Nepilnieji eksperimentų planai

Rūta Levulienė

Nepilnieji eksperimentų planai

Pagrindiniai šaltiniai:

V. Bagdonavičius, J. Kruopis. Matematinė statistika. II dalis. 2015 (www.statistika.mif.vu.lt).
R. Levulienė. Statistikos taikymai naudojant SAS. 2009, VUL
SAS ir R Help

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

- Tarkime, kad kiekvienas faktoriaus B lygmuo gali dalyvauti eksperimente kartu tik su vienu faktoriaus A lygmeniu.
- Faktorius A vadinamas *grupuojančiuoju*, o faktorius B – *sugrupuotuoju* pagal faktorių A.
- Toks eksperimento planas vadinamas *dvifaktorine dispersine analize su grupuojančiuoju faktoriumi* (*nested factors*) arba hierarchiniu (*hierarchical*) eksperimento planu.
- Dažniausiai nagrinėjami modeliai, kai sugrupuotasis faktorius yra atsitiktinis, o grupuojantysis faktorius – atsitiktinis arba fiksuotasis.
- Pavyzdžiui, norime palyginti kelių miestų mokyklų moksleivių testo rezultatus. Šiame pavyzdyje miestas (faktorius A) yra grupuojantysis, o mokykla (faktorius B) sugrupuotasis pagal faktorių A.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Modelis, kai abu faktoriai yra fiksuotieji, vadinamas **fiksuotuoju dvifaktorinės dispersinės analizės modeliu su grupuojančiuoju faktoriumi**. Šiuo atveju stebėjimai užrašomi taip:

$$Y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_{(i)j} + e_{ijk}; \qquad (1)$$

čia $i = 1, \dots, I$, $j = 1, \dots, J_i$, $k = 1, \dots, K_{ij}$; indeksas i žymi faktoriaus A lygmenį; j – faktoriaus B lygmenį, kuris priklauso nuo indekso i; indeksas k žymi stebėjimų kartotinumą.
Nežinomų parametrų μ_{ij} skaičių pažymėkime $m = J_1 + \dots + J_I$, t. y. užpildytų stebėjimais langelių skaičių.
Tarkime, kad paklaidos e_{ijk} yra nepriklausomos ir turi normalųjį skirstinį su nuliniu vidurkiu ir dispersija σ^2.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

$$Y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_{(i)j} + e_{ijk};$$

Žymėjimai:

$$\alpha_i = \bar{\mu}_{i.} - \mu, \quad \beta_{(i)j} = \mu_{ij} - \bar{\mu}_{i.}, \tag{2}$$

$$\bar{\mu}_{i.} = \frac{1}{K_{i.}} \sum_j K_{ij}\mu_{ij} \quad K_{i.} = \sum_j K_{ij}, \quad \mu = \frac{1}{n} \sum_i K_{i.}\mu_{i.};$$

čia $n = \sum_i \sum_j K_{ij}$. Parametrai tenkina sąlygas

$$\sum_i K_{i.}\alpha_i = 0, \quad \sum_j K_{ij}\beta_{(i)j} = 0, \quad i = 1, ..., I. \tag{3}$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Parametro μ_{ij} mažiausiųjų kvadratų įvertinys:

$$\hat{\mu}_{ij} = \bar{Y}_{ij.} = \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} Y_{ijk}, \tag{4}$$

o liekamoji kvadratinė forma

$$SS_E = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 \sim \sigma^2 \chi^2(n-m), \quad n = \sum_i \sum_j K_{ij}. \tag{5}$$

Parametrų α_i, $\beta_{(i)j}$ MK įvertiniai:

$$\hat{\alpha}_i = \frac{1}{K_{i.}} \sum_j K_{ij} \bar{Y}_{ij.} - \frac{1}{n} \sum_i \sum_j K_{ij} \bar{Y}_{ij.}, \quad \hat{\beta}_{(i)j} = \bar{Y} ij. - \frac{1}{K_{i.}} \sum_j K_{ij} \bar{Y}_{ij.}. \tag{6}$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Faktorius	SS	ν	MS	$\mathbf{E}(MS)$
A	SS_A	$I-1$	MS_A	$\sigma^2 + \frac{1}{I-1}\sum_i K_{i.}\alpha_i^2$
B (sugrupuotas pagal A	$SS_{B(A)}$	$m-I$	$MS_{B(A)}$	$\sigma^2 + \frac{1}{m-I}\sum_i \sum_j K_{ij}\beta_{(i)j}^2$
E	SS_E	$n-m$	MS_E	σ^2

$$SS_A = \sum_i K_{i.}(\hat{\alpha}_i)^2, \quad SS_{B(A)} = \sum_i \sum_j K_{ij}(\hat{\beta}_{(i)j})^2,$$

