

Hottinen Aaro, Mehtonen Mikko

# **Samankaltaiset kunnat**

**Menetelmäkuvaus 7.4.2022**



© Suomen Kuntaliitto ry  
Helsinki 2021

Kuntaliitto  
Toinen linja 14  
PL 200, 00101 Helsinki  
Puhelin 09 7711  
[www.kuntaliitto.fi](http://www.kuntaliitto.fi)

# Sisältö

<b>Menetelmäkuvaus .....</b>	<b>4</b>
Teema 1: kunnan kustantama varhaiskasvatus.....	5
Teema 2: Peruopetus.....	10
Teema 3: Esiopetus.....	13

# Menetelmäkuvaus

SAKU samankaltaiset kunnat -hankkeessa etsitään kuntien kustannuseroja selittäviä tekijöitä. Löydettyjen tekijöiden perusteella jokaiselle kunnalle valitaan mahdollisimman samankaltaiset vertailukunnat, joihin kunta voi verrata omaa toimintaansa.

Tutkimus toteutetaan teemoittain. Alkuvaiheeseen valittuja teemoja ovat: varhaiskasvatus, esiopetus sekä perusopetus. Opetuspalveluiden teemoissa selvitetään, mitkä tekijät selittävät kustannuseroja.

Hanke aloitettiin haastattelemalla kunkin teeman, esimerkiksi varhaiskasvatuksen, asiantuntijoita. Haastattelujen avulla kerättiin mahdollisimman laaja lista tekijöistä, joilla saattaa olla yhteys kustannuksiin. Seuraavassa vaiheessa selvitettiin, mitkä listauksen tekijöistä on muutettavissa numeeriseen muotoon eli muuttujiksi ja joihin löytyy numeerista tilastoaineistoa. Kunnilta ei kerätty aineistoa, vaan hankkeessa käytettiin ainoastaan jo kerättyä tilastotietoa. Osa SAKU-malliin ehdolla olevista tekijöistä karsiutui pois sen vuoksi, ettei tilastoaineistoa ollut saatavilla. Tältä osin tutkimusta voisi kehittää tulevaisuudessa.

Muuttujiksi valitut tekijät jaoteltiin olosuhde- sekä päätösperäisiin muuttujiin. Olosuhdetekijät ovat sellaisia, joihin kunta ei voi itse merkittävästi vaikuttaa. Päätösperäiset tekijät taas ovat sellaisia, joihin kunta voi vaikuttaa. Samankaltaiset kunnat valitaan olosuhdetekijöiden perusteella.

SAKU-hankkeen tilastollisen tutkimuksen vaiheessa selvitettiin regressioanalyysin avulla tähän vaiheeseen valittujen muuttujien selitysvoimaa. Regressiomallin avulla muuttujalistasta valittiin ne muuttujat, jotka selittivät parhaiten kuntien välisiä kustannuseroja. Regressiomalleja testattiin useammilla eri muuttujakombinaatioilla muun muassa siten, että keskenään vahvasti korreloivista muuttujista valittiin malliin vain toinen.

Regressiomalleissa hyödynnettiin ajallista ja alueellista vaihtelua. Etenkin alueellinen vaihtelu selittää eroja, koska se on lähtökohtaisesti suurempaa. Aineisto koostuu kuntakohtaisesta aineistosta vuosilta 2015–2019, lukuun

ottamatta varhaiskasvatusta, jonka kuntakohtainen aineisto koostuu vuosien 2016–2019 tilastotiedoista.

Estimoiteja kokeiltiin myös kuntien kiinteillä vaikutuksilla. Ne eivät kuitenkaan vaikuttaneet tuloksiin (testattu järjestämistapaestimaatilla). Kiinteät vaikutukset poistavat alueellisen vaihtelun eli kuntien väliset erot, jolloin päädytään seuraamaan vain ajallista vaikutusta. Kiinteiden vaikutusten estimointia ei käytetty varsinaisessa SAKU-mallissa.

Muuttujien valinnassa huomioitiin myös toimialatuntemusta. Samankaltaisia kuntia etsittäessä, malliin valittuja muuttujia painotetaan regressiokertoimien perusteella. Painoarvot laskettiin muuttamalla malliin valittujen muuttujien regressiokertoimet yhteismitalliseksi.

Eräät muuttujat karsiutuivat SAKU-mallista sen vuoksi, että niiden painoarvo jäi edellä mainitulla laskentavalla hyvin pieneksi. Tämä johtui regressiokertoimen alhaisesta arvosta. Esimerkiksi varhaiskasvatuksen mallissa väkiluku selitti kustannuseroja tilastollisesti merkitsevästi, mutta niin vähän, että väkilukua ei käytetä samankaltaisten kuntien etsimisessä.

