LUT-kauppakorkeakoulu

Kauppatieteiden kandidaatintutkielma

Liiketoiminta-analytiikka

Sään vaikutus kysyntään: Tilastollinen analyysi suomalaisesta autopesulasta

Effect of weather on demand: Statistical analysis of customer groups in the Finnish car wash

8.05.2020

Tekijä: Amira Achek

Ohjaaja: Pontus Huotari

TIIVISTELMÄ

Tekijä:	Amira Achek
Tutkielman nimi:	Sään vaikutus kysyntään: Tilastollinen analyysi suomalaisesta autopesulasta
Akateeminen yksikkö:	LUT-kauppakorkeakoulu
Koulutusohjelma:	Kauppatieteet, Liiketoiminta-analytiikka
Ohjaaja:	Pontus Huotari
Hakusanat:	Asiakaskäyttäytyminen, sääolosuhteet, asiakasryhmät

Tämän tutkimuksen tarkoituksena oli selvittää, miten sää, päivät ja saman rakennuksen asiakasmäärät vaikuttavat autopesulan pesumääriin. Lisäksi tavoitteena oli selvittää, onko vaikutus erilaista eri asiakasryhmien välillä. Asiakasryhmät jaettiin tutkimuksessa kanta-asiakkaisiin, jotka maksavat kuukausimaksun pesulan palveluista, sekä satunnaisiin asiakkaisiin.

Tutkimuksen aineisto koostuu yhden autopesulan pesutiedoista, kauppakeskuksen kävijätiedoista sekä Ilmatieteenlaitoksen antamista säätiedoista. Yhteensä pesuista on 3895 havaintoa, kauppakeskuksen kävijätiedoista 365 havaintoa ja säätiedoista 365 sekä 8760 havaintoa. Havaintoaineistot ovat kerätty vuonna 2019, ja tutkimus käsittelee aineistoa aikasarjana ajalta 1.1.2019-31.12.2019. Tutkimus toteutettiin tilastollisena ja asiakasryhmien eroja vertailtiin laadullisesti.

Tutkimustulokset osoittavat, että sääolosuhteista suhteellinen kosteus ja lämpötila vaikuttavat kaikkiin asiakkaisiin laskien kysyntää. Satunnaisiin asiakkaisiin vaikutti myös viikonloppu, sademäärä sekä lumimäärä, mutta kanta-asiakkaisiin ei. Laadullisesti tarkasteltuna päädyttiin lopputulokseen, että kanta-asiakkaat eivät ole yhtä herkkiä sääolosuhteille kuin satunnaiset asiakkaat. Tutkimuksen valossa kohdeyrityksen tulisi pyrkiä lisäämään kanta-asiakkaidensa määrää, koska sää ei vaikuta heihin yhtä negatiivisesti.

ABSTRACT

Author:	Amira Achek
Title:	Effect of weather on demand: Statistical analysis of customer groups in the Finnish car wash
School:	LUT School of Business and Management
Degree programme:	Business Administration, Business Analytics
Supervisor:	Pontus Huotari
Keywords:	Customer behavior, Weather conditions

The purpose of this study was to find out how the weather, the days, and the number of clients in the same building affect the car wash demand. In addition, the aim was to determine whether the effect differs between different customer groups. In the study, customer groups were divided into regular and occasional customers.

The survey data consists of information of washed cars, shopping center visitor information, and weather information provided by the Finnish Meteorological Institute. The data of washed cars consists of 3895 observations, the shopping center visitor data consists of 365 observations and the weather data consists 365 and 8760 observations. The observations were collected in 2019 and the study deals with the data as a time series from 1.1.2019 to 31.12.2019. The study was conducted statistically and the differences between customer groups were qualitatively compared.

Research results show that relative humidity and temperature affect all customers by reducing demand. Occasional customers were also affected by the weekend, rainfall and snowfall, but regular customers weren't. In qualitative terms, it was concluded that loyal customers are not as sensitive to weather conditions as occasional customers. Conclusion of this study is that the target company should strive to increase the number of its regular customers, since they are not as negatively affected by the weather.

SISÄLLYSLUETTELO

LIITTEET

KUVALUETTELO

TAULUKKOLUETTELO

1. Johdanto

Sää on osa päivittäisiä tekemisiämme ja kulutuspäätöksiämme. Joskus sää koskettaa mielialaamme, kun taas toisinaan se vaikuttaa suoraan tehtävään asiaan. Sen voima ei katso yrityksen kokoa, tulosta, brändiä tai menestystä, mutta toimialaa kylläkin.

Säätutkimus on etenkin sähköntuotannon puolella vanha aihe, mutta ekonometriassa suhteellisen tuore, eikä siitä löydy paljon aiempaa tutkimusmateriaalia (Israelsson, Tammet 2001). Sään on todettu vaikuttavan merkittävästi joidenkin yritysten liiketoimintaan, mutta säähän itsessään ei voi kummemmin vaikuttaa (Buchheim, Kolaska 2017). Sen takia se on toisaalta loistava elementti selittämään tapahtumia. Selvitettäväksi enää jää, miten sen vaikutuksen alla olisi parasta harjoittaa liiketoimintaa.

Aiemmat tutkimukset säästä käsittelevät suurimmaksi osaksi kaikkia asiakkaita samalla tavalla. Sen takia tämän tutkimuksen tarkoituksena onkin syventää tietoa kuluttajien käyttäytymisestä sään mukaan, ja jakaa asiakkaat kanta-asiakkuuden perusteella kahteen eri ryhmään Moon, Kang et al. (2018) tapaan, jotta kohdeyritys voisi hahmottaa paremmin asiakasryhmiensä piirteet ja käyttäytymismallit. Kun saadaan selville, miten sää vaikuttaa asiakasryhmiin, voidaan tehdä päätöksiä esimerkiksi markkinoinnin sekä asiakashankinnan suhteen. (Schla- ger, de Bellis et al. 2020) Tutkimus tullaan suorittamaan autopesulan näkökulmasta, mutta tavoitteena on, että tutkimuksen tuloksia voidaan soveltaa muihinkin samankaltaisiin yrityksiin.

1.1 Työn tausta

Suomessa autopesuloiden määrä on jatkuvassa kasvussa, ja niiden hinnoittelutapa on kokemassa muutoksen. Autopesulat keräävät itselleen asiakkaita kanta-asiakasjärjestelmään, jossa asiakas maksaa kiinteän kuukausihinnan, jolla saa pestä autoaan rajattoman määrän. Kuukausihinnat vaihtelevat 20-70 euron välillä. Normaalisti yhtä autoa pestään Ylen uutisen mukaan

3-4 kertaa vuodessa, mutta kanta-asiakkaat pesettävät noin 3-4 kertaa kuukaudessa. Tutkimuksen kohdeyrityksen tapauksessa luku on kanta-asiakkaiden keskuudessa 3,2 pesukertaa kuukaudessa. Uutisen mukaan pesuloiden liikevaihdosta 10-15% koostuu kanta-asiakkaista. (Matson-Mäkelä Kirsi) Kohdeyrityksen tapauksessa luku on 12,9%, mutta kanta-asiakkaiden pesujen osuus koko pesumäärästä on 23%.

Tämä tutkimus on tehty pienelle suomalaiselle autopesulalle. Autopesula sijaitsee vilkkaan kauppakeskuksen yhteydessä, missä on yksi läpiajettava pesuautomaatti. Tutkimuksen inspiraationa toimi johdon omat pohdinnat säästä, ja sen vaikutuksesta pesumääriin. Yritys on suhteellisen tuore, joten tarkastelun alle voitiin ottaa vain vuosi 2019. Valitettavasti tämä tarkoittaa myös sitä, ettemme voi verrata nykytilannetta aiempaan. Lisäksi uuden yrityksen liikevaihto on kasvussa koko ajan, mikä osaltaan saattaa vaikuttaa tutkimuksen tuloksiin.

1.2 Tavoitteet, rajaukset ja tutkimuskysymykset

Rajallisten resurssien takia tutkimus rajataan koskemaan vain tätä pesulayritystä, sekä yrityksen ja Ilmatieteenlaitoksen antamia tietoja, jotka käsitellään tarkemmin myöhemmin. Tietoihin kuitenkin lukeutuu pesumäärät, ohjelmat, hinnat, kauppakeskuksen kävijätiedot sekä havaintoasemalta saadut säätiedot. Pesulayrittäjän mukaan lähellä ei ole juurikaan kilpailijoita, joten niihin ei tämän vuoksi oteta sen enempää tutkimuksessa kantaa. Pesujen kustannuksia ei ole paljastettu, mutta niiden on kuvailtu pysyvän samana, joten myöskään niitä ei käsitellä.

Tutkimus keskittyy kohdeyritykseltä saadun datan käsittelyyn suhteessa vallinneisiin sääolosuhteisiin. Yritykset ovat jo pitkään keränneet dataa, ja sen hyödyntäminen sekä tutkiminen on ollut runsaassa kasvussa viimevuosina. Pienillä yrityksillä datan käsittely on jäänyt kuitenkin vähäiseksi niukkojen resurssien takia. Tästä syystä myös sään vaikutuksen tutkiminen näinkin rajatulle alalle on erittäin pientä.

Aiempiin tutkimuksiin, kuten Zwebner, Lee, Glodberg (2014) ja Hong, Sun (2012), verrattuna tässä tutkimuksessa pyritään selvittämään myös asiakasryhmien eroavaisuuksia. Työn tarkoituksena on löytää tekijät, jotka vaikuttavat autopesulan pesumääriin sekä selvittää onko

kanta-asiakkaiden ja satunnaisten asiakkaiden käyttäytymisessä merkittäviä eroja. Tutkimuksessa keskitytään yhteen suurempaan päätutkimuskysymykseen:

Miten sää vaikuttaa pesumääriin eri asiakasryhmien kohdalla?

Laajan tutkimuskysymyksen takia, jaetaan se pienempiin alatutkimuskysymyksiin. Niiden avulla pyrimme muodostamaan vastauksen päätutkimuskysymykseen. Alakysymyksiksi valittiin seuraavat:

1. *Miten sää, kauppakeskuksen asiakkaiden määrä sekä päivät vaikuttavat pesumääriin?*
2. *Onko eri asiakasryhmien käytöksessä huomatavissa eroavaisuuksia?*

Ensimmäisen alatutkimuskysymyksen avulla selvitetään, vaikuttavatko sää tai kauppakeskuksen asiakkaiden määrä merkittävästi yrityksen pesumääriin, ja jos, niin miten ne vaikuttavat. Tämä tieto vaikuttaa merkittävästi siihen, onko seuraavaa alatutkimuskysymystä järkevä lähteä edes tutkimaan. Toisen alatutkimuskysymyksen avulla selvitetään, onko kanta-asiakkaiden ja satunnaisten asiakkaiden käyttäytymisessä eroa. Ensimmäistä alatutkimuskysymystä lähdetään selvittämään kvantitatiivisesti, ja toista kysymystä selvitetään laadullisella tarkastelulla. Tarkoituksena on erotella sään vaikutusta asiakasryhmien käytökseen, ja luoda sen perusteella johtopäätöksiä.

Tavoitteena on, että tutkimustuloksia voidaan soveltaa myös muille aloille, joiden toimintaan sää vaikuttaa. Soveltamistapauksessa yritysten kanta-asiakasjärjestelmän tulee olla mieluusti kuukausihinnoitteluperusteinen, eikä esimerkiksi pelkkä kanta-asiakaskortti. Jos yrityksellä ei ole kanta-asiakkuusjärjestelmää, voidaan tutkimustuloksia käyttää myös tukemaan päätöstä, tulisiko yrityksen hankkia tällainen järjestelmä, ja millä perustein. Mainittakoon mahdollisista toimialoista esimerkiksi kuntosalit, ratsastuskoulut, ulkoliikuntalajien tarjoajat sekä parturikampaamot, joihin tutkimustuloksia voitaisiin soveltaa.

2. Teoreettinen tausta ja käsitteet

Työhön liittyvä kirjallisuus muodostuu pääosin aiemmasta tutkimuksesta koskien sään vaikutusta kuluttajien toimintatapoihin. Autopesuloista ei löydy aiempaa tutkimusta aiheeseen liittyen, mutta saman suuntaisia tutkimustuloksia on käytetty tukemaan myös tätä työtä. Lisäksi teoriaosuus esittelee lyhyesti aikasarja-analyysia, ja sen erityispiirteitä, jotka tulee ottaa huomioon työtä tehdessä.

2.1 Kuluttajan käyttäytyminen sään mukaan

Sää on psykologinen tekijä, joka vaikuttaa kuluttajien ostopäätöksiin (Hong, Sun 2012, Busse, Pope et al. 2015, Zwebner, Lee et al. 2014, de Bellis, Schulte-Mecklenbeck et al. 2018). Vuonna 2010 suoritetun tutkimuksen mukaan auringonpaisteella on negatiivinen vaikutus kuluttajan maksuhalukkuuteen. (Murray, Di Muro et al. 2010) Kuitenkin valaistumisasteen (de Bellis, Schulte-Mecklenbeck et al. 2018) ja lämpötilan vaikutusta tulisi vielä tutkia lisää (Zwebner, Lee et al. 2014).

Sään vaikutusta ruokakaupoissa käyntiin, sekä ostoskorin suuruuteen on tutkittu jakamalla kuluttajat kahteen ryhmään: säännöllisesti ja epäsäännöllisesti käyviin. Kylmällä säällä ostoskori pieneni molempien tapauksessa, mutta vain säännöllisten kävijöiden käytös muuttui niin, etteivät he välttämättä menneet kauppaan lainkaan. Säännöllisesti kaupassa käyvillä ei ollut välitöntä tarvetta käydä kaupassa, joten he siirsivät kauppapäivää paremmalle päivälle. (Moon, Kang et al. 2018). Myös joukkoliikenteen kuluttajien käytöstä ja sen muutosta on tutkittu sään mukaan. Tuloksina on ollut matkustusmatkan pidentyminen ja säännöllistyminen. Kuitenkin sateisina päivinä matkustusmäärä laskee hieman. (Hofmann, O'Mahony 2005) Elokuvalippujen ostoon vaikuttaa myös sää, riippumatta oikean näytösajan säästä (Buchheim, Kolaska 2017).

Tuoreempi tutkimus sai selville, että mielikuva sääolosuhteista vaikuttaa pääasiassa kuluttajien ostopäätökseen, toisin kuin itse sääolosuhteet. Sää lisää niiden tuotteiden arvotusta, jotka

yhdistetään positiivisesti vallitsevaan säätilaan. Toisin sanoen, jos kuluttaja saadaan ajattelemaan haluttua säätilaa ennen ostotilannetta, oston tekeminen on todennäköisempää. (Schlager, de Bellis et al. 2020)

Schlager de Bellis et al (2020, s.34-36) ovat löytäneet neljä keinoa, miten yritykset voivat hyödyntää säätä. Ensimmäisenä on valikoiva säätietojen näyttäminen asiakkaille. Esimerkiksi rantalomien myyjät voivat näyttää aurinkoisia sääennusteita verkkosivuillaan vieraileville. Toisena tuotteita tulisi näyttää vallitsevan sään mukaisesti. Kolmantena mahdollisuutena on luoda mainoksia, joissa toivottu sääilmiö esiintyy. Lopuksi hyödyntämisvaihtoehtona on hinnan sääntely dynaamisesti sään mukaan. CocaCola otti jo vuonna 2000 käyttöön dynaamisen hinnan sääntelyn juoma-automaateissa (King III, Narayandas 2000) . Sään hyödyntämisessä on kuitenkin ehtona, että sillä täytyy olla havaittua vaikutusta kuluttajien ostopäätökseen.

