

Ilmastonmuutos ja vastuut

Yleistetyt lineaariset -mallit kurssin harjoitustyö, kevät 2018

Erik Manelius & Lasse Rintakumpu

Sisältö

1 Johdanto	1
2 Aineiston ja tutkimuskysymyksen kuvaus	1
3 Tutkimuskysymyksen mallintaminen ja mallien tulkinta	5

1 Johdanto

Analyysityössämme tarkastelemme *European Social Surveyn* 8. kierroksen aineiston (ESS 2016) pohjalta millaisena suomalaiset kokevat henkilökohtaisen vastuunsa ilmastonmuutoksen hillitsemisestä.

European Social Survey on tieteellisistä lähtökohdista toteutettu vertaileva kyselytutkimus, joka kattaa yli 30 Euroopan ja lähialueiden maata. ESS kartoittaa Euroopan maiden yhteiskunnallisen muutoksen sekä väestön asenteiden, uskomusten ja käyttäytymisen välisiä suhteita. Suomessa tutkimus tunnetaan myös nimellä *Arvot ja mielipiteet Suomessa*.

Tutkimus on toteutettu Tilastokeskuksen ja Turun yliopiston yhteistyönä. Tutkimukseen on poimittu satunnaisesti 3 400 yli 15-vuotiasta suomalaista. Jokainen haastateltava edustaa vastauksillaan noin 1 300 suomalaista.¹

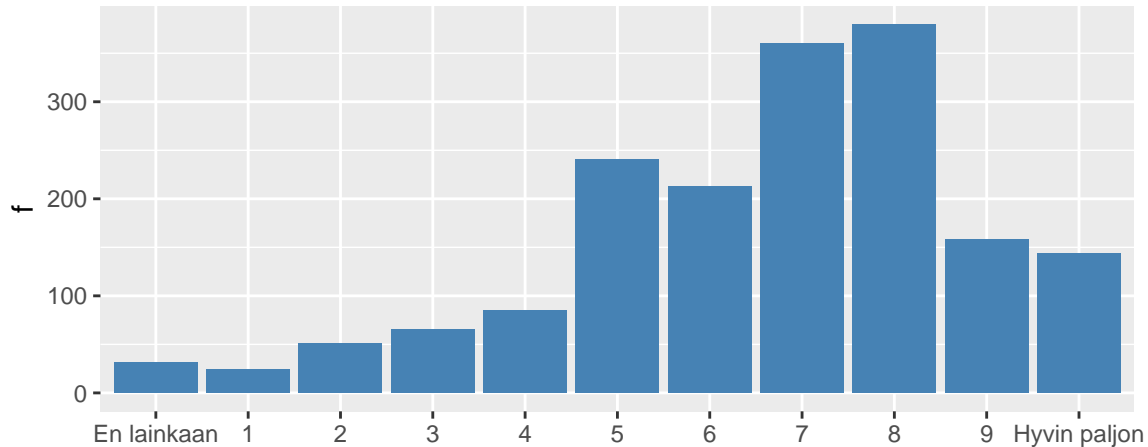
2 Aineiston ja tutkimuskysymyksen kuvaus

Suomen aineisto sisältää $n = 1925$ tilastoyksikköä, joilta on kerättyä tietoa 499 muuttujasta. Tutkiessamme suomalaisten henkilökohtaista ilmastovastuuta päämielenkiintomme kohdistuu muuttujaan D23. **To what extent do you feel a personal responsibility to try to reduce climate change?** eli “kuinka paljon tunnet henkilökohtaista vastuuta ilmastonmuutokset vähentämisestä?”. Muuttuja on 11-luokkainen ordinaalinen muuttuja, jossa luokka 1 vastaa vastausta “en lainkaan” ja luokka 11 vastausta “hyvin paljon”.

