbolj na vrhu, merjascem torej. Ker je to v modelu prvi vpliv, mo dobil indeks i. Oznaka vpliva merjasca je torej M_i. Naslednji vpliv je svinja, ki je ugnezdena znotraj merjasca, zato bo poleg indeksa merjasca, s katerim je bila pripuščena, nosila tudi svoj indeks. Označena bo torej S_{ij}. Z indeksom j označujemo vse svinje, ki smo jih pripustili oziroma osemenili z enim merjascem. Prav gotovo sta bili najmanj dve svinji pripuščeni z vsaj enim merjascem. Naslednji vpliv je gnezdo, ki pripada svinji. Tako bo dobil indekse od svinje, dodatni indeks k pa bo štel gnezda znotraj svinje. Oznaka za naključni vpliv skupnega okolja v gnezdu je g_{ijk}. Gnezdo je ugnezdeno tudi znotraj merjasca, ker je naveden tudi indeks i za merjasca. Vpliv pujska a je ugnezdeni znotraj gnezda, ker pujsek pripada enemu gnezdu in tako dobi vse indekse od gnezda. V poskusu je iz enega gnezda 10 pujskov, zato dobi vsak pujsek še dodaten indeks l. Pujska bomo pravilno opisali kot a_{ijkl}. Kot vidimo pujsek pripada merjascu i in svinji ij.

PRIMER 1: Vzemimo, da imamo pet merjascev, daje vsak merjasec parjen s sedmimi svinjami. Vsaka svinja ima po tri gnezda, iz vsakega gnezda je vzetih 10 pujskov. Vsakega pujska smo stehtali 8 krat. Narišimo strukturo podatkov in označimo merjasca in svinjo kot sistematska vpliva, gnezdo in pujska pa kot naključna vpliva. Rezultat je nakazan na sliki 2.1. Napišimo tudi osnovni in možni model! Preštejte število opazovanj! Določite število parametrov in stopinje prostosti!
PRIMER 2: Privzemimo informacije iz primera 1 in dodajmo, da je pol pujskov v vsakem gnezdu svinjk in druga polovica kastratov. Ponovite vajo!
PRIMER 3: Vzemimo drugi primer. V njem bomo spremenili le to, da so svinje praviloma parjenje z različnimi merjasci za posamezna gnezda. Ponovite vajo!



Slika 2.1: Hierarhična zgradba poskusa in ugnezdeni vplivi

Tabela 2.1: Struktura podatkov pri hierarhičnem modelu

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}

Tabela 2.2: Struktura podatkov pri križno klasificiranem modelu

Vpliv	Pasma 1	Pasma 2	Pasma 3
Krma 1	n_{11}	n_{12}	n_{13}
Krma 2	n_{21}	n_{22}	n_{23}
Krma 3	n_{31}	n_{32}	n_{33}
Krma 4	n_{41}	n_{42}	n_{43}
Krma 5	n_{51}	n_{52}	n_{53}
Krma 6	n_{61}	n_{62}	n_{63}

Tabela 2.3: Struktura podatkov pri križno-klasificiranem 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

2.4.2 Križno-klasificirani modeli

Pri križno klasificiranem modelu so vsi vplivi križno klasificirani. Pri dobrem poskusu imamo meritve pri vseh možnih kombinacijah glavnih vplivov, ki jo iz praktičnih razlogov poimenujemo celice. Število opazovanj je lahko različno, v vsaki celici pa mora biti najmanj eno opazovanje. To eno opazovanje ni dobro za kakovostno izveden poskus, a pomaha premostiti računske probleme. Pri poskusu moramo dobro vedeti, da so zaključki vredni le toliko kot najmanj točna informacija.

V preglednici 2.2 je prikazan enostaven križno klasificirani model z dvema vplivoma: pasmo in krmo. Ker ima pasma 3 nivoje, krma pa 6, je možnih 18 kombinacij. Pri vseh kombinacijah (celicah) imamo opazovanja, zato lahko rečemo, da je poskus dobro zasnovan. Vprašanje je le, zakaj smo hkrati preizkušali toliko krm.