Myös osakemarkkinoiden sijoittamista ja riskinottoa on pyritty selittämään sään mukaan. Vuonna 2003 tutkittiin kausittaisen mielialahäiriön vaikutusta osakemarkkinoiden tuottoihin. Tutkimuksessa todettiin häiriön vaikutuksen olevan suuri. (Kamstra, Kramer et al. 2003, s. 18-20) Cao ja Wei (2004) tutkivat myös osakemarkkinoiden muutosta perustuen sään psykologisiin vaikutuksiin. He perustivat tutkimuksensa sijoittajien aggressiivisuuden muutoksiin sään mukaan. Hypoteesina oli, että matala lämpötila vaikuttaa positiivisesti tuottoihin aggressiivisen riskien oton vuoksi. Kaksi aiempaa tutkimusta ovat hieman ristiriidassa keskenään, sillä kylmä sää johtaa usein myös pimeyteen, ja pimeys kausittaiseen mielialahäiriöön. (Cao, Wei 2005) Uudemman tutkimuksen mukaan sään psykologista vaikutusta ja osakemarkkinoiden tuottoa ei kuitenkaan voi yhdistää keskenään. Tutkimuksen mukaan ilmiötä tulisi kuvata sekä tutkia tarkemmin kuin vain tuottoon perustuen. (Jacobsen, Marquering 2008)

Sään vaikutuksen tutkiminen ympärillä tapahtuviin asioihin ei siis ole niin yksiselitteistä. Lisäksi sää vaikuttaa jokaiseen tapahtumaan hieman eri tavalla. Jokaisesta aihepiiristä on kuitenkin havaittavissa pääpiirteinä, että ei-toivottu säätila vaikuttaa negatiivisesti kysyntään sekä mielialaan. Ei-toivotulla säätilalla tarkoitetaan tiettyä toimintaa haittaavaa säätilaa. Se ei kuitenkaan tarkoita tietyn tapaista säätilaa jokaisen toiminnan kohdalla.

2.3 Muodostetut hypoteesit

Tutkimuksessa on muodostettu teorian valossa muutamia hypoteeseja, joita tutkimuksessa tullaan testaamaan. Tutkimuksen ensimmäisenä hypoteesina on, että sateinen sää vaikuttaa pesumääriin negatiivisesti. Sateisella säällä kura roiskuu ajaessa, joten asiakkaiden ei uskota käyttävän autoa pesussa, jos se tulee heti likaiseksi. Lisäksi aiempi tutkimusmateriaali puoltaa negatiivista vaikutusta maksuhalukkuuteen ei-toivotulla säällä (Moon, Kang et al. 2018, Buchheim, Kolaska 2017). Myös kohdeyrityksen mielestä sateisella säällä on vähemmän pesijöitä.

Oletamme myös, että asiakkaiden autolla ajomäärä, auton merkki, vuosimalli sekä ajomaasto vaikuttavat pesumääriin. Nämä seikat ovat kuitenkin rajattava tutkimuksen ulkopuolelle. Tästä saadaan myös johdettua toinen hypoteesi, että sää ja kauppakeskuksen vierailijamäärät eivät voi selittää täydellisesti pesumääriä, mikä tulee ottaa myös huomioon johtopäätöksiä laatiessa. Tätä emme voi todistaa todeksi, mutta huomioimalla toisen hypoteesimme, varaudumme siihen, etteivät tulokset ole välttämättä täydellisiä.

3. Tutkimusmenetelmä ja aineisto

Ensimmäisen alatutkimuskysymyksen ratkaisemiseen tullaan hyödyntämään määrällistä eli kvantitatiivista tutkimusta, koska tutkimusaineisto on suuri ja tilastollinen tutkimus on kaikista tehokkain sekä tarkin tapa sellaisen aineiston käsittelyyn. Kvantitatiivista tutkimusta varten tiedot voidaan kerätä erilaisista muiden keräämistä tilastoista, rekistereistä tai tietokannoista. Tiedot voidaan myös kerätä itse. (Heikkilä 2014) Tässä tutkimusaineistossa on käytetty yrityksen itse keräämää dataa tapahtuneista myynneistä. Lisäksi tutkimukseen on käytetty kauppakeskuksesta saatua vierailijadataa sekä ilmatieteen laitoksen erään havaintoaseman tietoja mahdollisimman lähellä kauppakeskusta. Saatuja tuloksia verrataan laadullisesti keskenään.

Tutkimus toteutetaan hyödyntäen ohjelmointikieltä R, joka on tehokas ympäristö tilastolliseen laskentaan ja grafiikan tuottamiseen. Suurin osa tutkimuksen kuvista on myös tehty R:llä, mutta jotkut kuviot sekä taulukot ovat tehty hyödyntäen Microsoft Exceliä tai MATLAB R2019b:tä.

3.1 Aikasarja-analyysi

Tässä luvussa käydään läpi muutamia aikasarjoihin kuuluvia käsitteitä, joita tullaan myöhemmin käyttämään analyysissa. Selvitetään, mitä aikasarjat ovat ja millaisia ominaisuuksia niillä kuuluu olla, jotta niitä voi käyttää analyysin työkaluina.

Aikasarja on havaintojono y_t, joka saa arvoja ajanhetkinä t. Diskreeteissä tapauksissa havaintoväli on jatkuvasti lähes yhtä pitkä. Pituus voi vaihdella minuuteista vuosiin. Tässä tutkielmassa havaintoväli on päivä. Aikasarjassa voi olla erotettavissa trendi, kausivaihtelu tai satunnainen vaihtelu. Joskus niitä voi olla useampikin. (Brockwell, Davis 2016, s.1-6)

Jotta aikasarjaa voidaan mallintaa autoregressiivisen liukuvan keskiarvon avulla, aikasarjan tulee olla stationaarinen, eli siitä ei saa olla erotettavissa trendiä tai kausivaihtelua. Tähän on kehitetty monia ratkaisuja, kuten residuaalien erotuksen ottaminen tai trendin sisällyttäminen aikasarja-aineistoon. (Palma 2016, s.53)

Stationaarisuus tulee aina testata. Yleisin käytetty testausmenetelmä on Dickey-Fuller ja Augmented Dickey-Fuller testi (ADF). Näistä jälkimmäinen ottaa huomioon aikasarjan autokorrelaation, minkä takia tulemme käyttämään sitä varmistaaksemme, etteivät residuaalit korreloi. (Cheung, Lai 1995)

Autokorrelaatio tarkoittaa, että havainnot korreloivat keskenään, mikä tarkoittaa, että myös residuaalit korreloivat keskenään. Autokorrelaatio tulee ottaa huomioon mallissa viivästyksillä, joita voi olla k:n verran. Muuttuja k mittaa, havaintojen etäisyydellä olevien havaintojen

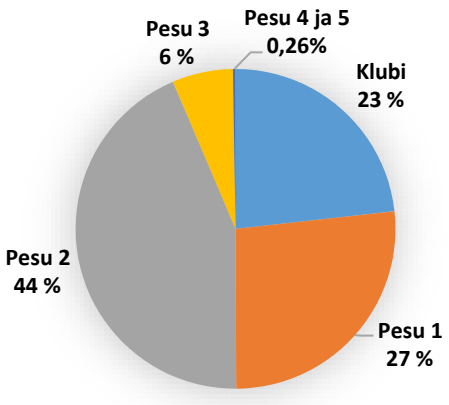
välistä riippuvuutta. (Palma 2016, s.91-93) Viivästyksien sijaan voidaan käyttää myös HAC korjausestimaattoria, joka ottaa huomioon sekä heteroskedastisuuden että autokorrelaation ja laskee korjatut keskivirheet. (Hill, Griffiths et al. 2012, s.357-358)

3.2 Pesuaineisto

Pesudata sisältää kaksi aineistoa. *Pesudata1* sisältää 3895 havaintoa pesuista aikavälillä 1.1.2019-31.12.2019. Havaintoihin kuuluu päivämäärä, kellonaika, pesu, sekä hinta. *Pesudata2* sisältää 365 havaintoa, joihin kuuluu päivämäärä, päivän pesujen määrä sekä myynti.

Pesuja on viittä eri tyyppiä. Niistä kuitenkin kaksi ensimmäistä on selvästi suositumpia. Pesu 1 kestää noin 12 minuuttia ja on hinnaltaan 28,90€, pesu 2 kestää noin 8 minuuttia ja on hinnaltaan 18,90€, pesu 3 on harjaton pesu ja hinnaltaan 17,90€, pesu 4 on suksiboksipesu ja hinnaltaan 24,90€ ja pesu 5 on avolavapesu ja hinnaltaan 24,90€.

Lisäksi yrityksellä on pesuklubi, joka toimii yrityksen kanta-asiakasjärjestelmänä. Puhuttaessa klubista tarkoitetaan siis yrityksen kanta-asiakkaita. Kanta-asiakkaat saavat itse valita pesuohjelman, mutta yrityksen mukaan he ottavat luonnollisesti yleensä pesun 1. Alla olevasta piirakkakuviosta näkee, että lähes neljäsosa yrityksen pesumääristä koostuu pesuklubin jäsenien pesuista.

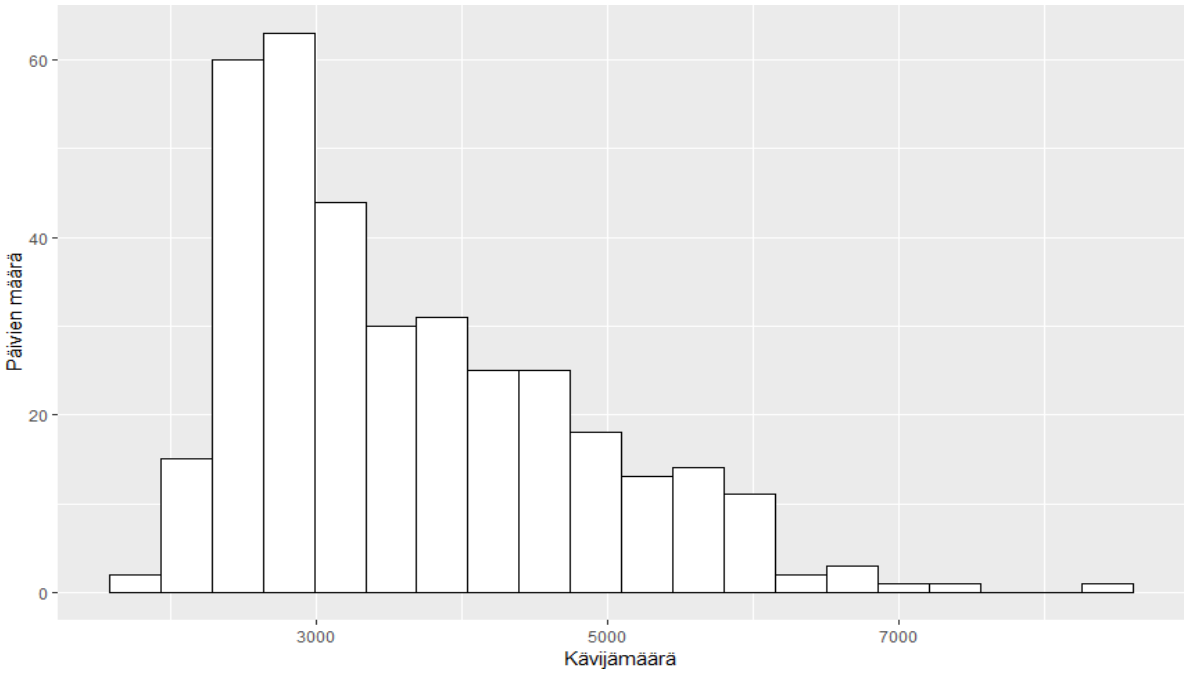


Kuva 1. Yrityksen tarjoamien pesujen jakautuminen käytön mukaan.

Yllä olevasta kuvasta 1 näkee, miten yrityksen tarjoamien pesuohjelmien kysyntä jakautuu. Huomioitavaa on, että pesujen 3, 4 ja 5 käyttömäärä on todella pieni.

3.3 Vierailija-aineisto

Vierailija-aineisto sisältää kauppakeskuksen kävijätietoja aikavälillä 1.1.2019-31.12.2019. Aineisto pitää sisällään päivämäärän, päivän, viikonpäivän, kolme keskeistä kauppakeskuksen paikkaa, parkkipaikan, vierailijat yhteensä sekä mahdolliset erikoispäivät.



Kuva 2. Kauppakeskuksen kävijämäärät vuonna 2019.

Yllä olevasta kuvasta 2 on tarkoitus saada kuva siitä, miten kauppakeskuksen kävijämäärät ovat jakautuneet. Huomataan, että jakauma ei aivan noudata normaalijakaumaa. Vierailijoiden määrä vaihtelee 1872 ja 8551 asiakkaan välillä. Keskiarvoltaan vierailijoita käy päivässä 3630 ja mediaani on lähes sama 3299.

Tutkimuksen toinen kiinnostava paikka *vierailija-aineistosta* on parkkihalli, koska yritys sijaitsee sen sisällä. Parkkihallin kävijät saavat arvoja 224 ja 4425 välillä. Keskiarvoltaan vierailijoita on 952 ja mediaani on 868.

3.4 Säähavainnot

Säähavainnot sisältävät myös kaksi eri aineistoa: vuorokautiset havainnot sekä hetkelliset havainnot. Kutsutaan näitä tiedostoja nimillä *vuorokausiaineisto* ja *hetkellinen aineisto*. Molempien havaintojen aikaväli on 1.1.2019-31.12.2019. *Vuorokausiaineisto* sisältää päivämäärän, sateen määrän (mm), lumimäärän (cm), lämpötilan, ylimmän lämpötilan sekä alimman lämpötilan.

Sademäärä saa arvoja väliltä 0 ja 42,2. Sademäärän keskiarvo on 1,52 ja mediaani vain 0,3. Suuri ero keskiarvon ja mediaanin välillä voi olla merkki outlier havainnoista, joihin puututaan seuraavassa luvussa. Lumimäärä saa arvoja välillä 0 ja 62, ja sen määrän keskiarvo on 8,76. Lämpötila sen sijaan saa havaintoja välillä -19,7 ja 23,1. Lämpötilan keskiarvo on 5,5 astetta ja mediaani 4,4 astetta.

Hetkellinen aineisto sisältää 8760 havaintoa, joihin sisältyvät vuosi, kuukausi, päivä, kellonaika, suhteellinen ilmankosteus sekä ilman lämpötila. Havaintoväli on aineistossa tunti. Työtä helpottaakseni vuodesta, kuukaudesta ja päivästä on yhdistetty myös päivämäärä. Suhteellinen kosteus saa arvoja välillä 15 ja 100. Sen keskiarvo on 83,7 ja mediaani on 92.

3.5. Muuttujat

Tässä osiossa on tarkoituksena käydä läpi muuttujat, joita tutkimuksessa tullaan käyttämään ja jotka luotiin tutkimusta varten. Lisäksi käydään läpi havaintoaineistoihin tehdyt rajaukset, kuten outliereiden poistot.

Pesuaineistosta 1 luotiin summamuuttuja *summapesu*, joka on pesujen määrä päivässä. Käytämme tätä regression selitettävänä muuttujana. *Summapesu* sisältää kaikkien maksettujen pesujen määrät. Tarkastelusta on siis jätetty pois aiemmin mainitun pesuklubin kävijät. Muuttuja ei ole normaalijakautunut (Liite 1). Muuttuja *summapesu* on vielä jaettu pesujen välillä muuttujiksi *summap1, summap2, summap3, summap4 ja summap5*. Vähennettyämme *pesuaineiston 2* pesujen määrät aineistosta summapesu, saimme selville pesuklubin käyttäjien määrän. Kutsutaan tätä muuttujaa nimellä *klubi. Klubista* luotiin myös logaritmiset havainnot *logklubi*.

