

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	5
4	Tulosten tulkinta ja johtopäätökset	7

1 Johdanto

Analyysityössämme tarkastelemme *European Social Survey*n 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-vuotiaasta suomalaista. Jokainen haastateltava edustaa vastauksillaan noin 1 300 suomalaista.¹

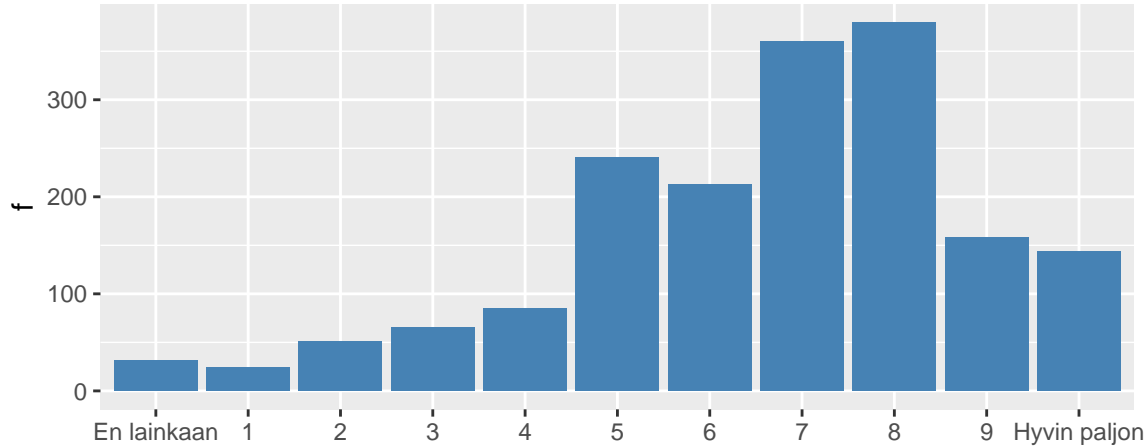
2 Aineiston ja tutkimuskysymyksen kuvaus

Suomen aineisto sisältää $n = 1925$ tilastoyksikköä, joista 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 myös 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. Käsitystä mitataan 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”.

Ilmastonmuutoksen syiden mittaamiseen on käytetty viisiluokkaista ordinaalista muuttujaa, jossa 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 (ika), jatkuva ja 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. Jatketaan pudottamalla nämä tilastoyksiköt tarkastelusta. 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.

Pitäksemme analyysin yksinkertaisena, rajoitamme tarkastelun näihin kahteen taustamuuttujaa. Lisäksi muuttujien välistä korrelaatiota (Taulukko 1) tarkastelemalla havaitsemme, ettei selittävien muuttujien välillä ole niin suurta korrelaatiota, 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.

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

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.

	D23: vastuu	D28: hallitus	D22: syy	sukupuoli	ika
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 riittävän yksinkertaisena.

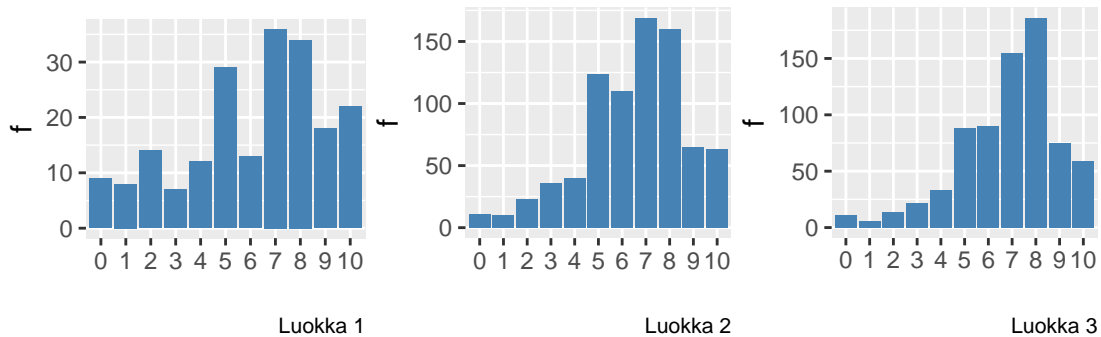
Ordinaalisten muuttujien D23 (vastuuta mittaava vastemuuttuja) ja D28 (“tarpeeksi moni hallitus”) kohdalla päädyimme mahdollisimman tasaiseen luokkajakoon, jossa luokat 0 – 3 yhdistetään alimmaksi luokaksi, luokat 4 – 6 keskimmäiseksi luokaksi ja luokat 7 – 10 ylimmäksi luokaksi.

Viisiluokkaisen (koska luokkaan “en usko ilmastonmuutokseen” ei kuulu yhtään tilastoyksikköä) ilmastonmuutoksen syytä 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.

