

# Rukoilututkimus 1998 -aineiston tilastollinen tutkimus

LuK-tutkielma  
Aapo Korhonen  
2553636  
Matemaattisten tieteiden laitos  
Oulun yliopisto  
Syksy 2021

# Sisällys

<b>1 Johdanto</b>	<b>2</b>
<b>Johdanto</b>	<b>2</b>
1.1 Aineisto . . . . .	2
1.2 Tutkimuskysymykset . . . . .	2
<b>2 Tutkimuksen muuttajat</b>	<b>3</b>
2.1 Mitkä asiat ovat yhteydessä henkilön rukoilumäärään? . . . .	3
2.1.1 Vastemuuttuja . . . . .	3
2.1.2 Vasemmisto-oikeisto mittari . . . . .	5
2.1.3 Bruttotulot . . . . .	5
2.1.4 Kuntaryhmä . . . . .	6
2.1.5 Syntymävuosi . . . . .	7
2.2 Muuttuvatko rukoilutottumukset iän myötä? . . . . .	8
2.3 Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa? . . . . .	9
<b>3 Menetelmät</b>	<b>9</b>
3.1 Lineaarinen regressio . . . . .	10
3.2 Logistinen regressio . . . . .	11
3.3 Ristiintaulukointi ja $\chi^2$ -testi . . . . .	12
<b>4 Tulokset</b>	<b>13</b>
4.1 Mitkä asiat ovat yhteydessä henkilön rukoilumäärään? . . . .	14
4.2 Muuttuvatko rukoilutottumukset iän myötä? . . . . .	15
4.3 Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa? . . . . .	16
<b>5 Pohdinta</b>	<b>17</b>
<b>Lähdeluettelo</b>	<b>19</b>

# 1 Johdanto

Kuinka paljon ihmiset rukoilevat? Miksi he rukoilevat ja minkä asioiden puolesta he rukoilevat? Nämä asiat pyörivät Kirkon tutkimuskeskuksen työntekijöiden mielessä, kun he alkoivat tekemään vuoden 1998 Rukoilututkimusta [1]. Tässä tutkielmassa lähdetään pohdiskelemaan samoja kysymyksiä ja analysoidaan vuoden 1998 rukoilututkimusta kolmella eri tutkimuskysymyksellä.

## 1.1 Aineisto

Aineistona tässä tutkielmassa on 1998 Rukoilututkimus. Sen on koontanut Kirkon tutkimuskeskus. Tutkimuksen tarkoituksena oli selvittää ihmisten rukouksen määrää, suhtautumista rukoilemiseen ja sitä, minkä asioiden puolesta he rukoilevat. Tutkimuksessa vastaajilta kysyttiin useita kysymyksiä ja mielipiteitä väittämiin.

Aineiston keräyksen ajankohta oli 27.3.1998-31.3.1998. Keruumenetelmänä oli tietokoneavusteinen kasvokkainen haastattelu. Aineiston perusjoukko oli 15-70 -vuotiaat suomalaiset, pois lukien Ahvenanmaalla asuvat.

Kysymykset voidaan jakaa karkeasti kolmeen erilaiseen ryhmään: perustieto- (39 kpl), rukoilutottumus- (45 kpl) ja rukouluun suhtautumiskysymykset (18 kpl). Tämä jako ei ole kuitenkaan tarkka rajainen, vaan moni kysymys jää näiden luokkien ulkopuolelle.

Perustietokysymykset käsittävät muun muassa iän, sukupuolen ja bruttotulot. Perustietokysymykset eivät siis suoranaisesti liity rukouluun itsessään, vaan antavat taustatietoa vastaajasta.

Rukoilutottumuskysymyksistä ensimmäinen on *Kuinka usein rukoilet?*. Sen jälkeen muissa kysymyksissä tiedustellaan, kuinka usein vastaaja rukoulee jonkin tietyn asian puolesta, esimerkiksi lapsien tai rakkauden. Näistä mielenkiintoisin on ensimmäinen kysymys.

Viimeisen ryhmän kysymykset, eli rukouluun suhtautumiskysymykset, poikkeavat toisistaan eniten ja tätä ryhmää on hieman vaikea määrittellä. Tässä ryhmässä on esimerkiksi väittämä *Rukoileminen on täysin turhaa puuhaa*.

## 1.2 Tutkimuskysymykset

Tässä tutkimuksessa tutkimuskysymyksiä on yhteensä kolme. Kysymykset ovat eriaiheisiin liittyviä ja niissä käydään läpi monia eri kyselyssä kysytyjä kysymyksiä ja aiheita. Tutkimuskysymykset järjestyksessä ovat: *Mitkä asiat ovat yhteydessä henkilön rukouksemäärään?*, *Muuttuvatko rukoilutottumukset*

*iän myötä?* ja *Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa?*

Ensimmäisen kysymyksen tarkoituksena on löytää asioita, jotka ovat yhteydessä rukoiluun määrään. Ideana on lähteä muodostamaan asetelmaa, jossa katsotaan, onko perustietokysymyksiä vastauksilla yhteyttä siihen, miten vastaaja on vastannut *Kuinka usein rukoilet?* -kysymykseen.

*Mitkä asiat vaikuttavat henkilön rukoilumäärään?* -tutkimuskysymys ottaa huomioon kaikki kysymykset, joissa käsitellään rukoilutottumuksia, ja yrittää löytää sieltä niitä, jotka ovat jollain tavalla yhteydessä vastaajan ikään.

Viimeisenä *Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa?* -tutkimuskysymyksellä halutaan selvittää, että onko löydettävissä yhteyttä poliittisella mielipiteellä ja lottovoiton puolesta rukoilun kanssa.

## 2 Tutkimuksen muuttujat

Tässä osiossa käydään läpi asetelmat, jotka muodostetaan kunkin tutkimuskysymyksen kanssa analyysia varten. Jokainen muuttuja on uudestaan luokiteltava tai muutettava sopivaan muotoon, jotta analyysit ovat mahdollista tehdä.

