Baysísk tölfræði og raðkvarðar

Helgi Tómasson, helgito@hi.is

Þjóðarspegillinn: Ráðstefna í Félagsvísindum IX

24. október 2008

Skipulag erindis

Skipulag erindis

Markmið greinar og fyrirlestrar

Skipulag erindis

Markmið greinar og fyrirlestrar
Nokkur orð um líkanasmíði

Skipulag erindis

Markmið greinar og fyrirlestrar
Nokkur orð um líkanasmíði
Bayesísk heimspeki og tækni

Skipulag erindis

Markmið greinar og fyrirlestrar
Nokkur orð um líkanasmíði
Bayesísk heimspeki og tækni
Líkön fyrir raðkvarða

Skipulag erindis

Markmið greinar og fyrirlestrar
Nokkur orð um líkanasmíði
Bayesísk heimspeki og tækni
Líkön fyrir raðkvarða
Repeated-measures/panel-data/longitudinal-data

Skipulag erindis

Markmið greinar og fyrirlestrar
Nokkur orð um líkanasmíði
Bayesísk heimspeki og tækni
Líkön fyrir raðkvarða
Repeated-measures/panel-data/longitudinal-data
Sýnidæmi, einvítt kennslubókardæmi, kennslukannanir

Markmið

-
-

Markmið

- Leggja áherslu á að tekið sé tillit til allra þátta í líkanasmíði
-

Markmið

- Leggja áherslu á að tekið sé tillit til allra þátta í líkanasmíði
- Sýna sveigjanleika bayesíkrar nálgunar í líkanasmíði

Markmið

- Leggja áherslu á að tekið sé tillit til allra þátta í líkanasmíði
- Sýna sveigjanleika bayesíkrar nálgunar í líkanasmíði
- Sýna dæmi um líkansmíði fyrir raðkvarða

Markmið

- Leggja áherslu á að tekið sé tillit til allra þátta í líkanasmíði
- Sýna sveigjanleika bayesíkrar nálgunar í líkanasmíði
- Sýna dæmi um líkansmíði fyrir raðkvarða
- Vekja umhugsun gagnsemi markaðskannana með sýnidæmi um kennslukannanir

2x2 töflur villandi

	starf=1	starf=2
karlar	154.000	241.429
konur	126.667	200.000

Tafla 1: Laun eftir starfi og kyni

- Lærdómur:

2x2 töflur villandi

	starf=1	starf=2
karlar	154.000	241.429
konur	126.667	200.000

Tafla 1: Laun eftir starfi og kyni

	aldur=1	aldur=2
karlar	145.000	235.000
konur	122.500	190.000

Tafla 2: Laun eftir aldri og kyni

- Lærdómur:

2x2 töflur villandi

	starf=1	starf=2
karlar	154.000	241.429
konur	126.667	200.000

Tafla 1: Laun eftir starfi og kyni

	aldur=1	aldur=2
karlar	145.000	235.000
konur	122.500	190.000

Tafla 2: Laun eftir aldri og kyni

- Lærdómur:Þetta er hlut af kennsluefni sem ég hef notað í 20 ár um að ekki megi sleppa mikilvægum breytum

2x2 töflur villandi

	starf=1	starf=2
karlar	154.000	241.429
konur	126.667	200.000

Tafla 1: Laun eftir starfi og kyni

	aldur=1	aldur=2
karlar	145.000	235.000
konur	122.500	190.000

Tafla 2: Laun eftir aldri og kyni

- Lærdómur:Þetta er hlut af kennsluefni sem ég hef notað í 20 ár um að ekki megi sleppa mikilvægum breytum
- Gögnin eru búin til með líkani sem mismunar konum í hag, en allar 2x2 töflur gefa karla með hærri laun

Það má ekki slá ólíkum hópum saman

Það má ekki slá ólíkum hópum saman

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 3: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

Það má ekki slá ólíkum hópum saman

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 3: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

	Há laun	Lág laun
Karlar	2	8
Konur	9	21

Tafla 4: Launadreifing í fyrirtæki B. 30% kvenna með há laun, 20% karla með há laun.

Það má ekki slá ólíkum hópum saman

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 3: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

	Há laun	Lág laun
Karlar	2	8
Konur	9	21

Tafla 4: Launadreifing í fyrirtæki B. 30% kvenna með há laun, 20% karla með há laun.

	Há laun	Lág laun
Karlar	20	20
Konur	16	24

Tafla 5: Launadreifing í fyrirtækjum A+B. 40% kvenna með há laun, 50% karla með há laun.

