Harjoitusten 2 ratkaisuehdotukset

1.

a) Tutkittavat matriisit ovat

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{kn} \end{bmatrix},$$

$$\mathbf{AB} = \begin{bmatrix} \sum_{i=1}^{k} a_{1i}b_{i1} & * & * \\ * & \vdots & * \\ * & * & \sum_{i=1}^{k} a_{ni}b_{in} \end{bmatrix},$$

ja

$$\mathbf{BA} = \begin{bmatrix} \sum_{i=1}^{n} b_{1i}a_{i1} & * & * \\ * & \vdots & * \\ * & * & \sum_{i=1}^{n} b_{ki}a_{ik} \end{bmatrix},$$

joissa kahdesta viimeisestä matriisista on esitetty vain kiinnostavat diagonaalialkiot. Nähdään, että

$$\mathsf{tr}(\mathbf{AB}) = \sum_{j=1}^{n} \sum_{i=1}^{k} a_{ji}b_{ij}$$

ja

$$\mathsf{tr}(\mathbf{BA}) = \sum_{j=1}^{k} \sum_{i=1}^{n} b_{ji}a_{ij} = \sum_{j=1}^{k} \sum_{i=1}^{n} a_{ij}b_{ji}.$$

Indeksejä vertaamalla huomataan, että

$$\mathsf{tr}(\mathbf{AB}) = \mathsf{tr}(\mathbf{BA}).$$

b)

$$\mathsf{tr}(\mathbf{P_X}) \equiv \mathsf{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \overset{\text{a)-kohta}}{=} \mathsf{tr}(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = \mathsf{tr}(\mathbf{I}_k) = k.$$

Toiseksi viimeinen yhtäsuuruus seuraa siitä, että $\mathbf{X}'$ on $k \times n$ -matriisi astetta k.

Tulkitaan tulos. Matriisialgebrasta tiedetään, että (vihjeet; esim. Patovaara (1982): Matriisilaskentaa tilastotieteilijöille, s:t 21–22):

1. Symmetrisen matriisin ominaisarvojen summa on matriisin jälki .
2. Projektiomatriisin ominaisarvot ovat ykkösiä tai nollia.
3. Symmetrisen matriisin aste on sen nollasta poikkeavien ominaisarvojen lukumäärä.

Kohdan 1. perusteella $\mathbf{P_X}$:n ominaisarvojen summa on k. Kohdasta 2. seuraa, että $\mathbf{P_X}$:n ominaisarvoista k on ykkösiä ja loput nollia. Kohdasta 3. saadaan, että $\mathbf{P_X}$:n aste on k. Koska matriisin välittämän lineaarikuvauksen kuva on aliavaruus, jonka aste on sama kuin matriisin aste, niin $\mathbf{P_X}$ projisoi $\mathbf{X}$:n sarakkeiden virittämään aliavaruuteen, joka on (vain) k-ulotteinen.

c)

$$\mathsf{tr}(\mathbf{M_X}) \equiv \mathsf{tr}(I_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathsf{tr}(I_n) - \mathsf{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \overset{\text{b)-kohta}}{=} n - k.$$

Tulkinta kuten b)-kohdassa: $\mathbf{M_X}$ projisoi avaruuteen, joka on (vain) $n-k$-ulotteinen.

Seuraus (yhtälö (3.48) sivulla 110):

$$\begin{aligned} \mathsf{E}(\mathsf{SSR}(\hat{\boldsymbol{\beta}})) &= \mathsf{E}(\mathsf{tr}(\mathbf{u}'\mathbf{M_X}\mathbf{u})) \\ &= \mathsf{E}(\mathsf{tr}(\mathbf{M_X}\mathbf{u}\mathbf{u}')) \\ &= \sigma_0^2\mathsf{tr}(\mathbf{M_X}\mathbf{I}_n) \\ &\overset{\text{yltä}}{=} \sigma_0^2(n-k). \end{aligned}$$

Jäännösneliösumman $SSR(\hat{\boldsymbol{\beta}})$ odotusarvo on siis pienempi kuin teoreettinen $SSR(\boldsymbol{\beta}) = n\sigma_0^2$. Estimoidut jäännökset (niiden neliöt) ovat ”liian pieniä”.