Muuttujasta *summapesu* poistettiin outlier-havaintoina pesut, joiden päiväkohtainen määrä menee yli 58. Yritys kuvaili, että heillä on joskus tempauksia, joiden takia pesujen määrä saattaa olla todella iso, joten tutkimuksen näkökulmasta havainnot olivat tarpeellista poistaa. Poistettuja havaintoja oli 2. Lisäksi yritys oli muutaman päivän tarkoituksella kiinni, jolloin ei taas ole yhtään pesua. Olemme poistaneet myös nämä havainnot havaintoaineistosta, etteivät ne vääristä tuloksia. Yhteensä aineistoon jäi siis 353 havaintoa.

Vierailija-aineistosta muuttujaa *erikoispäivä* muokattiin niin, että puuttuvat arvot saivat merkityksen *normaali päivä,* ja ne päivät, joista jokin arvo löytyi, saivat merkityksen *erikoispäivä.* Näin ollen saamme jaoteltua erikoispäivien merkityksen pesuihin. Jos dataa olisi monelta vuodelta, voisi erikoispäiviä tarkastella erikseen, mutta kun dataa on ainoastaan pelkästään vuoden ajalta verran, on aiheellista tarkastella erikoispäiviä yhtenä käsitteenä.

Hetkelliset aineistosta luotiin *avg_kosteus*, joka on yhden päivän suhteellisen kosteuden keskiarvo. Liitteestä 2 voi huomata, että *avg_kosteus* ei ole normaalijakautunut.

Vuorokausiaineiston muuttujasta *sade_mm,* joka kuvaa sateen määrää päivässä millimetreinä, poistettiin myös yksi outlier muuttuja. Poistettu muuttuja oli arvoltaan yli 40, kun kaikki muut olivat arvoltaan alle 25. Poisto tehtiin, siksi ettei yksittäinen suuri havainto vaikuttaisi liikaa tuloksiin.

Kaiken kaikkiaan tutkimuksessa käytettävä aikasarja-aineisto koostuu seuraavista havainnoista: *vuosi, kuukausi, päivä, päivämäärä, summapesu, summap1, summap2, summap3,*

summap4, summap5, klubi, logklubi, viikonpäivä, erikoispäivä, sade_mm, lumi_cm, lämpötila, avg_kosteus ja parkkipaikka.

4. Tulokset

Tässä tutkimuksessa käsitteiden välisiä yhteyksiä tutkitaan regressioanalyysillä. Pyrimme muodostamaan ratkaisun ensimmäiseen tutkimuskysymykseen. Tarkoituksena on löytää mahdollisimman hyvä yhtälö, joka selittää muuttujia *summapesu* ja *klubi.* Selitettävät muuttujat ovat luonteeltaan teoreettisesti jatkuvia, ja ne saavat ainoastaan positiivisia arvoja. Niin kuin uusien muuttujien luonnin yhteydessä todettiin, *summapesu* ja *klubi* eivät ole normaalijakautuneita, mutta pitkällä aikavälillä oletettavasti menevät sitä kohti. Kaikista sopivin tarkastelutapa on siis monimuuttujamenetelmänä lineaarinen regressio, jossa käytetään monen selittäjän malleja. *Summapesua* ja *klubia* tarkastellaan logaritmisina muuttujina kaikissa malleissa, koska suhdeasteikolla mitatut muuttujat vaihtelevat melko laajoissa rajoissa, ja oletamme että logaritminen muuttuja tuo täten tarkemmat tulokset.

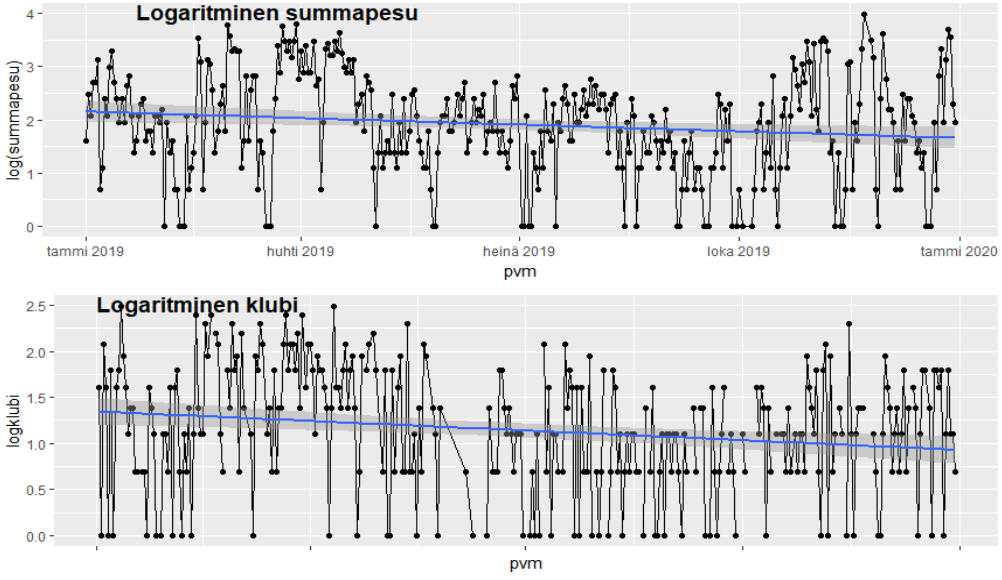
Regressioon on valittu seitsemän selittävää muuttujaa alussa muodostettujen hypoteesien mukaan: *erikoispäivä, viikonpäivä, avg_kosteus, lämpötila, sade_mm, lumi_cm* ja *parkkihalli.* Oletamme, että sää vaikuttaa pesuihin, koska ihmiset järkeilevät, milloin on hyvä aika pestä auto. Lisäksi oletamme, että pesuihin vaikuttaa viikonpäivä sekä erikoispäivä, sillä ihmisten vapaa-aika ja liikkuminen on erilaista päivän mukaan. Lopuksi oletamme, että parkkihallin kävijämäärä (*parkkihalli*) vaikuttaa pesumääriin, sillä pesuhalli on kauppakeskuksen parkkihallissa.

Parkkihallin kävijämäärät korreloivat voimakkaasti muiden vierailijatietojen kanssa. Siitä syystä regressiomalleihin otettiin *vierailija-aineistosta* vain parkkihallin kävijämäärät, koska sen ajateltiin olevan parhaiten yhteydessä pesumääriin.

4.1 Autokorrelaatio ja stationaarisuus

Koska kyseessä on aikasarja-aineisto, ensimmäisenä on hyvä tarkastaa sen stationaarisuus. Jotta pystyisimme huomioimaan mahdollisen autokorrelaation stationaarisuuden tarkastelussa, on luotu ensin mallit ja tutkittu niiden autokorrelaatiota. Autokorrelaatio on havaittu residuaalien välisenä korrelaationa, jossa nollahypoteesina toimii, että mallissa ei ole autokorrelaatiota. Liitteestä 3 näkee, että *summapesun* kohdalla p<0,05, joten autokorrelaatiota on. Lisäksi korrelogrammista voi havaita autokorrelaation olevan niin voimakasta, että viiveitä on haastava sisällyttää malliin. Liitteestä 4 sen sijaan huomaa, että *logklubin* tapauksessa p-arvo on 0,06, joka on vain hieman yli riskiarvon 0,05. Sen takia tarkastelemme vielä korrelogrammissa, onko havaittavissa viiveitä, joita tulisi ottaa malliin mukaan. Korrelogrammin mukaan (Liite 4) malliin tulisi ottaa yksi viive mukaan, mutta monen selittäjän regressiomallissa haasteena on oikean viiveen löytäminen.

Molempien regressiomallien tapauksessa viiveiden käyttö osoittautui haasteelliseksi, joten päädyimme käyttämään korjattuja kertoimia. Vaikka viiveitä ei voi sisällyttää malleihin, voimme ottaa ne kuitenkin huomioon aikasarjojen stationaarisuuden testaamisessa. Liitteiden 3 ja 4 korrelogrammeista voi nähdä, että tarkasteluun otetaan muuttujan *summapesu* tapauksessa kuusi viivettä ja *klubin* tapauksessa yksi.



Kuva 3. Aikasarjojen stationaarisuuden tarkastelua visuaalisesti.

Yllä olevasta kuvasta 3 nähdään, miten selitettävien aikasarjojen havainnot kohdistuvat. Sininen viiva on trendiviiva, ja sen ympärillä oleva harmaa alue sen keskivirhe. Yksinkertaisuuden vuoksi trendiviiva on ”pakotettu” lineaariseksi. Visuaalisesti tarkasteltuna aikasarja-aineisto näyttäisi olevan melko stationaarinen molempien muuttujien tapauksessa, vaikka trendiviiva meneekin hieman alaspäin. Aikasarjoissa ei ole myöskään havaittavissa selvää kausivaihtelua.

Tarkastetaan stationaarisuus vielä ADF-testillä. Nollahypoteesina on epästationaarisuus. Jokaisen testin p-arvo on kuitenkin 0.01, joten 5% riskitasolla nollahypoteesi hylätään (Liitteet 5 ja 6). Aikasarjat ovat siis stationaarisia ja voimme jatkaa niiden käsittelyä sellaisinaan.

4.2 Lineaarinen regressio muuttujalle summapesu

Tutkitaan ensin, millainen vaikutus säällä, viikonpäivällä tai kauppakeskuksen asiakkailla on pesumääriin päivässä. Selitettävänä muuttujana on *summapesu* ja selittävinä muuttujina *erikoispäivä, viikonpäivä, avg_kosteus, lämpötila, sade_mm, lumi_cm* ja *parkkihalli*. Aikasarjaan luotiin aikamuuttuja t, joka saa arvoja havaintojärjestyksen mukaan 1 – 353.

Taulukko 1. Ensimmäisen regressioanalyysin muuttujien kuvailevia tilastoja.

	n	avg	min	max
log(summapesu)	353	1,93	0	4,1
avg_kosteus	351	83,7	48,83	100
sade_mm	351	1,48	0	24
lumi_cm	353	8,89	0	62
lämpötila	353	5,51	-19,7	23,1
parkkihalli	353	943,1	224	3595

Yllä olevasta taulukosta 1 näkee käytettyjen muuttujien arvoja. Huomataan, että muuttujista vain lämpötila voi saada arvoja, jotka ovat alle 0. Suhteellisen kosteuden keskiarvo on melko korkea, eikä se havaintoaineistossa ole missään kohtaa 0, vaikka teoriassa se voisi olla myös 0. Lumimäärän keskiarvo taas on melko korkea suhteessa maksimiarvoon. Se ei kuitenkaan

ole poikkeava arvo, lumimäärä vain käyttäytyy todella erikoisesti (Liite 7). Muutamien outlier poistojen jälkeen on saatu havaintoaineisto, joka on kooltaan 353.

Taulukko 2. Ensimmäisen regressioanalyysin muuttujien korrelaatiomatriisi.

	log(summapesu)	avg_kosteus	sade_mm	lumi_mm	lampötila	parkkihalli
log(summapesu)	1,00					
avg_kosteus	-0,39	1,00				
sade_mm	-0,18	0,17	1,00			
lumi_mm	0,04	0,27	-0,02	1,00		
lampötila	-0,18	-0,50	0,00	-0,63	1,00	
parkkihalli	0,02	0,15	0,02	0,30	-0,23	1,00

Yllä olevasta korrelaatiomatriisista (Taulukko 2) voidaan todeta, että suhteellinen kosteus, lämpötila ja sademäärä korreloivat negatiivisesti *summapesun* kanssa. Kaikkein voimakkaimmin korreloi kuitenkin suhteellinen kosteus, sekin negatiivisesti. Parkkihalli ja lumimäärä korreloivat melko heikosti muuttujan *summapesu* kanssa, joten niillä ei välttämättä ole kovinkaan suurta merkitystä myöskään regressiossa, mutta niiden poisjättämisestä uskotaan olevan enemmän haittaa. Vahvimmin keskenään korreloivat lämpötila ja suhteellinen kosteus sekä lämpötila ja lumimäärä. Muuttujien välinen korrelaatio ei kuitenkaan ole niin voimakasta (yli 0,9), etteikö muuttujia voisi käyttää regressiossa. Tarkastamme kuitenkin vielä lopussa multikollineaarisuuden VIF-testillä. Korrealaatiomatriisin perusteella valitut muuttujat ovat sopivat selittämään muuttujaa *summapesu*, joten luodaan monen muuttujan lineaarinen regressiomalli *lm_pesut*:

$$\log(summapesu) = \beta_1 + \beta_2 avgkosteus + \beta_3 lämpötila + \beta_4 sade + \beta_5 lumi + \beta_6 parkkihalli + \beta_7 viikonpäivä + \beta_8 erikoispäivä \quad (1)$$

Estimoidessa mallia pienimmän neliösumman mukaan, heteroskedastisuus tulee usein ottaa huomioon. Estimointimenetelmä olettaa, että malli on homoskedastinen, eli residuaalien varianssi pysyy samana. Heteroskedastisuutta voi korjata esimerkiksi Whiten varianssikorjausestimaattorilla, painokertoimilla (WLS) tai poistamalla outlier havaintoja. (Chatterjee, Price 1977, s. 38) Liitteestä 8 voi huomata, että mallin *lm_pesut* residuaalien keskivirheet pysyvät melko samana, kun selitettävää muuttujaa *summapesu* käsitellään logaritmisena, toisin kuin

jos *summapesua* käsiteltäisiin normaalisti. Lisäksi heteroskedastisuus tarkastettiin vielä käyttäen Breusch Paganin testiä heteroskedastisuudelle. Testissä nollahypoteesina on homoskedastisuus, ja koska $p>0.05$, nollahypoteesi jää voimaan eli malli on homoskedastinen (Liite 9).

Koska mallissa huomattiin autokorrelaatiota, jota ei voida ottaa huomioon käyttämällä viiveitä, päädyttiin laskemaan oikeammat kertoimet Whiten varianssin - korjausestimaattorilla. Alla olevasta taulukosta 3 löytyy koonti estimoinnista käyttäen OLS: ia, sekä käyttämällä Whiten varianssin-korjausestimaattoria. Korjausestimaattorin käyttö ei kuitenkaan ole BLUE hengessä ”best”, ja se ei poista tosiasiaa siitä, että OLS on harhainen, eikä se täten sovellu parhaiten mallin määrittelyyn.

Taulukko 3. Koonti OLS:in tuloksista sekä Whiten varianssikorjausestimaattorin tuloksista mallille *lm_pesut*.

OLS	***Kulmakerroin***	***Std, Error***	***T-arvo***	***Prob > |t|***
Vakio	6,0761	0,4037	15,0500	**0,0000**
Normaali päivä	-0,0787	0,1599	-0,4920	0,6229
Lauantai	0,4688	0,1768	2,6510	**0,0084**
Maanantai	0,0112	0,1643	0,0680	0,9456
Perjantai	0,3831	0,1689	2,2680	**0,0240**
Sunnuntai	0,0764	0,1703	0,4490	0,6540
Tiistai	-0,0535	0,1680	-0,3180	0,7503
Torstai	0,0425	0,1640	0,2590	0,7958
avg_kosteus	-0,0427	0,0037	-11,6540	**0,0000**
lampötila	-0,0702	0,0074	-9,5410	**0,0000**
sade_mm	-0,0217	0,0108	-2,0140	**0,0449**
lumi_cm	-0,0093	0,0034	-2,7360	**0,0066**
parkkihalli	-0,0002	0,0001	-1,4800	0,1398
Residual se, 0,7889	Adjusted R-squared: 0,3557	F-statistic: 15,95	p-value: 0,000	
Whiten varianssikorjausestimaattori	***Kulmakerroin***	***Std, Error***	***T-arvo***	***Prob > |t|***
Vakio	6,1889	0,4250	14,5628	
Normaali päivä	-0,0806	0,1683	-0,4786	
Lauantai	0,4270	0,1861	2,2938	
Maanantai	-0,0506	0,1729	-0,2928	
Perjantai	0,3362	0,1778	1,8912	
Sunnuntai	0,0426	0,1792	0,2377	
Tiistai	-0,1182	0,1768	-0,6686	
Torstai	-0,0239	0,1727	-0,1385	
avg_kosteus	-0,0432	0,0039	-11,2109	
lampötila	-0,0705	0,0077	-9,0997	
sade_mm	-0,0220	0,0114	-1,9388	
lumi_cm	-0,0090	0,0036	-2,5217	
parkkihalli	-0,0002	0,0001	-1,4481	
Residual se, 0,8024				

Yllä olevasta taulukosta 3 näkee, että Whiten korjausestimaattorin tuottamien kertoimien keskivirheet ovat suurempia kuin alkuperäiset, joten alkuperäiset keskivirheet antavat kuvan tarkemmista kertoimista kuin mihin on aihetta. Selittävien muuttujien kulmakertoimien mer-

kit ovat odotetun mukaiset. Kulmakertoimien arvot eivät juurikaan muutu korjausestimaattoria käytettäessä, ja kertoimien merkit pysyvät samanlaisina. Moni muuttuja saa kuitenkin hieman suurempia arvoja kuin mitä OLS antoi. Koska malli ei muutu kovinkaan paljon, tarkastellaan hieman OLS mallin p-arvoja. Niistä voi huomata, että päivistä vain perjantailla ja lauantailla on tilastollisesti merkitsevä kerroin. Lisäksi säämuuttujat ovat 5% riskitasolla tilastollisesti merkitseviä tekijöitä pesumäärille. Tärkeimpänä huomataan, ettei parkkihallin kävijämäärällä tai erikoispäivillä ole mallissa tilastollisesti merkittävää vaikutusta pesumääriin.