Bayesísk hugsun

- Þess vegna getur Bayesisti sagt: Breytan X er $N(\mu, \sigma^2)$ og vissa mín um μ er $N(a, \tau)$.
  1. Set fram líkan $f(x|\theta)$
  2. Set fram „a priori" vitnseskju: $\pi(\theta)$
  3. Safna gögnum x.
  4. Set fram „a posteriori" vitneskju $\pi(\theta|x)$

$$\pi(\theta|x) \propto \pi(\theta) l(\theta|x), \quad l(\theta|x) = f(x|\theta)$$
$$l(\theta|x) = \text{sennileikafall=likelihood-fall}$$

Bayesísk hugsun

- Bayesistar leyfa að líkur séu túlkaðar sem mælikvarði á vissu Þess vegna getur Bayesisti sagt: Breytan X er $N(\mu, \sigma^2)$ og vissa mín um μ er $N(a, \tau)$.
  1. Set fram líkan $f(x|\theta)$
  2. Set fram „a priori" vitnseskju: $\pi(\theta)$
  3. Safna gögnum x.
  4. Set fram „a posteriori" vitneskju $\pi(\theta|x)$

$$\pi(\theta|x) \propto \pi(\theta) l(\theta|x), \quad l(\theta|x) = f(x|\theta)$$
$$l(\theta|x) = \text{sennileikafall=likelihood-fall}$$

Bayesísk hugsun

- Bayesistar leyfa að líkur séu túlkaðar sem mælikvarði á vissu
- Þess vegna getur Bayesisti sagt: Breytan X er $N(\mu, \sigma^2)$ og vissa mín um μ er $N(a, \tau)$.
  1. Set fram líkan $f(x|\theta)$
  2. Set fram „a priori" vitnseskju: $\pi(\theta)$
  3. Safna gögnum x.
  4. Set fram „a posteriori" vitneskju $\pi(\theta|x)$

$$\pi(\theta|x) \propto \pi(\theta) l(\theta|x), \quad l(\theta|x) = f(x|\theta)$$
$$l(\theta|x) = \text{sennileikafall=likelihood-fall}$$

Bayesísk hugsun

- Bayesistar leyfa að líkur séu túlkaðar sem mælikvarði á vissu
- Þess vegna getur Bayesisti sagt: Breytan X er $N(\mu, \sigma^2)$ og vissa mín um μ er $N(a, \tau)$.
- Gangurinn er eftirfarandi
  1. Set fram líkan $f(x|\theta)$
  2. Set fram „a priori" vitnseskju: $\pi(\theta)$
  3. Safna gögnum x.
  4. Set fram „a posteriori" vitneskju $\pi(\theta|x)$

$$\pi(\theta|x) \propto \pi(\theta) l(\theta|x), \quad l(\theta|x) = f(x|\theta)$$
$$l(\theta|x) = \text{sennileikafall=likelihood-fall}$$

Tæknatriði

- $\pi(\theta|x)$ er flókið fall
- Ef líkanið $f(x|\theta)$ og prior-dreifingin $\pi(\theta)$ mynda „conjugate"par þá eru til lokuð form á $\pi(\theta|x)$.
- Slíkt ekki í boði fyrir flókin líkön og þess vegna gengur tækni nútímans út á að herma dreifinguna $\pi(\theta|x)$.
- Hermunartækni byggir á fræðunum um Markov keðjur. Þ.e. að hönnuð er Markov keðja sem hefur $\pi(\theta|x)$ sem jafnvægisdreifingu.
- MCMC=Markov-Chain-Monte-Carlo er nafn á slíkum hermiaðferðum.
- Gibbs-sampling og Metropolis-Hastings eru tækni útfærslur á MCMC.

Um raðkvarða

Vinsæl hugmynd er að mælda breytan, y, sé klippt útgáfa af samfelldri ómældri (latent) breytu $y*$. Ef breytna y tekur gildi frá 1 til k þá,

$$y = \begin{cases} 1 & y^* < c_1 \\ 2 & c_1 < y^* < c_2 \\ \vdots & \vdots \\ k & c_{k-1} < y^* \end{cases} . \qquad (1)$$

Punktarnir $c_1, \ldots, c_{k-1}$ eru þröskuldar í breytunni y^* sem gefa röðuðu breytunni y gildi. Breytan y^* er ekki mæld en gert er ráð fyrir hún fylgi einhverri dreifingu, t.d.:

$$y^* \sim N(0, 1) \qquad (2)$$

Hægt er að bæta við skýribreytum:

$$y^* = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon \quad \varepsilon \sim N(0,1) \tag{3}$$

Ef fyrir hendi eru gögn á forminu:

$$\begin{array}{ccccc} y_1 & x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots & \\ y_n & x_{n1} & x_{n2} & \cdots & x_{np} \end{array}$$

þá er vandinn að meta $\boldsymbol{\beta}$ og $(c_1, \cdots, c_{k-1})$. Líkanið gefur þá mat á líkunum,

$$P(y \leq j | \boldsymbol{X} = \boldsymbol{x}) = \Phi(c_j - \boldsymbol{x}\boldsymbol{\beta}), \tag{4}$$

þar sem Φ er dreififall staðlaðrar normalhendingar. Hugmyndin er að mældi kvarðinn sé klippt útgáfa af ómældrar breytu, y_i^*. Til að stikar sé greinanlegir (identifiable) þarf því einhvers konar skorður. Dæmi um slíkar skorður er að setja staðalfrávik $\boldsymbol{y}^*$ jafnt 1. Jafna (4) sýnir að útreikningur sennileikafallsins (likelihood function) er einfaldur og því hægt að finna mat mesta sennileika (ML-mat) á $\boldsymbol{\beta}$ og $(c_1, \cdots, c_{k-1})$.