2.[1]

$$\underset{m\times n}{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{(1)} & \cdots & \mathbf{a}_{(n)} \end{bmatrix},\ \underset{m\times m}{\mathbf{D}_v} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & d_m \end{bmatrix},$$

$$\underset{n\times n}{\mathbf{D}_o} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & d_n \end{bmatrix}.$$

a)

$$\begin{aligned} \mathbf{D}_v\mathbf{A} &= \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & d_m \end{bmatrix} \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & d_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} d_1 a_{11} & \cdots & d_1 a_{1n} \\ \vdots & & \vdots \\ d_m a_{m1} & \cdots & d_m a_{mn} \end{bmatrix} = \begin{bmatrix} d_1\mathbf{a}_1' \\ \vdots \\ d_m\mathbf{a}_m' \end{bmatrix} \end{aligned}$$

b)

$$\begin{aligned} \mathbf{A}\mathbf{D}_o &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & d_n \end{bmatrix} = \begin{bmatrix} d_1 a_{11} & \cdots & d_n a_{1n} \\ \vdots & & \vdots \\ d_1 a_{m1} & \cdots & d_n a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} d_1\mathbf{a}_{(1)} & \cdots & d_n\mathbf{a}_{(n)} \end{bmatrix} \end{aligned}$$

3.

a) Koska matriisi $\mathbf{A}$ on singulaarinen, on olemassa k-vektori $\mathbf{b}$, joka toteuttaa yhtälöt $\mathbf{Ab} = \mathbf{0}$. Edellisestä seuraa, että sama vektori toteuttaa myös yhtälöt $\mathbf{XAb} = \mathbf{0}$. Määrittelemällä $n \times k$ -matriisi $\mathbf{X}^* = \mathbf{XA}$ saadaan yhtälöt $\mathbf{X}^*\mathbf{b} = \mathbf{0}$. Näin ollen matriisin $\mathbf{X}^* = \mathbf{XA} = [\mathbf{Xa}_1 \ldots \mathbf{Xa}_k]$ sarakkeet ovat lineaarisesti riippuvia.

Vastauksen voi myös päätellä säännöstä $r(\mathbf{BC}) \leq \min\{r(\mathbf{B}), r(\mathbf{C})\}$, jos se on tuttu ennestään. Edellä $r(\mathbf{B})$ on matriisin $\mathbf{B}$ aste.

b) Olkoon $\mathbf{z}$ sellainen n-vektori, joka on ilmaistavissa matriisin $\mathbf{XA}$ sarakkeiden lineaarikombinaationa: $\mathbf{z} = \mathbf{XA}\boldsymbol{\gamma}$ ($\boldsymbol{\gamma}$ on k-vektori). Määritellään $\boldsymbol{\gamma}^* = \mathbf{A}\boldsymbol{\gamma}$ ($\boldsymbol{\gamma}^*$ on k-vektori), jolloin vektori $\mathbf{z}$ voidaan kirjoittaa matriisin $\mathbf{X}$ sarakkeiden lineaarikombinaationa: $\mathbf{z} = \mathbf{X}\boldsymbol{\gamma}^*$. Mikä tahansa matriisin $\mathbf{XA}$ sarakeavaruuteen kuuluva vektori kuuluu siten myös matriisin $\mathbf{X}$ sarakeavaruuteen.

Toisaalta täytyy olla olemassa matriisin $\mathbf{X}$ sarakeavaruuteen kuuluva n-vektori $\mathbf{w} = \mathbf{X}\boldsymbol{\delta}$ ($\boldsymbol{\delta}$ on k-vektori), joka ei ole ilmaistavissa matriisin $\mathbf{XA}$ sarakkeiden lineaarikombinaationa. Näin täytyy olla, koska matriisin $\mathbf{XA}$ sarakkeet

[1] Vastaus on Jarkko Miettisen tekemä 2009.

ovat lineaarikombinaatioita matriisin $\mathbf{X}$ sarakkeista, joita rajoittaa a)-kohdan perusteella ehto $\mathbf{XAb} = \mathbf{Xb}^* = \mathbf{0}$, jossa $\mathbf{b}^* = \mathbf{Ab}$, mutta vektori $\mathbf{w} = \mathbf{X}\boldsymbol{\delta}$ voidaan valita täysin vapaana lineaarikombinaationa matriisin $\mathbf{X}$ sarakkeista. (Toinen tapa päätellä sama asia on seuraava: Jos pätisi sekä $\mathbf{w} = \mathbf{X}\boldsymbol{\delta}$ että $\mathbf{w} = \mathbf{XA}\boldsymbol{\delta}^*$, pitäisi päteä $\boldsymbol{\delta} = \mathbf{A}\boldsymbol{\delta}^*$. Nämä yhtälöt toteuttavaa vektoria $\boldsymbol{\delta}^*$ ei voida kuitenkaan ratkaista, koska matriisi $\mathbf{A}$ ei ole kääntyvä.)