Tarkastellaan nyt Whiten korjausestimaattorin antamia tuloksia huomioimatta sitä, oliko jokin muuttuja OLS:ssa tilastollisesti merkitsevä vai ei. Lauantaina, perjantaina ja sunnuntaina on enemmän pesuja verrattuna keskiviikkoon. Päivien suosiota tukee myös liite 10, jossa näkyy pesujen jakautuminen päivittäin. Muina päivinä taas kävijöitä on vähemmän, eli keskiviikko on arkipäivistä kiireisin pesupäivä. Normaalina päivänä on vähemmän pesijöitä verrattuna erikoispäiviin. Lisäksi kaikkien mallin sääolosuhteiden kasvaessa, pesumäärä vähenee. Kun haluamme tarkastella selittävien muuttujien vaikutusta oikeisiin pesumääriin, tulee regressiomallin tuloksia tarkastella seuraavasti:

$$summapesu = e^{\beta_1+\beta_2 avg-kosteus+\beta_3 lämpötila\ +\beta_4 sade+\beta_5 lumi+\beta_6 parkkihalli+\beta_7 viikonpäiväi+\beta_8 erikoispäivä} \quad (2)$$

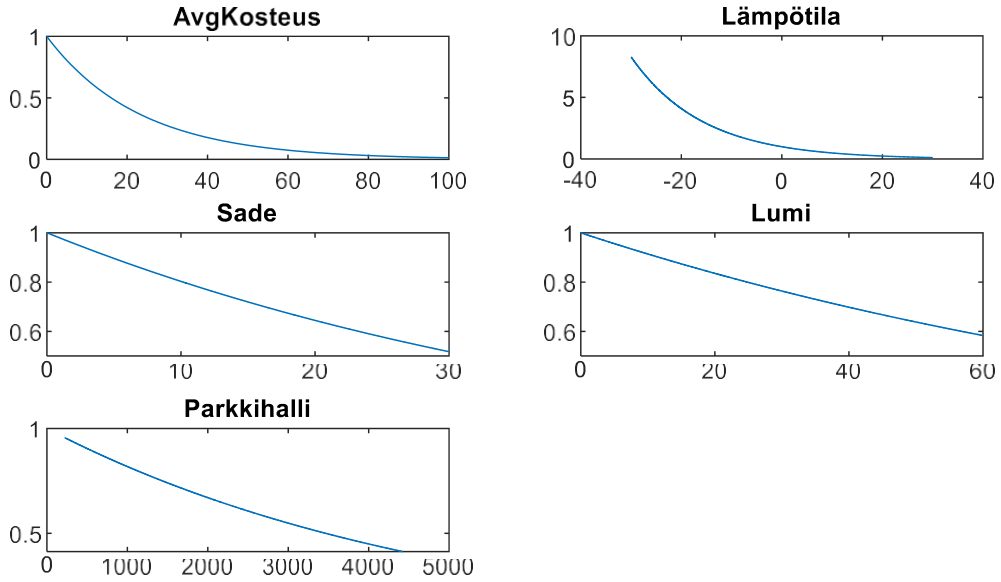
Yllä oleva kaava saadaan muokattua parempaan muotoon, jossa kertoimien yksittäisiä arvoja voidaan tarkastella paremmin:

$$summapesu = e^{\beta_1} * e^{\beta 2 avg-kosteus} * e^{\beta 3 lämpötila} * e^{\beta 4 sade} * e^{\beta 5 lum} * e^{\beta 6 parkkihalli} e^{\beta 7 viikonpäivä} * e^{\beta 8 erikoispäivä} \quad (3)$$

Huomioitavaa on, että selittävien muuttujien summaamisen sijaan niistä otetut eksponentit kuuluvat kertoa yhteen, jotta saadaan oikea pesumäärä. Nyt pesumäärä ei kasva, tai tässä tapauksessa vähene, lineaarisesti vaan eksponentiaalisesti, joten mallin tulkintakin muuttuu. Tulkittaessa mallia tulee muistaa, että pesumäärään vaikuttaa kaikkien havaintojen yhteisvaikutus eikä vain yksittäinen havainto. Lisäksi osa havainnoista sulkee toisensa pois. Esimerkiksi

ei voi olla tilannetta, jossa lämpötila on -30 ja sataisi vettä tai suhteellinen kosteus olisi 15 ja sataa vettä. Tarkastellaan kuitenkin aluksi vain yksittäisiä kertoimia.

Vakio saa nollatilanteessa todella ison arvon 487,3. Sitä kuitenkin kerrotaan selittävien tekijöiden vaikutuksella. Johtopäätöksenä voi kuitenkin jo vakion sekä residuaalien (liite. 9) tarkastelussa todeta, että mallin ääripäät voivat olla harhaisia. Suhteellinen kosteus voi saada teoriassa arvoja 0-100 välillä. Suhteellisen kosteuden muuttuessa vakiota tulee siis kertoa välillä 1 - 0.0133. Suhteellisen kosteuden vaikutus pesumääriin on todella suuri. Lämpötila saa isolla haarukalla arvoja noin välillä -30 - +30 astetta. Lämpötilan muuttuessa vakiota tulee siis kertoa edellä mainitun välin puitteissa 8.29 - 0.1206. Mallin mukaan matala lämpötila siis kasvattaa pesumäärää, mutta korkea lämpötila vähentää. Sademäärä millimetreinä vuorokaudessa saa *vuorokausiaineiston* mukaan arvoja nollan ja 30 millimetrin välillä. Tällöin sateen vaikutus vakioon on välillä 1 - 0.52, eli rankalla sateella pesumäärä vähenee noin puolella. Lumen määrä senttimetreinä voi saada arvoja *vuorokausiaineiston* mukaan nollasta 60:een. Lumen vaikutus vakioon on välillä 1 - 0.58, eli myös suuressa lumisateessa pesumäärät putoavat yli puolella. Kauppakeskuksen parkkihallin kävijöitä oli suurella haarukalla 224 – 4425 kappaletta. Parkkihallin vaikutus vakioon on välillä 0.96 - 0.41, eli kävijämäärän kasvaessa pesumäärät vähenevät. Alla olevasta kuvasta 5 näkyy vielä, miten jokaisen selittävän muuttujan arvon vaihtelu vaikuttaa pesumäärään. Kuvasta 4 huomaa, että vaikutus ei ole lineaarista ja käyrä vaihtelee jokaisen arvon mukaan. Numeerisesti tarkastelluista vaikutuksista sekä kuvasta 4 voi siis päätellä, että suurin vaikutus pesuihin on suhteellisella kosteudella sekä lämpötilalla.

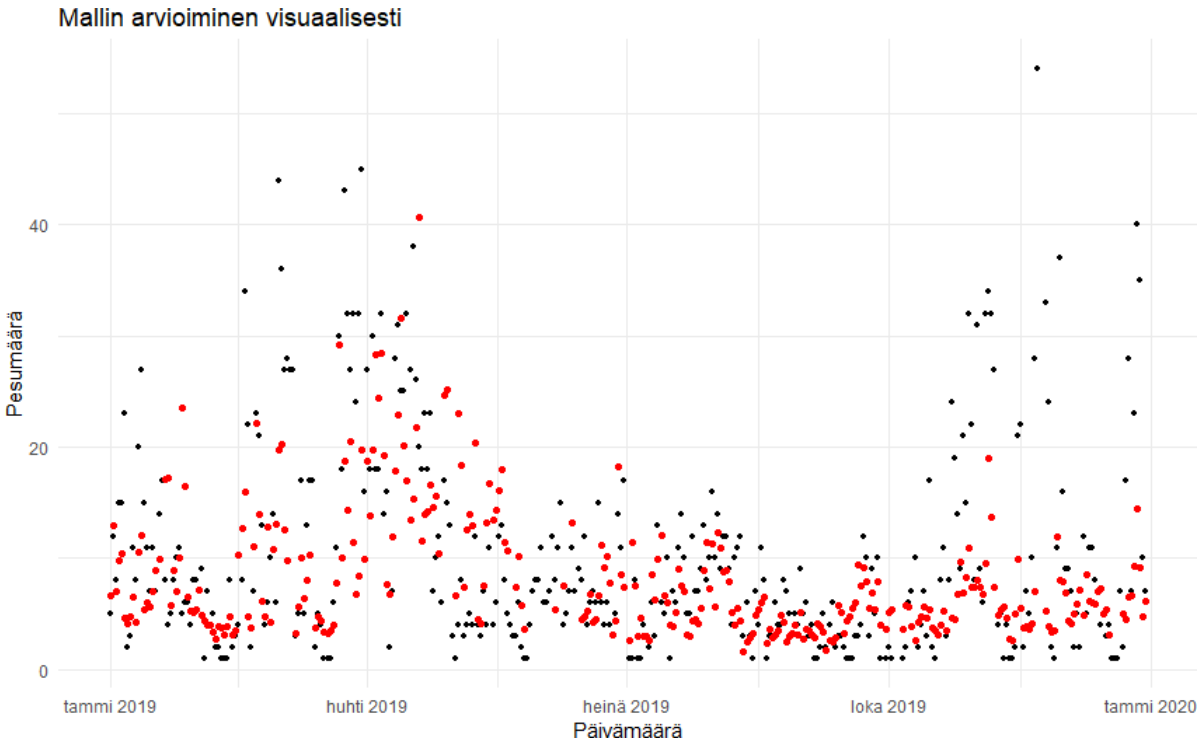


Kuva 4. Havainnollistava kuva, miten selittävien muuttujien arvojen vaihtelu (x) vaikuttaa satunnaisten asiakkaiden pesumäärään (y).

Haluamme vielä tarkastella, puuttuuko mallista jotain, tai onko sen funktiomuoto valittu väärin. Tähän käytämme RESET – testiä, joka luo selitettävästä muuttujasta *summapesu* polynomeja ja sovittaa niitä regressioyhtälöön. Jos tulos paranee, niin malli ei ole riittävän hyvä. Testin nollahypoteesi on, että mallissa ei ole puuttuvia arvoja ja sen funktiomuoto on oikea. ((Hill, Griffiths et al. 2012, s. 238-239) Testi on ainoastaan suuntaa antava ja sen tuloksia tulee tarkastella kokonaisuuden kannalta. RESET-testin mukaan mallimme nollahypoteesi jää voimaan ($p > 0,05$), eli mallistamme ei testin mukaan puutu mitään ja funktiomuoto on valittu oikein. (Liite 11) P-arvo on kuitenkin vain 0,06 eli mallia tulisi tarkastella kriittisesti.

Koska kyseessä on monen muuttujan regressioyhtälö, tarkistetaan vielä, ovatko selittävät muuttujat itsenäisiä vai riippuuko niiden arvot muista. Tähän käytämme VIF- testiä. Testi laskee regressiomallien kaikkien selittävien muuttujien variaatioinflaatiokertoimet. Nyrkkisääntönä on, että jos VIF- testin arvo on yli 1/(1-R2), jossa R2 on selityskerroin, niin selittävä muuttuja on riippuvainen jostain toisesta selittävästä muuttujasta. Tässä tapauksessa raja on

1.5521. VIF- testin arvoista *lämpötila*, *lumi_cm* ja *parkkihalli* ylittävät kyseisen arvon. Se ei kuitenkaan ylity niin paljon, että mallin voisi sanoa kärsivän multikollineaarisuudesta (yli 5-10), vaan kyseiset muuttujat eivät ole mallissa vain itsenäisiä. (Liite 12)



Kuva 5. *Lm_pesut* mallin laskemat pesumäärät (punainen) ja oikeat pesumäärät (musta).

Tarkastellaan vielä visuaalisesti, miten malli suoriutuu käytännössä ennustamaan vuoden 2019 pesumääriä. Yllä olevassa kuvassa 5 on esitetty oikeat pesumäärät mustalla ja mallin ennustamat pesumäärät punaisella. Kuvaa 5 katsomalla voisi todeta, että malli ennustaa melko hyvin pesumääriä normaalitilanteissa. Lopputalvesta malli ei onnistu ennustamaan pesuja niin hyvin, vaan ennusteet jäävät melko pieniin lukemiin.

4.2.1 Pesuohjelmakohtainen tarkastelu

Koska pesuohjelmista löytyy enemmänkin tietoa, syvennytään vielä lyhyesti siihen, miten ensimmäisessä regressiossa olevat muuttujat vaikuttavat eri pesuohjelmiin. Näin saamme tarkempaa tietoa sään ja muiden muuttujien vaikutuksesta. Tarkoituksena on tarkastella muuttuvatko kertoimet tai niiden merkitsevyys eri pesuohjelmien välillä.

Taulukko 4. Kuvailevia arvoja pesuohjelmien käytöstä.

	n	mean	min	max
summap1	343	3,825	0	28
summap2	343	6,23	0	33
summap3	343	0,881	0	6
summap4	343	0,068	0	1
summap5	343	0,043	0	1

Taulukosta 4 huomaa, että pesua 2 käytetään eniten. Taulukon 4 perusteella teemme myös johtopäätöksen, ettei muuttujia *summap4* ja *summap5* kannata käyttää lineaarisessa regressiossa, sillä ne saavat liian vähän nollasta poikkeavia arvoja. Jatkamme siis analyysin tekoa kolmella ensimmäisellä pesuohjelmalla. Käsittelemme näitä ilman logaritmimuunnosta, koska nolla-arvoja on niin paljon. Luodaan monen muuttujan lineaarinen regressiomalli *pesuohjelmat*, jossa on samat selittävät muuttujat kuin edellisessä regressiossa.

Liitteestä 13 voidaan todeta, että kyseessä on kaikkien muuttujien kohdalla stationaarinen aikasarja-aineisto. Lisäksi todetaan, että malli on heteroskedastinen, joten käsittelemme jälleen tuloksia Whiten varianssin-korjausestimaattorin avulla (Liite 14). Samalla pystytään ottamaan huomioon autokorrelaatio.

Taulukko 5. Pesuohjelmien kertoimien tarkastelua. Lihavoidut numerot saivat OLS:ia käyttäen tilastollisesti merkitsevän arvon.