Margvíð líkön (spurningalistar)

Með hugtakinu margvítt líkan (multivariate model) er átt við að fyrir hvern einstakling eru mörg svör. Þegar verið er að mæla fólk er oft nauðsynlegt að leiðrétta fyrir til kyni, aldri og hugsanlega fleiri breytum. T.d. margvítt normal aðhvarfslíkan:

$$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})' = \boldsymbol{x}_i \boldsymbol{B} + \boldsymbol{E}_i \quad (5)$$

$$\boldsymbol{B} = \begin{bmatrix} \beta_{11} & \cdots & \cdots & \beta_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \cdots & \cdots & \beta_{kp} \end{bmatrix} \quad (6)$$

$$V(\boldsymbol{E}_i) = V((\varepsilon_{i1}, \ldots, \varepsilon_{ip})') = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \vdots & \ddots & & \vdots \\ \sigma_{1p} & \cdots & \cdots & \sigma_{pp} \end{bmatrix} \quad (7)$$

Hér er $\boldsymbol{y}_i$ vektor að breytum sem skýra á hjá einstaklingi i. Skýristærðirnar eru í vektornum $\boldsymbol{x}_i$. Samdreifnistikanir (covariance), σ_{ij} í Σ fylkinu lýsa tengslum hnita i og j í í mælda vektornum þegar leiðrétt hefur verið fyrir $\boldsymbol{x}_i$.

Ef hnit í mældum vektor $\boldsymbol{y}_i$ samanstanda af gildum af raðkvörðum þá er breytan ekki samfelld og má ætla að normaldreifingin sé ekki góð nálgun. Vel má hugsa sér svipaða nálgun og í einvíða tilfellinu, þ.e. að raðkvarðinn sé klippt útgáfa að ómældri undirliggjandi breytu $\boldsymbol{y}^*$. Í slíku líkani yrðu þá þröskuldar í hverri hnit. Eins og í einvíða tilfellinu þarf að setja skorður (hliðstætt því að setja staðalfrávik ómældu stærðarinnar jafnt 1). Margvítt líkan er flóknara en einvítt og koma margar útfærslur af slíkum skorðum til greina. Mörgum mismunandi útfærslum er lýst af t.d. Agresti (2002), Congdon (2005) og Rossi, Allenby & McCulloch (2005). Þar er gengið út frá því að einstaklingur i svari J spurningum. Svör einstaklings i við spurningu j eru táknuð með breytunni y_{ij}, sem tekur gildi frá 1 upp í k. Hugmyndin er að hvert hnit í y_i sé skorin útgáfa af y_i^* sem fylgi margvíðri normaldreifingu:

$$y_i^* \sim N(\mu_i^*, \Sigma_i^*). \tag{8}$$

Líkanið leyfir því mismunandi kvarðanotkun einstaklinga með því að vera með einstaklingsbundin μ_i^* og Σ^*. Að sjálfsögðu þarf ýmis hliðarskilyrði til að hægt sé að meta svona.

Í spurningakönnum þar sem hver og einn kvarðar sitt svar á ekki að reikna með að

1.
2.

Í spurningakönnum þar sem hver og einn kvarðar sitt svar á ekki að reikna með að

1. Að hægt sé að draga „ratio-scale" ályktun (Rossi, Gilula & Allenby, 2001)
2.

Í spurningakönnum þar sem hver og einn kvarðar sitt svar á ekki að reikna með að

1. Að hægt sé að draga „ratio-scale" ályktun (Rossi, Gilula & Allenby, 2001)
2. Hver hefur sitt level og skrefstærð.

Í spurningakönnum þar sem hver og einn kvarðar sitt svar á ekki að reikna með að

1. Að hægt sé að draga „ratio-scale" ályktun (Rossi, Gilula & Allenby, 2001)
2. Hver hefur sitt level og skrefstærð.
   Samanber Celcius of Fahrenheit.

Í spurningakönnum þar sem hver og einn kvarðar sitt svar á ekki að reikna með að

1. Að hægt sé að draga „ratio-scale" ályktun (Rossi, Gilula & Allenby, 2001)
2. Hver hefur sitt level og skrefstærð.
   Samanber Celcius of Fahrenheit.

	Númer spurningar					
	1	2	3	4	5	6
Einstaklingur 1	4	5	4	5	4	5
Einstaklingur 2	2	3	2	3	2	5
Einstaklingur 3	1	5	1	5	1	5

Tafla 6: Sammála einstaklingar.