$$SS_E = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 \sim \sigma^2 \chi^2(n-m), \quad n = \sum_i \sum_j K_{ij}.$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Kriterijaus statistika tikrinti hipotezę

$$H_{B(A)} : \beta_{(i)j} = 0, \quad j = 1, ..., J_i, \quad i = 1, ..., I, \tag{7}$$

t. y. kad faktorius B įtakos neturi, yra

$$F_{B(A)} = \frac{MS_{B(A)}}{MS_E}. \tag{8}$$

Hipotezė atmetama, kai $F_{B(A)} > F_\alpha(m - I, n - m)$.
Hipotezė

$$H_A : \alpha_1 = ... = \alpha_I = 0, \tag{9}$$

t. y. kad faktorius A įtakos neturi, atmetama, kai

$$F_A = \frac{MS_A}{MS_E} > F_\alpha(I - 1, n - m). \tag{10}$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Modelis, kai sugrupuotasis faktorius B yra atsitiktinis, o grupuojantysis faktorius A – fiksuotasis, vadinamas **mišriuoju dvifaktorinės dispersinės analizės modeliu su grupuojančiuoju faktoriumi**. Šiuo atveju stebėjimai užrašomi:

$$Y_{ijk} = \mu + \alpha_i + b_{(i)j} + e_{ijk}; \tag{11}$$

čia parametrai α_i tenkina sąlygą $\alpha_1 + ... + \alpha_I = 0$, o $b_{(i)j}$ yra nepriklausomi tarpusavyje ir nuo e_{ijk} vienodai pasiskirstę atsitiktiniai dydžiai su nuliniu vidurkiu ir dispersija $\sigma^2_{B(A)}$.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Modelis, kai sugrupuotasis faktorius B ir grupuojantysis faktorius A yra atsitiktiniai, vadinamas **atsitiktiniu dvifaktorinės dispersinės analizės modeliu su grupuojančiuoju faktoriumi**. Šiuo atveju stebėjimai užrašomi:

$$Y_{ijk} = \mu + a_i + b_{(i)j} + e_{ijk}; \qquad (12)$$

čia a_i yra nepriklausomi tarpusavyje ir nuo $b_{(i)j}$ ir e_{ijk} atsitiktiniai dydžiai su nuliniu vidurkiu ir dispersija σ_A^2.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Tarkime, kad $J_1 = ... = J_I = J$ ir $K_{ij} = K$. Tada gausime, kad paskutiniame dispersinės analizės lentelės stulpelyje ir (11), ir (12) modelio atveju

$$\mathbf{E}(MS_{B(A)}) = \sigma^2 + K\sigma^2_{B(A)}. \tag{13}$$

Pirmoje paskutinio dispersinės analizės lentelės stulpelio eilutėje gausime:

$$\mathbf{E}(MS_A) = \sigma^2 + K\sigma^2_{B(A)} + \frac{JK}{I-1}\sum_i \alpha_i^2, \tag{14}$$

kai faktorius A yra fiksuotasis, ir

$$\mathbf{E}(MS_A) = \sigma^2 + K\sigma^2_{B(A)} + JK\sigma^2_A, \tag{15}$$

kai faktorius A – atsitiktinis.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Hipotezės $H_{B(A)}$ analogas yra tvirtinimas, kad $\sigma^2_{B(A)} = 0$. Ši prielaida yra atmetama, kai

$$F_{B(A)} = \frac{MS_{B(A)}}{MS_E} > F_\alpha(I(J-1), IJ(K-1)). \tag{16}$$

Hipotezė $H_A : \alpha_1 = ... = \alpha_I = 0$ (arba jos analogas $H_A : \sigma^2_A = 0$) atmetamama, kai

$$F_A = \frac{MS_A}{MS_{B(A)}} > F_\alpha(I-1, I(J-1)). \tag{17}$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Nepaslinktieji dispersijų $\boldsymbol{V}a_i = \sigma_A^2$ ir $\boldsymbol{V}(b_{(i)j}) = \sigma_{B(A)}^2$ įvertiniai atitinkamai yra:

$$\hat{\sigma}_A^2 = \frac{MS_A - MS_{B(A)}}{KJ} \quad \hat{\sigma}_{B(A)}^2 = \frac{MS_{B(A)} - MS_E}{K}. \qquad (18)$$

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

SAS naudojame procedūras `NESTED`, `GLM` arba `MIXED`. Procedūra `NESTED` skirta dispersinės anallizės modeliui su grupavimu, kai faktoriai yra atsitiktiniai.

```
PROC NESTED DATA=Lentelė;
CLASS Faktoriai;
VAR Kintamasis; RUN; QUIT;
```

Sakinyje `PROC` su `DATA=`*Lentelė* nurodome, kokios lentelės duomenis analizuosime. Sakinyje `VAR` nurodome priklausomą (tiriamą) kintamąjį. Sakinyje `CLASS` nurodome stulpelių, kurių reikšmės žymi faktorių lygmenis, vardus. **Išvardytų faktorių tvarka yra svarbi**. Tariama, kad antrasis nurodytas faktorius yra sugrupuotas pagal pirmąjį, trečiasis – pagal antrąjį ir t.t. Prieš naudojant procedūrą `NESTED` **būtina surūšiuoti** duomenų lentelę pagal `CLASS` sakinyje išvardytus kintamuosius (tokia pačia tvarka, kaip jie išvardyti `CLASS` sakinyje).