Regressiomalli toimii hyvin silloin, kun siinä on mukana kaikki selitettävään ilmiöön vaikuttavat selittävät tekijät ja niitä mittaavat muuttujat. Puuttuvan muuttujan harha syntyy silloin kun mallista puuttuu sellainen muuttuja, jolla olisi selitysvoimaa ja joka korreloi yhden tai useamman mallissa olevan selittävän muuttujan kanssa.

## **Teema 1: kunnan kustantama varhaiskasvatus**

Teemassa etsitään kuntien varhaiskasvatuksen kustannuseroja selittäviä tekijöitä. Kunnan kustantamalla varhaiskasvatuksella tarkoitetaan kunnan omaa perhepäivä- ja päiväkotihoidon, palvelusetelillä järjestettyä toimintaa, kunnan maksamaa yksityistä varhaiskasvatusta, kuntalisiä sekä muuta varhaiskasvatusta. Eli kustannukset muodostuvat varhaiskasvatuksen kokonaiskustannuksista.

Taloustietojen lähteenä käytetään Tilastokeskuksen taloustilastoa. Kustannuksina käytetään käyttökustannuksia eli niihin sisältyivät toimintakulut, poistot ja arvonalentumiset sekä sisäiset erät ja vyörytykset. Kustannuksista ei vähennetty asiakasmaksuja, mutta niistä vähennettiin muille ulosmyyty palvelu eli kustannusten pitäisi kuvata mahdollisimman hyvin oman kunnan asukkaille tarjottujen varhaiskasvatuspalveluiden kustannuksia. Näitä kustannustietoja on

mielekästä käyttää kun olosuhde- ja päätösmuuttajat kuvaavat myös kyseisen kunnan tilannetta.

Selitettävä muuttuja: varhaiskasvatuksen käyttökulut / kunnan 1-6 vuotias

Varhaiskasvatuksen viimeisimmän regressiomallin selitysaste on 68 %. Ilman mekaanisia kustannustekijöitä selitysaste on 63 %.

Selitettävän muuttujan laskennassa kunnan käyttökulut suhteutettiin aluksi suoraan varhaiskasvatuksen asiakasmäärään. Tämä laskentatapa ei kuitenkaan tuottanut regressioajoissa merkittäviä tuloksia. Mallien selitysaste nousi merkittävästi, kun kustannukset suhteutettiin palvelua käyttävään ikäluokkaan. Todennäköinen syy tälle on se, että julkisista lähteistä saadut asiakastiedot eivät olleet vertailukelpoisia kustannustietojen kanssa. Mallia voitaisiin jatkokehittää, jos selitettävä muuttuja saataisiin jatkossa laskettua luotettavasti muodossa euroa per asiakas.

Asiantuntijahaastatteluissa tunnistettiin yli 40 varhaiskasvatuksen kustannuksiin mahdollisesti vaikuttavaa tekijää. Kustannuserojen selittäjiksi näistä päästiin testaamaan yli 30 muuttujaa:

Kunnan kustantamaan varhaiskasvatukseen osallistuneiden (osallistumisaste*1-6 -vuotias väestö) osuus väestöstä, %	***
Kunnan kustantamaan varhaiskasvatukseen osallistuneiden (osallistumisaste*1-6 -vuotias väestö) vuosimuutos, %	***
Kunnan kustantamaan varhaiskasvatukseen osallistuneet 31.12., % 1-6-vuotiaista	***
Ansiotulot per asukas, eur./as.	***
Kelan tilastoissa positiivista kuntalisää maksaneet kunnat = 1, muut kunnat = 0	***
Koko- tai osa-aikaiseen päiväkotitai perhepäivähoidotoimintaan osallistuneiden lasten lukumäärän suhde lukuun, johon on laskettu edellisen lisäksi myös yksityisen hoidon tuella hoidetut lapset ja palveluseteliasiakkaat	***
Varhaiskasvatustoiminnan vuokratkustannukset (sisäiset+ulkoiset) per lapsi (= 1-6 v. väestö)	***
Taloudellinen huoltosuhte, %	***
Verotulot per asukas, eur./as.	***
Valtionosuuden osuus verorahoituksesta (verotulot + valtionosuudet), %	***
Väkiluku 31.12, lkm	***