Einfalt einvítt dæmi

Til að sýna að hvernig nota má nýtísku Bayes-aðferðafræði við útskýringu á raðkvörðum er hér tekið einfalt sýnidæmi úr R-tölfræðipakkanum R Development Core Team (2005). Gert er ráð fyrir að tvær mældar skýristærðir, X_1 og X_2. Gert er ráð fyrir að sanna líkanið sé á forminu:

$$y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_{i2} x_{i2} + \varepsilon_i \quad \varepsilon_i \sim N(0,1) \tag{9}$$

$$y_i = \begin{cases} 1 & -\infty < y_i^* < c_1 \\ 2 & c_1 < y_i^* < c_2 \\ \vdots & \vdots \\ 5 & c_4 < y_i^* < \infty \end{cases} \tag{10}$$

Skýristærðirnar X_1 og X_2 eru jafndreifðar $U(0,5)$. Vektorinn $\boldsymbol{\beta} = (0.5, 1, -0.5)$. Gert er ráð fyrir 300 mælingum og að y sé kvörðuð, 1,2,3,4,5. Gögn gætu til dæmis litið út eins og sýnt er í töflu 7.

i	y	x_1	x_2
1	5	4.95	2.26
2	4	4.15	2.66
$\vdots$	$\vdots$	$\vdots$	$\vdots$
299	2	2.42	4.31
300	1	1.14	1.00

Tafla 7: Dæmi um gögn í einföldu dæmi.

Hægt er að meta líkanið með hefðbundnum ML aðferð. Niðurstöðurnar eru sýndar í töflu 8.

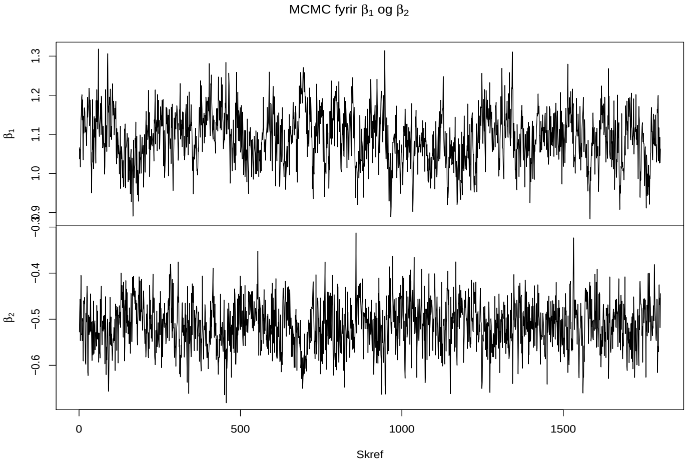
	$\hat{\beta}_{ML}$	s.e.-β	t-gildi
X_1	1.09	0.072	15.24
X_2	-0.51	0.054	-9.43

Tafla 8: Niðurstaða ML-mats í einföldu dæmi.

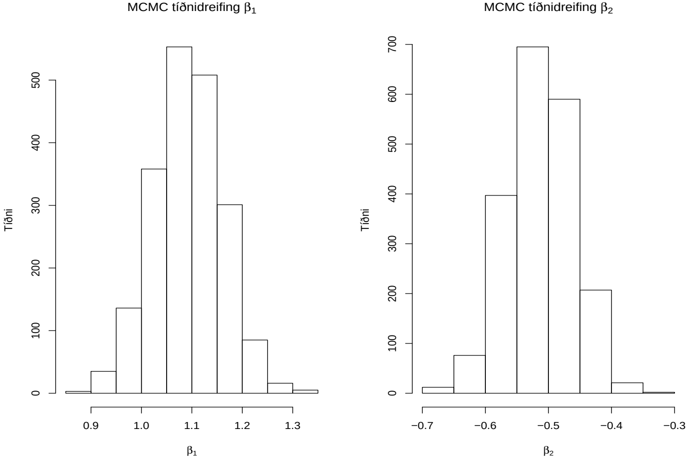
Með því að byrja með mjög veikar fyrirfram upplýsingar fæst með bayesískt mat með Gibbs-MCMC aðferð. Í töflu 9 er sýnt úrtaksmeðaltal og staðalfrávik Markov-keðjunnar. Reiknuð voru 2000 skref of fyrstu 200 hent.