Havainto kyseisestä muuttujasta puuttuu 34 tilastoyksiköltä. Kun aineistosta on pudotettu puuttuvat havainnot (havaintojen pudottamista tarkastellaan tarkemmin myöhemmin), havaitaan, että aineiston mukaan suomalaiset tuntevat keskimääräisesti (vastausten ka. 7.6) paljon vastuuta ilmastonmuutoksen ehkäisemisestä (kts. Kuva 1).

¹Tilastokeskus: "Arvot ja mielipiteet Suomessa -tutkimus (ESS)", <https://www.stat.fi/tup/htpalvelut/tutkimukset/arvot-ja-mielipiteet-suomessa-tutkimus-ess.html>, haettu 6.5.2018.

Kuva 1: Henkilökohtainen vastuu ilmastonmuutoksesta.



Analyysissa pyrimme ensin tarkastelemaan miten muuttuja D28. *How likely do you think it is that governments in enough countries will take action that reduces climate change?* eli se “kuinka todennäköisenä pitää sitä, että riittävän monen maan hallitus toimii ilmastonmuutoksen hillitsemiseksi” vaikuttaa vastaajan omaan vastuunottoon ilmastonmuutoksen vähentämisestä. Kyseinen muuttuja on 11-luokkainen ordinaalinen muuttuja, jossa luokka 1 vastaa vastausta “en lainkaan todennäköisenä” ja luokka 11 vastausta “hyvin todennäköisenä”.

Tämän jälkeen tarkastelemme, miten näiden kysymysten välinen yhteys muuttuu, kun vastaajan käsitys ilmastonmuutoksen syistä huomioidaan. Tätä käsitystä mitataan viisiluokkaisella ordinaalisella muuttujalla D22. *Do you think that climate change is caused by natural processes, human activity, or both?* eli “uskotko ilmastonmuutoksen aiheutuvan luonnollisista prosesseista, ihmisen toiminnasta vai molemmista”. Luokka 1 vastaa vastausta “kokonaan luonnollisista prosesseista”, luokka 3 vastausta “yhtä paljon luonnollisista prosesseista ja ihmisen toiminnasta” ja luokka 5 vastausta “kokonaan ihmisen toiminnasta”. Lisäksi muuttujassa on kuudes luokka “en usko ilmastonmuutokseen”, mutta tässä luokassa ei Suomen aineistossa ole yhtään vastausta (kun vastemuuttujan D23 puuttuvat havainnot poistetaan).

Lisäksi pyrimme tutkimaan taustamuuttujien

- ikä (ikä), jatkuva numeerinen, vaihteluväli 1 – 81,
- sukupuoli (sukupuoli), kategorinen, kaksiluokkainen,

vaikutusta henkilökohtaisen ilmastovastuun kokemiseen. Kun tarkastelemme kaikkia analyysiin valittuja muuttujia, havaitsemme, että puuttuvia havaintoja löytyy 24 eri luokkakombinaatiosta yhteensä 122 kappaletta. Jatkamme pudottamalla nämä tilastoyksiköt tarkastelusta, jolloin päädyimme tilanteeseen, jossa tilastoyksikköjä on jäljellä $n_{complete} = 1752$. Vaikka esimerkiksi Littlen MCAR-testin ² antaman χ^2 -testisuureen arvon 415.90 perusteella ei voida olettaa havaintojen puuttuvan täysin satunnaisesti, jatkamme kuitenkin aineistolla, josta on poistettu puuttuvat havainnot ja huomioimme tämä johtopäätöksissä.

Pitääksemme analyysin yksinkertaisena, rajoitamme tarkastelun kahteen taustamuuttujaa. Lisäksi muuttujien välistä korrelaatiota (Taulukko 1) tarkastelemalla havaitsemme, ettei selittävien muuttujien välillä ole niin suuria korrelaatioita, että jokin muuttujista olisi perustettua jättää tarkastelun ulkopuolelle multikollineaarisuuden välttämiseksi. Samalla huomaamme, että suurin korrelaatio löytyy vastemuuttujan D23 ja iän sekä ilmastonmuutoksen syyn välille. Muuttujien valinta vaikuttaa siis tässä suhteessa järkevältä.