V drugem poskusu (pregl. 2.3) sta vpliva tudi križno klasificirana, a načrt je nepopoln. Nobena od pasem ni dobila vseh krm. Vzroki so lahko različnji, med njimi bi lahko bili celo opravičljivi. Produktivne pasme ne moremo krmiti s skromnimi obroki, ki so dobri za stare, avtohtone pasme. Pri slednjih pasmah pa je obrok, naravnan na potrebe moderne pasme, preobilen in bi bil ekonomsko neupravičen, lahko pa bi povzročil tudi prehitro rast in zamastitev, kar je neugodno zlasti pri plemenski vzreji.

2.4.3 Kombinirani modeli

Modeli pogosto vsebujejo tako križno klasificirane vplive kot ugnezdene vplive, kar velja predvsem za podatke, ki jih beležimo v proizvodnih razmerah. Strukturo podatkov moramo poznati ne glede na to, da bi za samo obdelavo s statističnimi paketi to niti ne bilo potrebno. Dobro poznavanje podatkov pripomore k dobri interpretaciji.

2.5 Z ozirom na naravo parametrov (prisotnost vplivov)

2.5.1 Modeli s sistematskimi vplivi (*ang.* fixed model)

Model ima samo eno naključno spremenljivko - ostanek. Ostali vplivi so sistematski. Modele pogosto srečamo pri analizi podatkov načrtovanih poskusov, pri katerih smo se izognili koreliranim naključnim vplivom.

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

2.5.2 Modeli z naključnimi vplivi (*ang.* random model)

Model ima samo en sistematski vpliv - srednjo vrednost (mean). Vsi ostali vplivi so naključni. Model se je pogosto uporabljal pri nekaterih metodah analize variance (Henderson I). Podatki so se predhodno očistili sistematskih vplivov - korigirali. Sedaj pa modelov praktično ne srečamo.

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

2.5.3 Mešani modeli (*ang.* mixed model)

Pri mešanih modelih imamo prisotne tako sistematske kot naključne vplive. Na modele naletimo v živinoreji, ko obdelujemo podatke iz proizvodnje. Pri analizi fenotipske variance so praktično neizogibni, zlasti če obdelujemo podatke v populaciji podvrženi selekciji. Neizogibni so pri napovedovanju plemenskih vrednosti. Poleg vpliva živali (aditivni direktni genetski vpliv) bodo v mešanih modelih lahko prisotni še naslednji naključni vplivi: aditivni maternalni genetski vpliv, dominanca, skupno okolje v gnezdu ali v čredi, permanentno okolje. Med sistematskimi vplivi pogosto srečamo v teh modelih sezono, starost ali maso, pasmo oz. genotip, genetske skupine...

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

2.6 Z ozirom na število vplivov

2.6.1 Enorazsežni model

PRIMER:

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad i = 1, ...p. \qquad j = 1, ...n_i$$

Število parametrov:	$\mu,\ \alpha_1,\ \alpha_2,\ \ldots,\ \alpha_p$	$= 1 + p$
Število zasedenih celic:		$s \leq p$

2.6.2 Dvorazsežni model brez interakcije

PRIMER:

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \qquad i = 1, ...p; \qquad j = 1, ...q; \qquad k = 1, ...n_{ij}$$

Število parametrov:	$\mu,\ \alpha_1,\ \alpha_2,\ \ldots \alpha_p,\ \beta_1,\ \beta_2,\ \ldots,\ \beta_q,\ \ldots,$	$= 1 + p + q$
Število zasedenih celic:		$s \leq p + q$

2.6.3 Dvorazsežni model z interakcijo

PRIMER:

$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$ $\quad i = 1, \ldots, p; \quad j = 1, \ldots, q; \quad k = 1, \ldots, n_{ij}$

Število parametrov: $\mu, \alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q, \alpha\beta_{11}, \alpha\beta_{1q}, \alpha\beta_{21}, \alpha\beta_{2q}, \alpha\beta_{p1}, \alpha\beta_{pq}$ $\underline{= 1 + p + q + pq}$

Število zasedenih celic: $s \leq pq$

Ocenimo lahko največ toliko parametrov, kot je zasedenih celic.

