DAUGAVPILS UNIVERSITĀTE
Dabaszinātņu un matemātikas fakultāte
Matemātikas katedra

Studiju kurss

Eksperimentu plānošana un analīze

4.lekcija

Docētājs: Dr. P. Daugulis

2008./2009.studiju gads

Saturs

Lekcijas mērķis:

- apgūt faktoriālo eksperimentu pamatjēdzienus.

1. Faktoriālie eksperimenti

1.1. Ievads

1.1.1. Motivācija

Eksperimentā var būt divi vai vairāk faktori - svarīgi parametri, kuri ir jāpēta.

Ja ir n faktori, katrs no kuriem var pieņemt m vērtības, tad kopējais faktoru vērtību skaits ir vienāds ar nm.

1.1. piemērs. Pieņemsim, ka

- ir divi faktori X un Y,
- katrs no faktoriem vai pieņemt divas vērtības - maksimālo un minimālo,
- $f(X_{\min}, Y_{\min}) = 20$, $f(X_{\min}, Y_{\max}) = 50$, $f(X_{\max}, Y_{\min}) = 30$, $f(X_{\max}, Y_{\max}) = 70$.

$X = \frac{70+30}{2} - \frac{50+20}{2} = 15$ - faktora X *efekts*, vidējo vērtību starpība.

$Y = \frac{70+50}{2} - \frac{30+20}{2} = 35$ - faktora Y efekts.

1.2. piemērs. Pieņemsim, ka

- $f(X_{\min}, Y_{\min}) = 20$, $f(X_{\min}, Y_{\max}) = 50$, $f(X_{\max}, Y_{\min}) = 40$, $f(X_{\max}, Y_{\max}) = 10$.

$XY = \frac{(10-50)-(50-20)}{2} = 15$ - *faktoru mijiedarbības efekts.*

Var atlikt punktus Dekarta plaknē, ja mijiedarbība ir vāja, tad nogriežņi $[(X_{\min}, Y_{\min}), (X_{\max}, Y_{\min})]$ un $[(X_{\min}, Y_{\max}), (X_{\max}, Y_{\max})]$ nekrustojas.

Ir svarīgi pētīt faktoru mijiedarbību, jo tā var slēpt faktoru iedarbību.

1.1.2. Modeļa piemērs - regresijas modelis

Ja ir divi faktori ar divām kvantitatīvām līmeņu vērtībām, tad var piedāvāt *regresijas modeli:*

$$x = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_{12} t_{12} + \epsilon,$$

kur

- x ir mērāmais izvadlielums,
- β ir modeļa parametri,
- $t_i \in [-1, 1]$,
- ϵ ir kļūdas gadījuma lielums.

1.1.3. Mērījumu efektivitāte

Faktoriālie eksperimenti ļauj samazināt mērījumu skaitu.

1.3. piemērs. Ir doti divi faktori X, Y un divi līmeņi $\min, \max$. Lai daudz maz ticami noteiktu faktoru iedarbību veicot mērījumus

$(X_{\min}, Y_{\min})$, $(X_{\max}, Y_{\min})$, $(X_{\min}, Y_{\max})$, ir vajadzīgi $6 = 3 \cdot 2$ eksperimenti. To pašu var noteikt, veicot 4 mērījumus ar visām faktoru kombinācijām.

1.2. Divu faktoru faktoriālais dizains

1.2.1. Ievads un modelis

Ir divi faktori X un Y, kas var pieņemt, attiecīgi, n un m fiksētas vērtības - kvalitatīvas (diskrētas) vai kvantitatīvas (nepārtrauktas).

1.4. piemērs. Detaļas darba laika atkarība no ārējās temperatūras un tās materiāla.