	Summap1 kerroin	*Summap2* kerroin	*Summap3* kerroin
Vakio	**13,11450**	**25,4881**	**3,6474**
Lauantai	**2,03080**	1,297	**0,4116**
Maanantai	0,29040	-0,3847	-0,2361
Perjantai	**1,45020**	1,2091	0,0798
Sunnuntai	**0,99260**	-0,0275	-0,0129
Tiistai	0,24070	-0,4126	-0,2622
Torstai	0,39380	-0,8175	-0,1163
Normaali päivä	0,03050	-0,172	-0,2845
avg_kosteus	**-0,11250**	**-0,194**	**-0,0255**
lampötila	**-0,15670**	**-0,4233**	**-0,036**
sade_mm	**-0,09080**	-0,1073	-0,0075
lumi_mm	**-0,03050**	**-0,0511**	-0,0031
parkkihalli	0,00010	-0,0012	**-0,0003**

Yllä olevasta taulukosta 5 huomaa, että kaikkien pesuohjelmien tapauksessa lauantai vaikuttaa positiivisesti pesumääriin. Pesuohjelmien 1 ja 3 tapauksessa kerroin on ollut tilastollisesti merkitsevä. Pesuohjelman 1 tapauksessa myös sunnuntai ja perjantai vaikuttavat tilastollisesti merkitsevästi pesuihin. Viikonpäivät vaikuttavat siis voimakkaimmin pesuohjelmaan 1. Erikoispäivien kohdalla normaali päivä vaikuttaa negatiivisesti pesuihin 2 ja 3, mutta pesuun 1 positiivisesti.

Kaikkiin pesuohjelmiin sää vaikuttaa niitä laskien. Suhteellinen kosteus, lämpötila, sademäärä sekä lumimäärä vaikuttavat voimakkaimmin pesuun 2, sitten pesuun 1 ja lopuksi pesuun 3. Yllätyksellisesti pesuohjelmaan 3 vaikuttaa tilastollisesti merkitsevästi parkkihallissa olevien määrä. Kertoimien mukaan viikonpäivät vaikuttavat voimakkaimmin pesuohjelmaan 1 ja sää pesuohjelmaan 2.

4.3 Lineaarinen regressio muuttujalle logklubi

Seuraavaksi tutkitaan, miten viikonpäivä, kauppakeskuksen asiakkaiden määrä ja sää vaikuttavat pesuklubin pesumääriin päivässä. Selitettävänä muuttujana on *logklubi* ja selittävinä muuttujina samat kuin *lm_pesut* regressiossa, eli *erikoispäivä, viikonpäivä, avg_kosteus, lämpötila, sade_mm, lumi_cm* ja *parkkihalli*. Myös tähän aikasarjaan luotiin aikamuuttuja t, joka saa arvoja havaintojärjestyksen mukaan 1 – 325.

Taulukko 6. Toisen regressioanalyysin muuttujien kuvailevia tilastoja.

	n	avg	min	max
logklubi	301	1,15	0	2,48
avg_kosteus	325	83,4	48,83	100
sade_mm	323	1,51	0	24
lumi_cm	323	9,23	0	62
lämpötila	325	5,2	-19,7	23,1
parkkihalli	301	952	224	3595

Yllä olevasta taulukosta 6 näkee käytettyjen muuttujien arvoja. Muutamien outlier poistojen jälkeen sekä nolla-arvojen takia saadaan havaintoaineisto, joka on kooltaan 325. Arvot ovat lähes samat kuin Taulukossa 1, mutta hieman pienemmän havaintoaineiston takia ne poikkeavat toisistaan jonkin verran.

Taulukko 7. Korrelaatiomatriisi toisen regression muuttujista.

	logklubi	*avg_kosteus*	*sade_mm*	*lumi_mm*	*lampötila*	*parkkihalli*
logklubi	1,00					
avg_kosteus	-0,19	1,00				
sade_mm	-0,10	0,17	1,00			
lumi_mm	0,14	0,27	-0,02	1,00		
lampötila	-0,20	-0,50	0,01	-0,62	1,00	
parkkihalli	0,03	0,15	0,05	0,32	-0,25	1,00

Yllä olevasta korrelaatiomatriisista (Taulukko 7) voidaan todeta, että suhteellinen kosteus, lämpötila ja sademäärä korreloivat negatiivisesti *logklubin* kanssa. Nyt lämpötila korreloi kuitenkin kaikkein voimakkaimmin *logklubin* kanssa verrattuna Taulukkoon 2, ja lumimäärä korreloi yllättävän positiivisesti muuttujan *logklubi* kanssa. Vahvimmin keskenään korreloivat jälleen lämpötila ja suhteellinen kosteus sekä lämpötila ja lumimäärä, joiden takia tässäkin regressiossa tarkistetaan multikollineaarisuus VIF-testillä. Valitut muuttujat ovat korrelaatiomatriisin perusteella sopivat kuvaamaan muuttujaa *logklubi*, joten luodaan monen muuttujan lineaarinen regressiomalli *lm_klubi*:

$$\text{logklubi} = \beta_1 + \beta_2 avgkosteus + \beta_3 \text{lämpö}tila + \beta_4 sade + \beta_5 lumi + \beta_6 parkkihalli + \beta_7 viikonp\ddot{a}iv\ddot{a} + \beta_8 erikoisp\ddot{a}iv\ddot{a} \quad (4)$$

Tarkastellaan myös *lm_klubin* kohdalla ensin heteroskedastisuutta. Liitteestä 15 voi huomata, että mallin *lm_klubi* residuaalien keskivirheet pysyvät melko samana, kun selitettävää muuttujaa *klubi* käsitellään logaritmisena, toisin kuin jos *klubia* käsiteltäisiin normaalisti. Tarkastetaan heteroskedastisuus vielä käyttäen Breusch Paganin testiä heteroskedastisuudelle. Koska p>0.05, nollahypoteesi jää voimaan, eli malli on homoskedastinen (Liite 16).

Koska mallissa huomattiin autokorrelaatiota, jota ei voida ottaa huomioon käyttämällä viiveitä, päädyttiin laskemaan oikeammat kertoimet jälleen Whiten varianssin - korjausestimaattorilla. Alla olevasta taulukosta 8 löytyy koonti estimoinnista käyttäen OLS:ia sekä käyttämällä Whiten varianssin-korjausestimaattoria.

Taulukko 8. Koonti OLS:in tuloksista sekä Whiten varianssikorjausestimaattorin tuloksista mallille *lm_klubi*.

OLS	***Kulmakerroin***	***Std. Error***	***T-arvo***	***Prob > |t|***
Vakio	2,9020	0,3405	8,5240	**0,0000**
Normaali päivä	-0,1829	0,1351	-1,3540	0,1770
Lauantai	0,0218	0,1465	0,1490	0,8820
Maanantai	0,0952	0,1403	0,6790	0,4980
Perjantai	0,1528	0,1396	1,0950	0,2750
Sunnuntai	-0,0372	0,1457	-0,2550	0,7990
Tiistai	0,0925	0,1433	0,6460	0,5190
Torstai	0,1027	0,1367	0,7510	0,4530
avg_kosteus	-0,0174	0,0031	-5,6500	**0,0000**
lampötila	-0,0297	0,0061	-4,8490	**0,0000**
sade_mm	-0,0038	0,0089	-0,4270	0,6700
lumi_cm	0,0013	0,0028	0,4650	0,6420
parkkihalli	-0,0001	0,0001	-0,6890	0,4920
Residual se. 0,6312	Adjusted R-squared: 0,1249	F-statistic: 4,568	p-value: 0,000	
Whiten varianssikorjausestimaattori	***Kulmakerroin***	***Std.Error***	***T-arvo***	***Prob > |t|***
Vakio	3,4931	0,4232	232874,0000	
Normaali päivä	-0,2575	0,1399	-1,8398	
Lauantai	0,1312	0,1683	0,7795	
Maanantai	0,6040	0,1504	0,4017	
Perjantai	0,1929	0,1569	144638,0000	
Sunnuntai	-0,0152	0,1571	-0,0969	
Tiistai	0,0390	0,1540	0,2534	
Torstai	0,0965	0,1462	0,6598	
avg_kosteus	-0,0199	0,0033	-5,9696	
lampötila	-0,0254	0,0080	-3,1916	
sade_mm	-0,0025	0,0095	-0,2631	
lumi_cm	0,0009	0,0029	0,3090	
parkkihalli	-0,0001	0,0001	-1,7856	
Residual se. 0597				

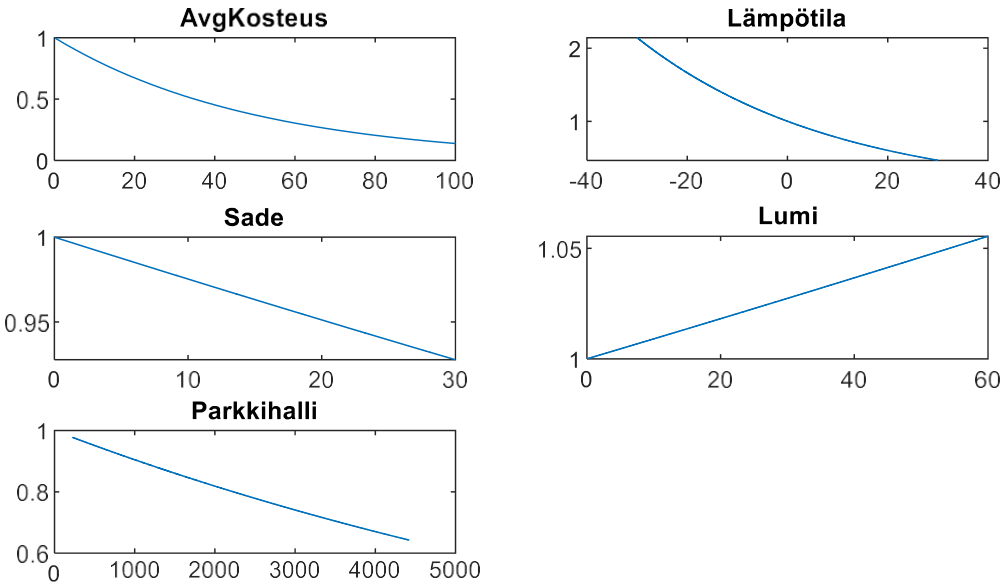
Yllä olevasta taulukosta 8 näkee, että Whiten korjausestimaattorin tuottamien kertoimien keskivirheet ovat kaikkien estimoitujen kertoimien tapauksessa hieman suurempia kuin OLS:issa, joten alkuperäiset keskivirheet antavat tässäkin tapauksessa kuvan tarkemmista kertoimista kuin mihin on aihetta. Estimoidut kertoimet muuttuvat hieman, mutta jokaisen etumerkki on odotettu ja pysyy samana molemmissa estimointimenetelmissä lukuun ottamatta

muuttujaa *lumi_cm*, joka saa positiivisen kertoimen. Koska malli ei juurikaan muutu, tarkastellaan hieman OLS mallin p-arvoja. Niistä voi huomata, ettei päivistä yhdelläkään ole tilastollisesti merkittävää arvoa verrattuna keskiviikkoon. Lisäksi säämuuttujista vain *avg_kosteus* ja *lämpötila* ovat 5% riskitasolla tilastollisesti merkitseviä tekijöitä pesumäärille.

Tarkastellaan nyt Whiten korjausestimaattorin antamia tuloksia. Kaikkina muina päivinä paitsi sunnuntaina on oletettavasti enemmän pesuja verrattuna keskiviikkoon. Toisaalta päivien suosion jakautumista ei tue liite 17, jossa näkyy pesujen jakautuminen päivittäin. Liitteen 17 mukaan tiistai olisi kaikkein hiljaisin päivä. Niin kuin OLS:sta huomattiin, pesupäivien kertoimet eivät ole tilastollisesti merkittäviä, ja niiden luottamusvälit menevät nollan molemmille puolille. Myös *normaalin_päivän, lumi_cm*, *sade_mm* sekä *parkkihallin* tapauksessa on näin. (Liite 18) Tarkastelemme kuitenkin niidenkin muuttujien kertoimien arvot ja vaikutuksen läpi. Kun haluamme tarkastella selittävien muuttujien vaikutusta oikeisiin pesumääriin, tulee regressiomallin tuloksia tarkastella seuraavasti:

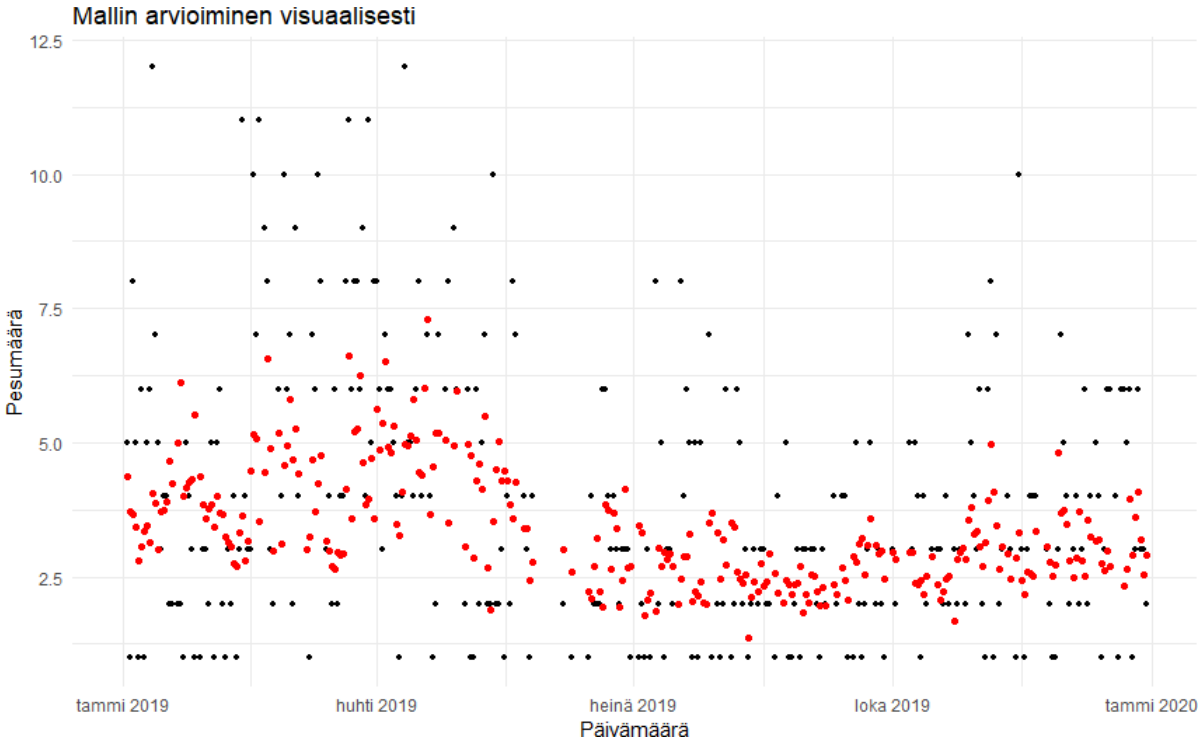
$$logklubi = e^{\beta 1} * e^{\beta 2 avg-kosteus} * e^{\beta 3 lämpötila} * e^{\beta 4 sade} * e^{\beta 5 lum} * e^{\beta 6 parkkihalli}$$
$$e^{\beta 7 viikonpäivä} * e^{\beta 8 erikoispäivä} \quad (5)$$

Verrattuna *lm_pesut* mallin vakioon, vakio saa nollatilanteessa melko pienen arvon 32,89. Käytetään kertoimien tulkintaan samoja määreitä kuin mallin *lm_pesut* tapauksessa. Suhteellisen kosteuden muuttuessa vakiota tulee siis kertoa välillä 1 – 0,1367. Suhteellisen kosteuden vaikutus pesumääriin on täten melko suuri, muttei läheskään niin suuri kuin mallin *lm_pesut* tapauksessa. Lämpötilan muuttuessa vakiota tulee kertoa edellä mainitun välin puitteissa 2,14 - 0,467. Lämpötilankaan tapauksessa muutos ei ole niin iso kuin mallissa *lm_pesu.* Mallin mukaan matala lämpötila kuitenkin edelleen kasvattaa pesumäärää, mutta korkea lämpötila vähentää. Sateen vaikutus vakioon on välillä 1 – 0,93, eli rankkasade ei paljoa klubilaisia haittaa. Lumen vaikutus vakioon on välillä 1 – 1,06, eli suuressa lumimäärässä klubilaiset lisäisivät auton pesua 6%. Positiiviseen kertoimeen voi osittain vaikuttaa myös lämpötila. Parkkihallin vaikutus vakioon on välillä 0.98 - 0.64, eli kävijämäärän kasvaessa pesumäärät vähenevät. Alla olevasta kuvasta 6 näkyy vielä, miten jokaisen selittävän muuttujan arvon vaihtelu vaikuttaa *klubin* pesumäärään. Kuvasta huomaa, että vaikutus ei ole tässäkään tapauksessa lineaarista ja se vaihtelee eri tavalla jokaisen arvon mukaan.