PRIMER:

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}	

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

μ

$A_1 A_2$

$B_1 B_2 B_3$

$AB_{11} AB_{12} AB_{13} AB_{21} AB_{22}$ $\quad$ **AB$_{23}$**

V tem primeru imamo samo 9 opazovanj, radi pa bi ocenili kar 12 parametrov in pet neodvisnih enačb-stopinj prostosti za model (za μ 1, za A 1, za B 2, za interakcije AB pa 1, skupaj 5 stopinj prostosti). Nemogoče je oceniti vse parametre zaradi manjkajočega razreda (AB_{23}). Število neodvisnih funkcij je enako številu polnih celic.

Če so prisotne manjkajoče celice, je potrebno pazljivo uporabljati statistične pakete (SAS). "Rezultati" so lahko odvisni od vrstnega reda vplivov v modelu, celo vrstnega reda nivojev. To pa je seveda povsem nezaželeno, saj bi lahko eden lahko dobil značilne razlike, drugi pa bi model ali podatke malo pomešal in dobil povsem drugačne zaključke. Odgovoriti si moramo na vprašanje, kaj je ocenljivo. Pri ocenljivih parametrih pa bodo rezultati vedno enaki.

Vedno je ocenljivo poprečje celic, saj so modeli z najmanjšimi celicami polnega ranga. Tako lahko model napišemo:

$$y_{ij} = \mu_i + e_{ij}$$

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

2.6.4 Kompleksni modeli

Modeli so redko tako preprosti - samo v skrbno načrtovanih poskusih. Praviloma pa imamo v živinoreji opravka z več vplivi in nam enostavni modeli služijo kot učni primeri. Poleg tega pa smo se zgoraj naučili, da sta lahko enorazsežni in dvorazsežni model z interakcijo ekvivalentna - omogočata povsem iste zaključke. Delitev je bila v navadi v statistiki, ki so uporabljali skalarno algebro in ročno računanje. Pri uporabi matrik in statističnih paketov v statistiki pa ta delitev izgubi na pomenu.

2.7 Pogojno linearni modeli

Pri pogojno linearnih modelih lahko osnovni model transformiramo tako, da prvi odvodi odvisne spremenljivke po parametrih ne vsebujejo parametrov (neznank).

Inverzni model

$$y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i)^{-1} \quad [2.40]$$

Je model linearen?

$$\frac{\partial y_i}{\partial \beta_0} = \frac{\partial \left[(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i)^{-1}\right]}{\partial \beta_0} \quad [2.41]$$

NE, še vedno vsebuje parametre.

Poskusimo transformirati model:

$$y_i^* = \frac{1}{y_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \quad [2.42]$$

$$\frac{\partial y_i^*}{\partial \beta_0} = 1; \qquad \frac{\partial y_i^*}{\partial \beta_1} = x_{1i}; \qquad \frac{\partial y_i^*}{\partial \beta_2} = x_{2i} \quad [2.43]$$

V transformiranemu modelu smo pri vseh prvih parcialnih odvodih izgubili parametre. Torej, model je pogojno linearen. Transformirane podatke bomo lahko obdelali po običajnih metodah. Pri interpretaciji moramo paziti: rezultate ne dobimo na normalni (originalni) skali, ampak transformirani. Predno poskusimo rezultate prevesti na originalno skalo, moramo preveriti, če je transformacija nazaj matematično sploh mogoča.

Logit-model

$$y_i = \frac{1}{1 + exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i\}} \quad [2.44]$$

Model je nelinearen (e_i predstavlja napako v eksponentu). Poskusimo poiskati transformacijo.

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

$$ln \underbrace{\left(\frac{1}{y_i} - 1\right)}_{logit} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \quad [2.45]$$

Zgornji model je linearen po transformaciji, torej je pogojno linearen.