Datus var sakārtot tabulā:

	1	2	...	m
1	$z_{111}, z_{112}, ..., z_{11k}$	$z_{121}, z_{122}, ..., z_{12k}$	...	$z_{1m1}, z_{1m2}, ..., z_{1mk}$
2	$z_{21}, z_{212}, ..., z_{21k}$	$z_{221}, z_{222}, ..., z_{22k}$	...	$z_{2m1}, z_{2m2}, ..., z_{2mk}$
...	...	...	...	...
n	$z_{n1}, z_{n12}, ..., z_{n1k}$	$z_{n21}, z_{n22}, ..., z_{n2k}$	...	$z_{nm1}, z_{nm2}, ..., z_{nmk}$

Uzkatot, ka ir starp faktoriem ir mijiedarbība, var pieņemt atbilstošo efektu modeli:

$$z_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk},$$

kur

- μ ir visu mērījumu vidējā vērtība,
- τ_i ir faktora X i-tā līmeņa efekts, $\sum_{i=1}^{n} \tau_i = 0$,
- β_j ir faktora Y j-tā līmeņa efekts, $\sum_{j=1}^{m} \beta_j = 0$,
- $(\tau\beta)_{ij}$ ir faktoru mijiedarbības efekts, $\sum_{i=1}^{n}(\tau\beta)_{ij} = \sum_{j=1}^{m}(\tau\beta)_{ij} = 0$,
- $\epsilon_{ijk} \sim N(0, \sigma^2)$ ir kļūdas gadījuma lielums.

Var definēt arī vidējo vērtību modeli:

$$z_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

kur

$$\mu_{ij} = \mu + \tau_i + \beta_j.$$

Var lietot ari regresijas tipa modeļus, ja faktoru līmeņi ir kvantitatīvi.

Hipotēzes:

- X faktora efektu vienādība:
  - H_0: $\tau_1 = \tau_2 = ... = \tau_n = 0$,
  - H_1: $\tau_k \neq 0$ vismaz vienam k.
- Y faktora efektu vienādība:
  - H_0: $\beta_1 = \beta_2 = ... = \beta_m = 0$,
  - H_1: $\beta_k \neq 0$ vismaz vienam k.
- faktoru mijiedarbības pastāvēšana:
  - H_0: $(\tau\beta)_{ij} = 0$, visiem pāriem i, j,
  - H_1: $(\tau\beta)_{ij} \neq 0$ vismaz vienam pārim i, j.

1.2.2. Statistiskā analīze

Definēsim šādas vidējās vērtības:

- $\overline{z} = \frac{\sum_{i,j,k} z_{ijk}}{nmk}$ - visas mērījumu kopas vidējā vērtība;
- $\overline{z}_{i**} = \frac{\sum_{j,k} z_{ijk}}{mk}$ - faktora X i-tā līmeņa mērījumu vidējā vērtība;

- $\overline{z}_{*j*} = \frac{\sum_{i,k} z_{ijk}}{nk}$ - faktora Y j-tā līmeņa mērījumu vidējā vērtība;
- $\overline{z}_{ij*} = \frac{\sum_{k} z_{ijk}}{k}$ - i-tā un j-tā līmeņa mērījumu vidējā vērtība;

Var pierādīt, ka pilnā dispersija sadalās šādā veidā:

$$\underbrace{\sum_{i,j,k}(z_{ijk} - \overline{z})^2}_{=DP} = \underbrace{mk\sum_{i}(\overline{z}_{i**} - \overline{z})^2}_{=DX} + \underbrace{nk\sum_{j}(\overline{z}_{*j*} - \overline{z})^2}_{=DY} +$$

$$\underbrace{k\sum_{i,j}(\overline{z}_{ij*} - \overline{z}_{i**} - \overline{z}_{*j*} + \overline{z})^2}_{DM} + \underbrace{\sum_{i,j,k}(z_{i,j,k} - \overline{z}_{ij*})^2}_{=DI}$$

Brīvības pakāpju skaiti ir šādi:

- DP - $nmk - 1$, $MP = \frac{DP}{nmk-1}$,
- DX - $n - 1$, $MX = \frac{DX}{n-1}$,
- DY - $m - 1$, $MY = \frac{DY}{m-1}$,

- DM - $(n-1)(m-1)$, $MM = \frac{DM}{(n-1)(m-1)}$,
- DI - $nm(k-1)$, $MI = \frac{DI}{nm(k-1)}$.