Taajama-aste, %	**
Maanteiden yhteispituus, km	**
Keskiasteen koulutuksen saaneet, % 20 vuotta täyttäneistä	**
Korkea-asteen koulutuksen saaneet, % 20 vuotta täyttäneistä	**
Perusasteen jälkeen suoritetun korkeimman koulutuksen keskimääräinen pituus henkeä kohti	**
Kunnan järjestämään kokoaikaiseen varhaiskasvatukseen osallistuneiden 1-2 -vuotiaiden osuus vastaavalla tavalla luokitelluista 1-5 -vuotiaista, %	*
Kehyskunta = 1, muu kunta = 0	*
Työttömyysaste, %	*
Keski-ikä, vuotta	ei selitysastetta
Väkiluvun vuosimuutos, %	ei selitysastetta
Maapinta-ala, km <sup>2</sup>	ei selitysastetta
Saaristokunta = 1, muu kunta = 0	ei selitysastetta
Saaristo-osakunta = 1, muu kunta = 0	ei selitysastetta
Liikennekäyttöisten autojen lukumäärä per asukas	ei selitysastetta
Vasemmistoenemmistöinen (SDP/Vasemmistoliitto) valtuusto = 1, muu valtuusto = 0	ei selitysastetta
Toimeentulotukea lyhytaikaisesti saaneet yksinhuoltajakotitaloudet, % toimeentulotukea saaneista kotitalouksista	ei selitysastetta
Huostassa vuoden aikana olleet 3-6-vuotiaat, % vastaavanikäisistä	ei selitysastetta
Koko- tai osa-aikaiseen päiväkotitoimintaan osallistuneiden lasten lukumäärän (oma toiminta) suhde lukuun, johon on laskettu edellisen lisäksi myös koko- tai osa-aikaiseen perhepäivähoitoon osallistuneiden lasten lukumäärä (oma toiminta), %	ei selitysastetta
Pidennetyn oppivelvollisuuden oppilaat, % perusopetusoppilaista	ei selitysastetta
Ruotsinkielisten osuus väestöstä, %	ei selitysastetta
Vieraskielisten osuus väestöstä, %	ei selitysastetta
Toimialarakenne	ei selitysastetta

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Ikärakennetta kokeiltiin aluksi eri ikäryhmien väestöosuuksina. Kunnan ikärakennetta kuitenkin yksinkertaistettiin, jolloin päädyttiin kunnan keski-ikää kuvaavaan muuttujaan. Keski-ian ei kuitenkaan havaittu selittävän kustannuseroja.

Edellä mainituista muuttujista varhaiskasvatuksetn SAKU-malliin valittiin seuraavat muuttujat:

- Taloudellinen huoltosuhte (Painoarvo 24 %)

- Maanteiden pituus: (Painoarvo 4 %)
- Kuntalisä: (Painoarvo 5 %)
- Osallistumisaste: (Painoarvo 27 %)
- Omatoiminnan osuus: (Painoarvo 7 %)
- Varhaiskasvatuksen asiakasmäärän osuus väestöstä: (Painoarvo 19 %)
- Ansiotulot asukasta kohden: (Painoarvo 14 %)

Kaikki edellä mainitut muuttajat selittivät kustannuseroja tilastollisesti merkitsevästi ja merkitsevyys säilyi hyvänä eri muuttujakombinaatioilla. Prosentit kuvaavat muuttujalle laskettua painoarvoa.

Regressioanalyysi ei kerro syytä siihen, miksi jokin asia selittää kuntien välisiä kustannuseroja. Muuttujien yhteyttä kustannuksiin voidaan kuitenkin arvioida toimialatuntemuksen avulla.

Taloudellinen huoltosuhde ja osallistumisaste saivat laskelmissa korkeimman painoarvon ja ne siis selittävät mallissa testatuista muuttujista parhaiten kustannuseroja. Molempien muuttujien tulkinta tuntuu hyvin intuitiiviselta. Kun selitettävässä muuttujassa kustannukset on suhteutettu ikäluokkaan, liittyy ainakin osallistumisaste suoraan palvelun käyttöön ja siten kustannuksiin. Taloudellinen huoltosuhde liittyy myös ainakin palveluiden käyttöön, mutta mahdollisesti myös johonkin muuhun selittävään tekijään. Molemmat muuttajat pidettiin mallissa, koska ne eivät korreloineet vahvasti keskenään (0.18).

Maanteiden pituus näytti selittävän kustannuseroja ja sen voidaan olettaa liittyvän palveluverkkokysymyksiin. Pitkä maantieverkosto kuvaa laajalle levittäytynyttä asukasverkostoa paremmin kuin kunnan pinta-ala, joka ei saanut tilastollista merkitsevyyttä. Maantiet mahdollistavat laajasti asumisen ja maanteiden ympärille muodostuu taajamia. Laajalle levinnyttä palveluverkkoa on vaikeampi järjestää tehokkaasti, mikä näkyy kustannuksissa.

Kuntalisän vaikutus kustannuksiin tuntuu lähes itsestäänselvältä tulokselta. Kuntalisän merkitys on tässä kohden kuitenkin monitulkintaisempi. Kuntalisän maksamisella tavoitellaan yleensä alhaisempaa osallistumisastetta, jolloin kuntalisän aiheuttamien lisäkustannusten pitäisi kompensoitua alhaisempina oman toiminnan kustannuksina. Kyseiset muuttajat eivät kuitenkaan korreloineet keskenään (0.11) ja myös kuntalisä sekä osallistumisaste päätettiin pitää mallissa. Kuntalisää maksaneilla kunnilla kustannukset olivat lasta kohden keskimäärin noin 400 euroa korkeammat.