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

Pavyzdžiui, dvifaktorinės dispersinės analizės su grupuojančiuoju faktoriumi (faktorius A – grupuojantysis, faktorius B – sugrupuotasis) atveju rašytume:

```
PROC SORT DATA=Lentelė; BY A B; RUN;
PROC NESTED DATA=Lentelė;
CLASS A B;
VAR Kintamasis; RUN; QUIT;
```

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi

- Procedūra `NESTED` nepritaikyta modeliams su kovariantėmis, reikėtų naudoti `GLM` arba `MIXED` procedūrą.
- Jeigu eksperimente naudojama ne tik hierarchinė klasifikacija (grupavimas), bet ir kryžminė klasifikacija, tai procedūros `NESTED` naudoti negalime. Analizę turime atlikti su `GLM` arba `MIXED` procedūra.

Jeigu norime nurodyti, kad faktorius B sugruotas pagal faktorių A (B *nested within* A), tai rašome: *B(A)*, t. y. nurodome sugrupuotąjį faktorių, o skliaustuose – grupuojantįjį.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

Pavyzdys. Eksperimentiškai tirtas kalcio kiekis ropių lapuose. Atsitiktinai atrinkus keturis augalus (kintamasis *Augalas*), nuo kiekvieno iš jų atsitiktinai buvo paimti trys lapai (kintamasis *Lapas*). Iš kiekvieno lapo buvo paimti du 100 miligramų mėginiai ir tam tikru būdu nustatytas kalcio kiekis (kintamasis *Kalcis*).

Augalas	Lapas											
	1	2	3	1	2	3	1	2	3	1	2	3
1	3.28 3.09	3.52 3.48	2.88 2.80									
2				2.46 2.44	1.87 1.92	2.19 2.19						
3							2.77 2.66	3.74 3.44	2.55 2.55			
4										3.78 3.87	4.07 4.12	3.31 3.31

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

- Šį eksperimentą galime aprašyti atsitiktiniu dvifaktorinės dispersinės analizės modeliu su grupuojančiuoju faktoriumi.
- Faktorius *Augalas* yra grupuojantysis, o faktorius *Lapas* – sugrupuotasis pagal faktorių *Augalas*, nes kiekvienas faktoriaus *Lapas* lygmuo dalyvauja eksperimente kartu tik su vienu faktoriaus *Augalas* lygmeniu.
- Abu faktoriai yra atsitiktiniai, nes faktoriaus *Augalas* lygmenys ir faktoriaus *Lapas* lygmenys buvo atrinkti atsitiktinai.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

```
DATA Duomenys; INPUT Augalas Lapas Kalcis @@;
DATALINES;
1 1 3.28 1 1 3.09 1 2 3.52 1 2 3.48 1 3 2.88 1 3 2.80
2 1 2.46 2 1 2.44 2 2 1.87 2 2 1.92 2 3 2.19 2 3 2.19
3 1 2.77 3 1 2.66 3 2 3.74 3 2 3.44 3 3 2.55 3 3 2.55
4 1 3.78 4 1 3.87 4 2 4.07 4 2 4.12 4 3 3.31 4 3 3.31
;RUN;
PROC NESTED DATA=Duomenys;
CLASS Augalas Lapas;
VAR Kalcis; RUN; QUIT;
```

Jeigu duomenų lentelė nebūtų surūšiuota pagal faktorius, nurodytus `CLASS` sakinyje (tokia tvarka, kaip jie išvardyti), tai reikėtų ją surūšiuoti su procedūra `SORT`.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

The NESTED Procedure

Nested Random Effects Analysis of Variance for Variable Kalcis

Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Square	Variance Component	Percent of Total
Total	23	10,270396				0,446539	0,532938	100,0000
Augalas	3	7,560346	7,67	0,0097	Lapas	2,520115	0,365223	68,5302
Lapas	8	2,630200	49,41	<,0001	Error	0,328775	0,161060	30,2212
Error	12	0,079850				0,006654	0,006654	1,2486

Kalcis Mean	3,01208333
Standard Error of Kalcis Mean	0,32404445

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

Procedūra pateikia dispersinės analizės lentelę lentelę. Joje nurodytas dispersijos šaltinis (Variance Source); eilutė *Augalas* atitinka lentelės eilutę A; eilutė *Lapas* atitinka lentelės eilutę B; *Error* – eilutę E); laisvės laipsnių skaičius (stulpelis *DF*); kvadratų suma (stulpelis *Sum of Squares*); vidutinė kvadratų suma (stulpelis *Mean Square*).