1.1. teorēma. Ir spēkā šādas vienādības:

1. $E(MX) = \sigma^2 + \frac{mk\sum_i \tau_i^2}{n-1}$,
2. $E(MY) = \sigma^2 + \frac{nk\sum_j \beta_j^2}{m-1}$,
3. $E(MM) = \sigma^2 + \frac{k\sum_{i,j} (\tau\beta)_{ij}^2}{(n-1)(m-1)}$,
4. $E(MI) = \sigma^2$.

Tādējādi hipotēžu pārbaudei var izmantot F-testu, ievērojot pareizo brīvības pakāpju skaitu.

1.5. piemērs.

1.2.3. Modeļa atbilstības pārbaude

Ir nepieciešams pārbaudīt, vai datu sadalījums ir normāls.

Ir lietderīgi pētīt atlikumus

$$e_{ijk} = z_{ijk} - \overline{z}$$

veicot normālā sadalījuma pieņēmuma pārbaudi.

1.2.4. Modeļa parametru novērtēšana

Modeļa parametrus μ, τ_i, β_j, $(\tau\beta)_{ij}$ var noteikt izmantojot mazāko kvadrātu metodi.

Sastādām summu

$$L = \sum_{i,j,k} \epsilon_{ijk}^2 = \sum_{i,j,k} (z_{ijk} - \mu - \tau_i - \beta_j - (\tau\beta)_{ij})^2.$$

Meklēsim tādas μ, τ_j vērtības, kas minimizē L.

Ir jāatrisina vienādojumu sistēma

$$\begin{cases} \frac{\partial L}{\partial \mu} = 0, \\ \frac{\partial L}{\partial \tau_i} = 0, \ \forall\ i, \\ \frac{\partial L}{\partial \beta_j} = 0, \ \forall\ j, \\ \frac{\partial L}{\partial (\tau\beta)_{ij}} = 0, \ \forall\ i, j. \end{cases}$$

Iegūtās sistēmas vienādojumi nav lineāri neatkarīgi, tāpēc tai ir bezgalīgi daudz atrisinājumu. Lai iegūtu vienu atrisinājumu, uzliekam papildus nosacījumus

$$\begin{cases} \sum_i \tau_i = 0, \\ \sum_j \beta_j = 0, \\ \sum_i (\tau\beta)_{ij} = 0, \\ \sum_j (\tau\beta)_{ij} = 0. \end{cases}$$

Atrisinot sistēmu, iegūsim

$$\begin{cases} \mu = \overline{z}, \\ \tau_i = \overline{z}_{i**} - \overline{z}, \\ \beta_j = \overline{z}_{*j*} - \overline{z}, \\ (\tau\beta)_{ij} = (\overline{z}_{ij*} - \overline{z}) - \underbrace{(\overline{z}_{i**} - \overline{z})}_{=\tau_i} - \underbrace{(\overline{z}_{*j*} - \overline{z})}_{=\beta_j} = \overline{z}_{ij*} - \overline{z}_{i**} - \overline{z}_{*j*} + \overline{z}. \end{cases}$$

1.2.5. Izlases apjoma noteikšana

Tiek izmantotas speciālu funkciju vērtību tabulas.

1.2.6. $n = 1$

1.3. Vispārīgais faktoriālais dizains

Ir r faktori $X_1, ..., X_r$, kas var pieņemt, attiecīgi, $n_1, ..., n_r$ fiksētas vērtības - kvalitatīvas (diskrētas) vai kvantitatīvas (nepārtrauktas).

Ja $r = 3$, tad var pieņemt šādu efektu modeli:

$$y_{ijkl} = \mu + \underbrace{\tau_i + \beta_j + \gamma_k}_{\text{faktoru ietekme}} + \underbrace{(\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk}}_{\text{faktoru pāru mijiedarbības ietekme}} + \underbrace{(\tau\beta\gamma)_{ijk}}_{\text{visu faktoru mijiedarbības ietekme}} + \epsilon_{ijkl}$$

1.4. Blokošana faktoriālajos dizainos

Ja faktoriālajā eksperimentā ir blakusfaktori, tad var būt nepieciešams apvienot eksperimentālās vienības blokos.