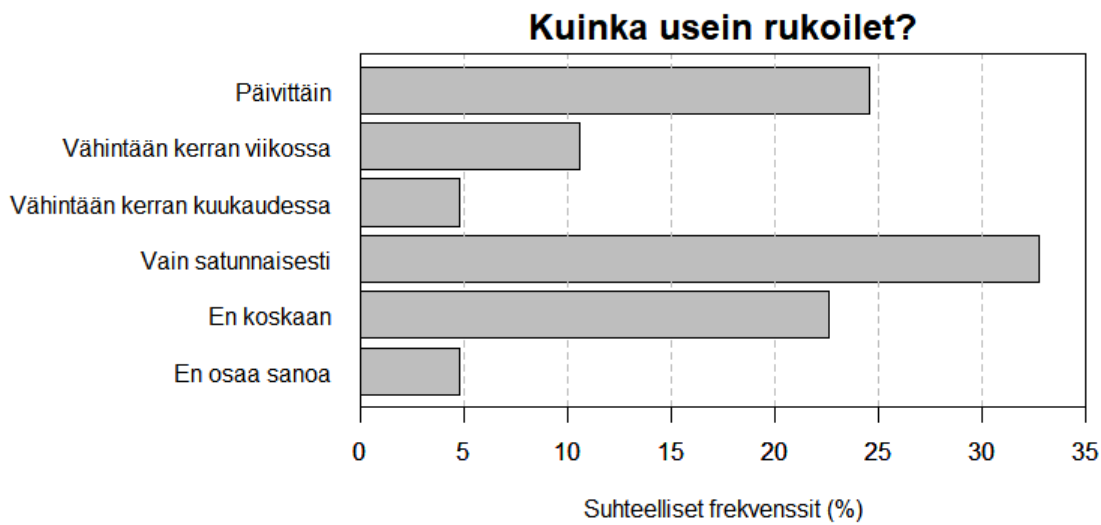
### 2.1 Mitkä asiat ovat yhteydessä henkilön rukoilumäärään?

Ensimmäisen tutkimuskysymyksessä halutaan selvittää, mitkä perustietokysymykset ovat yhteydessä rukoilumäärään. Otetaan perustietokysymyksistä selittäviksi muuttujiksi viisi eri kysymystä, jotka ovat *Sukupuoli*, *Syntymävuosi*, *Bruttotulot mk/vuosi*, *Vasemmisto-oikeisto mittari* ja *Kuntaryhmä*. Vastemuuttujaksi asetetaan aineistosta löytyvä kysymys *Kuinka usein rukoilet?*

#### 2.1.1 Vastemuuttuja

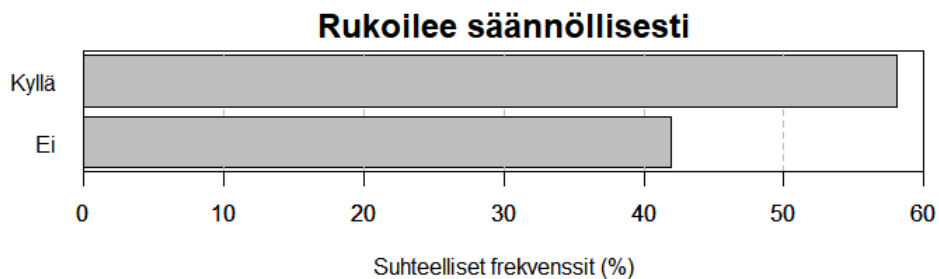
Vastemuuttajan, eli *Kuinka usein rukoilet?* -kysymyksen, alkuperäinen luokittelu on nähtävissä kuvassa 1.

Vastemuuttuja muutetaan kaksiluokkaiseksi siten, että muutetaan kysymys muotoon: *Rukoilee säännöllisesti*. Sitten luokat *Päivittäin*, *Vähintään*



Kuva 1: Alkuperäisen *Kuinka usein rukoilet?* -kysymyksen vastausten jakauma

*kerran viikossa* ja *Vähintään kerran kuukaudessa* yhdistetään uudeksi luokaksi *Kyllä*. Seuraavaksi luokat *Vain satunnaisesti* ja *En koskaan* yhdistetään toiseksi uudeksi luokaksi *Ei*. *En osaa sanoa* vastaukset jätetään pois analyysistä. Sopiva uudestaan luokittelu näkyy kuvassa 2.



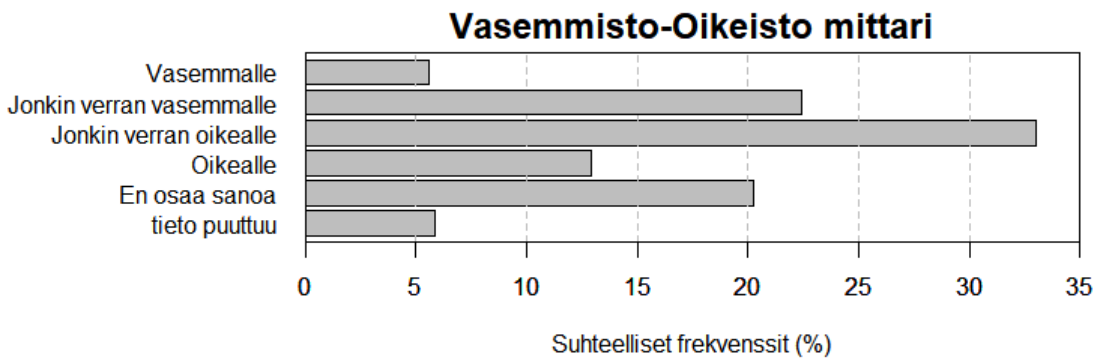
Kuva 2: *Kuinka usein rukoilet?* -kysymys kaksiluokkaisena

Uuden luokittelu takia menetetään tutkittavasta aineistosta 58 havaintoyksikköä.