Stulpelyje *F value* pateikiamos įgytos statistikų (16) ir (17) reikšmės (eilutės B ir A atitinkamai), o stulpelyje *Pr>F* – atitinkamos p reikšmės.

Stulpelyje *Error Term* nurodoma, kuri vidutinė kvadratų suma (stulpelis *Mean Square*) imama vardiklyje apskaičiuojant įgytą kriterijaus statistikos reikšmę.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

Tikrinant hipotezę apie grupuojančiojo faktoriaus *Augalas* įtaką, vardiklyje imama vidutinė kvadratų suma, atitinkanti sugrupuotąjį faktorių *Lapas* (17), o tikrinant hipotezę apie sugrupuotojo faktoriaus *Lapas* įtaką, vardiklyje imama vidutinė kvadratų suma, atitinkanti atsitiktines paklaidas (16). Patikrinę hipotezę (16) gavome, kad faktorius *Lapas* yra statistiškai reikšmingas esant reikšmingumo lygmeniui ne mažesniam už 0,0097. Patikrinę hipotezę (17) gavome, kad faktorius *Augalas* yra statistiškai reikšmingas esant reikšmingumo lygmeniui ne mažesniam už 0,0001.

Stulpelyje *Variance Component* pateikti dispersijų (18) įverčiai: dispersijos σ_A^2 įvertis yra 0,3652, o dispersijos $\sigma_{B(A)}^2$ įvertis – 0,1611 (šiame pavyzdyje $I = 4$, $J_1 = \ldots = J_4 = J = 3$, $K = 2$). Atsitiktinių paklaidų dispersijos σ^2 įvertis yra 0,0067, o bendrojo vidurkio μ įvertis – 3,0121 (eilutė *Kalcis Mean*). Stulpelyje *Percent of Total* pateikiama, kokią dalį bendrosios dispersijos paaiškina kiekvienas dispersijos šaltinis.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

```
PROC GLM DATA=Duomenys;
CLASS Augalas Lapas;
MODEL Kalcis = Augalas Lapas(Augalas);
RANDOM Augalas Lapas(Augalas) / TEST;
RUN; QUIT;
```

Sakinyje `CLASS` nurodome stulpelių, kurių reikšmės žymi faktorių lygmenis, vardus. Sakinio `MODEL` lygybės dešinėje pusėje nurodome efektus (šiuo atveju du faktorius), o kairėje – priklausomą kintamąjį. Sakinyje `RANDOM` išvardijame atsitiktinius efektus. Užrašas `Lapas(Augalas)` žymi, kad faktorius *Lapas* yra sugrupuotasis pagal faktorių *Augalas*.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

The GLM Procedure

Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: Kalcis

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Augalas	3	7,560346	2,520115	7,67	0,0097
Error	8	2,630200	0,328775		
Error: MS(Lapas(Augalas))					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Lapas(Augalas)	8	2,630200	0,328775	49,41	<,0001
Error: MS(Error)	12	0,079850	0,006654		

Šio pavyzdžio eksperimentas aprašomas atsitiktiniu dispersinės analizės modeliu su grupuojančiuoju faktoriumi, todėl skaičiavimo rezultatus imame iš lentelės *Tests of Hypotheses for Random Model Analysis of Variance*. Joje pateikiamos tokios pačios charakteristikos kaip ir dispersinės analizės lentelėje, kurią pateikia procedūra NESTED.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

Procedūra GLM nepateikia dispersijos įverčių. Juos galime apskaičiuoti naudodami procedūrą VARCOMP:

```
PROC VARCOMP DATA=Duomenys ;
CLASS Augalas Lapas;
MODEL Kalcis = Augalas Lapas(Augalas);
RUN; QUIT;
```

```
     MIVQUE(0) Estimates

Variance Component      Kalcis
Var(Augalas)           0,36522
Var(Lapas(Augalas))    0,16106
Var(Error)           0,0066542
```

Gavome tokius pačius dispersijų įverčius kaip ir su procedūra NESTED (žr. procedūros NESTED pateikiamos dispersinės analizės lentelės stulpelį *Variance Component*).