Vienkāršākajā gadījumā ir 2 faktori X (n līmeņi) un Y (m līmeņi). Apvienosim blokos nm eksperimentus ar dažādām faktoru vērtībām tā, lai katra faktoru kombinācija ir tieši vienu reizi.

Efektu modelis:

$$z_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk},$$

kur

- μ ir visu mērījumu vidējā vērtība,
- τ_i ir faktora X i-tā līmeņa efekts, $\sum_{i=1}^{n} \tau_i = 0$,
- β_j ir faktora Y j-tā līmeņa efekts, $\sum_{j=1}^{m} \beta_j = 0$,

- $(\tau\beta)_{ij}$ ir faktoru mijiedarbības efekts -

$$\sum_{i=1}^{n}(\tau\beta)_{ij} = \sum_{j=1}^{m}(\tau\beta)_{ij} = 0,$$

- δ_k ir k-tā bloka efekts,
- $\epsilon_{ijk} \sim N(0, \sigma^2)$ ir kļūdas gadījuma lielums.

Šajā modelī tiek pieņemts, ka nav mijiedarbības starp bloku un faktoriem.

2. 2^k faktoriālais dizains

2.1. Ievads

Var būt nepieciešams veikt eksperimentus, kuros ir k faktori, katrs no kuriem var pieņemt 2 vērtības -

- maksimālo un minimālo,
- normālo un uzlaboto/nobīdīto,
- normālo un ar kādu papildus efektu.

Pieņemam, ka

- faktoru vērtības ir fiksētas,
- dizaini ir pilnīgi randomizēti,
- kļūdas apmierina normālā sadalījuma likumus.

Šādus eksperimentus bieži veic pētījuma sākuma posmā, kurā ir liels skaits parametru, kuru ietekme ir jāpēta. Šajā posmā ir jāveic *skrīnings* - jānoskaidro, kādi parametri un parametru mijiedarbības visvairāk ietekmē interesējošos parametrus.

Faktoriālie eksperimenti ļauj veikt skrīningu ar vismazāko iespējamo eksperimentu skaitu.

2.1. piemērs. Tiek pētīti kāda organisma gēni - DNS apgabali, kas kodē proteīnu struktūru. Šādu gēnu skaits ir liels pat pēc atlases un ierobežojumiem. Var veikt eksperimentus, kuros tikai noteikta gēnu kopa ir "ieslēgta", ieslēgto gēnu kopu var mainīt. Veicot skrīningu - faktoriālos 2^k eksperimentus, var noteikt, kuriem gēniem ir lielāks efekts.

Skrīninga un citu 2^k eksperimentus ērtāk analizēt pētot ortogonālos kontrastus - faktoru efektu lineāras kombinācijas.

2.2. 2^2 dizains

2.2.1. Ievads un apzīmējumi

Ir 2 faktori X un Y, katrs no kuriem pieņem 2 vērtības - $-$ un $+$. Ar katru no četrām kombinācijām $--$, $+-$, $-+$, $++$ tiek veikti n eksperimenti.

Faktoru vērtības var attēlot Dekarta plaknē.

Apzīmēsim

- ar (1) visu mērījumu summu ar faktoru vērtībām $--$,
- ar x visu mērījumu summu ar faktoru vērtībām $+-$,
- ar y visu mērījumu summu ar faktoru vērtībām $-+$,
- ar xy visu mērījumu summu ar faktoru vērtībām $++$.

2.2.2. Efekti

Definēsim svarīgākos *efektu kontrastus*:

- faktora X efektu

$$X = \frac{\frac{xy-y}{n} + \frac{x-(1)}{n}}{2} = \frac{\Big(xy - y\Big) + \Big(x - (1)\Big)}{2n} = \frac{\Big(xy + x - y - (1)\Big)}{2n},$$

- faktora Y efektu

$$Y = \frac{\frac{xy-x}{n} + \frac{y-(1)}{n}}{2} = \frac{\Big(xy - x\Big) + \Big(y - (1)\Big)}{2n} = \frac{\Big(xy + y - x - (1)\Big)}{2n},$$

- faktoru mijiedarbības efektu

$$XY = \frac{\frac{xy-y}{n} - \frac{x-(1)}{n}}{2} = \frac{\frac{xy-x}{n} - \frac{y-(1)}{n}}{2} =$$

$$\frac{\Big(xy-y\Big) - \Big(x-(1)\Big)}{2n} = \frac{\Big(xy-x-y+(1)\Big)}{2n}.$$

2.2.3. Dispersiju analīze

Efektu kontrasti ir ortogonāli.