Kuva 6. Havainnollistava kuva, miten selittävien muuttujien arvojen vaihtelu (x) vaikuttaa kanta-asiakkaiden pesumäärään (y).

Tutkitaan jälleen mallin sopivuutta RESET-testillä, josko mallista puuttuu jotakin tai funktiomuoto olisi väärin. RESET-testin tulokseksi saadaan p>0.05, jonka seurauksena nollahypoteesi jää voimaan (Liite 19). Malli ei siis sisällä ylimääräisiä muuttujia tai siitä ei puutu muuttujaa. Myös sen funktiomuoto on testin mukaan oikein. Lisäksi testattiin VIF- testillä muuttujien itsenäisyys. Nyt rajana on 1,1427, ja liitteestä 20 huomaa, että raja ylittyy kaikkien muuttujien paitsi sateen tapauksessa. Vaikka muuttujat eivät ole itsenäisiä, eivät ne kuitenkaan kärsi haitallisesti multikollineaarisuudesta, koska mikään ei saa arvoa 5-10.



Kuva 7. Mallin *lm_klubi* ennustamat arvot (punainen) ja oikeat arvot (musta).

Kuvasta 7 voi havaita, että *lm_klubi* ei ennusta kovin hyvin toteutuneita pesuja verrattuna malliin *lm_pesu.* Kun vertaa *klubin* toteutuneita pesuja *summapesun* toteutuneisiin pesuihin, niin huomataan että myös pesujen jakauma on todella erilainen. Kuvasta voi kuitenkin todeta, että säällä pystyy hieman ennustamaan *klubin* pesuja, koska ennusteet eivät mene aivan väärin. Ennuste on kuitenkin varsin maltillinen.

4.4 Asiakasryhmien eroavaisuuksia

Seuraavaksi vertaamme malleja *lm_pesut* ja *lm_klubi* keskenään. Suhteellisen kosteuden ollessa 100, *lm_pesut* mallissa vakiota tulee kertoa luvulla 0.0133 ja *lm_klubi* mallissa vain 0,1367. Suhteellisen kosteuden vaikutus on siis huomattavasti suurempi satunnaisten asiakkaiden kohdalla verrattuna kanta-asiakkaisiin. Lämpötilan ollessa -30 *lm_pesut* mallissa vakiota kerrotaan luvulla 8,29 ja *lm_klubi* mallissa luvulla 2,14. Kun lämpötila on taas +30, *lm_pesut* mallissa vakiota kerrotaan luvulla 0,1206 ja *lm_klubi* mallissa luvulla 0,467. Kanta-asiakkai-

den voidaan siis sanoa olevan vähemmän herkkiä lämpötilan vaihtelulle kuin satunnaisten asiakkaiden. Toisaalta viileällä säällä todennäköisesti satunnaiset asiakkaat pesevät autoaan enemmän kuin kanta-asiakkaat. Regressioyhtälön kertoimien vaikutuksesta sekä niiden merkitsevyydestä voi siis päätellä, että sääolosuhteet vaikuttavat huomattavasti vähemmän kanta-asiakkaiden käytökseen, kuin satunnaisten asiakkaiden.

Tarkastellaan vielä molempien ryhmien eroavaisuuksia kauppakeskuksen asiakasmäärän sekä erikoispäivien suhteen, vaikka ne eivät olleet kummassakaan mallissa tilastollisesti merkitsevällä tasolla. Kun parkkihallissa on 4425 autoa, *lm_pesut* mallin mukaan vakiota kerrotaan luvulla 0,41 ja *lm_klubi* mallin mukaan luvulla 0,64. Kun parkkihallissa on 224 autoa, *lm_pesut* mallin mukaan vakiota kerrotaan luvulla 0,96 ja *lm_klubi* mallin mukaan luvulla 0,98. Vilkkaampina päivinä kanta-asiakkaat pesevät enemmän kuin satunnaiset asiakkaat verrattuna hiljaisempiin päiviin. Toisaalta hiljaiset päivät eivät vaikuta kovinkaan paljoa kumpaankaan asiakasryhmään. Normaaleina päivinä kanta-asiakkaiden pesumäärän oletetaan olevan 16,7% vähemmän verrattuna erikoispäiviin, kun taas satunnaisten asiakkaiden vain 7,5% vähemmän.

Tarkastelussa voidaan todeta, että asiakasryhmien käyttäytymiset eroavat toisistaan. Kanta-asiakkaat suosivat sekä vilkkaita, että erikoispäiviä satunnaisasiakkaita enemmän. Lisäksi kanta- asiakkaisiin ei vaikuta yhtä vahvasti sääilmiöt toisin kuin satunnaisiin asiakkaisiin. Toisin sanoen kanta-asiakkaisiin vaikuttaa vahvemmin tutkimuksesta pois jätetyt selittävät tekijät kuten esimerkiksi ajomaasto tai auton merkki verrattuna satunnaisiin asiakkaisiin. Toisaalta satunnaiset asiakkaat ovat viileällä säällä otollisempia asiakkaita.

5. Pohdinta ja johtopäätökset

Tämän tutkimuksen tarkoituksena oli selvittää sään vaikutusta kysyntään eri asiakasryhmien kohdalla. Ensimmäinen alatutkimuskysymys käsitteli aihetta ” Miten sää ja kauppakeskuksen asiakkaiden määrä vaikuttavat pesumääriin?”. Ensimmäisen regressioanalyysin tarkoituksena oli testata satunnaisia asiakkaita, ja toisen regressioanalyysin tarkoituksena oli testata kanta-asiakkaita eli *klubilaisia.*

Ensimmäisestä regressioanalyysista saatiin selville, että kaikki regressiossa käytetyt sääolosuhteet vaikuttivat tilastollisesti merkittävästi pesumäärään niitä laskien. Näihin kuuluivat suhteellinen kosteus, sademäärä, lumimäärä sekä lämpötila. Parkkihallin kävijät ja erikoispäivät eivät sen sijaan olleet tilastollisesti merkitseviä tekijöitä. Viikonpäivistä vain lauantai ja perjantai olivat tilastollisesti merkitseviä vaikuttaen pesumäärään positiivisesti. Voidaan siis todeta, että satunnaisten asiakkaiden käyttäytymiseen vaikuttaa erittäin vahvasti sää sekä viikonloppu. Sään vaikutus on negatiivista ja viikonlopun positiivista. Eniten sääilmiöistä vaikuttaa suhteellinen kosteus. Kuitenkaan muu kauppakeskuksen asiakasmäärä parkkipaikkoina mitattuna ei vaikuta merkittävästi satunnaisten asiakkaiden pesumäärään.

Pesuohjelmiin sään, vierailijoiden ja päivien vaikutus oli melko samanlaista. Kaikkiin pesuohjelmien kysyntään sää vaikutti negatiivisesti. Pesuohjelmaan 1 vaikuttaa pesuohjelmista voimakkaimmin viikonpäivät ja pesuohjelmaan 2 säätekijät. Erikoispäivät vaikuttavat pesuohjelmien 2 ja 3 kysyntään positiivisesti, mutta pesuohjelmaan 1 negatiivisesti. Pesuohjelman 1 kysyntä erikoispäivänä yllättää, sillä muuten sekä klubiin, että kaikkiin muihin pesuihin erikoispäivä nostattaa kysyntää.

Toisessa regressioanalyysissa saatiin selville, että vain lämpötila ja suhteellinen kosteus vaikuttivat tilastollisesti merkitsevästi kanta-asiakkaiden pesuihin. Tärkeänä havaintona tuli esille, etteivät viikonpäivät, erikoispäivät tai sademäärä olleet tilastollisesti merkitsevällä tasolla. Voidaan siis todeta, että kanta-asiakkaiden käytökseen vaikuttaa osa sääilmiöistä, mutta eivät kaikki. Vahvin vaikutus myös kanta-asiakkaiden kohdalla on kuitenkin suhteellinen kosteus. Myöskään päivillä tai kauppakeskuksen muulla asiakasmäärällä ei ole tarpeeksi vahvaa vaikutusta kanta-asiakkaiden pesukäyttäytymiseen.

Toinen alatutkimuskysymys käsitteli aihetta ”Onko eri asiakasryhmien käytöksissä eroavaisuuksia?”. Tähän etsittiin vastausta vertaamalla laadullisesti muodostettuja regressiomalleja. Vertailussa huomattiin, että asiakasryhmien käyttäytymiset eroavat toisistaan paljon. Kanta-asiakkaat suosivat kauppakeskuksen vilkkaita päiviä sekä erikoispäiviä enemmän kuin satunnaiset asiakkaat. Lisäksi satunnaisiin asiakkaisiin vaikuttaa huomattavasti enemmän sääilmiöt

verrattuna kanta-asiakkaisiin. Kuitenkin molempiin asiakkaisiin sää vaikuttaa samansuuntaisesti.

Tutkimuksessa esitettiin kaksi hypoteesia, joita tutkimuksen aikana testattiin. Ensimmäisenä hypoteesina oli, että sää vaikuttaa kysyntään negatiivisesti, mikä oli johdettu teoreettisen taustan mukaan. Kolmantena hypoteesina oli johdettu, että muuttujat eivät kykene selittämään täysin pesumääriä. Molemmat hypoteesit jäävät voimaan.

Työn teoreettinen tausta tukee tutkimuksen tuloksia siltä osin, että asiakkaiden määrä molempien asiakasryhmien tapauksissa ei-toivotulla säällä vähenee, mitä väitti myös Buchheim & Kolaska (2017). Lisäksi Murray, Di Muro, Finn ja Leszczyz (2010) tukee myös havaintoa, että lämpötilan ollessa korkea, asiakkaiden maksuhalukkuus vähenee. Toisaalta tässä tutkimuk- sessa huomattiin, että myös kanta-asiakkaiden tapauksessa pesumäärä väheni, vaikka he eivät maksa samalla tavalla pesukerroista kuin satunnaiset asiakkaat. Schlager, de Bellis ja Hoegg (2020) mukaan säällä voisi ennustaa asiakkaiden kysyntää, joka myös tässä tutkimuksessa osoittautui melko onnistuneeksi. Kun asiakkaiden kysyntää voi ennustaa, pystytään myös markkinointia kohdentamaan paremmin oikeaan ajankohtaan. Moon et al (2018) tutkimus tukee tätä tutkimusta havainnolla, että molempien asiakasryhmien tapauksessa sää vaikutti maksuhalukkuuteen negatiivisesti. He kuitenkin totesivat ruokakaupassa säännöllisesti kävijöiden olevan juuri herkempiä ei-toivotulle säätilalle, mikä on siltä osin ristiriidassa tämän tutkimuksen kanssa, jossa todettiin, että kanta-asiakkaat eivät ole yhtä herkkiä. Toisaalta ruoka on välttämätön hyödyke ja sitä voi varastoida pahan päivän varalle toisin kuin puhdasta autoa. Tältä osin tapausta voisi silti tutkia vielä tarkemmin.

Johtopäätöksenä yrityksen tulisi pyrkiä lisäämään kanta-asiakkaidensa määrää, koska sää ei vaikuta heihin yhtä paljon. Näin yritys voi käytännössä turvata liikevaihtonsa määrän riippumatta sään vaikutuksista, jotka ovat vahvasti havaittavissa satunnaisten asiakkaiden kohdalla. Koska kohdeyritykseen vaikuttaa säätilat vahvasti, voisi yrityksen toimintaa kehottaa ohjaamaan Schlager, de Bellis ja Hoegg (2020) mukaan: Kohdeyritys voisi näyttää asiakkaille suotuisia säätiedotuksia ja säännellä hintaa dynaamisesti. Kohdeyrityksen tapauksessa olisi hyvä laskea asiakkaiden jousto kanta-asiakkaaksi liittymiselle, jotta yritys voisi maksimoida kanta-asiakkaidensa määrän. Huomioon tulee tällöin ottaa myös pesukoneen rajallinen pesukapasiteetti

5.1 Rajoitteet ja jatkotutkimusaiheita

Tässä osiossa pyritään huomioimaan tutkimusta rajoittaneet useat tekijät. Osa tekijöistä voi vaikuttaa myös tutkimuksen luotettavuuteen, joten tarkoituksena on tarkastella tutkimusta kriittisesti.

Tutkimuksessa muodostetut regressiomallit eivät ole täydellisiä, ja niin kuin jo alussa todettiin, niistä jäi puuttumaan moni merkittävä tekijä. Havaintoaineiston muuttujien määrä rajoitti siis tutkimusta merkittävästi. Tutkimuksessa on myös tehty karkea yleistys kanta-asiakkaista ja niiden luonteesta. Tässä tutkimuksessa kanta-asiakkaat ovat maksaneet etukäteen kiinteän hinnan. Tutkimuksen tuloksia ei siis voi luotettavasti soveltaa ainakaan muunlaisiin kanta-asiakkaisiin tai vakioasiakkaisiin.

Tutkimuksessa on käytetty vain yhtä yritystä, joka on vielä melko pieni ja jonka data sijoittuu yhteen paikkaan Suomea. Ei siis voida tarkalleen sanoa, miten kyseistä tutkimusta voidaan soveltaa toisiin yrityksiin, koska tutkimuksessa luotuja malleja ei ole verrattu mihinkään muuhun yritykseen. Havaintoaineistoa on kuitenkin kerätty sen verran pitkältä ajalta, että mallin oletetaan olevan melko luotettava kohdeyrityksen kohdalla. Lisäksi rajoittaviin ja luottamusta heikentäviin tekijöihin pitää laskea myös kohdeyrityksen nuoruus ja jatkuva kasvu. Klubiasiakkaiden sekä satunnaisten asiakkaiden määrä on yrityksen mukaan jatkuvasti kasvanut, vaikka kumpikaan havainto ei aikasarjana osoittanut epästationaarisuutta. Yritys on kuitenkin siis vielä muutosvaiheessa.

Jatkotutkimusaiheina voisikin olla laajempi tutkimus eri yrityksiltä sekä useammalta asiakasryhmältä, koska tämä tutkimus viittaa merkittävästi siihen, että sää vaikuttaa asiakasryhmiin eri tavalla. Samalla saisi tietoon, miten yleistettävissä tutkimuksessa todettu ilmiö on. Lisäksi voisi tutkia, miten yritykset jo ottavat sään huomioon liiketoiminnassaan ja miten se vaikuttaa yrityksen tuloksiin. Tuloksia voisi verrata saman alan yrityksiin, jotka eivät ota säätä huomioon. Tutkimuksella voisi ottaa selvää, voiko sään aiheuttamaan kysynnän muutokseen todellisuudessa puuttua.

Lähdeluettelo

BROCKWELL, P.J. and DAVIS, R.A., 2016. Introduction to time series and forecasting. springer.

BUCHHEIM, L. and KOLASKA, T., 2017. Weather and the psychology of purchasing outdoor movie tickets. Management Science, 63(11), pp. 3718-3738.

BUSSE, M.R., POPE, D.G., POPE, J.C. and SILVA-RISSO, J., 2015. The psychological effect of weather on car purchases. The Quarterly Journal of Economics, 130(1), pp. 371-414.