²Little, Roderick J. A., 1988: "A Test of Missing Completely at Random for Multivariate Data with Missing Values", *Journal of the American Statistical Association*, Vol. 83, No. 404 (Dec., 1988), pp. 1198-1202.

Taulukko 1: Muuttujien väliset korrelaatiot.

	D23: vastuu	D28: hallitus	D22: syy	sukupuoli	ika
D23: vastuu	1.000	0.100	0.244	0.199	-0.092
D28: hallitus	0.100	1.000	-0.073	-0.001	0.131
D22: syy	0.244	-0.073	1.000	0.024	-0.197
sukupuoli	0.199	-0.001	0.024	1.000	0.056
ika	-0.092	0.131	-0.197	0.056	1.000

Koska pyrimme mallintamaan ordinaalisten muuttujien välistä yhteyttä erilaisilla yleistetyillä lineaarisilla logit- ja probit-malleilla, vähennämme alkuperäisten muuttujien luokkia, jotta mallien sovittaminen pysyy sekä laskennallisesti että tulkinnallisesti hallittavana.

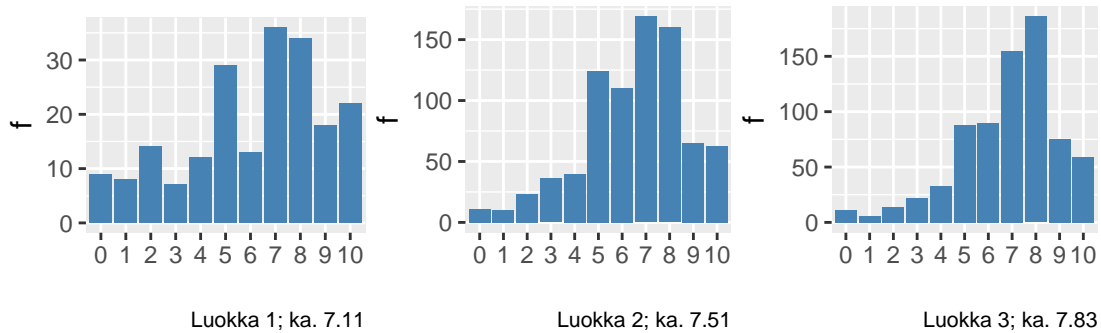
Ordinaalisten muuttujien D23 (vastuuta mittaava vastemuuttuja) ja D28 (“tarpeeksi moni hallitus”) kohdalla päädyimme mahdollisimman tasaiseen luokkajakoon, jossa luokat 0 – 3 yhdistetään alimmaksi luokaksi, luokat 4 – 6 keskimmäiseksi luokaksi ja luokat 7 – 10 ylimmäksi luokaksi.

Viisiluokkaisen (koska luokkaan “en usko ilmastonmuutokseen” ei kuulu yhtään tilastoyksikköä) ilmastonmuutoksen syitä mittaavan muuttujan D22 kohdalla päädyimme myös kolmiluokkaiseen jakoon, jossa luokka 1 sisältää vastaukset, joiden mukaan ilmastonmuutos aiheutuu joko kokonaan tai pääosin luonnollisista prosesseista, luokka 2 sisältää vastauksen, jonka mukaan ilmastonmuutos on yhtä paljon luonnon ja ihmisen aiheuttamaa ja luokka 3 vastaukset, joiden mukaan ilmastonmuutos on pääasiassa tai kokonaan ihmisen aiheuttamaa.