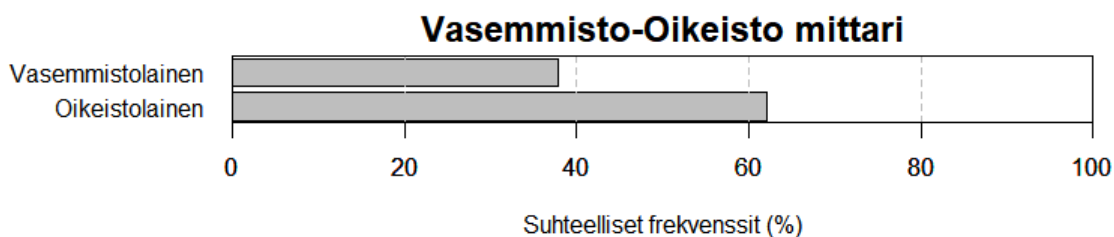
### 2.1.2 Vasemmisto-oikeisto mittari

*Vasemmisto-oikeisto mittari* -kysymyksen tarkoituksena oli selvittää vastaajien poliittista suuntausta. Kuvassa 3 nähdään, mitä vastausvaihtoehtoja oli vastajille annettu.

Tämä kysymys joudutaan niin ikään muuttamaan kaksi arvoiseksi. Se suoritetaan siten, että asetetaan kaksi uutta luokkaa, joista ensimmäinen on *Vasemmistolainen* ja toinen *Oikeistolainen*. Nyt ensimmäiseen luokkaan laitetaan kaikki *Vasemmalle* tai *Jonkin verran vasemmalle* vastanneet vastaajat ja jälkimmäiseen luokkaan *Jonkin verran oikealle* tai *Oikealle* vastanneet. Muut jätetään analyysistä pois.



Kuva 3: Alkuperäisen *Vasemmisto-oikeisto mittari* -kysymyksen vastausten jakauma



Kuva 4: *Vasemmisto-oikeisto mittari* -kysymys kaksiluokkaisena

### 2.1.3 Bruttotulot

Tutkimuksessa kysyttiin vastaajilta heidän bruttotulojaan. Kysymyksessä ei tosin suoraan kysytä absoluuttista rahamäärää tuloista, vaan vastaajille an-

nettiin 14 eri tuloluokkaa, joista heidän piti valita se, mihin heidän tulonsa asettuvat

Tulojen luokittelu on kaikista muuttujista haastavin. Aineistossa *Bruttotulot* muuttujan mahdolliset arvot kuvaavat siis jotain tulojen väliä. Nyt tulot täytyy luokitella uudestaan siten, että muuttujaa voidaan käsitellä analyysissä jatkuvana muuttujana. Luokitellaan jokainen havaintoyksikön arvo sen väliä vastaavaan keskiarvoon. Tällöin esimerkiksi *100.000-125.000* -luokan vastaukset asetetaan arvoon 112.500 markkaa, joka on kyseisen välin keskiarvo. Tämän käsittelyn ansiosta saadaan muutettua kaikki *Bruttotulot* -muuttujan arvot vastaamaan joitain tiettyä markoissa mitattavaa tuloa. Reunaluokkien uudestaan luokittelu on tehty siten, että *<50.000* -luokka olisi oikeasti *0-50.000* -luokka ja *>350.000* -luokka olisi *350.000 - 375.000* -luokka. Viimeisen luokan, eli *En osaa sanoa*, havaintoyksiköt on jätetty analyysistä kokonaan pois.

Bruttotulot Vanhat luokat	Bruttotulot Uudet arvot (mk/vuosi)	n	%
<50.000	25.000	35	3,0
50.000 - 75.000	62.500	62	5,3
75.001 - 100.000	87.500	87	7,5
100.001 - 125.000	112.500	125	10,8
125.001 - 150.000	137.500	112	9,7
150.001 - 175.000	162.500	102	8,8
175.001 - 200.000	187.500	149	12,8
200.001 - 225.000	212.500	147	12,7
225.001 - 250.000	237.500	97	8,4
250.001 - 275.000	262.500	91	7,8
275.001 - 300.000	287.500	28	2,4
300.001 - 325.000	312.500	51	4,4
325.001 - 350.000	337.500	17	1,5
>350.000	362.500	57	4,9
Yhteensä		1160	100

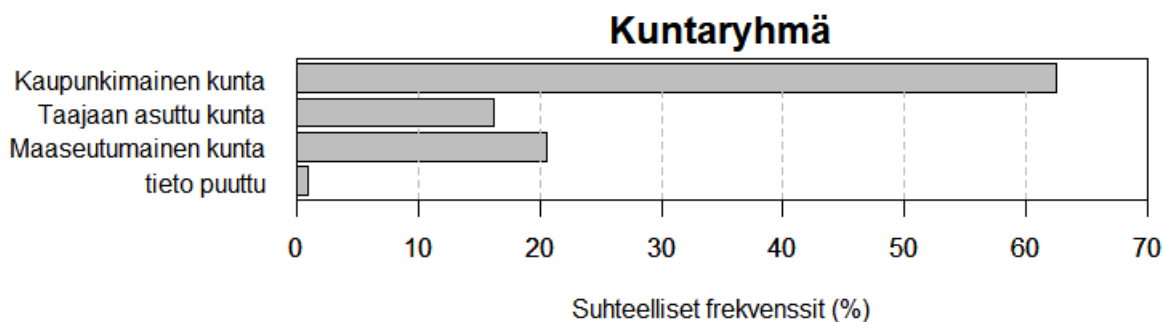
Taulukko 1: Uusiin arvoihin muokatut *Bruttotulot*-kysymyksen vastaukset. Uudet arvot vastaavat nyt vanhoja arvoja vastaavan välin keskiarvoa.

#### 2.1.4 Kuntaryhmä

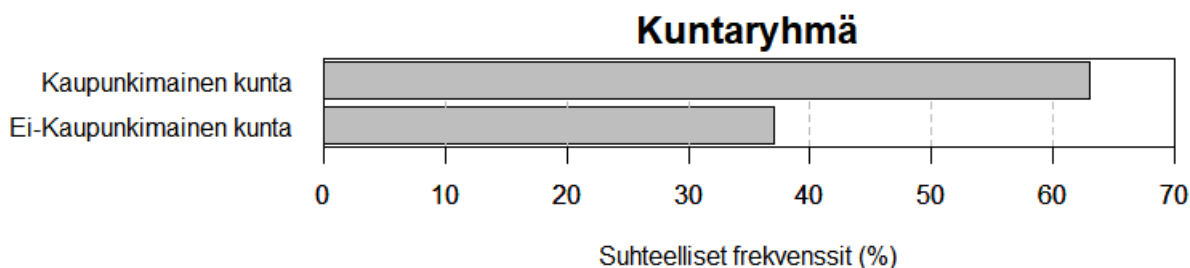
Viimeisenä tarkistellaan vielä *Kuntaryhmä* -muuttujaa. Tässä kysymyksessä kysyttiin vastaajan asuinseutua, tarkemmin asuinkunnan kuntaryhmää.