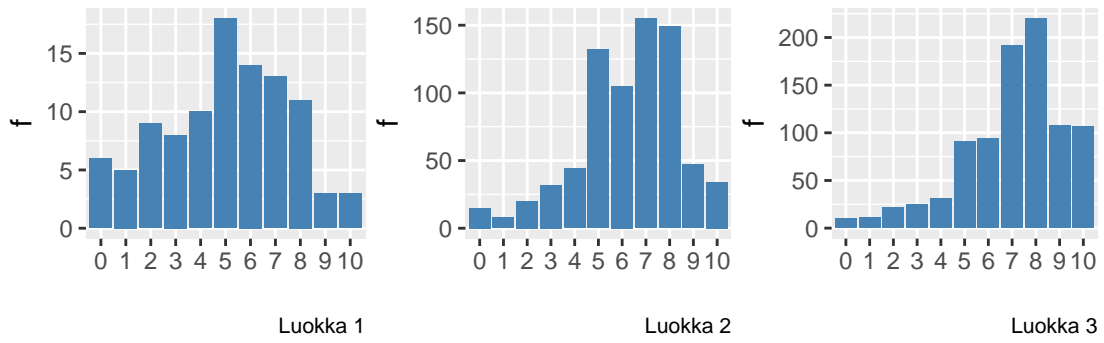
Lisäksi jaamme jatkuvan ikämuuttujan neljään luokkaan niin, että ensimmäiseen luokkaan sijoittuvat alle 31-vuotiaat, toiseen luokkaan 31-44-vuotiaat, kolmanteen luokkaan 45-64-vuotiaat ja neljänteen luokkaan yli 64-vuotiaat.

Kun tarkastellaan henkilökohtaista ilmastovastuuta muiden muuttujien suhteen eri luokkien suhteen ennen kuin myös vastemuuttujan luokkia vähennetään (Taulukot) huomataan että HALLITUS ja

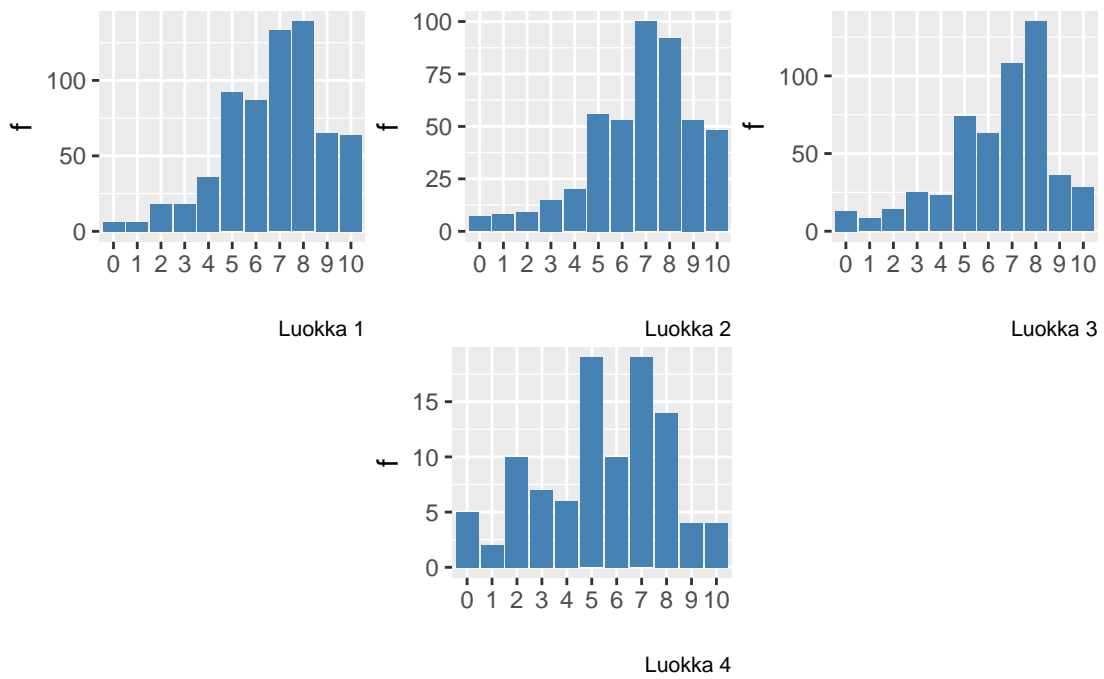
Kuva 2: D23 vastuu vs. D28 hallitus.



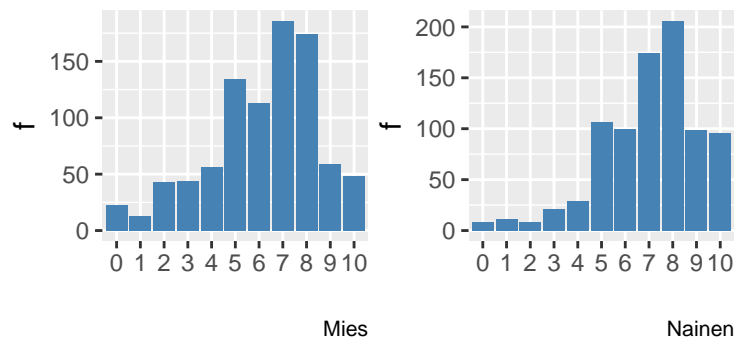
Kuva 3: D23 vastuu vs. D22 syy.



