

# **Automaattinen analyysityökalu**



Ammattikorkeakoulututkinnon opinnäytetyö

Tieto- ja viestintäteknikka, Riihimäki

Kevät 2021

Mervi Saarinen

---

## TIIVISTELMÄ

Opinnäytetyön tavoitteena oli kehittää Murata Electronics Oy:lle automaattinen analyysityökalu anturielementtien mittaustulosten analysointiin, keskittyen ryhmien välisten erojen analysointiin. Käyttökohde analyysityökalulle on tyypillisesti koejärjestelmien tulosten analysointi tuotekehityksen tai prosessimuutosten yhteydessä. Analysoinnin automatisoinnilla tavoiteltiin aikasäästöä ja analysointiprosessin yhdenmukaistamista. Analyysien tekeminen on aikaa vievää, suuren osan ajasta vie tiedon hankinta ja esikäsitteily.

Analyysityökalu suunniteltiin siten, että kaikki tarvittava tieto vertailtavista tuotteista haetaan automaattisesti tiedostoista tai tietokannoista, jolloin käyttäjän vastuulle jää vain antaa tiedoksi anturielementtejä sisältävien koottujen kiekkojen yksilöivät tunnisteet, tieto siitä mihin ryhmään kiekot kuuluvat sekä valita halutut visualisoinnit ja analyysit. Analyysityökalu koostaa tarvittavat tiedot ja tarjoaa valmiit analyysit ja visualisoinnit.

Työ toteutettiin JMP-analytiikkaohjelmistolla, hyödyntäen JSL-skriptikieltä. Ohjelmoinnin pääasiallinen painopiste oli analyysejä varten tarvittavien datasettien koostaminen siten, että samalla ohjelmakoodilla saadaan haettua kaikki erilaiset tuoteversiot. Myös haettavan datan oikeellisuuteen tuli kiinnittää huomiota. Analyysityökaluun toteutettiin helppokäyttöinen käyttöliittymä ja tulosten esittämisen selkeyteen kiinnitettiin huomiota. Analyysityökalu saatiin onnistuneesti toteutettua ja sen käyttöopastus lisättiin yrityksen koulutuskalenteriin, koulutuksen pääpainon ollessa analyysitulosten tulkinnessa.

Avainsanat JMP, JSL, varianssianalyysi, data.

Sivut 40 sivua ja liitteitä 0 sivua

---

Author Mervi Saarinen

Year 2021

Subject Automatic analysis tool

Supervisors Jari Mustajärvi

---

## ABSTRACT

The purpose of the thesis was to develop an automatic analysis tool for Murata Electronics Oy for the analysis of the measurement results of sensor elements, focusing on the analyses of differences between groups. The typical use of the analysis tool is to analyse the results of experiments related to product development or process changes. The objective for automating the analysis is to save time and to standardize the analysis process. Performing analyses is time consuming, mostly due to acquiring and pre-processing the data.

The analysis tool was designed so that all the necessary information about the wafers selected for comparison is automatically retrieved from files or databases, the user only has to provide the unique identifiers, lotnumbers, of the assembled wafers, information on which group the wafers belong to, and to select the desired visualizations and analyses. The analysis tool compiles the necessary data and provides analyses and visualizations.

The work was implemented with JMP analytics software, utilizing the JSL scripting language. The focus of the software development was to compile the data sets needed for the analyses, so that all the different product versions could be handled with the same program code. Quality of the data retrieved is essential. A user interface was created, and attention was paid to the clarity of the presentation of the results. The analysis tool was successfully implemented, and a training session was added to the company's training calendar, with the focus of the training being the interpretation of the analysis results.

Keywords JMP, JSL, analysis of variance, data.

Pages 40 pages and appendices 0 pages

## Sisälllys

1	Johdanto .....	1
1.1	Yrityksestä .....	1
1.2	Tuotteista .....	2
2	Datatiede .....	4
2.1	Datatieteen prosessi .....	6
2.2	Tutkivaa data-analyysiä.....	8
2.2.1	Pistekaavio .....	8
2.2.2	Pylväskaavio, histogrammi ja pareto .....	9
2.2.3	Laatikko-jana-kuvio .....	11
2.2.4	Aikasarja .....	12
2.3	Hypoteesien testaus.....	12
2.4	Ryhmiä keskiarvojen vertailu, Student's t-test ja Oneway Anova.....	13
3	Analyysityökalun kehitysprojekti.....	15
3.1	Tavoitteet automaattiselle analyysityökalulle .....	16
3.2	Jira .....	16
3.3	Projektin aikajana.....	16
4	JMP ohjelmisto .....	17
5	Tietolähteet .....	18
5.1	Standard data.....	18
5.2	Tietokannat .....	19
6	Analyysityökalun testaus kehitysprojektin aikana .....	19
7	Automaattinen analyysityökalu.....	20
7.1	Analyysityökalun käyttökohteet .....	20
7.2	Rajoitukset .....	21
7.3	Käyttäjän antamat lähtötiedot ja valmiit asetukset .....	21
7.4	Käyttöliittymä.....	22
7.5	Ylätason logiikka.....	23
7.6	Tulosten esittäminen JMP projektinäkyvässä .....	23
7.7	Datasettien muodostus.....	25
7.7.1	Wafer info .....	25
7.7.2	Yield datasetti ja FC datasetti.....	26
7.7.3	RAW datasetti ja FCO RAW datasetti .....	27
7.8	Visualisoinnit ja analyysit .....	29