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

```
PROC MIXED DATA=Duomenys METHOD=TYPE3;
CLASS Augalas Lapas;
MODEL Kalcis = ;
RANDOM Augalas Lapas(Augalas); RUN; QUIT;
```

Sakinyje `CLASS` nurodome stulpelių, kurių reikšmės žymi faktorių lygmenis, vardus. Sakinio `MODEL` lygybės dešinėje pusėje reikia nurodyti fiksuotuosius efektus (šiuo atveju fiksuotųjų efektų nėra, nes abu faktoriai atsitiktiniai, todėl po lygybės nieko nenurodome), o kairėje – priklausomą kintamąjį. Sakinyje `RANDOM` išvardijame atsitiktinius efektus. Užrašas `Lapas(Augalas)` žymi, kad faktorius *Lapas* yra sugrupuotasis pagal faktorių *Augalas*.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 1 pavyzdys

The Mixed Procedure

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Error Term	Error DF	F Value	Pr > F
Augalas	3	7,560346	2,520115	MS(Lapas(Augalas))	8	7,67	0,0097
Lapas(Augalas)	8	2,630200	0,328775	MS(Residual)	12	49,41	<,0001
Residual	12	0,079850	0,006654	.	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate
Augalas	0,3652
Lapas(Augalas)	0,1611
Residual	0,006654

Lentelė *Type 3 Analysis of Variance* atitinka dispersinės analizės lentelę. Tokios pačios charakteristikos kaip ir su procedūra *NESTED*. Lentelėje *Covariance Parameter Estimates* procedūra `MIXED` pateikia dispersijų (18) įverčius.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

Pavyzdys. Eksperimentiškai tirta trijų modelių staklių (faktorius A) ir darbuotojų (faktorius B) įtaka produktyvumui. Atsitiktinai atrinktiems dvylikai darbuotojų buvo priskirtos vienos iš trijų staklių. Eksperimento metu buvo dirbama pagal šešių darbo valandų grafiką. Savaitę buvo fiksuojamas pagamintų detalių skaičius. Lentelėje pateiktas pagamintų kiekvieną dieną detalių skaičius per valandą (analizuojamas kintamasis Y).

Staklės	1				2				3			
Darbuotojas	1	2	3	4	1	2	3	4	1	2	3	4
Diena $k = 1$	65	68	56	45	74	69	52	73	69	63	81	67
$k = 2$	58	62	65	56	81	76	56	78	83	70	72	79
$k = 3$	63	75	58	54	76	80	62	83	74	72	73	73
$k = 4$	57	64	70	48	80	78	58	75	78	68	76	77
$k = 5$	66	70	64	60	68	73	51	76	80	75	70	71

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

- Šį eksperimentą galime aprašyti mišriuoju dvifaktorinės dispersinės analizės modeliu su grupuojančiuoju faktoriumi.
- Faktorius A (staklių modelis) yra fiksuotasis, grupuojantysis, o faktorius B (darbuotojas) yra atsitiktinis, sugrupuotasis pagal faktorių A, nes kiekvienas faktoriaus B lygmuo dalyvauja eksperimente kartu tik su vienu faktoriaus A lygmeniu.
- Darbuotojai atrenkami atsitiktinai, todėl tariame, kad faktorius B atsitiktinis.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

Procedūrą `NESTED` galima naudoti ir mišriojo dispersinės analizės su grupavimu modelio atveju, nes kvadratų sumų, statistikų, skirtų patikrinti hipotezėms apie faktorių įtaką, išraiškos yra tokios pačios kaip ir tuomet, kai faktoriai yra atsitiktiniai.
Skaičiavimus galima atlikti naudojant procedūrą `GLM` arba `MIXED`.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

```
DATA Duomenys; INPUT A B Y @@;
DATALINES;
1 1 65 1 2 68 1 3 56 1 4 45 2 1 74 2 2 69 2 3 52 2 4 73 3 1 69 3 2 63
3 3 81 3 4 67 1 1 58 1 2 62 1 3 65 1 4 56 2 1 81 2 2 76 2 3 56 2 4 78
3 1 83 3 2 70 3 3 72 3 4 79 1 1 63 1 2 75 1 3 58 1 4 54 2 1 76 2 2 80
2 3 62 2 4 83 3 1 74 3 2 72 3 3 73 3 4 73 1 1 57 1 2 64 1 3 70 1 4 48
2 1 80 2 2 78 2 3 58 2 4 75 3 1 78 3 2 68 3 3 76 3 4 77 1 1 66 1 2 70
1 3 64 1 4 60 2 1 68 2 2 73 2 3 51 2 4 76 3 1 80 3 2 75 3 3 70 3 4 71
;RUN;
PROC GLM DATA=Duomenys; CLASS A B;
MODEL Y = A B(A);
RANDOM B(A) / TEST; RUN; QUIT;
```