Luokkia on yhteensä kolme ja ne ovat *Kaupunkimainen kunta* , *Taajaan asuttu kunta* ja *Maaseutumainen kunta*.

Kaksi jälkimmäistä luokkaa on yhdistetty uuteen *Ei-kaupunkimainen kunta* -luokaan. Ensimmäinen *Kaupunkimainen kunta* on pidetty ennallaan. Havaintoyksiköt, joilta puuttuu tieto tästä muuttujasta, on jätetty pois analyysistä. Vanha ja uusi luokittelu on nähtävissä kuvissa 5 ja 6



Kuva 5: Alkuperäisen *Kuntaryhmä* -kysymyksen vastausten jakauma.



Kuva 6: *Kuntaryhmä* -kysymys kaksiluokkaisena

### 2.1.5 Syntymävuosi

Aineistossa ollaan kysytty vastaajilta heidän syntymävuottaan. Muutetaan syntymävuodet vastaamaan vastaajan ikää vuonna 1998. Vastaajan ikää käytetään sitten tutkimuksessa muuttujan arvona, mikä helpottaa siihen saatavien tulosten tulkintaa.



## 2.2 Muuttuvatko rukoilutottumukset iän myötä?

Tämän kysymyksen analysoimiseksi tutkimus asetetaan muotoon, jossa vastemuuttujana on vastaajan ikä ja selittävinä muuttujina ovat kaikki tutkimuksen 45 kysymystä, joilla selvitetään vastaajan rukoilutottumuksia. Tutkimus kulkee tästä eteen päin niin, että selittävästä muuttujasta poistetaan yksi kerrallaan sellaiset muuttujat pois, joilla ei näytä olevan yhteyttä vasteeseen. Lopulliseen malliin ei siis jää kaikki 45 kysymystä, vaan vain ne, joilla on yhteys vasteeseen. Tätä sanotaan askeltavaksi regressio menetelmäksi.

Käydään esimerkkinä yhden kysymyksen uudestaan luokittelu läpi, koska jokaisen 45 kysymyksen kohdalla on toimittu samalla tavalla.

Kuinka usein rukoilet lastesi puolesta?	n	%
Vähintään kerran viikossa	312	25,6
Vähintään kerran kuukaudessa	84	6,9
Vain satunnaisesti	236	19,4
En koskaan	37	3,0
Ei sovi minuun/minulla ei ole	236	19,4
En osaa sanoa	43	3,5
tieto puuttuu	270	22,2
Yhteensä	1218	100

Taulukko 2: Alkuperäisen *Kuinka usein rukoilet lastesi puolesta?* -kysymyksen vastausten jakauma

Otetaan käsittelyyn *Kuinka usein rukoilet lastesi puolesta?* -kysymys. Taulukossa 2 nähdään sen alkuperäinen luokittelu. Muutetaan kysymys muotoon *Rukoilee säännöllisesti lastensa puolesta*. Alkuperäisistä luokista kaksi ensimmäistä yhdistetään uudeksi *Kyllä* -luokaksi ja luokat *Vain satunnaisesti*, *En koskaan* ja *Ei sovi minuun/minulla ei ole* toiseksi luokaksi *Ei*. *En osaa sanoa* ja *tieto puuttuu* vastaukset jätetään pois analyysistä. Uusi luokittelu on nähtävissä taulukossa 3.

Kaikissa kysymyksissä ei ollut vastausvaihtoehtoa *Ei sovi minuun/minulla ei ole*. Niiden kysymysten osalta, joissa ei ollut edellä mainittua vastausvaihtoehtoa, toimittiin muiden luokkien suhteen samalla tavalla.

Rukoilee säännöllisesti lastensa puolesta	n	%
Kyllä	396	43,8
Ei	509	56,2
Yhteensä	905	100

Taulukko 3: *Kuinka usein rukoilet lastesi puolesta?* -kysymys kaksiluokkaisen

## 2.3 Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa?

Viimeisen tutkimuskysymyksen tutkintaan tarvitaan aineistosta kolmea muuttujaa. Vasteena on kysymys *Kuinka usein rukoilet lottovoittoa?*. Muut ovat *Mitä puoluetta tai ryhmittymää äänestit vuoden 1996 lokakuussa järjestetyissä kunnallisvaaleissa ?* ja *Vasemmisto-oikeisto mittari*, jota käsiteltiin ensimmäisen tutkimuskysymyksen parissa. Tutkitaan näiden kahden viimeksi mainitun kysymyksen suhdetta lottovoittokysymykseen ristiintaulukoiden ja  $\chi^2$ -testin avulla.

*Kuinka usein rukoilet lottovoittoa?* -kysymys luokitellaan samalla tavalla, kuin toisen tutkimuskysymyksen kaikki rukoilutottumuskysymykset, eli tehdään siitä kaksi luokkainen. Luokkien jako on sama kuin taulukossa 3.

*Mitä puoluetta tai ryhmittymää äänestit vuoden 1996 lokakuussa järjestetyissä kunnallisvaaleissa ?* -kysymyksen kanssa otetaan käsittelyyn vain kolme suosituinta vastausta, eli puolueet SDP, Keskusta ja Kokoomus. Vastausvaihtoehtoina oli useita muitakin puolueita, mutta kovinkaan moni ei ollut valinnut niitä.

*Vasemmisto-oikeisto mittari* -kysymys luokitellaan samalla tavalla kaksiarvoiseksi kuin aiemmin ensimmäisen tutkimuskysymyksen tilanteessa (kuva 3).

## 3 Menetelmät

Tässä luvussa esitellään kolme tilastollista menetelmää, joita käytetään tässä tutkielmassa. Lineaarista regressiota käytetään *Muuttuvatko rukoilutottumukset iän myötä?*-tutkimuskysymyksen analysoinnissa, logistista regressiota *Mitkä asiat ovat yhteydessä henkilön rukoilumäärään?* -tutkimuskysymyk-

sen kanssa, ja viimeisenä ristiintaulukointia käytetään *Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa?* -tutkimuskysymyksen analyysissä.