Yhdistämällä edellinen päättely ja a)-kohta saadaan $\mathsf{S}(\mathbf{XA}) \subset \mathsf{S}(\mathbf{X})$.[2]

c) Regression sovitteet ja residuaalit eivät ole yleensä samoja, kun regression selittäjämatriisi on $\mathbf{X}$ tai $\mathbf{XA}$, kun $\mathbf{A}$ on singulaarinen matriisi. Jälkimmäisessä tilanteessa residuaalivektori on yleensä pidempi ja sovitevektori eri.

Erityistilanteissa on mahdollista, että residuaalit ja sovitteet eivät eroa. Esimerkki: $\mathsf{S}(\mathbf{X})$ on taso ($k = 2$), mutta selitettävän vektorin ($\mathbf{y}$) ortogonaalinen projektio tasoon osuu toiselle tason virittävistä matriisin $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2]$ sarakkeista (esim. $\mathbf{x}_1$:lle). Mikäli kyseinen tasovektori sisältyy $\mathsf{S}(\mathbf{XA})$:han (esim. $\mathbf{A} = [\boldsymbol{\iota} \quad \mathbf{0}]$), residuaalit ja sovitteet ovat samat regressioissa.

4.

a) Normaaliyhtälöt ovat $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. Tehtävässä

$$\begin{aligned}
\mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \underset{2\times 2}{\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}}, \\
\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n x_i \\ \hat{\beta}_1 \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2 \end{bmatrix} \text{ ja} \\
\mathbf{X}'\mathbf{y} &= \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}.
\end{aligned}$$

Normaaliyhtälöt ovat siis

$$\begin{aligned}
& \begin{bmatrix} \sum_{i=1}^n y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = \hat{\beta}_1 \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2 \end{bmatrix} \\
\Leftrightarrow & \begin{bmatrix} \hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x} \\ \sum_{i=1}^n x_i y_i = \hat{\beta}_1 \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2 \end{bmatrix}.
\end{aligned}$$

Sijoittamalla johdettu tulos

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

alempaan yhtälöön saadaan

$$\begin{aligned}
\sum_{i=1}^n x_i y_i &= (\overline{y} - \hat{\beta}_2 \overline{x}) \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2 \\
&= \overline{y} \sum_{i=1}^n x_i - \hat{\beta}_2 \overline{x} \sum_{i=1}^n x_i + \hat{\beta}_2 \sum_{i=1}^n x_i^2.
\end{aligned}$$

[2] Tehtävä on Davidsonin ja MacKinnonin HT 2.14.

Ratkaistaan $\hat{\beta}_2$:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.\end{aligned}$$

Estimaattoreiden kaavat lienevät tutut Tilastollisen päättelyn kurssilta (Pekka Nieminen ja Pentti Saikkonen (2007): Tilastollisen päättelyn kurssi, s. 15) ja mahdollisesti Lineaarisen mallin kurssilta 2009 (HT 3.4).

Vielä pitää johtaa $\hat{\beta}_2$:lle kysytty muoto. Ilmaistaan otoskorrelaatio (r) $\hat{\beta}_2$:n avulla:

$$\begin{aligned}r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})}} \\ &= \hat{\beta}_2 \frac{s_x}{s_y}.\end{aligned}$$

Todistettava tulos seuraa:

$$\hat{\beta}_2 = r\frac{s_y}{s_x}.$$

b) Kysytty tulos saadaan sijoittamalla PNS-estimaattorien kaavat sovitteen kaavaan:

$$\begin{aligned}\hat{y}_t &= \hat{\beta}_1 + \hat{\beta}_2 x_t \\ &= \bar{y} - r\frac{s_y}{s_x}\bar{x} + r\frac{s_y}{s_x}x_t \\ &\Longleftrightarrow \\ \frac{\hat{y}_t - \bar{y}}{s_y} &= r\frac{x_t - \bar{x}}{s_x}.\end{aligned}$$