CAO, M. and WEI, J., 2005. Stock market returns: A note on temperature anomaly. Journal of Banking & Finance, 29(6), pp. 1559-1573.

CHATTERJEE, S. and PRICE, B., 1977. Regression analysis by example. New York (NY): Wiley.

CHEUNG, Y. and LAI, K.S., 1995. Lag order and critical values of the augmented Dickey–Fuller test. Journal of Business & Economic Statistics, 13(3), pp. 277-280.

DE BELLIS, E., SCHULTE-MECKLENBECK, M., BRUCKS, W., HERRMANN, A. and HERTWIG, R., 2018. Blind haste: As light decreases, speeding increases. PLoS one, 13(1)

HEIKKILÄ, T., 2014. Tilastollinen tutkimus. 9., uud. p. edn. Helsinki: Edita.

HILL, R.C., GRIFFITHS, W.E. and LIM, G.C., 2012. Principles of econometrics. 4th ed edn. Hoboken (NJ): Wiley.

HOFMANN, M. and O'MAHONY, M., 2005. The impact of adverse weather conditions on urban bus performance measures, Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005. 2005, IEEE, pp. 84-89.

HONG, J. and SUN, Y., 2012. Warm it up with love: The effect of physical coldness on liking of romance movies. Journal of Consumer Research, 39(2), pp. 293-306.

ILMATIETEENLAITOS, 2019-last update, Tietokanta säähavannoista. Available: https://www.ilmatieteenlaitos.fi/havaintojen-lataus#!/.

ISRAELSSON, S. and TAMMET, H., 2001. Variation of fair weather atmospheric electricity at Marsta Observatory, Sweden, 1993–1998. *Journal of atmospheric and solar-terrestrial physics*, *63*(16), pp.1693-1703.

JACOBSEN, B. and MARQUERING, W., 2008. Is it the weather? Journal of Banking & Finance, 32(4), pp. 526-540.

KAMSTRA, M.J., KRAMER, L.A. and LEVI, M.D., 2003. Winter blues: A SAD stock market cycle. American Economic Review, 93(1), pp. 324-343.

KING III, C. and NARAYANDAS, D., 2000. Coca-Cola's New Vending Machine (A): Pricing to Capture Value, or Not?

MATSON-MÄKELÄ KIRSI. Auton pesukerrat lisääntyvät, kun kuukausimaksulliset autopesulat valtaavat markkinoita – ympäristöasiantuntija toppuuttelee pesuintoa. Yle.fi.

MOON, S., KANG, M.Y., BAE, Y.H. and BODKIN, C.D., 2018. Weather sensitivity analysis on grocery shopping. International Journal of Market Research, 60(4), pp. 380-393.

MURRAY, K.B., DI MURO, F., FINN, A. and LESZCZYC, P.P., 2010. The effect of weather on consumer spending. Journal of Retailing and Consumer Services, 17(6), pp. 512-520.

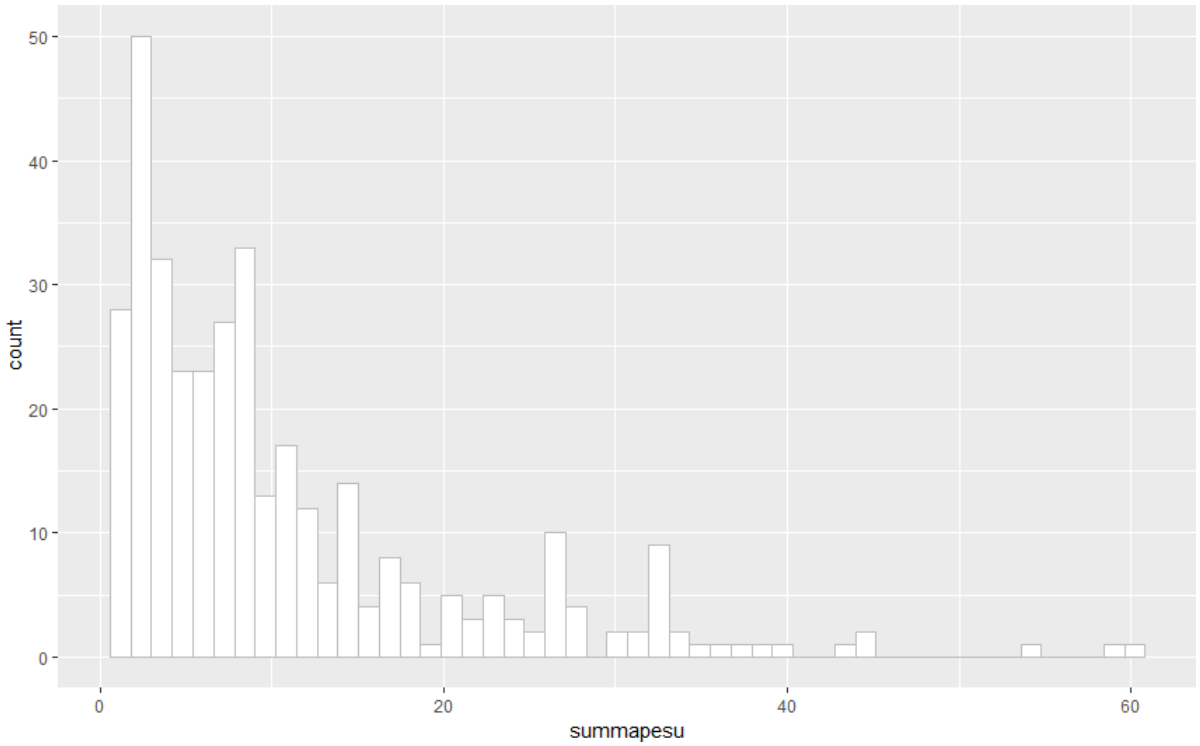
PALMA, W., 2016. Time series analysis. Hoboken: John Wiley & Sons.

SCHLAGER, T., DE BELLIS, E. and HOEGG, J., 2020. How and when weather boosts consumer product valuation. Journal of the Academy of Marketing Science.

ZWEBNER, Y., LEE, L. and GOLDENBERG, J., 2014. The temperature premium: Warm temperatures increase product valuation. Journal of Consumer Psychology, 24(2), pp. 251-259.

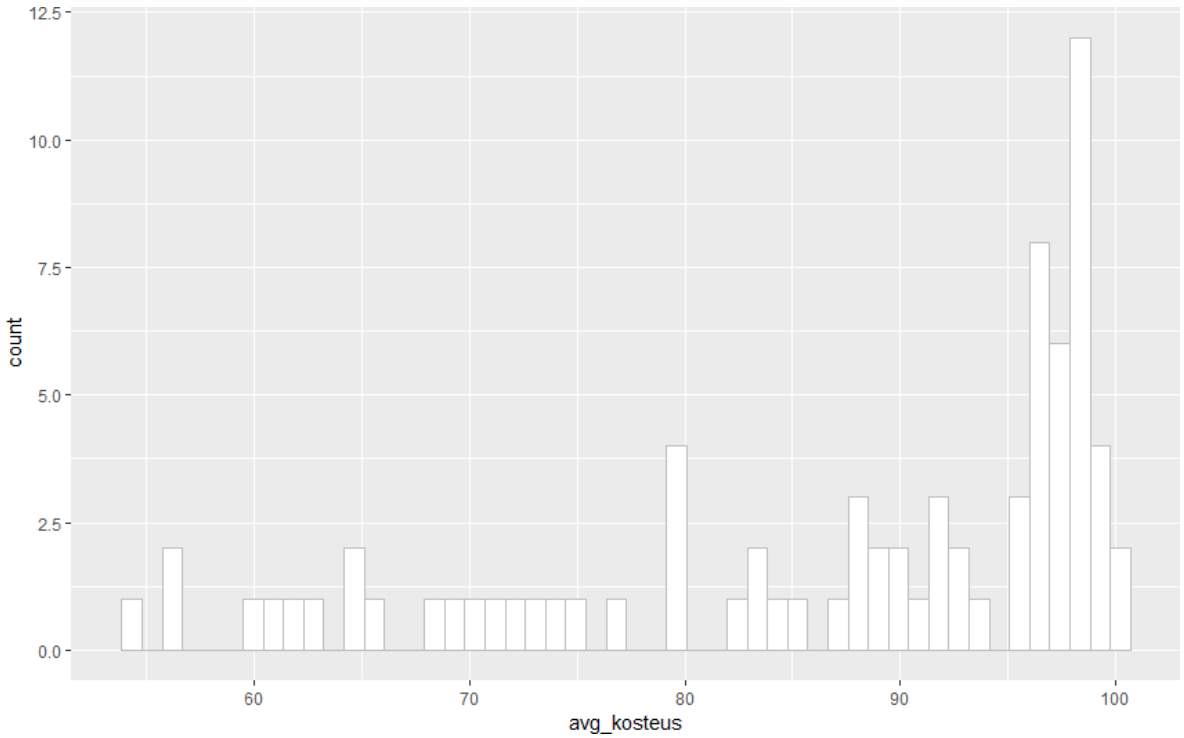
LIITTEET

Liite 1. *Summapesut* muuttujan jakautuminen.



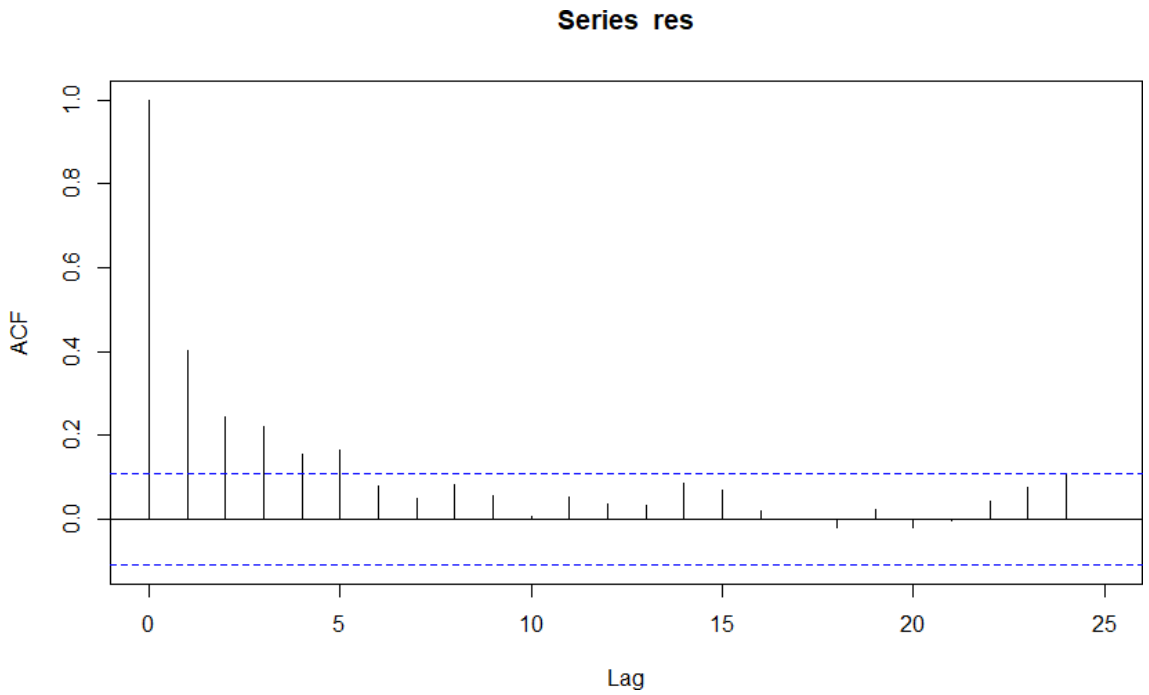

```
        Shapiro-Wilk normality test

data:  kaikki3$summapesu
W = 0.80529, p-value < 2.2e-16
```

Liite 2. Suhteellisen kosteuden keskiarvon jakautuminen.



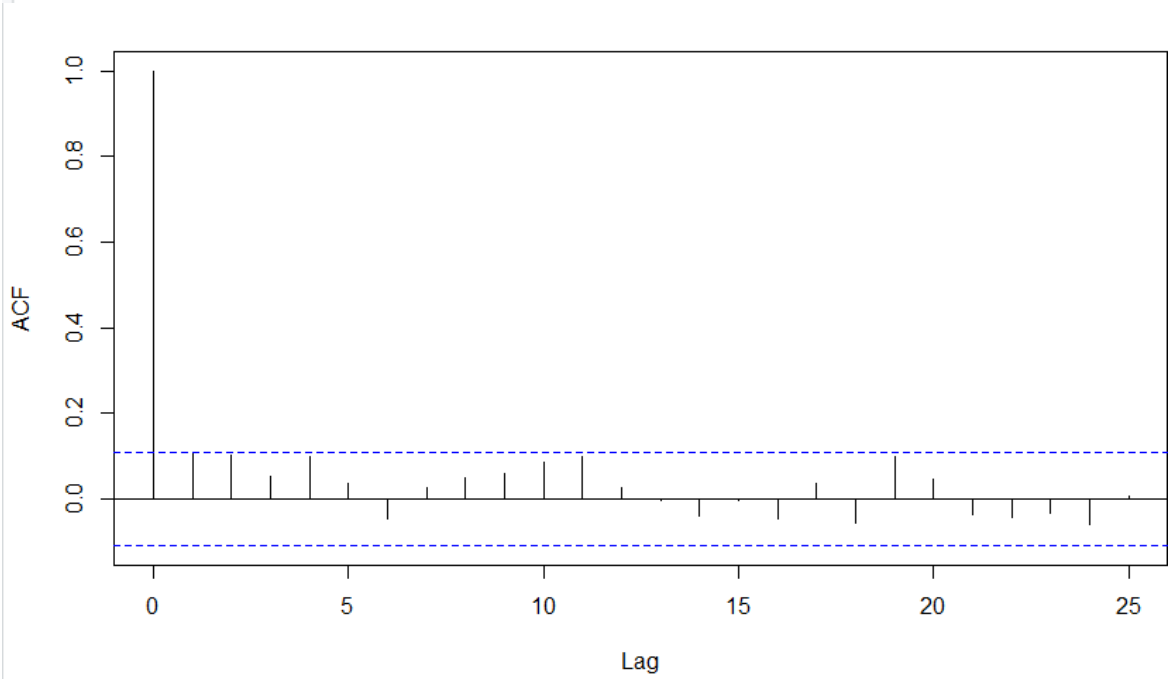
Liite 3. *Summapesu* autokorrelaation havaitseminen.

```
Call:
lm(formula = res[-n] ~ res[-1])

Residuals:
     Min       1Q   Median       3Q      Max
-2.49462 -0.47551  0.08352  0.48445  2.08857

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0006864  0.0394308  -0.017    0.986
res[-1]      0.4012214  0.0510586   7.858  5.9e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7098 on 322 degrees of freedom
Multiple R-squared:  0.1609,    Adjusted R-squared:  0.1583
F-statistic: 61.75 on 1 and 322 DF,  p-value: 5.9e-14
```



Liite 4. *Klubi* autokorrelaation havaitseminen.

```
Call:
lm(formula = res[-n] ~ res[-1])

Residuals:
    Min      1Q  Median      3Q     Max
-5.3342 -1.5886 -0.1398  1.2594  7.4354

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.004534   0.127725   0.035   0.9717
res[-1]     0.105033   0.055737   1.884   0.0604 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.285 on 318 degrees of freedom
Multiple R-squared:  0.01104,   Adjusted R-squared:  0.007934
F-statistic: 3.551 on 1 and 318 DF,  p-value: 0.06042
```