### 3.1 Lineaarinen regressio

Lineaarinen regressio soveltuu hyvin tilanteisiin, joissa halutaan selvittää, onko jollakin vastemuuttujalla yhteys johonkin toiseen muuttujaan, eli selittäjään. Esimerkki tutkimuskysymyksistä, jossa lineaarinen regressio on luonteva menetelmä, on: onko vanhempien pituuksilla yhteyttä lapsen pituuteen? Tässä esimerkiksi otetaan selittäväksi muuttujaksi vanhempien pituudet ja vastemuuttujaksi lapsen pituus.

Tällä menetelmällä on yksi selkeä rajoite. Tutkittavan vastemuuttujan on oltava luonteeltaan jatkuva. Selittävien muuttujien ei tarvitse olla jatkuvia ja niitä voi olla useita.

Lineaarissa regressiossa analysointi ja mallinnus pyörii hyvin vahvasti regressiokertoimien ympärillä. Nämä kertoimet estimoidaan analyysin alussa aineiston pohjalta jokaiselle selittävällä muuttujalle. Kerroin kertoo, miten selittävämuuttuja on yhteydessä vasteeseen.

Selittäväille muuttujalle muodostettu malli lineaarisessa regressiossa on seuraava:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon \quad (1)$$

, jossa  $Y$ -arvot vastaavat vasteen arvoja,  $X_i$  ovat selittävien muuttujien arvoja,  $\beta_i$  ovat regressiokertoimia ja  $\varepsilon$  on virhetermi.

Regressiokertoimet estimoidaan pns-, eli pienimmän neliösumman -menetelmällä. Käytännössä tämä tarkoittaa sitä, että regressiokertoimet valitaan siten, että mallin antamat sovitetut vastemuuttujan arvot olisivat mahdollisimman lähellä oikeita vastemuuttujan arvoja.

Kun regressiokertoimet on estimoitu, niitä voidaan alkaa tutkimaan tilastollisen päättelyn menetelmillä. Yleisin tapa on tehdä nollahypoteesi, että regressiokerroin olisikin arvoltaan 0, ja verrata tätä havaittuun arvoon. Tämä vertailu tehdään studentin t-jakauman avulla, jossa otetaan huomioon tilanteeseen liittyvät vapausasteet. Tämän prosessin avulla voidaan lopulta määrittää nollahypoteesille  $P$ -arvo.

$P$ -arvon tulkinta menee erittäin yksinkertaistettuna siten, että se kertoo todennäköisyyden sille, että regressiokertoimelle on satunnaisuudesta johtuen havaittu tutkimuksessa havaittu arvo, tai vielä harvinaisempi arvo, vaikka to-

dellisuudessa sen arvo on 0. Pieni  $P$ -arvo tarkoittaa sitä, että aineiston perusteella ollaan saatu näyttöä siitä, että regressiokerrointa vastaava muuttuja on tavalla tai toisella yhteydessä vasteeseen.

$P$ -arvon lisäksi on hyvä muodostaa regressiokertoimista luottamusväli. Luottamusväli on väli, jonka sisällä todellinen regressiokertoimen arvo todennäköisesti on mallioletusten vallitessa. Usein käytetään 95 % luottamustasoa luottamusväleille.

### 3.2 Logistinen regressio

Logistinen regressio on hyvin samankaltainen kuin edellä käyty lineaarinen regressio. Tässä muodostetaan samankaltaiset regressiokertoimet, jotka niin ikään kertovat, miten selittävät muuttujat ovat yhteydessä vasteeseen.

Selkeä ero lineaariseen regressioon on se, että tässä menetelmässä vastemuuttuja ei ole jatkuva, vaan kaksi arvoinen luokkamuuttuja. Usein vastemuuttuja koodataan arvoilla 1 ja 0.

Mallissa ei käsitellä suoraan vastemuuttujan arvoja. Lähdetään käsittelemään alkuun todennäköisyyttä, että vastemuuttuja saa arvokseen arvon 1.

$$P(Y = 1) = \pi \tag{2}$$

$$P(Y = 0) = 1 - P(Y = 1) = 1 - \pi \tag{3}$$

Kuten yllä olevista yhtälöistä huomataan, niin samalla saadaan laskettua todennäköisyys sille, että vastemuuttuja saa arvokseen arvon 0. Tämän todennäköisyyden arvoksi voidaan ottaa tilanteesta riippuen joko ennakkotietona tiedetty todennäköisyys tai se voidaan estimoida aineiston perusteella. Estimointi lasketaan yleensä näin:

$$\pi = \frac{y_1}{N} = \frac{\text{Vasteen arvon 1 saaneiden lkm}}{\text{Aineiston havaintoyksiköiden lkm}} \tag{4}$$

Todennäköisyyden vaihteluväli on  $\pi \in [0 : 1]$ . Lähdetään seuraavaksi käsittelemään vedonlyöntikerrointa, joka lasketaan seuraavalla kaavalla:

$$\omega = \frac{\pi}{1 - \pi} \tag{5}$$

Vedonlyöntikerroimen vaihteluväli on suurempi kuin todennäköisyyden. Tarkemmin sen vaihteluväli on:  $\omega \in [0 : \infty[$ . Tätä voidaan jälleen laajentaa ottamalla tästä vielä luonnollinen logaritmi. Logaritmoitu vedonlyöntikerroin on:

$$\lambda = \ln(\omega) = \ln\left(\frac{\pi}{1 - \pi}\right) \tag{6}$$

Lopulta tämän arvon vaihteluväli on  $\lambda \in ] - \infty : \infty [$ . Logaritmoitua vedonlyöntikerrointa voidaan käsitellä kuten se olisi jatkuva muuttuja. Tämän jälkeen voidaan logaritmoidulle vedonlyöntikertoimelle muodostaa samanlainen malli kuin lineaarisessa regressiossa.

Logistisessa regressiossa muodostettu malli on muotoa:

$$\lambda = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon \quad (7)$$

, jossa  $\lambda$  on logaritmoitu vedonlyöntikerroin,  $X_i$  on selittävän muuttujan arvo,  $\beta_i$  on regressiokerroin ja  $\varepsilon$  on virhetermi.