Oman toiminnan osuus selittää myös kustannuseroja ja muuttuja valittiin osaksi SAKU-mallia. Oman toiminnan osuus sai regressiossa negatiivisen kertoimen eli ulkoistaminen näyttäisi korottavan kustannuksia. Tuloksista ei voida kuitenkaan



suoraan päätellä ulkoistamisen olevan kalliimpaa. Yksi selitys ilmiölle voisi olla se, että ulkoistamisella on nouseva rajakustannus. Aluksi ulkoistetaan yksinkertaista päiväkotitoimintaa ja kuntiin jää vaikeammat tapaukset, kuten erityisopetus ja muu kalliimpi päivähoito. Sitten kun ulkoistetaan enemmän, alkaa vaativampaa toimintaa siirtymään myös ulkoistamisen piiriin. SAKU-mallin kannalta tärkeä havainto on kuitenkin se, että oman toiminnan osuudella on yhteys kustannuksiin ja samankaltaisia kuntia valittaessa muuttuja on syytä pitää mallissa.

Varhaiskasvatuksen skaalaetuja testattiin väkimäärään suhteutetulla asiakasosuudella sekä asiakasosuuden muutoksella. Skaalaetuja ei voitu testata absoluuttisella asiakasmäärällä koska se korreloisi liikaa regressioajoissa olevan väkiluvun kanssa. Sekä asiakasosuus että asiakasosuuden muutos saivat negatiivisen regressiokertoimen. Tulosten perusteella varhaiskasvatuksessa näyttäisi siten olevan skaalaetuja, mikä selittyy sillä, että kun verrataan väestöltään, osallistumisasteeltaan ja palveluverkoltaan (sekä muut selittävät muuttujat) samankaltaisia kuntia, näkyy asiakasosuuden kasvu laskevin kustannuksina. SAKU-malliin valittiin näistä kahdesta muuttujasta asiakasmäärän osuus väestöstä, koska se sai suuremman painoarvon isomman regressiokertoimen ansiosta.

Kunnan asukkaiden ansiotulot asukasta kohden on toistaiseksi viimeisin SAKU-malliin valittu muuttuja. Ansiotulojen hyvä selitysaste ja positiivinen regressiokerroin antavat viitteitä siitä, että hyvän taloudellisen tilanteen kunnissa varhaiskasvatuksen palveluihin panostetaan, mikä näkyy korkeampina kustannuksina. Korkean tulotason kunnissa ei useinkaan ole maan korkein kunnallisten sote-palveluiden kysyntä, joka taas antaa kunnalle taloudellista liikkumavaraa panostaa esimerkiksi opetuksen palveluihin. Ansiotulojen pitämistä SAKU-mallissa tulisi kuitenkin pohtia jatkokehittämisen yhteydessä, koska kyseinen muuttuja korreloi jossain määrin taloudellisen huoltosuhteen kanssa (-0.64).

## Teema 2: Peruopetus

Perusopetuksen kustannuserojen regressioestimoinneissa selitetään perusopetuksen käyttökustannuksia per oppilas. Kustannuksiin sisältyy koko perusopetuksen kustannukset sisältäen myös pidennetyn oppivelvollisuuden oppilaiden opetuksen kustannukset.

Liikkeelle lähdettiin yli 30 selittävästä muuttujasta, joista useat selittivät samoja olosuhdetekijöitä, mutta hieman eri kanteilta. Mallin selkeyden kannalta selittävien muuttujien määrä täytyi rajata maltilliseksi. Näistä poimittiin tilastollisesti merkitsevimmät ja jätettiin pois liiallista kovarianssia aiheuttaneet muuttajat (varianssin inflaatiotekijä yli 10). Lisäksi poistettiin joitakin hankalasti opetuskustannuksiin yhdistettäviä muuttujia, vaikka lineaarinen korrelaatio oli merkittävä (esimerkiksi ajoneuvokanta).

Perusopetuksen regressioajoihin sisällytettiin aluksi myös päätösperäisiä muuttujia, kuten ryhmäkoko, kuljetuskustannukset ja ateriakustannukset. Päätösperäiset muuttajat päätettiin kuitenkin poistaa regressiomallista, koska mallin tarkoitus on selittää olosuhteita. Toisaalta huomattiin, että useat malliin päätyneet olosuhdemuuttajat selittivät sekä ryhmäkokoja että kuljetuskustannuksia. Esimerkiksi taajama-aste mittaa epäsuorasti kouluverkon kokoa ja selittää siten kuljetuskustannuksia. Toisaalta ryhmäkokoon ja kuljetuskustannuksiin liittyy päätösperäisyyttä, joten niitä on mielekkäämpää vertailla SAKU-mallin tulosten perusteella siten, ettei niitä kuitenkaan sisällytetä itse malliin.