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

$$y_i = \frac{1}{1 + exp\{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\} + e_i} \quad [2.46]$$

$$ln \underbrace{\left(\frac{1 - y_i}{y_i}\right)}_{logit} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ln\, e_i \quad [2.47]$$

$$ln \underbrace{\left(\frac{1 - y_i}{y_i}\right)}_{logit} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \underbrace{ln\, e_i}_{novi\, ostanek} \quad [2.48]$$

Pri modelih je zelo dobrodošla lastnost, ko je ostanek normalno porazdeljen. Slednji model bi bil dobrodošel v primerih, ko se variabilnost povečuje s povprečjem.

PRIMER: Količina mleka in dnevni prirast:

$\bar{x}$	300	500	900
σ	30	50	100

Če obravnavamo vse tri priraste kot isto lastnost (prirast), jih bomo transformirali.

Log-transformirani modeli

Log-transformirani model (2.49) je v našem primeru linearen, saj v prvih odvodih ni več parametrov. Logaritem spremenljivke x_i je le transformacija neodvisne spremenljivke, ki smo jo ob izvedbi poskusa izmerili. Oceniti želimo parametra α in γ.

$$ln\,(y_i) = \alpha + \gamma ln\,(x_i) + e_i \qquad i = 1, \ldots, n \qquad [2.49]$$

$$\frac{\partial\,(ln(y_i))}{\partial\alpha} = 1 \qquad i = 1, \ldots, N \qquad [2.50]$$

$$\frac{\partial\,(ln(y_i))}{\partial\gamma} = ln\,(x_i) \qquad i = 1, \ldots, N \qquad [2.51]$$

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.

2.8 (Pogojno) nelinearni modeli

Prvi odvodi vsebujejo parametre in hkrati ni transformacije, s katero bi dobili linearni model. Sem sodijo nekatere proizvodne funkcije, ki opisujejo prirejo v reprodukcijskem ciklusu ali življenju.

Rastne krivulje

Model:

$$y_i = \underbrace{A\left(1 - Be^{-kt_i}\right)^{-1}}_{\eta\,(\boldsymbol{\beta})} + e_i \qquad [2.52]$$

kjer pomeni:

y_i	- opazovanje	
t_i	- čas, starost	$i = 1, 2, ..., n$
B	- masa ob rojstvu	
A	- odrasla velikost	
k	- parameter, ki je povezan z ukrivljenostjo	
e_i	- ostanek	
e	- ... konstanta	

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

$$\frac{\partial y_i}{\partial A} = \left(1 - Be^{-kt_i}\right)^{-1} \qquad [2.53]$$

$$\frac{\partial y_i}{\partial B} = A\frac{e^{-kt_i}}{\left(1 - Be^{-kt_i}\right)^2} \quad [2.54]$$

$$\frac{\partial y_i}{\partial k} = AB\frac{e^{-kt_i}}{\left(1 - Be^{-kt_i}\right)^2} \quad [2.55]$$

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.

2.9 Pseudo-linearni model (linearna aproksimacija)

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

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.

2.10 Ekvivalentni modeli

Ekvivalentni modeli so različice modela, ki pa popolnoma enakovredno opisujejo podatke. Iščemo jih lahko samo med modeli z istim številom ocenljivih parametrov. Modele preoblikujemo zaradi interpretacije ali numerične stabilnosti sistema enačb. Zaradi računalniške nenatančnosti predstavitve realnih števil, pa tudi numeričnih napak pri računskih operacijah, ki so še numerične težave nekoliko bolj verjetne pri neuravnoteženih poskusih, računalnik ne more zanesljivo proglasiti, katere enačbe so linearno odvisne od drugih.

Preoblikovanje modela imenujemo reparametrizacija. Kot ime samo pove, poskusimo najti parametre, ki bodo parametre v starem modelu združili tako, da bomo odstranili čim več "neocenljivih" parametrov, ali razdružili tako, da bomo model prilagodili interpretaciji.