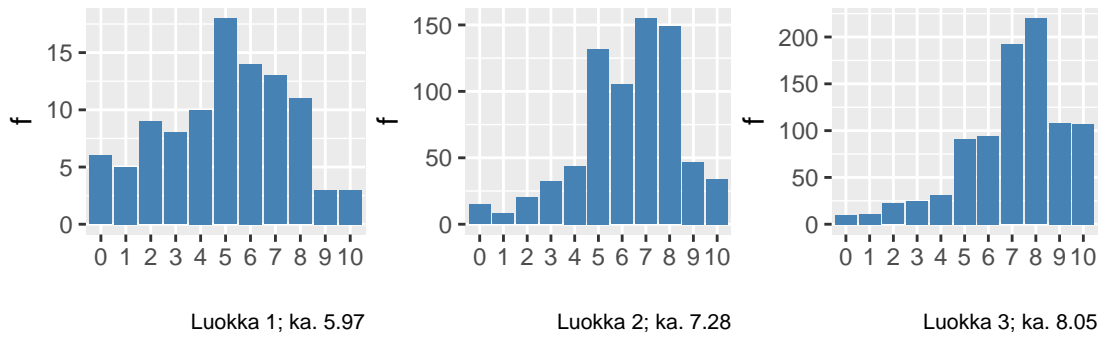
Lisäksi jaamme jatkuvan ikämuuttujan neljään luokkaan niin, että ensimmäiseen luokkaan sijoittuvat alle 31-vuotiaat, toiseen luokkaan 31-44-vuotiaat, kolmanteen luokkaan 45-64-vuotiaat ja neljänteen luokkaan yli 64-vuotiaat.

Kun tarkastelemme henkilökohtaisen ilmastovastuun jakaumaa muiden muuttujien suhteen ennen vastemuuttujan luokkien yhdistämistä (Kuvat 2 – 5) huomaamme, että henkilökohtainen vastuu ilmastonmuutoksen vähentämisestä näyttää olevan suurinta niiden vastaajien keskuudessa, jotka uskovat riittävän monen hallituksen tekevän riittävästi ilmastonmuutoksen hillitsemiseksi sekä niiden vastaajien keskuudessa, jotka uskovat ilmastonmuutoksen olevan osittain tai kokonaan ihmisen aiheuttamaa. Lisäksi nuoremmat ikäluokat sekä naiset näyttävät kokevan enemmän henkilökohtaista vastuuta ilmastonmuutoksen ehkäisemisestä.

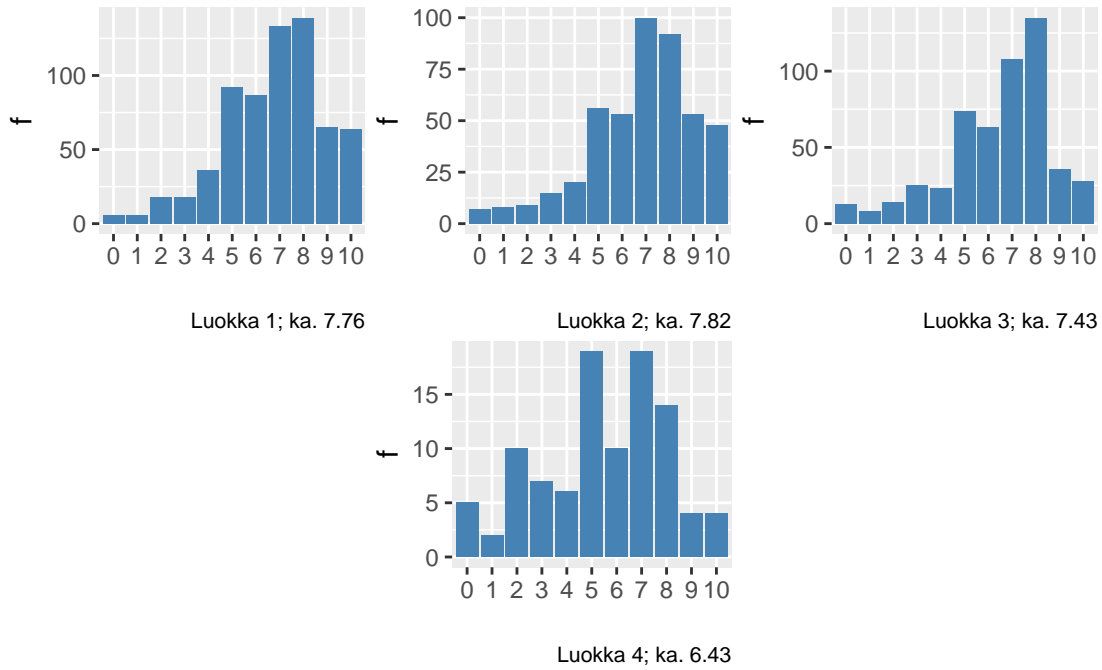
Kuva 2: D23 vastuu vs. D28 hallitus.