Jos halutaan selvittää todennäköisyys sille, että vaste saa arvokseen arvon 1, niin on käytettävä seuraavaa yhtälöä:

$$P(Y = 1 | X_1, X_2 \dots) = \frac{e^\lambda}{1 + e^\lambda} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots}} \quad (8)$$

Mielekkäämpi tulkittava arvo on *OR*-tunnusluku, joka kertoo vedonlyöntikertoimien suhdeluvun selittävän muuttujan arvoilla 0 ja 1. Usein tämä tarkoittaa eroa kahden eri ryhmän välillä.

$$OR = e^{\beta_i} \quad (9)$$

Logistisella regressiolla saaduille regressiokertoimille voidaan lopuksi tehdä samanlainen nollahypoteesi kuin lineaarisessa regressiossa tehtiin. *OR*-tunnusluvulle voidaan muodostaa myös luottamusväli.

### 3.3 Ristiintaulukointi ja $\chi^2$ -testi

Ristiintaulukointia käytetään hyväksi tilanteissa, joissa halutaan selvittää, onko jokin ominaisuus jakautunut erillä tavalla kahden tai useamman ryhmän välillä. Esimerkki ristiintaulukosta on nähtävissä taulukossa 4.

Jotta voidaan testata, onko ryhmien välillä eroa, pitää selvittää riippumattomuus oletuksen mukaiset odotetut arvot jokaiselle taulukon solulle. Satunnaiselle solulle  $ij$  odotettu arvo lasketaan seuraavasti:

$$e_{ij} = \frac{N_{i.}}{N} N_{.j} \quad (10)$$

Odotettujen arvojen ja havaittujen arvojen avulla voidaan laskea  $\chi^2$ -arvo seuraavalla kaavalla:

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (11)$$

	Ryhmät				
Ryhmät	$X_1$	$X_2$	$\dots$	$X_k$	Yhteensä
$Y_1$	$n_{11}$	$n_{21}$	$\dots$	$n_{k1}$	$N_{.1}$
$Y_2$	$n_{12}$	$n_{22}$	$\dots$	$n_{k2}$	$N_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$Y_l$	$n_{1l}$	$n_{2l}$	$\dots$	$n_{kl}$	$N_{.l}$
Yhteensä	$N_{.1}$	$N_{.2}$	$\dots$	$N_{.k}$	$N$

Taulukko 4: Esimerkki ristiintaulukosta.  $X_i$  = sarakeluokka,  $Y_j$  = riviluokka,  $n_{ij}$  = solun havaittu arvo,  $N_{.j}$  = rivisumma,  $N_{i.}$  = sarakesumma,  $N$  = havaintoyksiköiden lukumäärä.

Tämä on suoraan hieman harhainen arvo. Tulosta voidaan korjata harhattomampaan suuntaan Yatesin jatkuvuuskorjauksella.

Jatkuvuuskorjatun  $\chi^2$ -arvon yhtälöksi tulee:

$$\chi^2 = \sum_{(i,j)} \frac{(|n_{ij} - e_{ij}| - 0,5)^2}{e_{ij}} \quad (12)$$

Edellä laskettu  $\chi^2$  arvo noudattaa  $\chi^2$  jakaumaa (vapausasteluvuilla  $(k - 1) \cdot (l - 1)$ ), kun nollahypoteesi on voimassa. Nollahypoteesina on se, että vertailtavien luokkien välillä ei ole eroa. Tämän jakauman perusteella saadaan laskettua  $P$ -arvo, jonka avulla voidaan päätellä, onko löydetty näyttöä siitä, että ryhmien välillä on eroa.

## 4 Tulokset

Seuraavaksi käydään läpi tilastollisesta analyysistä saadut tulokset. Alla oleviin taulukoihin on kirjattu kaikki analyysin ja tulosten tulkinnan kannalta oleelliset arvot R-ohjelmalla saadusta tulosteesta.

## 4.1 Mitkä asiat ovat yhteydessä henkilön rukoilumäärään?

Selittävä muuttuja	Regressiokerroin		OR-tunnusluku	
	$\hat{\beta}$	<i>P</i> -arvo	Estim	95% -luottamusväli
Tulot	-4,62e-7	0,0218	0,9999995	[ 0,9999993 : 0,9999999 ]
Ikä	5,04e-3	5,92e-5	1,005	[ 1,003 : 1,008 ]
Sukupuoli (mies)	-1,89e-1	1,39e-8	0,827	[ 0,775 : 0,883 ]
Kunta (Kaupunki)	-1,04e-1	0,00297	0,901	[ 0,841 : 0,965 ]
Poliittinen suuntaus (vasen)	-1,59e-1	3,79e-6	0,853	[ 0,798 : 0,912 ]

Taulukko 5: Logistinen regressio: Vastemuuttujaan *Rukoilee säännöllisesti* (0 = ei, 1 = kyllä) yhteydessä olevat tekijät. Tulot ja ikä ovat jatkuvia muuttujia. Sukupuoli (0 = nainen, 1 = mies), kunta (0 = Ei-kaupunkimainen kunta, 1 = kaupunkimainen kunta) ja poliittinen suuntaus (0 = oikeistolainen, 1 = vasemmistolainen) ovat kaksiarvoiset muuttujat.

Taulukosta 5 on hyvä huomata, että jokaisen regressiokertoimen kohdalla määritetyt *P*-arvot ovat hyvin pieniä. Lisäksi minkään *OR*-tunnusluvun luottamusvälille ei kuulu arvoa 1. Voidaan sanoa, että ollaan löydetty näyttöä sille, että jokaisella mallin selittäväällä muuttujalla on yhteys vasteeseen.

Millainen yhteys sitten muuttujilla on vasteeseen? Koska analyysissä on käytetty logistista regressiota, niin käytetään tulkinnoissa *OR*-tunnuslukuja apuna. Alhaalle on listattu jokaisen muuttujan yhteys vasteeseen.