Perusopetuksen kustannuseroja testattiin yli 30 muuttujalla:

keski_ikä	valtuusto_vas
väkiluku	yksinhuoltaja_tuki
väkiluku_muutos_prosentti	k_keskiaste20_osuus
ruotsi_osuus	k_korkeaaste20_osuus
vieraskieliset_osuus	koulutustasomittain
taajama_aste	oppilaat_11ov_osuus
pinta_ala	po_oppilaat_osuus
kehyskunta_ind	po_oppilaat_muutos_prosentti
saaristo_ind	huostassa_7_12
saaristo_osa_ind	huostassa_13_15
ajoneuvokanta_as	oppilaat_per_peruskoulu
maantie_pituus	ryhmäkoko_po
ansiotulo_as	po_erityinen_tuki_osuus
verotulot_as	po_tehostettu_tuki_osuus
vos_suhde	po_valmistava_osuus
huoltosuhde_taloudellinen	kust_kuljetus_po
työttömyysaste	kust_ruokailu_po

Muuttujien valintaan käytettiin pienimmän neliösumman menetelmää selittämään perusopetuksen käyttökustannuksia per oppilas. Regressiomallissa pidettiin aluksi mukana myös päätösperäiset muuttujat: ryhmäkoko, kuljetuskustannukset sekä ruokailukustannukset. Olosuhde- ja päätösperäiset muuttujat selittivät yhdessä suurimman osan perusopetuksen käyttökustannuksista (korjattu selitysaste 70,2 %).

Regressioajo olosuhde- ja päätösperäisillä muuttujilla on nähtävissä seuraavissa taulukoissa:

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change
						F Change	df1	df2	
1	,839 <sup>a</sup>	,704	,702	1243,080397	,704	384,525	9	1455	,000

a. Predictors: (Constant), Oppilasmäärä, ruotsi\_osuus, po\_oppilaat\_muutos\_prosentti, ryhmäkoko\_po, ansiotulo\_as, kust\_ruokailu\_po, keski\_ikä, kust\_kuljetus\_po, taajama\_aste

b. Dependent Variable: kust\_po\_op\_uusi

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	6288,282	642,181		9,792	<,001		
	ryhmäkoko_po	-335,156	17,439	-,360	-19,218	<,001	,581	1,722
	kust_kuljetus_po	1,906	,112	,437	16,950	<,001	,307	3,261
	kust_ruokailu_po	3,117	,211	,269	14,754	<,001	,613	1,632
	keski_ikä	62,898	10,938	,123	5,750	<,001	,446	2,244
	taajama_aste	16,860	3,112	,144	5,418	<,001	,289	3,460
	ansiotulo_as	,061	,015	,071	3,953	<,001	,625	1,600
	po_oppilaat_muutos_prosentti	-33,772	5,779	-,086	-5,844	<,001	,929	1,076
	ruotsi_osuus	8,147	1,983	,061	4,108	<,001	,916	1,091
	Oppilasmäärä	,054	,010	,094	5,554	<,001	,706	1,417

a. Dependent Variable: kust\_po\_op\_uusi

Varsinaista SAKU-mallia varten päätösperäisistä muuttujista luovuttiin, koska mallin tarkoitus on löytää olosuhteiltaan samankaltaisia kuntia. Päätösperäiset muuttujat kuitenkin esitetään SAKU-mallin tulosteessa.

Lopullinen regressioajo pelkillä olosuhdemuuttujilla on nähtävissä seuraavissa taulukoissa:

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics			Sig. F Change	
					R Square Change	F Change	df1		df2
1	,685 <sup>a</sup>	,469	,467	1662,958671	,469	214,796	6	1458	<,001

a. Predictors: (Constant), Oppilasmäärä, ruotsi\_osuus, po\_oppilaat\_muutos\_prosentti, ansiotulo\_as, keski\_ikä, taajama\_aste

b. Dependent Variable: kust\_po\_op\_uusi

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1101,196	728,182		-1,512	,131		
	keski_ikä	258,335	12,146	,505	21,269	<,001	,647	1,546
	taajama_aste	-40,112	3,278	-,342	-12,239	<,001	,466	2,145
	ansiotulo_as	,080	,021	,093	3,873	<,001	,627	1,594
	po_oppilaat_muutos_prosentti	-30,762	7,679	-,079	-4,006	<,001	,941	1,062
	ruotsi_osuus	18,839	2,611	,142	7,215	<,001	,946	1,057
	Oppilasmäärä	,062	,013	,108	4,764	<,001	,713	1,403

a. Dependent Variable: kust\_po\_op\_uusi

Regressiotulosteista nähdään, että selitysaste tippui huomattavasti ensimmäisestä mallista (korjattu selitysaste 46,7 %), eli päätösperäiset muuttujat kattavat merkittävän osan opetuskustannuksista. Samalla olosuhdemuuttujien regressiokertoimet kasvoivat. Nyt esimerkiksi taajama-asteen vaikutus vaikuttaa loogiselta, eli korkeampi taajama-aste laskee perusopetuksen käyttökustannuksia, kun kuljetuskustannukset eivät ole mukana mallissa. Lähes kaikissa mal-leissa kunnan keski-ikä on ollut merkittävä vaikutus perusopetuksen kustan-nuksiin.