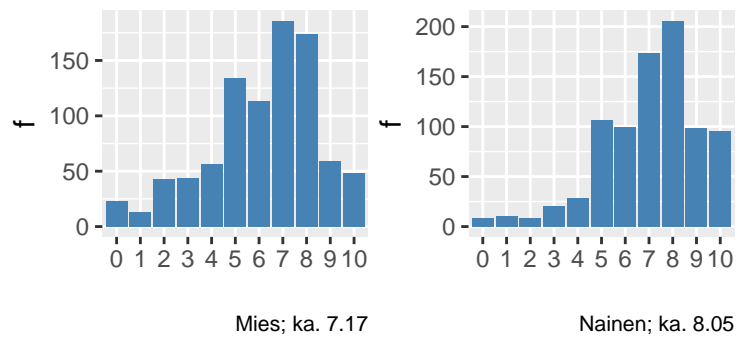
Kuva 3: D23 vastuu vs. D22 syy.



Kuva 4: D23 vastuu vs. Ikä.



Kuva 5: D23 vastuu vs. Sukupuoli.



Seuraavaksi siirrymme mallintamaan tilannetta sekä tarkastelemaan tässä havaittujen erojen tilastollista merkitsevyyttä (sekä mahdollisia eroja muuttujien yhdysvaikutuksissa).

3 Tutkimuskysymyksen mallintaminen ja mallien tulkinta

Ensimmäiseksi mallinamme henkilökohtaisen ilmastovastuun sekä käsittelyn hallitusten toiminnasta välistä yhteyttä. Käytämme myös ikää ja sukupuolta selittäjinä.

Koska tarkoituksenamme on mallintaa ordinaalista vastetta, vertailemme useampaa erilaista ordinaalisen aineiston mallia keskenään. Vertailuun käytämme AIC- ja BIC-informaatiokriteereitä, jäännösdevianssia sekä log-uskottavuutta. Kaikkien muiden vertailuarvojen kohdalla pienempi arvo viittaa parempaan malliin, mutta log-uskottavuuden tapauksessa suurempi arvo on parempi. Vertailussa käytämme hyväksi myös täyttä mallia, joka ei välttämättä ole sellaisenaan kiinnostava, mutta toimii hyvin mallien vertailussa.

Taulukko 2: Mallien vertailua.

	AIC	BIC	Devianssi	Log-uskottavuus
Kumulatiivinen logit	240.3783	269.8297	26.47027	-95.18915
Viereisten kategorioiden logit	231.2810	260.7324	17.37301	-90.64052
Continuation-Ratio logit	241.9010	271.3523	27.99296	-95.95050
Kumulatiivinen probit	233.5566	263.0080	19.64857	-91.77830
Kumulatiivinen clog-log	245.3469	274.7983	31.43889	-97.67347

Kaikkien testistatistiikkojen perusteella paras malli on viereisten kategorioiden logistinen regressiomalli (Taulukko 2). Malli on muotoa

$$\text{logit}(\gamma_j) = \alpha_j + x\beta \quad (1)$$

missä

$$\gamma_{ij} = \frac{\pi_{ij+1}}{\pi_{ij} + \pi_{ij+1}}. \quad (2)$$

Tästä saadaan

$$\text{logit}(\gamma_{ij}) = \log\left(\frac{\pi_{ij+1}}{\pi_{ij}}\right) \quad (3)$$

joten

$$\log\left(\frac{\pi_{ij+1}}{\pi_{ij}}\right) = \alpha_j + x_i^T \beta. \quad (4)$$