Sakinyje `CLASS` nurodėme stulpelių, kurių reikšmės žymi faktorių lygmenis, vardus. Sakinio `MODEL` lygybės dešinėje pusėje nurodėme efektus (šiuo atveju du faktorius), o kairėje – priklausomą kintamąjį. Sakinyje `RANDOM` nurodėme atsitiktinį efektą (šiuo atveju sugrupuotąjį faktorių B). Užrašas `B(A)` žymi, kad faktorius B yra sugrupuotasis pagal faktorių A.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

```
                    The GLM Procedure
   Tests of Hypotheses for Mixed Model Analysis of Variance
                  Dependent Variable: Y

Source              DF   Type III SS   Mean Square   F Value   Pr
A                    2   1695,633333    847,816667      3,36   0,(
Error: MS(B(A))      9   2272,300000    252,477778

Source              DF   Type III SS   Mean Square   F Value   Pr
B(A)                 9   2272,300000    252,477778     10,70   <,(
Error: MS(Error)    48   1132,800000     23,600000
```

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

Lentelė *Tests of Hypotheses for Mixed Model Analysis of Variance* atitinka dispersinės analizės lentelę. Tikrinant hipotezę apie grupuojančiojo faktoriaus A (staklių modelio) įtaką, vardiklyje imama vidutinė kvadratų suma, atitinkanti sugrupuotąjį faktorių B (darbuotoją) (žr. (17)), o tikrinant hipotezę apie sugrupuotojo faktoriaus B įtaką, vardiklyje imama vidutinė kvadratų suma, atitinkanti atsitiktines paklaidas (žr. (16)). Hipotezė (16) atmetama (faktorius B yra statistiškai reikšmingas) remiantis kriterijais, kurių reikšmingumo lygmuo didesnis už 0,0001. Hipotezė (17) neatmetama (faktorius A yra statistiškai nereikšmingas), remiantis kriterijais, kurių reikšmingumo lygmuo mažesnis už 0,0814.

Dvifaktorinė dispersinė analizė su grupuojančiuoju faktoriumi. 2 pavyzdys

```
PROC SORT DATA=Duomenys; BY A B; RUN;

PROC NESTED DATA=Duomenys; CLASS A B;
VAR Y; RUN; QUIT;
```

Pradinė duomenų lentelė nebuvo surūšiuota pagal stulpelius, kurių reikšmės žymi faktorių lygmenis, todėl naudojome procedūrą `SORT`.

- Tarkime, kad faktoriai A, B ir C turi po tokį patį lygmenų skaičių.
- Pažymėkime lygmenų skaičių m.
- Tada vietoje kryžminės klasifikacijos, pagal kurią reikėtų m^3 stebėjimų, galima atlikti m^2 stebėjimų, išdėsčius juos pagal **lotyniškųjų kvadratų** (*latin square*) schemą.
- Lotyniškuoju kvadratu vadinama tokia lentelė, turinti m eilučių ir m stulpelių, kuriuose įrašyti skaičiai nuo 1 iki m. Skaičiai išdėstyti taip, kad kiekvienoje eilutėje ir kiekviename stulpelyje kiekvienas skaitmuo parašytas po vieną kartą.

Pavyzdžiui, keturių lygmenų atveju, t. y. $m = 4$, vienas iš galimų išdėstymų:

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

Skaičius nurodo faktoriaus A lygmenį, eilutės numeris – faktoriaus B lygmenį, o stulpelio numeris – faktoriaus C lygmenį.

Lotyniškieji kvadratai

- Sukeitus lotyniškojo kvadrato bet kurias eilutes arba bet kuriuos stulpelius vietomis, vėl gaunamas lotyniškasis kvadratas.
- Keiskime stulpelius taip, kad pirmoje eilutėje gautume iš eilės surašytus skaičius $1, 2, ..., m$. Tada keiskime vietomis eilutes, išskyrus pirmąją, kol gausime, kad pirmame stulpelyje būtų iš eilės parašyti skaičiai $1, 2, ..., m$. Toks lotyniškasis kvadratas vadinamas *standartiniu*.
- Iš bet kurio standartinio lotyniškojo kvadrato galima gauti $m!(m-1)!$ skirtingų lotyniškųjų kvdratų.
- Pagal randomizacijos principą, atliekant eksperimentą, konkretų lotyniškąjį kvadratą reikia pasirinkti atsitiktinai.
- Galime taikyti tokį būdą: atsitiktinai parenkame standartinį lotyniškąjį kvadratą. Tada atsitiktinai parenkame lotyniškąjį kvadratą iš dydžio $m!(m-1)!$ aibės, kuri gali būti gaunama iš šio standartinio lotyniškojo kvadrato keičiant vietomis stulpelius ir eilutes.