Kuva 4: D23 vastuu vs. Ikä.



Kuva 5: D23 vastuu vs. Sukupuoli.



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

3 Tutkimuskysymyksen mallintaminen

- käytettyjen menetelmien kuvauksen
- menetelmien teoreettista taustaa
- mallien rakentamisen periaatteet
- mallien diagnostiikan

Pieni teksti: Tarkastellaan erilaisia ordinaalisen aineiston malleja, vertaillaan keskenään. Tarkastellaan saturoituja malleja (sopivat hyvin vertailuun)

Kaikkien testistatistiikoiden perusteella paras malli on viereisten kategorioiden logistinen regressiomalli. Malli on muotoa

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

,
missä

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

.

	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

selitys valitusta mallista, vertailee rinnakkaiseen luokkaan. Linkkifunktio. Malliyhtälö. Kerrotaan miten mallia haettiin. Aloitetaan täydestä mallista, edetään ylhäältä alaspäin. Verrataan mallia χ^2 -testillä.

```
##
## Call:
## vglm(formula = cbind(Freq.1, Freq.2, Freq.3) ~ government_action *
##   gender * age - government_action:gender:age - government_action:gender -
##   gender:age, family = acat(parallel = TRUE), data = round8)
##
##
## Pearson residuals:
##               Min         1Q   Median       3Q      Max
## loge(P[Y=2]/P[Y=1]) -1.808 -0.6272  0.1422  0.5068  1.159
## loge(P[Y=3]/P[Y=2]) -1.080 -0.5847 -0.0755  0.6096  1.398
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1      1.02047    0.19158   5.327 1.00e-07 ***
## (Intercept):2       0.65790    0.17256   3.813 0.000137 ***
## government_action2    0.34302    0.19312   1.776 0.075704 .
## government_action3    0.40046    0.20271   1.976 0.048201 *
## gender2              0.63513    0.08956   7.091 1.33e-12 ***
```

```
## age2                -0.23089    0.25852   -0.893  0.371777
## age3                -0.72954    0.26765   -2.726  0.006416 **
## age4               -1.73498    0.52346   -3.314  0.000918 ***
## government_action2:age2  0.12617    0.30460    0.414  0.678707
## government_action3:age2  0.44946    0.32364    1.389  0.164905
## government_action2:age3  0.38066    0.30530    1.247  0.212465
## government_action3:age3  0.77286    0.31897    2.423  0.015394 *
## government_action2:age4  0.73718    0.57276    1.287  0.198070
## government_action3:age4  0.96029    0.57448    1.672  0.094607 .
## ---
## 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: 29.4483 on 34 degrees of freedom
##
## Log-likelihood: -96.6782 on 34 degrees of freedom
##
## Number of iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
```

Mallin tulkintaa: Vakiotermien tulkinta, ovat merkitseviä eli luokkien välillä on eroa. Government action 3 eroaa luokasta 2 (ero $\exp(0.400) = 1.5$ odds verrattuna kakkoseen => Tulkinta, mitä enemmän uskoo että riittävän moni hallitus tekee ilmastonmuutoksen ehkäisemiseksi sen todennäköisemmin kokee myös henkilökohtaista vastuuta ilmastonmuutoksen ehkäisystä. Samoin naiset kokevat miehiä enemmän henkilökohtaista vastuuta ja nuoret enemmän kuin vanhat. Kuitenkin kun otetaan mukaan malliin käsitys riittävän monen hallituksen toimista, iän vaikutus vähenee (tilastollisesti merkitsevästi ikäluokkien kolme ja kaksi välillä).

	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

Samat jutut tästä. Adjacent jälleen paras. Edetään nyt alhaalta ylös.

```
##
## 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

```

Tulkintaa: Syyn ottaminen mukaan malliin => Vakiotermiit eivät ole merkitseviä. Eli responsibility-luokan vaikutus oddsiin voi johtua sattumasta. Caused-by muuttujan tulkinta erilainen, koska ei ordinaalinen. Caused_by vaikuttaa merkitsevästi niin, että mitä enemmän uskoo ihmisen aiheuttavan ilmastonmuutosta sitä enemmän henkilökohtaista vastuuta kokee vastuuta ilmastonmuutoksen ehkäisemisestä. Sukupuoli ja ikä kuten aikaisemmassa mallissa, mutta caused_by

4 Tulosten tulkinta ja johtopäätökset

Mallien tulkinta tänne.