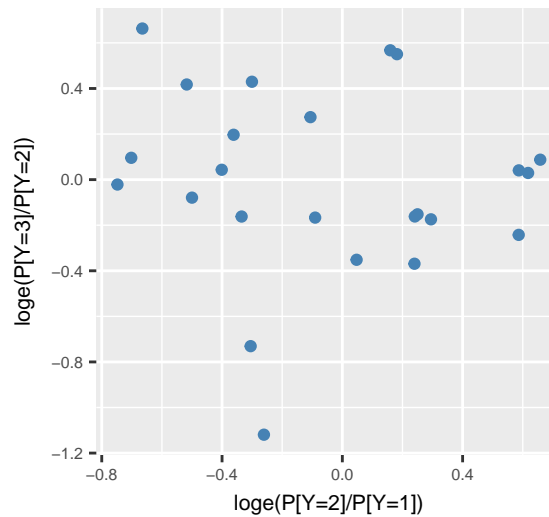
Mallin linkkifunktio on siis muotoa

$$\log\left(\frac{\pi_{ij+1}}{\pi_{ij}}\right). \quad (5)$$

Malli siis vertaa seuraavan kategorian todennäköisyyttä sitä edeltäneeseen kategoriaan.

Etsimme parhaan selittävien muuttujien yhdistelmän aloittamalla täydestä mallista ja poistamalla siitä termejä yksi kerrallaan ("ylhäältä alas"). Jokaisen poistetun termin jälkeen vertaamme uutta mallia edelliseen χ^2 -testillä parametreina devianssien erotus ja vapausasteiden erotus. Etenemme näin, kunnes löydämme mallin, josta ei voida enää poistaa termejä ilman mallin merkittävää huonontumista. Päädyimme malliin, jonka selittäjinä ovat vain päävaikutukset eli ikä, sukupuoli ja hallituksien toiminta.

Kuva 6: Mallin residuaalit.



Mallin residuaalit (Kuva 6) ovat jakautuneet tasaisesti nollan molemmin puolin ilman havaittavaa rakennetta, mutta yksi residuaali poikkeaa selkeästi muista. Tämä poikkeama johtuu havaintojen puutteesta kombinaatiolla D28 hallitus = 1, sukupuoli = mies, ikä = yli 64 ja D23 vastuu korkea verrattuna keskinäiseen. Yleisesti voi kuitenkin sanoa, että malli sopii aineistoon hyvin.

```
##
## Call:
## vglm(formula = cbind(Freq.1, Freq.2, Freq.3) ~ government_action +
##       gender + age, family = acat(parallel = TRUE), data = round8)
##
##
## Pearson residuals:
##               Min      1Q  Median      3Q      Max
## loge(P[Y=2]/P[Y=1]) -1.093 -0.5604 -0.2221  0.5540  1.219
## loge(P[Y=3]/P[Y=2]) -1.736 -0.7206 -0.1136  0.8067  1.963
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1      0.81260    0.14946   5.437 5.42e-08 ***
## (Intercept):2      0.46671    0.13070   3.571 0.000356 ***
## government_action2  0.49701    0.12138   4.095 4.23e-05 ***
## government_action3  0.78678    0.12851   6.122 9.22e-10 ***
## gender2            0.63248    0.08910   7.099 1.26e-12 ***
## age2              -0.02698    0.11210  -0.241 0.809827
## age3              -0.27201    0.10424  -2.609 0.009071 **
## age4             -1.00867    0.16044  -6.287 3.24e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors: 2
##
## Names of linear predictors: loge(P[Y=2]/P[Y=1]), loge(P[Y=3]/P[Y=2])
##
## Residual deviance: 37.6601 on 40 degrees of freedom
##
```

```
## Log-likelihood: -100.7841 on 40 degrees of freedom
##
## Number of iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
```