Var pierādīt, ka

$$DP = DX + DY + DM + DI, \text{ kur}$$

- $$DX = \frac{\Big(xy+x-y-(1)\Big)^2}{4n},$$

- $$DY = \frac{\Big(xy - x + y - (1)\Big)^2}{4n},$$

- $$DM = \frac{\Big(xy - x - y + (1)\Big)^2}{4n},$$

DX, DY, DM brīvības pakāpju skaits ir 1.

Lai pārbaudītu hipotēzes par kontrastu vienādību nullei, ir jāizmanto F-sadalījums.

2.2.4. Regresijas analīze

Ja pieņem, ka faktoru līmeņi ir kvantitatīvi, var atrast regresijas funkcijas izmantojot mazāko kvadrātu metodi.

2.3. 2^3 dizains

2.3.1. Efekti

Apzīmēsim faktorus ar x, y, z.
Definēsim šādus efektus (efektu kontrastus) :

- faktora X efektu

$$X = \frac{\Big(xyz + xy + xz - yz + x - y - z - (1)\Big)}{4n},$$

- faktora Y efektu

$$Y = \frac{\Big(xyz + xy - xz + yz - x + y - z - (1)\Big)}{4n},$$

- faktora Z efektu

$$Z = \frac{\Big(xyz - xy + xz + yz - x - y + z - (1)\Big)}{4n},$$

- faktoru X un Y mijiedarbības efektu

$$XY = \frac{\Big(xyz + xy - xz - yz - x - y + z + (1)\Big)}{4n},$$

- faktoru X un Z mijiedarbības efektu

$$XY = \frac{\Big(xyz - xy + xz - yz - x + y - z + (1)\Big)}{4n},$$

- faktoru Y un Z mijiedarbības efektu

$$XY = \frac{\Big(xyz - xy - xz + yz + x - y - z + (1)\Big)}{4n},$$

- visu faktoru mijiedarbības efektu

$$XY = \frac{\Big(xyz - xy - xz - yz + x + y + z - (1)\Big)}{4n},$$

2.3.2. Dispersiju analīze

Ir jāsadala pilnā dispersija un jāveic dispersiju analīze izmantojot F-sadalījumu.

2.4. 2^k dizains

Ir izstrādāta teorija visiem k.

3. 4.mājasdarbs

4.1 (Montgomery, DAO, lpp. 276) Darbgalda darba (dzīves) ilgums tiek pētīts atkarībā no darba intensitātes (X), griezošās detaļas tipa (Y) un griešanas leņķa (Z). Katram faktoram ir 2 līmeņi, tiek veikti 3 eksperimenti ar katru kombināciju. Rezultāti ir doti tabulā:

X	Y	Z	Apz.	1.mēr	2.mēr	3.mēr
-	-	-	(1)	22	31	25
+	-	-	x	32	43	29
-	+	-	y	35	34	50
+	+	-	xy	55	47	46
-	-	+	z	44	45	38
+	-	+	xz	40	37	36
-	+	+	yz	60	50	54
+	+	+	xyz	39	41	47

(a) Novērtējiet visu faktoru un to mijiedarbību efektus. Kādi no tiem šķiet svarīgākie?
(b) Veiciet dispersiju analīzi, pārbaudiet hipotēzes.
(c) Veiciet regresijas analīzi, izmantojot mazāko kvadrātu metodi.
(d) Izpētiet atlikumus.
(e) Kādus faktoru līmeņus var rekomendēt, lai palielinātu darbgalda darba ilgumu?