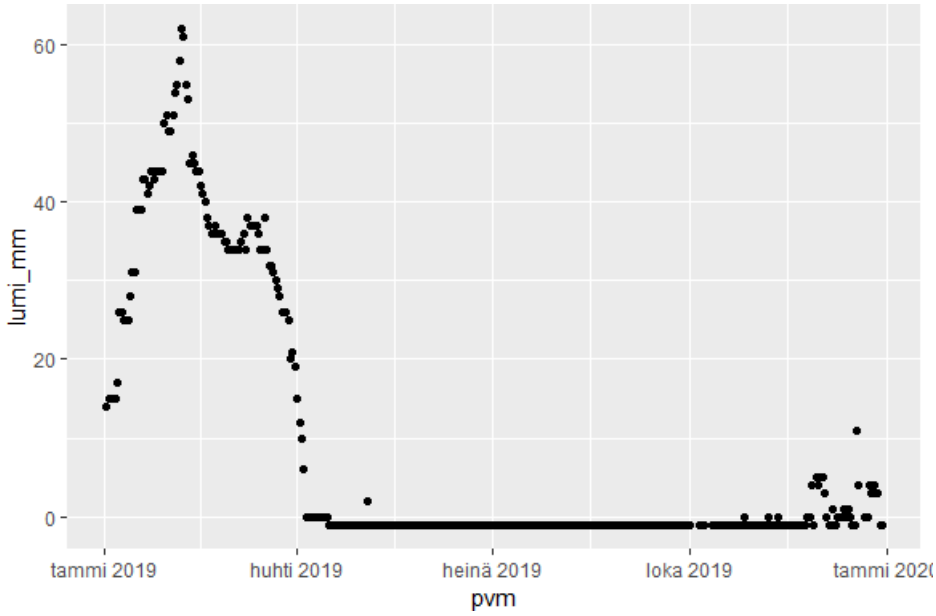
Liite 5. *Summapesu* ADF-testi.

```
        Augmented Dickey-Fuller Test

data:  na.omit(log(kaikki3$summapesu))
Dickey-Fuller = -4.8889, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

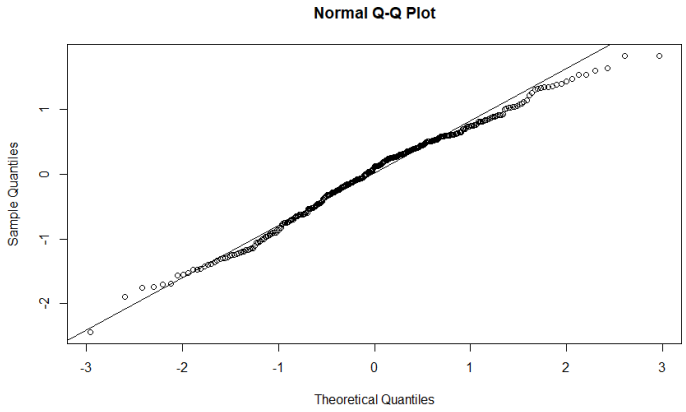
Liite 6. *Klubi* ADF-testi.

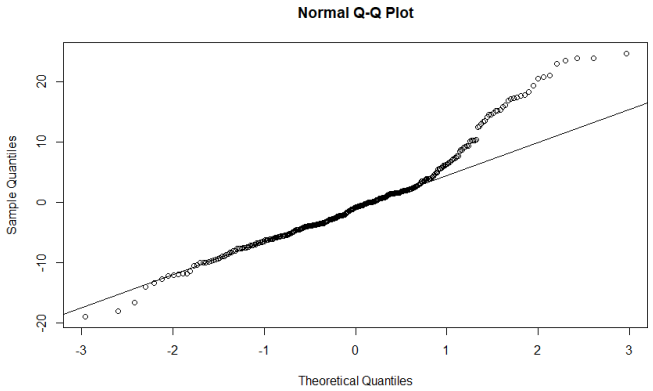
```
        Augmented Dickey-Fuller Test

data:  na.omit(kaikki4$logklubi)
Dickey-Fuller = -10.507, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary
```

Liite 7. Lumimäärän jakautuminen.



Liite 8. *lm_pesut* residuaalit, kun *summapesu* logaritminen (vasen) versus normaalisti (oikea).

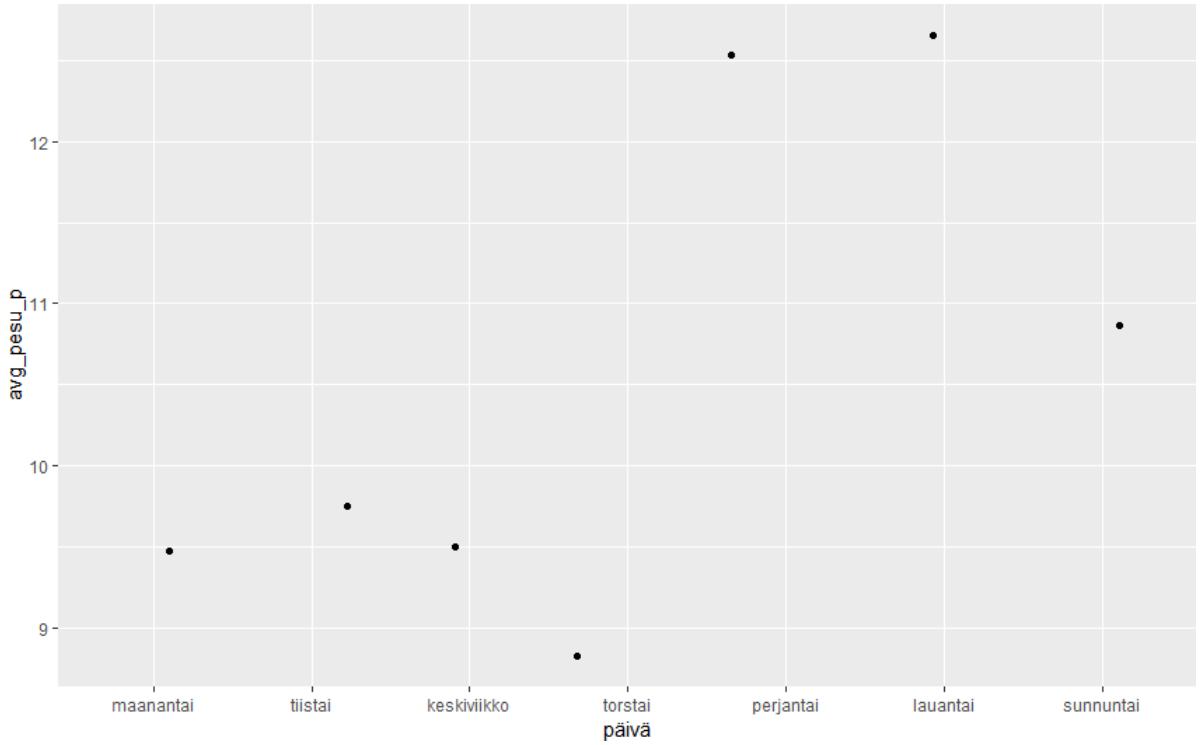




Liite 9. Breusch Pagan Testi heteroskedastisuudesta: *lm_pesut*

```
Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

                  Data
-----------------------------------------
Response : log(summapesu)
Variables: fitted values of log(summapesu)

       Test Summary
---------------------------
DF            =    1
Chi2          =    1.285235
Prob > Chi2   =    0.2569279
```

Liite 10. Pesujen jakautuminen päivittäin.



Liite 11. RESET-testin tulokset mallille *lm_pesut.*

```
        RESET test

data:  lm_pesut
RESET = 2.7775, df1 = 2, df2 = 311, p-value = 0.06373
```

Liite 12. VIF- testi mallille *lm_pesut.*

```
                GVIF Df GVIF^(1/(2*Df))
viikonpaiva 1.389989  6        1.027821
erikoispaiva 1.379411  1        1.174483
avg_kosteus  1.405098  1        1.185368
lampötila    2.070244  1        1.438834
sade_mm      1.089040  1        1.043571
lumi_mm      1.756839  1        1.325458
parkkihalli  1.555722  1        1.247286
```

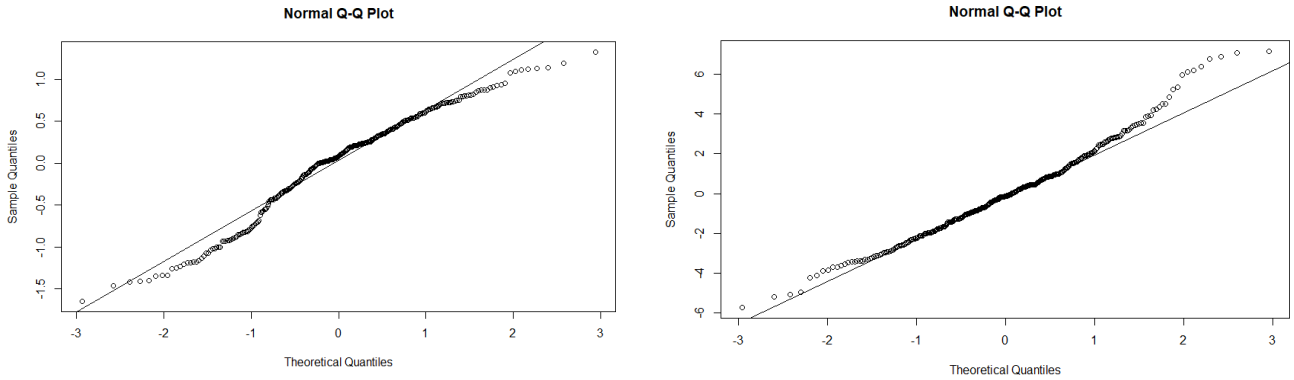
Liite 13. Pesuohjelmien summien stationaarisuuden testaus.

```
        Augmented Dickey-Fuller Test

data:  na.omit(kaikki3$summap1)
Dickey-Fuller = -4.6748, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(na.omit(kaikki3$summap1)) :
  p-value smaller than printed p-value
> adf.test(na.omit(kaikki3$summap2))

        Augmented Dickey-Fuller Test

data:  na.omit(kaikki3$summap2)
Dickey-Fuller = -4.2055, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(na.omit(kaikki3$summap2)) :
  p-value smaller than printed p-value
> adf.test(na.omit(kaikki3$summap3))

        Augmented Dickey-Fuller Test

data:  na.omit(kaikki3$summap3)
Dickey-Fuller = -5.0201, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(na.omit(kaikki3$summap3)) :
  p-value smaller than printed p-value
```

Liite 14. Heteroskedastisuuden testaaminen *pesuohjelmat* regression yhteydessä.

```
Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

              Data
----------------------------------
Response : summap1
Variables: fitted values of summap1

        Test Summary
-------------------------------
DF            =    1
Chi2          =    96.4209
Prob > Chi2   =    9.288142e-23


Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

              Data
----------------------------------
Response : summap2
Variables: fitted values of summap2

        Test Summary
-------------------------------
DF            =    1
Chi2          =    88.58981
Prob > Chi2   =    4.857851e-21


Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

              Data
----------------------------------
Response : summap3
Variables: fitted values of summap3

        Test Summary
-------------------------------
DF            =    1
Chi2          =    30.91268
Prob > Chi2   =    2.69902e-08
```

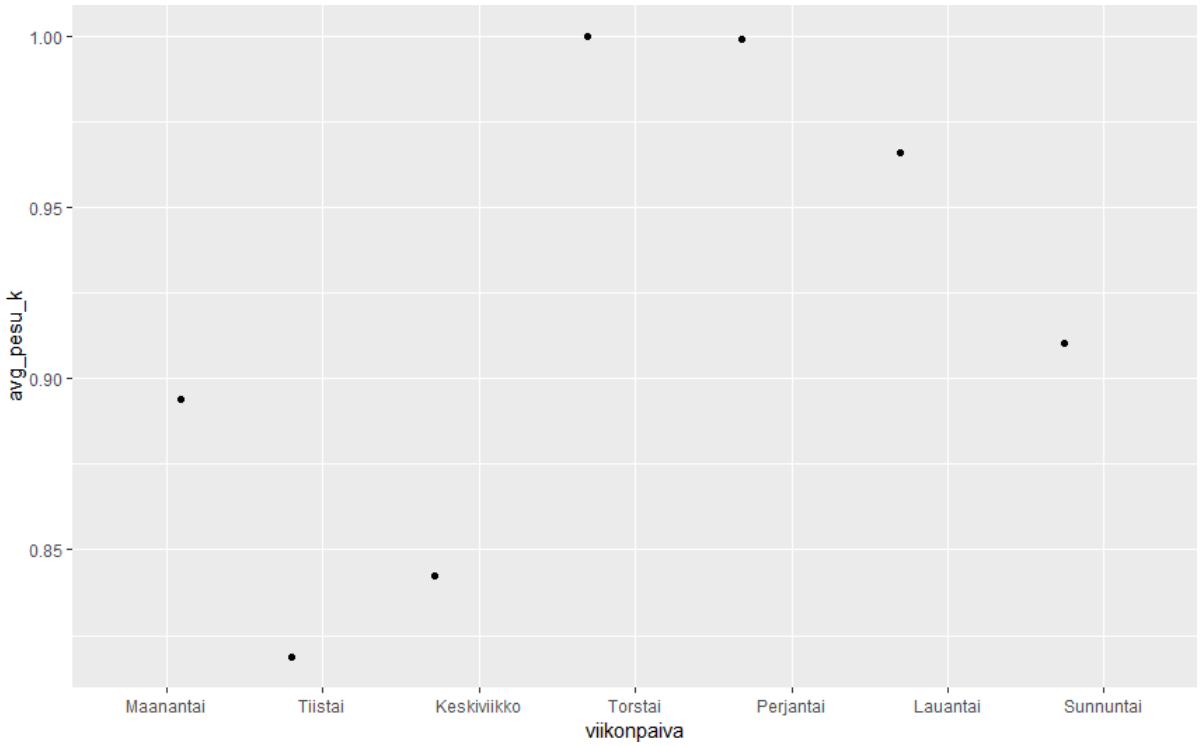
Liite 15. *lm_klubi* residuaalit, kun *klubi* logaritminen (vasen) versus normaalisti (oikea).



Liite 16. Breusch Pagan Testi heteroskedastisuudesta mallille *lm_klubi.*

```
Breusch Pagan Test for Heteroskedasticity
-----------------------------------------
Ho: the variance is constant
Ha: the variance is not constant

                Data
-------------------------------------
Response : logklubi
Variables: fitted values of logklubi

        Test Summary
----------------------------
DF            =    1
Chi2          =    2.303358
Prob > Chi2   =    0.1290947
```

Liite 17. *Klubin* myynnin jakautuminen päivittäin.



Liite 18. *Lm_klubi* luottamusvälit.

	2.5 %	97.5 %
(Intercept)	2.2321200800	3.572460421
erikoispaivaNormaali päivä	-0.4487780959	0.082925446
viikonpaivaLauantai	-0.2664552651	0.310059385
viikonpaivaMaanantai	-0.1808807950	0.371283698
viikonpaivaPerjantai	-0.1220050750	0.427692945
viikonpaivaSunnuntai	-0.3240866851	0.249650089
viikonpaivaTiistai	-0.1895078910	0.374565145
viikonpaivaTorstai	-0.1663455798	0.371810805
avg_kosteus	-0.0234438621	-0.011329844
lampötila	-0.0417680396	-0.017651067
sade_mm	-0.0212152897	0.013657442
lumi_mm	-0.0041919367	0.006785578
parkkihalli	-0.0003028505	0.000145881

Liite 19. RESET-testi mallille *lm_klubi.*

```
        RESET test

data:  lm_klubi
RESET = 0.11683, df1 = 2, df2 = 286, p-value = 0.8898
```

Liite 20. VIF- testi *lm_klubille.*

	GVIF	Df	GVIF^(1/(2*Df))
erikoispaiva	1.332921	1	1.154522
viikonpaiva	2.041091	6	1.061260
avg_kosteus	1.483827	1	1.218124
lampötila	3.077907	1	1.754397
sade_mm	1.095741	1	1.046776
lumi_mm	1.641459	1	1.281195
yht	3.261325	1	1.805914