Vakiotermien merkitsevyydestä näemme, että eri luokkien välillä on eroa. Luokkaan kaksi on vastattu 2.25 kertaa suhteessa luokkaan yksi ja luokkaan kolme on vastattu 1.59 kertaa suhteessa luokkaan kaksi. Mitä enemmän uskoo riittävän monen hallituksen tekevän tarpeeksi ilmastonmuutoksen estämiseksi, sitä todennäköisemmin kokee myös henkilökohtaista vastuuta ilmastonmuutoksen ehkäisystä. Naiset kokevat merkitsevästi todennäköisemmin vastuuta ilmastonmuutoksesta kuin miehet ja yli 64 vuotiaat kokevat merkitsevästi vähemmän vastuuta kuin 44-64 vuotiaat. Samoin 45-64 vuotiaat kokevat merkitsevästi vähemmän vastuuta kuin 30-44 vuotiaat.

Seuraavaksi mallinamme miten henkilökohtaisen vastuun sekä hallituksien toiminnan keskinäinen yhteys muuttuu, kun otamme mukaan muuttujaksi kysymyksen ilmastonmuutoksen syystä. Lisäksi käytämme jälleen ikää ja sukupuolta selittäjinä. Käytämme mallin valintaan edellä kuvattua metodologiaa.

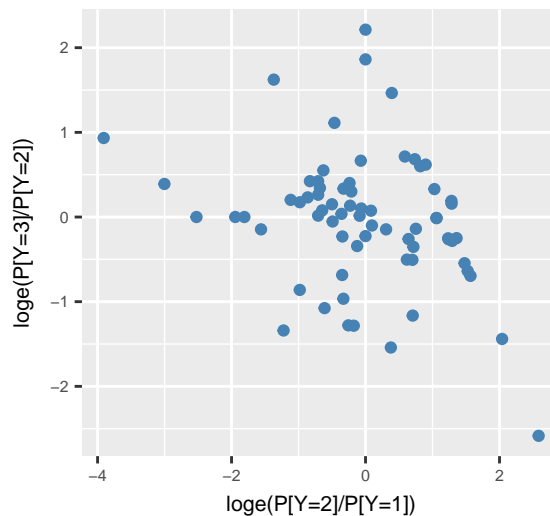
Taulukko 3: Mallien vertailua.

	AIC	BIC	Devianssi	Log-uskottavuus
Kumulatiivinen logit	456.2725	478.6135	157.1930	-218.1362
Viereisten kategorioiden logit	451.7171	474.0582	152.6377	-215.8586
Continuation-Ratio logit	458.1428	480.4838	159.0633	-219.0714
Kumulatiivinen probit	452.9032	475.2442	153.8237	-216.4516
Kumulatiivinen clog-log	462.0807	484.4218	163.0013	-221.0403

Päädymme jälleen viereisten kategorioiden logit-malliin (Taulukko 3). Nyt etenemme mallin parametrien valinnassa “alhaalta ylöspäin” eli lähdemme liikkeelle nollamallista ja kokeilemme lisätä yhdysvaikutuksia yksi kerrallaan ja testaamme χ^2 -testillä mallin merkitsevyyden muutosta. Huomioitava muutos aiempaan mallintamiseen on testaaminen huonommasta parempaan eikä paremmasta huonompaan.

Saamme parhaimman mallin valikoimalla parametreiksi muuttujat D28: hallitus, D22: syy, sukupuoli, ikä sekä muuttujien ikä ja D22: syy yhdysvaikutuksen.

Kuva 7: Mallin residuaalit.