	$E(\beta)$	$\sqrt{V(\beta)}$
X_1	1.08	0.072
X_2	-0.51	0.054

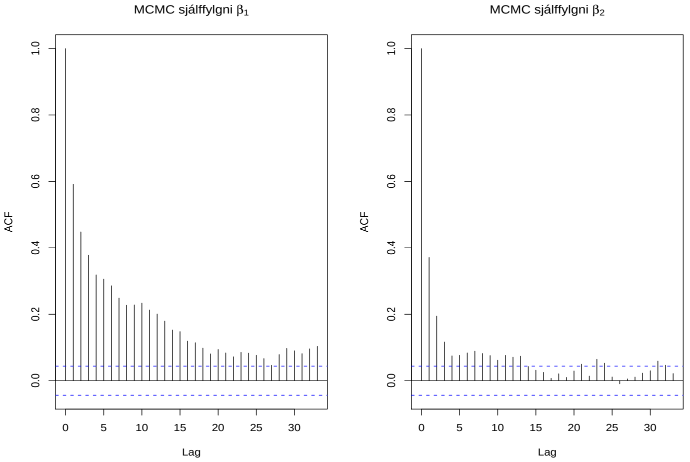
Tafla 9: Meðaltal og staðalfrávik bayesísks mats.



Mynd 1: MCMC keðjur fyrir β_1 og β_2.



Mynd 2: Tíðnidreifing MCMC fyrir β_1 og β_2.

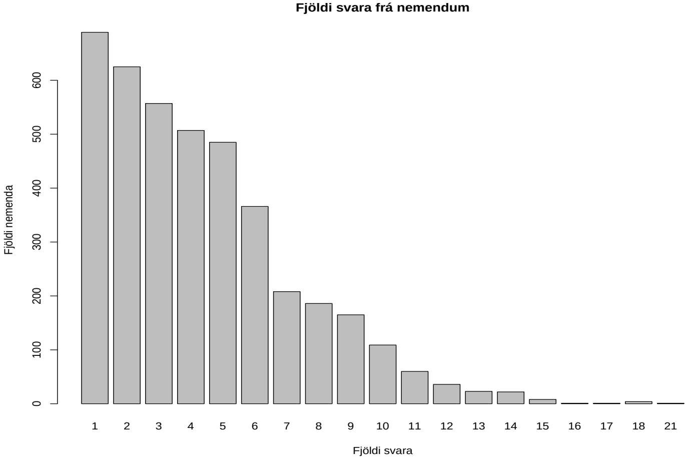


Mynd 3: Sjálffylgni MCMC-keðju fyrir β_1 og β_2.

Gögn úr kennslukönnun

Í markaðskönnun er það tíðkað að svarendur eru látnir kvarða ýmis atriði með spurningalista. Dæmi um slíkt eru kennslukannanir sem framkvæmdar eru í Háskóla Íslands og eiga sér hliðstæður í öðrum háskólum víða um heim. Að áliti höfundar er notkun viðeigandi líkana til ályktana út frá slíkum mælingum vanþróuð. Kannanirnar fara þannig fram að safnað er svörum við ákveðnum spurninarlista frá nemendum. Hverjum nemanda býðst að svara slíkum lista fyrir hvert námskeið sem viðkomandi er skráður í. Það er ekki skylda að svara og ekkert eftirlit er með því hvort nemandinn svarar einn og óháður eða hvernig hann ber sig að við svörin. Ljóst er því að gæði mælinganna er mjög misjöfn.

Höfundur hefur undir höndum gögn um könnunina haustið 2005. Henni svöruðu 4053 nemendur frá einu svari per nemanda upp í 21 svör per nemanda. Fjöldi svara per nemenda er sýndur í mynd 4. Þar sést að í kringum 400-500, þ.e. samtals um 20%, svara 4 eða 5 námskeiðum. Svarendur voru sjálfum sér samkvæmir með kyn sitt, þ.e. innan hvers einstaklings þá var kynið óbreytt. Hugsanlega er þetta leiðrétt með vélrænum hætti. Karlkyns svarendur voru 1384, kvenkyns svarendur voru 2603 og kyn óuppgefið 66. Alls bárust 17723 svör.



Mynd 4: Dreifing svarfjölda eftir nemendum.

Kyn er möguleg skýribreyta og sömuleiðis er hugsanlegt að nota spurninguna: Hvað einkunn telur þú að þú fáir í námskeiðinu?, sem skýrirbreytu. Þýðing þessara breyta var metin fyrir eina spurningu hjá í einu námskeiði. Alls 96 svör. Niðurstöður úr ML-mati eru sýndar í tölfu 10. Sömu gögn voru metin með bayesískri aðferð þar sem veikar fyrirfram upplýsingar voru notaðar. Matið sem sýnt er í töflu 11 er reiknað með 10000 umferðum í Gibbs-sampler.

	$\hat{\beta}_{ML}$	s.e.-β	t-gildi
kyn	0.15	0.22	0.67
einkunn	0.31	0.11	2.89

Tafla 10: Niðurstaða ML-mats fyrir eina spurningu í einu námskeiði.

	$E(\beta)$	$\sqrt{V(\beta)}$
kyn	0.15	0.22
einkunn	0.31	0.11

Tafla 11: Meðaltal og staðalfrávik bayesísks mats fyrir eina spurningu í einu námskeiði.