- Ikä: Iän kasvaessa rukoilun määrä myös kasvaa (*OR* = 1,005, 95% lv. [ 1,003 : 1,008 ]).
- Tulot: Tulojen kasvaessa rukoilun määrä vähenee (*OR* = 0,9999995, 95% lv. [ 0,9999993 : 0,9999999 ]).
- Sukupuoli: Miehet rukoilevat vähemmän kuin naiset (*OR* = 0,827, 95% lv. [ 0,775 : 0,883 ]).
- Kunta: Kaupungissa asuvat rukoilevat vähemmän kuin muualla asuvat (*OR* = 0,901, 95% lv. [ 0,841 : 0,965 ]).
- Poliittinen suuntaus: Vasemmistolaiset rukoilevat vähemmän kuin oikeistolaiset (*OR* = 0,853, 95% lv. [ 0,798 : 0,912 ]).

## 4.2 Muuttuvatko rukoilutottumukset iän myötä?

Rukoilun aihe	$\hat{\beta}$	$P$ -arvo	95% -luottamusväli
Rakkaus	-4,8	0,00093	[ -7,65 : -1,97 ]
Terveys	4,1	0,0061	[ 1,17 : 6,98 ]
Neuvoa käytännön ratkaisuihin	-3,6	0,014	[ -6,42 : -0,72 ]
Jumalan armoa	5,4	0,00051	[ 2,35 : 8,35 ]
Lapset	10,1	1,7e-15	[ 7,71 : 12,58 ]
Vanhemmat	-13,2	<2e-16	[ -16,25 : -10,19 ]
Muut sukulaiset	3,3	0,046	[ 0,05 : 6,49 ]
Työttömät	5,0	0,017	[ 0,91 : 9,04 ]
Vammaiset	4,0	0,033	[ 0,32 : 7,73 ]
Vakiotermi	39,9	<2e-16	[ 38,53 : 41,30 ]

Taulukko 6: Lineaarinen regressio: Vastaaajan ikään ( $v$ ) yhteydessä olevat tekijät (0 = ei, 1 = kyllä)

Taulukon 6 tuloksista huomataan, että  $P$ -arvot ja luottamusvälit antavat hyvää näyttöä sille, että jokaisella malliin jääneellä selittäjällä on yhteys vasteeseen. Tarkastellaan listattuna jokaisen selittävän muuttujan yhteyttä vasteeseen. Huomataan, että koska käytetty menetelmä oli lineaarinen regressio, niin vakiotermillä ja regressiokertoimilla on ymmärrettävä tulkinta.

- Positiivisesti iän kanssa korreloivat rukoilun aiheet:
  - Terveys: Tämän aiheen puolesta säännöllisesti rukoileva on keskimäärin 4,1 vuotta vanhempi kuin sellainen vastaaaja, joka ei rukoile säännöllisesti tämän puolesta ( $\hat{\beta} = 4,1$ , 95% lv. [ 1,17 : 6,98 ]).
  - Jumalan armo: 5,4 vuotta vanhempi keskimäärin ( $\hat{\beta} = 5,4$ , 95% lv. [ 2,35 : 8,35 ]).
  - Lapset : 10,1 vuotta vanhempi keskimäärin ( $\hat{\beta} = 10,1$ , 95% lv. [ 7,71 : 12,58 ]).
  - Muut sukulaiset (ei vanhemmat tai lapset) : 3,3 vuotta vanhempi keskimäärin ( $\hat{\beta} = 3,3$ , 95% lv. [ 0,05 : 6,49 ]).
  - Työttömät: 5,0 vuotta vanhempi keskimäärin ( $\hat{\beta} = 5,0$ , 95% lv. [ 0,91 : 9,04 ]).



- Vammaiset: 4,0 vuotta vanhempi keskimäärin ( $\hat{\beta} = 4,0$ , 95% lv. [ 0,32 : 7,73 ]).
- Negatiivisesti iän kanssa korreloivat rukoilun aiheet:
  - Rakkaus: Tämän aiheen puolesta säännöllisesti rukoileva on keskimäärin 4,8 vuotta nuorempi kuin sellainen vastaaja, joka ei rukoile säännöllisesti tämän puolesta ( $\hat{\beta} = -4,8$ , 95% lv. [ -7,65 : -1,97 ]).
  - Käytännön ratkaisuihin neuvoa: 3,6 vuotta nuorempi keskimäärin ( $\hat{\beta} = -3,6$ , 95% lv. [ -6,42 : -0,72 ]).
  - Vanhemmat: 13,2 vuotta nuorempi keskimäärin ( $\hat{\beta} = -13,2$ , 95% lv. [ -16,25 : -10,19 ]).
- Vakiotermin tulkinta: Vastaaja, joka ei rukoile minkään yllä mainitun asian puolesta, on keskimäärin 39,9-vuotias ( $\hat{\beta} = 39,9$ , 95% lv. [ 38,53 : 41,30 ]).

### 4.3 Onko poliittisella mielipiteellä yhteyttä siihen, kuinka usein ihminen rukoilee lottovoittoa?

Kolmannen tutkimuskysymyksen tulokset. Alla on nähtävissä molemmat ristiintaulukot  $\chi^2$ - ja  $P$ -arvoineen.

Puoluekanta	Rukoilee säännöllisesti lottovoittoa (%)		n	Kuinka moni rukoilee ylipäättänsä säännöllisesti (%)
	Ei	Kyllä		
KESK	95,5	4,5	157	62,0
KOK	90,0	10,0	160	46,2
SDP	86,1	13,9	151	41,1
Yhteensä	90,6	9,4	468	42,0

$\chi^2 = 8,1697$ , Vapausasteet=2,  $P$ -arvo=0,01683

Taulukko 7: Ristiintaulukointi: Vastemuuttujan *Rukoilee säännöllisesti lottovoittoa* ja puoluekannan ristiintaulukoinnin tulokset. Lisätietona säännöllisen rukoilun esiintyvyyys puolueittain.

Poliittinen suuntaus	Rukoilee säännöllisesti lottovoittoa (%)		n	Kuinka moni rukoilee ylipäättänsä säännöllisesti (%)
	Ei	Kyllä		
Oikeistolainen	93,1	6,9	436	49,4
Vasemmistolainen	91,1	8,9	225	38,5
Yhteensä	92,4	7,6	661	42,0

$\chi^2 = 0,59285$ , Vapausasteet=1,  $P$ -arvo=0,4413

Taulukko 8: Ristiintaulukointi: Vastemuuttujan *Rukoilee säännöllisesti lottovoittoa* ja poliittisen suuntauksen ristiintaulukoinnin tulokset. Lisätietona säännöllisen rukoilun esiintyvyys eri poliittisilla suuntauksilla.