Regressiomallissa on joitakin puutteita. Yksikkökustannukset eivät täytä ho-moskedastisuuden oletusta, koska kustannuksien varianssi kasvaa kustannus-ten kasvaessa. Sama ilmiö näkyy myös residuaalien normaalijakautuneisuu-nessa, eli nollan lähellä olevien arvojen määrä painottuu. Päätimme kuitenkin pitäytyä pienimmän neliösumman menetelmässä, koska heteroskedastisuuden määrä on rajallinen ja emme halunneet häivyttää suuria kustannuksia aiheutta-vien muuttujien painoja. Kokeilimme myös painotettua regressiomallia, mutta malli toimi huonommin korkeiden kustannusten kunnille.

Lopulliseen perusopetuksen SAKU-malliin valittiin seuraavat selittävät muuttajat:

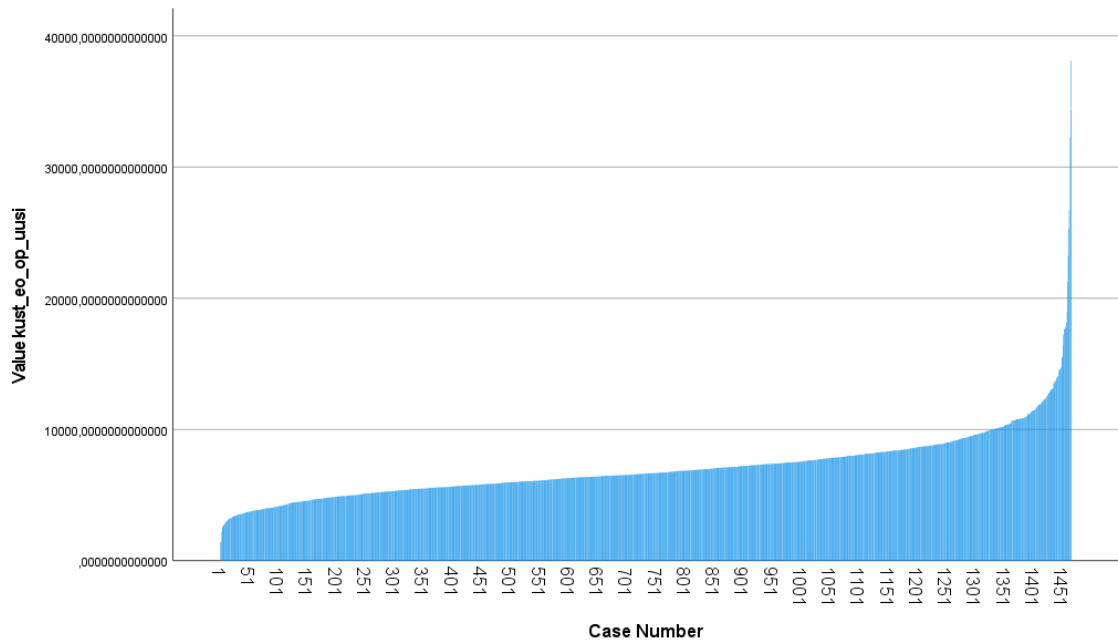
- Kunnan keski-ikä: painoarvo 30 %
- Taajama-aste: painoarvo 14 %
- Oppilasmäärän muutos (keskiarvo 2015–2019): painoarvo 24 %
- Ansiotulot asukasta kohden: painoarvo 11 %
- Ruotsinkielisten osuus väestöstä: painoarvo 8 %
- Oppilasmäärä: painoarvo 13 %

Painoarvot laskettiin kertomalla muuttujan regressiokerroin vaihteluvälin pituudella. Muuttujakohtainen tulos suhteutettiin kaikkien malliin valittujen muuttujien yhteenlaskettuun tulokseen ja näin saatiin selville kunkin muuttujan prosentuaalinen painoarvo. Edellä mainittu laskentatoimenpide toteutettiin sen vuoksi, että muuttajat eivät ole samalla asteikolla.