Stebėjimai Y_{ijk} aprašomi modeliu

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk}, \quad (i, j, k) \in \mathcal{D}; \tag{19}$$

čia indeksai i, j ir k atitinkamai žymi faktorių A, B ir C lygmenis; $\mathcal{D}$ – dydžio m^2 indeksų aibė; atsitiktinės paklaidos e_{ijk} nepriklausomos ir $e_{ijk} \sim N(0, \sigma^2)$. Parametrai α_i, β_j ir γ_k tenkina sąlygas

$$\sum_{i=1}^{m} \alpha_i = 0, \ \sum_{j=1}^{m} \beta_j = 0, \ \sum_{k=1}^{m} \gamma_k = 0.$$

Lotyniškieji kvadratai

Parametrų μ, α_i, β_j ir γ_k MK įvertiniai:

$$\hat{\mu} = \bar{Y}_{...} = \frac{1}{m^2} \sum_{(i,j,k)\in\mathcal{D}} Y_{ijk}, \tag{20}$$

$$\hat{\alpha}_i = \frac{1}{m} \sum_{(j,k)\in\mathcal{D}_i} Y_{ijk} - \bar{Y}_{...} = \bar{Y}_{i..} - \bar{Y}_{...}, \quad i = 1, \dots, m. \tag{21}$$

$$\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}, \quad j = 1, \dots, m; \quad \hat{\gamma}_k = \bar{Y}_{..k} - \bar{Y}_{...} \quad k = 1, \dots, m. \tag{22}$$

Liekamoji kvadratinė forma

$$SS_E = \sum_{(i,j,k)\in\mathcal{D}} (Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k)^2 =$$

$$= \sum_{(i,j,k)\in\mathcal{D}} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})^2 \sim \sigma^2\chi^2((m-1)(m-2)). \tag{23}$$

Faktorius	SS	ν	MS	$\mathbf{E}(MS)$
A	SS_A	$m-1$	MS_A	$\sigma^2 + m\sigma_A^2$
B	SS_B	$m-1$	MS_B	$\sigma^2 + m\sigma_B^2$
C	SS_C	$m-1$	MS_C	$\sigma^2 + m\sigma_C^2$
E	SS_E	$(m-1)(m-2)$	MS_E	σ^2

$$SS_A = m\sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2,$$

$$SS_B = m\sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2, \quad SS_C = m\sum_k (\bar{Y}_{..k} - \bar{Y}_{...})^2, \tag{24}$$

$$\sigma_A^2 = \frac{1}{m-1}\sum_i \alpha_i^2, \quad \sigma_B^2 = \frac{1}{m-1}\sum_j \beta_j^2, \quad \sigma_C^2 = \frac{1}{m-1}\sum_k \gamma_k^2. \tag{25}$$

Hipotezės

$$H_A : \alpha_1 = ... = \alpha_m = 0,$$

$$H_B : \beta_1 = ... = \beta_m = 0, \;\; H_C : \gamma_1 = ... = \gamma_m = 0 \qquad (26)$$

atmetamos, kai atitinkamai statistikos

$$F_A = \frac{MS_A}{MS_E}, \;\; F_B = \frac{MS_B}{MS_E}, \;\; F_C = \frac{MS_C}{MS_E} \qquad (27)$$

viršija Fišerio skirstinio su $m-1$ ir $(m-1)(m-2)$ laisvės laipsniais lygmens α kritinę reikšmę $F_\alpha(m-1,(m-1)(m-2))$.

Pavyzdys. Eksperimento metu buvo tiriama prekės kainos įtaka jos pardavimui. Atsižvelgiant į tai, kad kainų keitimas toje pačioje parduotuvėje kelis kartus gali būti painus, buvo nuspręsta parduotuvėje taikyti šešis mėnesius tik vieną nustatytą kainą. Tyrime dalyvavo 16 parduotuvių. Kad sumažintume paklaidą, parduotuvės buvo parinktos taip, kad atstovautų kiekvienam iš keturių geografinių rajonų (kintamasis B) ir kiekvienai iš keturių klasių parduotuvių (kintamasis A; suskirstyta atsižvelgiant į pardavimo apimtis). Tyrimo metu buvo taikyti keturi kainų lygiai (kintamasis C). Pardavimas buvo stebimas šešis mėnesius (kintamasis Y).

```
DATA Duomenys; INPUT Y A B C @@;
DATALINES;
1.2 1 1 2 1.5 1 2 3 1.0 1 3 1 1.7 1 4 4
1.4 2 1 1 1.9 2 2 4 1.6 2 3 2 1.5 2 4 3
2.8 3 1 3 2.1 3 2 2 2.7 3 3 4 2.0 3 4 1
3.4 4 1 4 2.5 4 2 1 2.9 4 3 3 2.7 4 4 2
;RUN;

PROC GLM DATA=Duomenys;
CLASS A B C; MODEL Y = A B C; RUN; QUIT;
```

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	7,24062500	0,80451389	40,65	0,0001
Error	6	0,11875000	0,01979167		
Corrected Total	15	7,35937500			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	5,98187500	1,99395833	100,75	<,0001
B	3	0,12187500	0,04062500	2,05	0,2081
C	3	1,13687500	0,37895833	19,15	0,0018

Hipotezė H_C atmetama (faktorius C yra statistiškai reikšmingas), remiantis kriterijais, kurių reikšmingumo lygmuo didesnis už 0,0001. Hipotezė H_C atmetama (faktorius C yra statistiškai reikšmingas; kainos įtaka pardavimui statistiškai reikšminga) remiantis kriterijais, kurių reikšmingumo lygmuo didesnis už 0,0018.