5.

a) Satunnaismuuttujan y_y odotusarvo riippuu regressiomallin tilanteessa kiinteäksi oletetusta selittäjästä:

$$\begin{aligned} \mathsf{E}(y_t|x_t) &= \mathsf{E}[(\beta_1 + \beta_2 x_t + u_i)|x_t] \\ &= \mathsf{E}(\beta_1|x_t) + \mathsf{E}(\beta_2 x_t|x_t) + \mathsf{E}(u_t|x_t) \\ &= \beta_1 + \beta_2 x_t. \end{aligned}$$

Odotusarvo on tässä yhteydessä ehdollinen, eli se on odotusarvo ehdolla selittävän muuttujan arvo havainnon t kohdalla. Koska selittävän muuttujan arvo vaihtelee havainnoittain, selitettävän muuttujan ehdollinen) odotusarvo ylipäänsä ei ole sama eri havainnoille.

b) Tarkastellaan esimerkiksi Galtonin havaitsemaa yhteyttä vanhempien pituuksien (mahdollisesti painotetun) keskiarvon (mid-parent[3]; jatkossa "keskipituus") ja heidän lastensa pituuden (children) välillä. Kuviosta nähdään, että vanhempien keskipituus on ollut keskimäärin runsas 68 tuumaa. "Midparents"-suora kuvaa vanhempien keskipituuden poikkeamaa keskimääräisestä pituudesta (suoran kulmakerroin on yksi). Kuvion mukaan keskimääräistä

- pidempien vanhempien lapsi on myös keskimääräistä pidempi muttei yhtä paljon kuin vanhempansa (suoran "children" kulmakerroin on 0:n ja 1:n välillä).
- lyhyempien vanhempien lapsi on myös keskimääräistä lyhyempi muttei yhtä paljon kuin vanhempansa.

Tässä mielessä pituus "regressoituu" eli pyrkii palaamaan kohti keskiarvoansa (yllä runsas 68 tuumaa).

Ohessa on myös Galtonin alkuperäinen kuvio vuodelta 1886. Se on paljon epäselvempi ja esitetään kuriositeettina.[4]

c) Regressiomallissa ei aina tapahdu palautumista kohti keskiarvoa (tai odotusarvoa). Ensinnäkin käsite olettaa, että selitettävä ja selittävä muuttuja ovat "samantapaisia" muuttujia (edellä vanhempien keskipituus ja lapsen pituus), joilla on sama keskiarvo (tai odotusarvo). Näin ei tietenkään aina ole. Toiseksi, regressiosuoran kulmakerroin ei ole välttämättä 0:n ja 1:n välissä. Esimerkiksi oletetaan, että regressiosuoran kulmakerroin on 3, selittävän muuttujan keskiarvo on nolla ja että regressiosuora kulkee (x, y)-koordinaatistossa origon kautta. Tällöin selitettävän muuttujan poikkeamat keskiarvostaan pyrkivät moninkertaistumaan selittävän muuttujan poikkeamiin keskiarvostaan nähden.

[3] Midparent-käsitteen määritelmä löytyy Wikipediasta (http://en.wikipedia.org/wiki/Midparent; viitattu 16.4.2011).

[4] Galton kertoi tyttölasten pituudet aineistossaan 1,08:lla, mistä kerrotaan kuviossa englanniksi. Lähde: A. Wachsmuth, L. Wilkinson ja G.E. Dallal (2003): Galton's Bend: A Previously Undiscovered Nonlinearity in Galton's Family Stature Regression Data. *American Statistician*, 57, 190–192.

Hieman haetummassa mielessä regressiomallissa voidaan aina kuitenkin ajatella tapahtuvan regressoitumista. Edellisessä tehtävässä osoitettiin, että

$$\frac{\hat{y}_t - \bar{y}}{s_y} = r \frac{x_t - \bar{x}}{s_x}.$$

Koska otoskorrelaatiokertoimelle pätee $-1 < r < 1$, niin regressiomallin sovite (sen itseisarvo) poikkeaa selitettävän keskiarvosta vähemmän kuin selittäjä poikkeaa keskiarvostaan (erotuksen itseisarvosta), kun poikkeamat on standardoitu vastaavilla otosvariansseilla. Tässä mielessä regressiota on aina regressiomallissa.