Poglejmo spodnje modele in jih primerjajmo. Vzemimo, da je prvi model (2.56) pravilen.

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{ijk} \quad [2.56]$$

kjer pomeni:

y_{ijk}	-	odvisna spremenljivka / lastnost
μ	-	srednja vrednost za lastnost
A_i	-	sistematski vpliv A, $i = 1, 2, . 3$
B_j	-	sistematski vpliv B, $j = 1, 2, 3, 4$
AB_{ij}	-	interakcija med vplivoma A in B
e_{ijk}	-	naključna napaka

Tabela 2.4: Primerjava števila parametrov in števila stopinj prostosti med šestimi modeli iz tega poglavja

Model	2.56	2.57	2.58	2.59	2.60	2.61
Vplivi	Število parametrov					
μ	1	1	1	1	-	1
A_i	3	3	3	-	-	3
B_j	4	-	-	-	-	4
AB_{ij} ali B_{ij}	12	12	12	12	12	-
Število stopinj prostosti						
μ	1	1	1	1	-	1
A_i	2	2	2	-	-	2
B_j	3	-	-	-	-	3
AB_{ij}ali B_{ij}	6	9	9	11	12	-
Skupaj s.p.	12	12	12	12	12	6

V modelu 2.56 izpustimo vpliv B z indeksom j. Drugi model (2.57) je nekoliko "nerodno" napisan, lahko pa bomo izpeljali povsem iste ocenljive funkcije. Povsem jasno je, da so interakcije ugnezdene znotraj vpliva A, so pa prav tako ugnezdene znotraj vpliva B. V novem modelu 2.57 je ostala interakcija AB, a jo moramo sedaj obravnavati kot ugnezden vpliv.

$$y_{ijk} = \mu + A_i + AB_{ij} + e_{ijk} \qquad [2.57]$$

Če bi se za tak model odločili od vsega začetka, bi vpliv AB poimenovali z eno črko (npr.B), indeksa ij pa bi obdržal. Bolj sprejemljiv je zapis tretjega model (2.58). Statistično gledano so model 2.56 in 2.58 ekvivalentna, različna pa je interpretacija (vsebina). Tudi modela 2.57 in 2.58 sta ekvivalentna, tudi interpretacija rezultatov je podobna.

$$y_{ijk} = \mu + A_i + B_{ij} + e_{ijk} \qquad [2.58]$$

Četrti (2.59) in peti (2.60) pa se ujemata s prvim tudi v interpretaciji, znebimo pa se linearno odvisnih vrstic.

$$y_{ijk} = \mu + AB_{ij} + e_{ijk} \qquad [2.59]$$

$$y_{ijk} = AB_{ij} + e_{ijk} \qquad [2.60]$$

Do sedaj smo bili torej vsi modeli ekvivalentni in so nas lahko pripeljali do istih zaključkov. Naslednji model (2.61) pa ni ekvivalenten predhodnim modelom, saj ima manj parametrov. Po izgledu je model 2.61 skoraj enak modelu 2.58, vendar nikakor nista ekvivalentna, saj se razlikujeta v številu parametrov.

$$y_{ijk} = \mu + A_i + B_j + e_{ijk} \qquad [2.61]$$

Modelom (pregl. 2.4) določimo še število zahtevanih (želenih) parametrov, število ocenljivih parametrov (stopinj prostosti), potem pa poiščimo linearne kombinacije parametrov v manjšem modelu za izpuščene parametre iz prvega modela.

$$B_j = \frac{1}{3}\sum_{i=1}^{3} AB_{ij} = \frac{1}{3}\sum_{i=1}^{3} B_{ij}, \qquad j = 1,2,3,4 \qquad [2.62]$$

$$A_i = \frac{1}{4}\sum_{j=1}^{4} AB_{ij}, \qquad i = 1, 2, 3 \qquad [2.63]$$

Za parameter μ v modelu 2.60 imamo tri možnosti. Lahko ga izračunamo kot povprečje iz parametrov za vpliv A, za vpliv A ali za interakcijo AB.