Nepilnieji subalansuotieji blokai

- Nagrinėdami randomizuotus blokus tarėme, kad bloko dydis sutampa su faktoriaus A lygmenų skaičiumi I.
- Kartais tenka nagrinėti atvejį, kai bloko dydis yra mažesnis už faktoriaus A lygmenų skaičių.
- Toks eksperimentas, kai bloko dydis k yra mažesnis už faktoriaus A lygmenų skaičių I, vadinamas nepilnųjų blokų schema (*incomplete block design*).

Nepilnieji subalansuotieji blokai. Pavyzdys

- Tarkime, kad kiekvienas faktoriaus A lygmuo kartojasi r kartų, o visų blokų dydžiai vienodi ir lygūs k, nėra tokio faktoriaus A lygmens, kuris du ar daugiau kartų pasikartotų tame pačiame bloke.
- Tarkime, kad blokų yra J. Tada bendras stebėjimų skaičius $n = Jk = Ir$.
- Nepilnųjų blokų schema vadinama *subalansuotąja*, jeigu skaičius tokių blokų, kuriuose pasirodo tam tikra faktoriaus A dviejų lygmenų pora, yra tas pats bet kuriai porai.
- Šią eksperimento schemą galime interpretuoti kaip dvifaktorinės dispersinės analizės su skirtingais stebėjimų skaičiais langeliuose atvejį: $K_{ij} = 1$, kai faktoriaus i-asis lygmuo pasirodo j-ajame bloke, ir $K_{ij} = 0$ priešingu atveju.

Pavyzdys. Eksperimentiškai buvo tirta keturių gumos mišinių įtaka automobilių padangų atsparumui. Gamybos procese vienai padangai galima panaudoti iki trijų mišinių. Padanga buvo suskirstyta į tris dalis ir skirtingi mišiniai naudoti skirtingose padangos dalyse. Padanga šiame pavyzdyje – blokas. Bloko dydis $k = 3$. Eksperimentui naudotos keturios padangos (t. y. $J = 4$). Tam tikru būdu buvo išmatuotas padangos nusidėvėjimo lygis.

Nepilnieji subalansuotieji blokai. Pavyzdys

Šiame pavyzdyje faktorius A – gumos mišinys. Jis turi keturis lygmenis, t. y. $I = 4$. Kiekvienas faktoriaus A lygmuo kartojasi $r = 3$. Bendras stebėjimų skaičius $n = Jk = 4 * 3 = Ir = 4 * 3 = 12$.

1	1	1	2
2	2	3	3
3	4	4	4

Šioje lentelėje stulpelis – blokas (padangos numeris), skaičius langelyje atitinka gumos mišinio numerį (faktoriaus A lygmuo). Eksperimento planas subalansuotasis, nes kiekviena faktoriaus A lygmenų pora sutinkama vienodą skaičių $\lambda = 2$ kartus. Parametras λ tenkina sąryšį: $\lambda = \frac{r(k-1)}{I-1}$.

Šią eksperimento schemą galima interpretuoti kaip dvifaktorinės dispersinės analizės (faktorius A – gumos mišinys, faktorius B – padanga, analizuojamas kintamasis Y – padangos nusidėvėjimo lygis) su skirtingais stebėjimų skaičiais langeliuose atvejį.

Nepilnieji subalansuotieji blokai. Pavyzdys

```
DATA Duomenys; INPUT B A Y @@;
DATALINES;
1 1 238 1 2 238 1 3 279
2 1 196 2 2 213 2 4 308
3 1 254 3 3 334 3 4 367
4 2 312 4 3 421 4 4 412
;RUN;

PROC GLM DATA=Duomenys;
CLASS A B; MODEL Y = A B;
RUN; QUIT;
```

Nepilnieji subalansuotieji blokai. Pavyzdys

The GLM Procedure

Dependent Variable: Y

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	20729,08333	6909,69444	19,73	0,0034
B	3	21037,75000	7012,58333	20,03	0,0032

Hipotezė: faktorius A yra statistiškai nereikšmingas, atmetama remiantis kriterijais, kurių reikšmingumo lygmuo didesnis už 0,0034. Gumos mišinio įtaka padangų atsparumui yra statistiškai reikšminga (p reikšmė 0,0034).