Vastaus perustuu Sheldon Rossin (2010) kirjan *Introductory Statistics*, 3. laitos jaksoon "Regression to the mean" ja Wikipedian artikkeliin http://en.wikipedia.org/wiki/Regression_toward_the_mean (viitattu 6.4.2011).

6. Ensimmäisessä mallissa on pientä viitettä siitä, että residuaalit (estimoidut jäännökset) eivät olisi riippumattomia: Negatiivista residuaalia näyttää seuraavan (kuviota vasemmalta oikealle katseltaessa) toinen negatiivinen residuaali ja positiivista residuaalia toinen positiivinen. Jäännösten riippumattomuutta olisi syytä testata sopivilla testeillä (onko vaikutelma satunnaisvaihtelun puitteissa olevaa riippumattomuuden pätiessä), mihin ei mennä tässä.

Teollisuus kritisoi Yhdysvalloissa oikeudessa, että mallista puuttuu tärkeitä muuttujia. Oikeus arvioi, että teollisuus ei ole näyttänyt, että puuttuvat muuttujat vaikuttaisivat tuloksiin, ja ympäristönsuojeluvirasto voitti oikeuskiistan. Edellisen kappaleen perusteella teollisuudella saattoi olla perusteltu syy kritiikkiin.

Toisessa eli kuljetusturvahallinnon mallissa on huomionarvoisinta kaksi seikkaa:

- Kirjan regressiomalli on lineaarinen parametrien suhteen. Selittäjät voivat olla kuitenkin alkuperäisten selittäjien epälineaarisia muunnoksia ja sovitekin voi siten olla epälineaarinen funktio alkuperäisestä selittäjästä!
- Oikeanpuoleisin havainto ilmeisesti vaikuttaa muita havaintoja merkittävämmin mallin kertoimien PNS-estimaatteihin. Lisäksi tehtävässä kerrottiin sen olevan laskettu keskiarvona tietyistä havainnoista. Se herättää kysymyksen, voidaanko havainto sisällyttää malliin muiden havaintojen tavoin.

Kuljetusturvahallinto halusi mallillaan osoittaa, että onnettomuuksien suuri kasvu 10–11-tuntien kohdalla johtui pienestä havaintomäärästä 11-tunnin kohdalla. (Mallin sovite 11-tunnin kohdalla on paljon pienempi kuin havaittu kolaririski.) Oikeus Yhdysvalloissa hylkäsi mallin muun muassa laitimmaiseen havaintoon liittyvien seikkojen takia.

Kolmannen kuvion aineisto on vaikea kuvitella kirjassa kuvatun regressiomallin tuottamaksi, sillä selitettävän muuttujan havainnot rajoittuvat viiteen arvoon $1, \ldots, 5$, vaikka regression selittävä muuttuja saa hyvin vaihtelevia arvoja. Ongelma on, että mallin selitettävä muuttuja on diskreettiarvoinen, vaikka lineaarisen mallin perusoletusten mukaan sen tulee olla jatkuva-arvoinen.

Selitettävän ja selittäjän roolin vaihtaminen tuottaisi kuvion perusteella mahdollisen mallin. Tällöin selittävä muuttuja olisi indikaattorimuuttuja, joka saisi arvoja $1, \ldots, 5$. Selitettävä muuttuja saisi hyvin vaihtelevasti arvoja kuten regressiomallissa kuuluukin. Kuvion perusteella muuttujien välillä vallitsisi tällöin negatiivinen suhde.

Myös regressiomallin erikoistapaus

$$Y_j = 3 + \varepsilon_j,$$

jossa jäännös ε_j (hyvin poikkeuksellisesti) voisi saada vain arvoja $-2, -1, 0, 1, 2$ (tietyillä todennäköisyyksillä), voisi olla sopusoinnussa havaintojen kanssa. (Luku $3 = 3 \times 1$ on muotoa kerroin kertaa muuttuja, joka on tässä yhteydessä vakio 1.) Tässä mallissa etnisellä heterogeenisuudella ei olisi roolia.