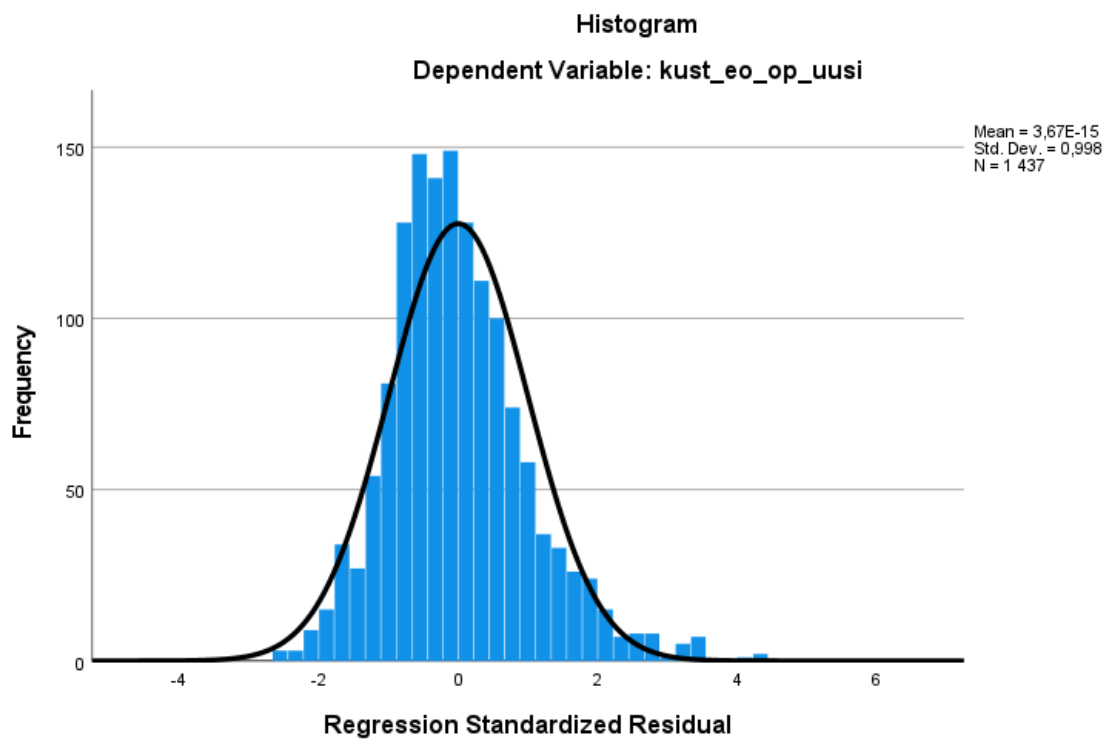
### **Teema 3: Esiopetus**

Teimme lineaarisen regressiomallin, jossa selitimme esiopetuksen käyttökustannuksia per oppilas vuosina 2015–2019. Koronavuosi 2020 jätettiin tarkoituksella pois analyysistä. Lähdimme liikkeelle noin 25 selittävästä muuttujasta, joista karsimme liiallista kovarianssia aiheuttaneet muuttujat. Tämän jälkeen tiiputimme muuttujia merkitsevyysjärjestyksessä SAKU-malliin soveltuvaan noin seitsemään muuttujaan.

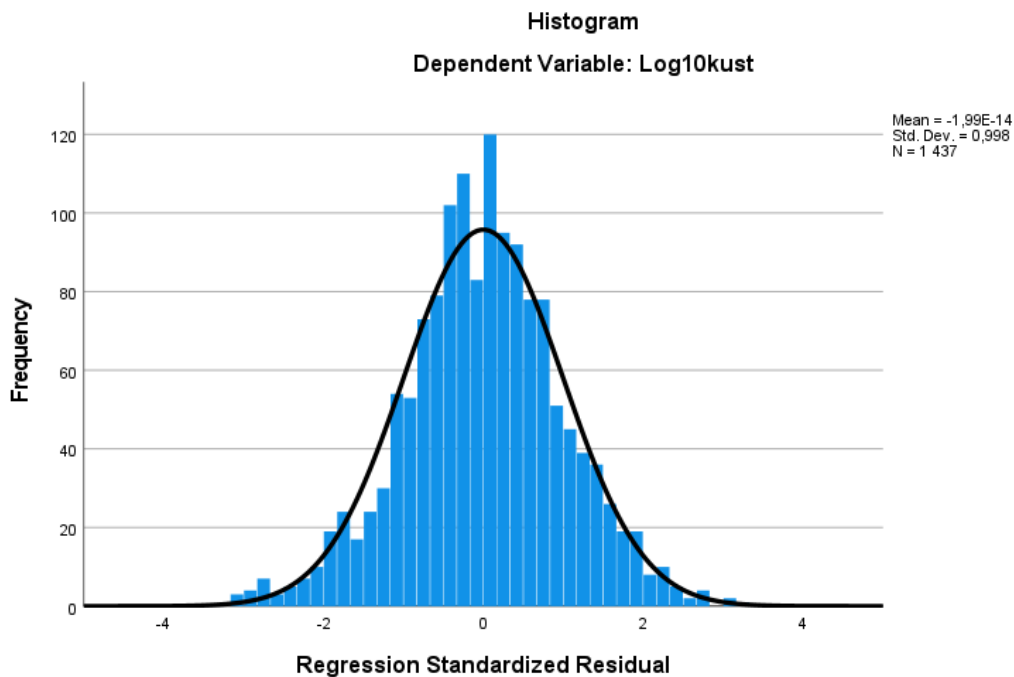
Havaintojen karsimiseen käytimme analyttisen ja stepwise-menetelmän yhdistelmää. Yksikkökustannusten ääripäissä oli todella poikkeuksellisia havaintoja, jotka poistimme virheellisinä. Kustannusten jakautuminen on nähtävissä kuvassa 1. Residuaalit eivät olleet täysin normaalijakautuneita eivätkä läpäisseet Kolmogorov-Smirnovin testiä, joten teimme lisäksi logaritmisen mallin. Pidimme myös alkuperäisen mallin tulosten tulkinnan selkeyden vuoksi. Normaalijakautuneisuus on nähtävissä kuvissa 2 ja 3.