Eðlilegt er að gera rá fyrir að nemendur noti kvarðann á mismunandi hátt. Heitin „repeated-measures", „panel-data" og „longitudinal-data" eru samheiti á því fyrirbæri að margar mælingar eru á hverjum einstaklingi. Í þeim fræðum er mikilvægur þáttur að leiðrétta fyrir einstaklingsáhrifum. Hugtök „first-difference"(FD) og „fixed-effecs"(FE) eru tæknilega nálganir sem hafa það að markmiði að leiðrétta fyrir einstaklingsáhrifum. Um þær aðferðir má lesa í kennslubókum eins og t.d. Cameron & Trivedi (2005); Wooldridge (2002); Pinheiro & Bates (2000). Einfalt og algengt er að notast við einstaklingsstöðluð svör,

$$(x_{ij} - \bar{x}_i)/s_i, \qquad (11)$$

þar sem x_{ij} er svar einstaklings i við spurningu j, $\bar{x}_i$ er meðaltal hjá einstakling númer i og s_i er staðalfrávik hjá einstakling númer i. Það að draga $\bar{x}_i$ frá leiðréttir fyrir því að einstaklingar noti skalann misnunandi hátt á þann hátt að þeir staðsetji sig á mismunandi hátt. Það að deila með s_i er viðleitni til að leiðrétta fyrir því að skrefstærðin í kvarðanum sé mismunandi eftir einstaklingum, t.d. svipað og að sumir mældu hitann á Fahrenheit og aðrir á Celius.

Til að sýna hvernig nota megi aðferðafræði Rossi, Gilula & Allenby (2001) var valið eitt námskeið með 96 gildum svörum við 16 spurningum.

$$y_{ij} = \begin{cases} 1 & -\infty < y_{ij}^* < c_1 \\ \vdots & \vdots \end{cases} \tag{12}$$

$$y_i^* = (y_{i1}, \ldots, y_{iJ})', \quad J = 16. \tag{13}$$

$$y_i^* \sim (\mu_i^*, \Sigma_i^*). \tag{14}$$

Til að gera stika greinanlega eru settar ýmsar skorður, t.d.:

$$\mu_i^* = \mu + \tau_i + \sigma_i z_i, \quad z_i \sim N(0, \Sigma) \tag{15}$$

þar sem z_i er 16 víð ómæld breyta með samdreifnifylki Σ. Gibbs sampler er settur upp í nokkrum þrepum. Þrepin eru valin þannig að skilyrtar dreifingar verði þægilegar.

1	2	3	4	5	6	7	8	9	10	11
1.00	0.06	0.39	0.01	0.10	0.24	0.04	−0.07	0.10	0.05	0.02
0.06	1.00	0.26	0.14	0.26	0.22	0.29	0.24	0.07	−0.09	−0.01
0.39	0.26	1.00	0.05	0.12	0.35	0.12	0.02	0.06	−0.03	−0.10
0.01	0.14	0.05	1.00	0.09	−0.02	0.13	0.21	0.18	0.23	0.20
0.10	0.26	0.12	0.09	1.00	0.17	0.33	0.32	0.04	−0.11	0.12
0.24	0.22	0.35	−0.02	0.17	1.00	0.40	0.11	−0.05	−0.09	0.02
0.04	0.29	0.12	0.13	0.33	0.40	1.00	0.48	0.05	−0.13	0.01
−0.07	0.24	0.02	0.21	0.32	0.11	0.48	1.00	0.03	0.01	0.06
0.10	0.07	0.06	0.18	0.04	−0.05	0.05	0.03	1.00	0.11	−0.05
0.05	−0.09	−0.03	0.23	−0.11	−0.09	−0.13	0.01	0.11	1.00	0.21
0.02	−0.01	−0.10	0.20	0.12	0.02	0.01	0.06	−0.05	0.21	1.00
−0.14	0.09	−0.15	0.22	0.10	−0.10	0.02	0.12	−0.04	0.14	0.40
−0.05	0.19	−0.11	0.26	0.12	−0.00	0.26	0.25	−0.03	0.14	0.36
0.03	−0.18	0.01	−0.06	−0.21	−0.06	−0.27	−0.17	−0.17	−0.00	−0.05
−0.03	−0.24	−0.02	−0.07	−0.23	−0.19	−0.33	−0.19	−0.07	0.01	−0.06
−0.00	−0.16	0.02	−0.11	−0.27	−0.08	−0.21	−0.15	−0.11	0.03	−0.10

Tafla 12: Einstaklingshreinsað mat á fylgni svara.