Taulukosta 7 saadaan selville, miten lottovoitosta rukoilu poikkeaa eri puolueiden kannattajien keskuudessa. Keskustaa äänestäneet rukoilivat suhteessa vähintään lottovoiton puolesta. SDP:tä äänestäneet rukoilivat suhteessa eniten lottovoiton puolesta. Kokoomuslaiset ovat näiden välissä.  $P$ -arvon tulkinna perusteella voidaan sanoa, että on näyttöä, että jotain yhteyttä puoluekannalla on lottovoiton rukoiluun.

Taulukon 8 perusteella ei ole havaittavissa eroa lottovoiton rukoilun määrässä oikeistolaisten ja vasemmistolaisten välillä. Huomattavaksi asiaksi voidaan sanoa, että oikeistolaiset kuitenkin yleisesti ottaen rukoilevat enemmän kuin vasemmistolaiset.

## 5 Pohdinta

Tutkimuksessa oli useita vaiheita, joissa olisi voinut toimia eri tavalla. Jokaisen kysymyksen kohdalla jouduttiin valitsemaan käsittelyyn sopivat muuttajat.

Viimeisen tutkimuskysymyksen ristiintaulukossa 7 olevan puoluekantamuuttajan kohdalle oli mahdollista valita myös toinen muuttuja. Analyysissä ollut muuttuja oli kysymys siitä, mitä puoluetta vastaaja oli äänestänyt edellisissä vaaleissa. Toinen vaihtoehto tähän oli kysymys, jossa vastaajalta tiedusteltiin, mitä puoluetta hän äänestäisi, jos nyt olisi vaalit. Viimeksi mainittua muuttujaa ei valittu tutkimuksiin, koska siihen oli hyvin moni jättänyt vastaamatta.

Muuttujien uudestaan luokittelussa oli paljon valinnan varaa ja mahdol-

lisuuksia toimia toisin. Jokaisen muuttujan kohdalla yritettiin valita sellainen uudestaan luokittelu, että sen tulkinta ja käyttö olisi mahdollisimman helppoa ja selkeää.

*Kuntaryhmä*-muuttujalla oli alun perin kolme luokkaa, jotka olisi voinut ottaa kaikkineen mukaan mallinnukseen. Ongelmaksi olisi koitunut mahdollisesti tulkinta. Yhdistämällä kaksi viimeistä luokkaa yhdeksi, saatiin muuttujan tulkintaan selkeä kaupungissa asuvien ja ei kaupungissa asuvien vastakkain asettelu.

*Bruttotulot*-muuttujan uudestaan luokittelu oli haastavaa. Lopullinen luokittelu, mihin päädyttiin (taulukko 1), on ongelmallinen erityisesti viimeisen luokan suhteen. Siinä asetetaan viimeiseen luokkaan kuuluvien vastaajien tulot arvoon 362.500, vaikka heidän tulonsa olisivat voineet olla vaikka kuinka paljon suuremmat. Tämä voi aiheuttaa harhaa tuloksiin. Toinen vaihtoehtoinen luokittelutapa olisi ollut ensin selvittää aineiston vastaajien keskitulo. Sen jälkeen luokiteltaisiin jokainen havaintoyksikkö joko keskivertoa parempi tai huonompi tuloiseksi. Tällainen kaksi arvoinen luokittelu olisi ollut helppo tulkita, mutta keskitulojen määrittäminen ei olisi ollut yksinkertaista, koska ei tiedetä suoraan vastaajien tuloja.

Tutkimuksissa ei tullut vastaan mitään oudoksuttavia tuloksia tai kummallisuuksia vastaan. Itse tutkimus oli suoraviivaista ja ongelmatonta.

Mahdollisia harhoja tuloksiin on voinut tulla siitä, että lähes jokaisen muuttujan kohdalla on ollut melko paljon katoa. Esimerkiksi viimeisen tutkimuskysymyksen ristiintaulukoista on huomattavasti vähemmän havaintoyksiköitä kuin mitä koko aineistossa on mukana. Ristiintaulukossa 7 on vain 468 havaintoyksikköä ja ristiintaulukossa 8 havaintoyksikköjä oli 661, eli vähän yli puolet kaikista vastaajista. Tämä ei välttämättä ole ongelma, mutta suurella kadolla voi olla vaikutusta tulosten tarkkuuteen ja lisätä niiden epävarmuutta.

Aineistosta oli löydettävissä paljon mielenkiintoisia korrelaatioita eri muuttujien välillä ja kiinnostavia tutkimuskysymyksiä olisi voinut tehdä vaikka kuinka paljon. Esimerkiksi viimeiseen tutkimuskysymykseen liittyen aineiston vastaajien puoluekannatuksilla ja poliittisella suuntauksella on löydettävissä selkeä korrelaatio. Keskustaa ja kokoomusta äänestäneistä vastaajista noin 90 % luokitteli itsensä oikeistolaisiksi ja SDP:tä äänestäneistä noin 80 % luokitteli itsensä vasemmistolaisiksi. Valitettavasti tämä tutkielma ei ollut tarpeeksi laaja syvempään aineiston analyysiin.

## Lähdeluettelo

- [1] Kirkon tutkimuskeskus: Rukoilututkimus 1998 [sähköinen tietoaineisto]. Versio 1.0 (2002-01-24). Yhteiskuntatieteellinen tietoaarkisto [jakaaja]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD1127>
- [2] *Johdatus monimuuttuja menetelmiin* -kurssin kurssimateriaali Moodlessa, Oulun yliopisto, 2020.
- [3] *Johdatus regressio- ja varianssianalyysiin* -kurssin kurssimateriaali Moodlessa, Oulun yliopisto, 2020.
- [4] Tietoaarkisto, *Ristiintaulukointi*, <https://www.fsd.tuni.fi/fi/palvelut/metelmaopetus/kvanti/ristiintaulukointi/ristiintaulukointi/>, Viittauspäivämäärä 10.12.2021