Kuva 1. Esiopetusten käyttökustannusten jakautuminen Manner-Suomen kunnissa 2015-2019 €/opp



Kuva 2. Residuaalien normaalijakautuneisuus



Kuva 3. Residuaalien normaalijakautuneisuus logaritmissen muunnoksen jälkeen

Mallin selitysaste pelkillä olosuhdemuuttujilla jäi 30 prosentin tasolle (alkuperäinen malli 30,6 % ja logaritminen malli 30,2 %), joten päätimme olla toteuttamatta varsinaista esiopetuksen SAKU-mallia. Kuntien vertailusta olosuhdemuuttujien perusteella ei ole suurta hyötyä, mikäli olosuhdemuuttujat eivät selitä esiopetuksen kustannuksia riittävän vahvasti. Malli päätösperäisillä ja olosuhdemuuttujilla selitti 56,4 % esiopetuksen kustannuksista.

Regressiotulosteet ovat kuitenkin nähtävissä taulukoissa 1 ja 2. Selittäviksi muuttujiksi valikoituivat keski-ikä, pinta-ala, muuttovoitto, oppilaiden muutosprosentti, valtionosuuksien suhde verorahoitukseen, ruotsinkielisten osuus ja erityisen tuen oppilaiden osuus. Logaritmissen mallin parhaat muuttujat erosivat hieman alkuperäisestä mallista.

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,556 <sup>a</sup>	,309	,306	1694,186329

a. Predictors: (Constant), eo\_erityinen\_tuki\_osuus, vos\_suhde, ruotsi\_osuus, eo\_oppilaat\_muutos\_prosentti, pinta\_ala, muuttovoitto, keski\_ikä

b. Dependent Variable: kust\_eo\_op\_uusi

Taulukko 1. Mallin yhteenveto

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1612,138	503,019		-3,205	,001		
	keski_ikä	126,499	12,715	,276	9,949	<,001	,628	1,592
	pinta_ala	,206	,029	,160	7,102	<,001	,953	1,049
	muuttovoitto	,500	,144	,080	3,479	<,001	,908	1,101
	eo_oppilaat_muutos_prosentti	3,323	,892	,083	3,725	<,001	,967	1,034
	vos_suhde	40,532	4,285	,265	9,458	<,001	,615	1,627
	ruotsi_osuus	12,034	2,717	,100	4,428	<,001	,955	1,047
	eo_erityinen_tuki_osuus	10,744	1,428	,173	7,522	<,001	,917	1,090

a. Dependent Variable: kust\_eo\_op\_uusi

Taulukko 2. Regressiokertoimet ja kovarianssi

Esiopetuksen käyttökustannukset aiheutuvat pääosin muista tekijöistä, kuin tässä analyysissä löydetystä olosuhdemuuttujista. Yksi selitys tälle voi olla opetuskustannusten suuri osuus kokonaiskustannuksista ja näiden kustannusten suuri vaihtelu. Kuntaa kuvaavat olosuhdemuuttujat eivät pystyneet selittämään opetuskustannusten vaihtelua, eikä saatavilla ollut laadukasta opetuskustannuksia selittävää muuttujaa.

Tämän lisäksi uskomme ongelmiin esiopetuksen kustannusten vyörytyksissä. Esiopetus toteutetaan varhaiskasvatuksen tai peruskoulun alaisuudessa, ja vaikuttaa siltä, että kulut eivät ole riittävän tarkasti vyörytetty eri toiminnoille. Tämä näkyy liian halpojen kustannuksien kuntien kohdalla. Raportoidulla kustannustasolla tuskin pystyisi järjestämään esiopetusta, eli kustannukset sekoituvat varhaiskasvatuksen tai perusopetuksen kustannuksiin.