Hlutverk	Námskeiðið er bæði örvandi og krefjandi
Kennari er áhugasamur um kennsluna	Kennari heldur áhuga þínum í kennslustun
Markmið og námskeiðslýsing er í samræmi við kennslu	Það er auðvelt að skrifa niður glósur í fyrir
Kennari hvetur til spurninga/umræðna um efni námskeiðsins	Kennari er jákvæður gagnvart hugmyndum
Viðmót kennara gagnvart stúdentum er vinsamlegt	Stúdentar hafa nægan aðgang að kennara
Próf og verkefni í námskeiðinu endurspegla áherslur í kennslu	Námsefni og verkefni ýta undir skilning á e
Framsetning er skýr og skilmerkileg	Námskeiðið er vel skipulagt
Aðstaða til kennslunnar er fullnægjandi	Tæki/áhöld eru í góðu ásigkomulagi
Fjöldi tækja/áhalda er nægjanlegur	Stúdentar hafa nægan aðgang að tækjum/
Stúdentar fá nægar leiðbeiningar/aðstoð um notkun tækja/áhalda	Námskeiðið er, í samanburði við önnur H.Í
Kennarinn er, í samanburði við aðra H.Í.-kennara?	Vinnuálag í námskeiði miðað við önnur ná
Hraði yfirferðar í námskeiðinu er?	Að jafnaði var vinna á viku vegna námskei
Hvert er kyn þitt?	Deildarnúmer
Hversvegna tókstu þetta námskeið?	Hvaða einkunn býst þú við að fá í námskei
Áhugi á efni námskeiðsins við upphaf þess?	Áhugi á efni námskeiðs við lok þess?
Kennitala nemanda	

Tafla 13: Spurningar í kennslukönnun.

Tilvitnanir

Tilvitnanir

Hér hefur verið rakið hvernig setja megi upp bayesískt líkan fyrir einvíða og margvíða raðkvarðahendingu. Augljóslega er hægt að bæta við skýristærðum, t.d. aldri, kyni o.s.frv. Í markaðskönnunum er gögnum stundum safnað með spurningalistum eru svarendur stundum látnir kvarða svör sín. Ef svarendur hafa ekki samræmt sig fyrirfram er eðlilegt að gera ráð fyrir að þeir noti kvarðann mismunandi.

Tilvitnanir

Hér hefur verið rakið hvernig setja megi upp bayesískt líkan fyrir einvíða og margvíða raðkvarðahendingu. Augljóslega er hægt að bæta við skýristærðum, t.d. aldri, kyni o.s.frv. Í markaðskönnunum er gögnum stundum safnað með spurningalistum eru svarendur stundum látnir kvarða svör sín. Ef svarendur hafa ekki samræmt sig fyrirfram er eðlilegt að gera ráð fyrir að þeir noti kvarðann mismunandi. Rossi, Gilula & Allenby (2001) segja: „*we focus on two major uses of customer-statisfaction-measurement ratings data: (1) measurement of the relationship between overall satisfaction with the specific product attributes and (2) identification of customers with extreme views. Scale usage heterogeneity can substantially bias analysis aimed at either use*". Því er oft nauðsynlegt að leiðrétta fyrir því að einstaklingar noti kvarðann á mismunandi hátt.

Slík leiðrétting er þó alls ekki allra meina bót. Það má ekki slá misleitum hópum saman. Í kennslubók í markaðsfræðum segir Churchill (1995) „*it has been observed that there are large cultural or cross-country differences in scale usage, making it difficult to combine data across cultural or international boundaries*". Hæfni, vilji og smekkur svarenda ræður gæðum gagna í spurningakönnun.

Slík leiðrétting er þó alls ekki allra meina bót. Það má ekki slá misleitum hópum saman. Í kennslubók í markaðsfræðum segir Churchill (1995) „*it has been observed that there are large cultural or cross-country differences in scale usage, making it difficult to combine data across cultural or international boundaries*". Hæfni, vilji og smekkur svarenda ræður gæðum gagna í spurningakönnun.

Morris (1978) lýsir líkindafræðilegu líkani sem meta á hæfni einstaklinga til að greina skemmt epli frá óskemmdu. Hann nefnir einnig vínsmökkun eins og Randall (1989). Ljóst er að sumir einstaklingar geta ekki greint skemmt frá óskemmdu og sumir sem geta greint á milli gætu metið skemmt epli betur.

Ef vit á að vera í markaðskönnunum verður að nota líkan sem ræður við svona atriði. Það má ekki slá saman hópum sem hugsanlega snúa kvarðanum sitt í hvora áttina, samanber dæmið með fyrirtæki A og fyrirtæki B. Í því dæmi virtist sambandið í summunni vera öðruvísi en í hvoru fyrir sig.

Ef vit á að vera í markaðskönnunum verður að nota líkan sem ræður við svona atriði. Það má ekki slá saman hópum sem hugsanlega snúa kvarðanum sitt í hvora áttina, samanber dæmið með fyrirtæki A og fyrirtæki B. Í því dæmi virtist sambandið í summunni vera öðruvísi en í hvoru fyrir sig.