Residuaalit ovat jälleen jakautuneet kohtalaisen tasaisesti (Kuva 7), mutta aineistosta löytyvien nolla-frekvenssien vuoksi havaitsemme useamman ison residuaalin.

```
##
## Call:
## vglm(formula = cbind(Freq.1, Freq.2, Freq.3) ~ government_action +
##      caused_by + gender + age + caused_by * age, family = acat(parallel = TRUE),
##      data = round8_2)
##
##
## Pearson residuals:
##              Min        1Q      Median        3Q       Max
## loge(P[Y=2]/P[Y=1]) -2.067 -0.7845 -0.131787 0.7441 1.460
## loge(P[Y=3]/P[Y=2]) -1.633 -0.5621  0.008601 0.5024 2.242
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept):1      0.12379   0.27683   0.447   0.6548
## (Intercept):2     -0.37806   0.27732  -1.363   0.1728
## government_action2  0.61035   0.12764   4.782 1.74e-06 ***
## government_action3  0.90618   0.13474   6.725 1.75e-11 ***
## caused_by2         0.26687   0.27279   0.978   0.3279
## caused_by3         1.15020   0.27351   4.205 2.61e-05 ***
## gender2            0.58087   0.09151   6.348 2.18e-10 ***
## age2              -0.64393   0.36317  -1.773   0.0762 .
## age3              -0.16184   0.36194  -0.447   0.6548
## age4              -0.66648   0.44663  -1.492   0.1356
## caused_by2:age2     0.96545   0.40192   2.402   0.0163 *
## caused_by3:age2     0.58133   0.40450   1.437   0.1507
## caused_by2:age3     0.28219   0.39118   0.721   0.4707
## caused_by3:age3    -0.33751   0.39882  -0.846   0.3974
## caused_by2:age4     0.40382   0.50961   0.792   0.4281
## caused_by3:age4    -1.00637   0.50877  -1.978   0.0479 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  2
##
## Names of linear predictors: loge(P[Y=2]/P[Y=1]), loge(P[Y=3]/P[Y=2])
##
## Residual deviance: 127.2337 on 122 degrees of freedom
##
## Log-likelihood: -203.1566 on 122 degrees of freedom
##
## Number of iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
```

Mallissa vakiotermit eivät ole tilastollisesti merkitsevästi eroavia eli vastuuluokan vaikutus mallin vedonlyöntisuhteeseen (*odds*) voi johtua sattumasta. Hallituksen toiminnan vaikutus lisää vedonlyöntisuhdetta suurempaan henkilökohtaiseen vastuuseen. Mallissa D22: *syy* -muuttujan tasojen yksi ja kaksi ero ei ole merkitsevä, mutta tasojen kaksi ja kolme ero on eli kun uskoo ihmisen oleva syypää ilmastonmuutokseen ottaa myös enemmän henkilökohtaista vastuuta ilmastonmuutoksen hillitsemisestä (vastausluokat ovat 1 = luonto, 2 = ihminen ja luonto yhtä paljon ja 3 = ihminen). Naiset taas kokevat tilastollisesti merkitsevästi enemmän vastuuta ilmastonmuutoksesta kuin miehet.

Iällä yksin ei ole merkittävästi yhteyttä vastuun tuntemiseen. Mutta kun katsomme iän ja syyn yhdysvaikutusta merkittävästi enemmän vastuuta kokevat ne nuoret, jotka uskovat luonnon olevan ilmastonmuutoksen aiheuttaja kuin nuoret, jotka uskovat aiheuttajan olevan sekä luonto että ihminen. Merkittävästi enemmän vastuuta kokevat myös keski-ikäiset, jotka uskovat sekä luonnon että ihmisen olevan ilmastonmuutoksen syytä kuin vanhat ihmiset, jotka uskovat ihmisten olevan syytä ilmastonmuutokseen.

Mallinnuksen perusteella voimme todeta, ettei D22: **syy** ja D28: **hallitus** -muuttujien välinen yhdysvaikutus merkittävästi vaikuta muutoksiin henkilökohtaista vastuuta mittaavassa vastemuuttujassa D23: **vastuu**.

3.1 Liitteet

Analyysissä käytetty R-koodi on katsottavissa ja ladattavissa omalla [GitHub-sivullaan](#).