Jafnvel þó að allar forsendur séu uppfylltar þarf að túlka útkomur með varúð. Höfundur þessarar greinar er algerlega sammála Parducci (1968) og Rossi, Gilula & Allenby (2001), „*we do not believe that even properly adjusted ratings data can provide **ratio scale information.** For example, if a respondent gives only ratings at the top end of the scale, we cannot infer that he or she is extremely satisfied*", þ.e. ekki má álykta að sá sem svarar öllum spurningum með hæsta gildi kvarðans afar ánægður. Einstaklingsleiðréttingin býður upp á einstaklingshreinsað mat á fylgnifylki (eða samdreifnifylki) margvíðu, ómældu hendingarinnar sem liggur að baki kvörðuðum svörum.

Að ofansögðu ætti að vera ljóst að ég tel ekki rétt að framkvæma þáttagreiningu á óleiðréttum gögnum. Það má aldrei sleppa mikilvægum skýristærðum. Rampichini, Grilli & Petrucci (2004) segja: ... *it should be recognized that the satisfaction of a student, as expressed by the ratings, depends not only on the course characteristics of interest, but also on the student's traits and expectations. Therefore a fair comparison among courses requires the calculation of net measures that* **adjust for individual characterisics. Such measures can be obtained, among others, by means of multilevel models** ...

og síðar í sömu grein:

The descriptive indicators proposed in the previous section are mainly useful for the analysis of a single course. However, when it comes to comparison among courses, the are not appropriate since the evaluations expressed by the students are influenced by individual expectations and traits of the student themselves. Any comparison among courses that is based on the descriptive indicators, could be misleading. .. **To make fair comparison among courses it is necessary to isolate the influence of these components on the expressed evaluations. In particular, it is important to net out the effects related to the individual characteristics of the students in order to obtain fair rankings of courses, schools, universities, etc.**

Lokaorð

-

Lokaorð

- Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði.

Lokaorð

- Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði.
- Í útfærslunni hér var sú hæpna forsenda notuð að einstaklingar hefðu sameiginlegt fylgnifylki þegar búið væri að hreins einstaklinsáhrif frá.

Lokaorð

- Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði.
- Í útfærslunni hér var sú hæpna forsenda notuð að einstaklingar hefðu sameiginlegt fylgnifylki þegar búið væri að hreins einstaklinsáhrif frá.
- Hugsanlega snúa einhverjir kvarðanum öfugt, t.d. vilja heldur skemmd epli

Lokaorð

- Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði.
- Í útfærslunni hér var sú hæpna forsenda notuð að einstaklingar hefðu sameiginlegt fylgnifylki þegar búið væri að hreins einstaklinsáhrif frá.
- Hugsanlega snúa einhverjir kvarðanum öfugt, t.d. vilja heldur skemmd epli
- Slíkt kallar á að meta mörg fylgnifylki, þ.e. „mixture-modelling". Í stærðfræði I gæti ég trúað að væru 3 hópar.

Lokaorð

- Hugleiðingar þessara höfunda um kennslukannanir eiga við um markaðskannanir og alla aðra hagnýta tölfræði.
- Í útfærslunni hér var sú hæpna forsenda notuð að einstaklingar hefðu sameiginlegt fylgnifylki þegar búið væri að hreins einstaklinsáhrif frá.
- Hugsanlega snúa einhverjir kvarðanum öfugt, t.d. vilja heldur skemmd epli
- Slíkt kallar á að meta mörg fylgnifylki, þ.e. „mixture-modelling”. Í stærðfræði I gæti ég trúað að væru 3 hópar.
- Bayesísk aðferðafræði opnar möguleika á vinnu með flókin líkön án þess að rannsakendur þurfi að skuldbindast bayesísku heimspekinni. Ef „diffuse”-prior er notaður er útkoman oft svipuð og ML.

Agresti, A. (2002). *Categorical Data Analysis, 2nd edition*. John Wiley & Sons: New York.

Cameron, A. & Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Churchill, G. (1995). *Marketing Research: Methodological Foundations* (6th ed.). Orlando: Dryden Press.

Congdon, P. (2005). *Bayesian Models for Categorical Data*. John Wiley & sons.

Morris, D. (1978). A probability model for forced binary choice. *The American Statistician*, *32*(1), 23–25.

Parducci, A. (1968). The relevatism in absolute judgement. *Scientific American*, *219*, 84–90.

Pinheiro, J. & Bates, D. (2000). *Mixed-effects models in S and S-Plus*. Springer.

R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rampichini, C., Grilli, L., & Petrucci, A. (2004). Analysis of university course evaluations: from descriptive measures to multilevel models. *Statistical Methods & Applications*, *13*, 357–373.

Randall, J. (1989). The analysis of sensory data by generalized linear models. *Biometrical Journal*, *31*, 783–791.

Rossi, P., Allenby, G., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley & sons.

Rossi, P., Gilula, Z., & Allenby, G. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20–31.

Wooldridge, J. (2002). *Econometric Analysis of cross section and panel data*. MIT press.