7.8.1	Wafer level yields .....	29
7.8.2	Yields by group .....	30
7.8.3	Yield graph.....	30
7.8.4	FC Pareto .....	31
7.8.5	Errorcode %.....	32
7.8.6	Errorcode pareto .....	33
7.8.7	Parameter wafermaps.....	33
7.8.8	Time series .....	35
7.8.9	ANOVA.....	35
7.8.10	Superimposed distribution.....	36
8	Yhteenveto .....	37
	Lähteet.....	39

## Kuvat, taulukot ja kaavat

Kuva 1. Murata ja Murata Electronics. (Murata Electronics Oy, 2021 sisäinen materiaali)	2
Kuva 2. Anturikomponentti. (Murata Electronics Oy, 2021, sisäinen materiaali)	3
Kuva 3. Esimerkki anturin rakenteesta. (Murata Electronics Oy, 2021, sisäinen materiaali)	3
Kuva 4. Anturielementin valmistusprosessi. (Murata Electronics Oy, 2021, sisäinen materiaali)	4
Kuva 5. Erilaisia anturielementtejä. (Murata Electronics Oy, 2021, sisäinen materiaali)	4
Kuva 6. Datatieteen Venn-diagrammi. (Conway, 2010)	5
Kuva 7. Datatieteen prosessi. (Harvard, 2015, 1. luento)	7
Kuva 8. Pistekaavio	9
Kuva 9. Pylväskaavio	10
Kuva 10. Histogrammi.	10
Kuva 11. Pareto-kaavio	11
Kuva 12. Laatikko-jana-kuvio. (Sas Institute Inc, n.d.-d)	12
Kuva 13. Aikasarja	12
Kuva 14. Hypoteesit	13
Kuva 15. Oneway Anova raportti.	15

Kuva 16. Projektin aikajana. ....	17
Kuva 17. Käyttöliittymä. ....	22
Kuva 18. Analyysityökalun ylätasen logiikka. ....	23
Kuva 19. JMP projektinäkökulma. ....	24
Kuva 20. JSL koodia.....	25
Kuva 21. JSL ohjelmakoodia tyhjien kolumnien poistamiseen.....	26
Kuva 22. Yield datasetin muodostus. ....	27
Kuva 23. Speksirajojen asetus kolumnin ominaisuuksiin. ....	28
Kuva 24. Speksirajojen asettaminen kolumneille.....	29
Kuva 25. Kiekkotason saannot taulukkomuodossa. ....	30
Kuva 26. Ryhmäkohtaiset mittapistesäännöt.....	30
Kuva 27. Graph Builder näkymä mittapistesäännöistä. ....	31
Kuva 28. FC Pareto.....	32
Kuva 29. Vikakoodien prosenttiosuudet ryhmittäin. ....	32
Kuva 30. Vikakoodipareto.....	33
Kuva 31. Testiparametrien tarkastelu. ....	34
Kuva 32. Näytettävän asteikon valinta.....	34
Kuva 33. Testiparametrin aikasarjadata.....	35
Kuva 34. Anova. ....	36
Kuva 35. Vaihtelun visualisointia ryhmittäin. ....	37

## Lyhenteet ja termistö

ANOVA	Analysis of Variance, varianssianalyysi
ASIC	Application Specific Integrated Circuit, mikropiiri
BOXPLOT	Laatikkokuvaio
Eräkoodi	Erän tai prosessin loppuvaiheessa olevan kiekon uniikki tunnus tuotantojärjestelmässä ja mittaustestikannassa
FC	Fault Code, vikakoodi
FCO RAW DATA	Hyväksytyjen elementtien testaustieto
JMP	Analytiikkaohjelmisto
JSL	JMP ohjelmiston skriptikieli
Kiekko	Piikiekko, myös koottu kiekko, tarkoitus riippuu asiayhteydestä
MEMS	Micro-Electro-Mechanical System, mikrosysteemi
MFI	Murata Finland, Murata Electronics Oy
RAW DATA	Raakadata, elementtikohtainen testaustieto
UI	User Interface, käyttöliittymä

# 1 Johdanto

Opinnäytetyön toimeksiantajana toimii työnantajani Murata Electronics Oy. Työskentelen yrityksessä Data Integration-insinöörinä, työnkuvaani kuuluu esimerkiksi standardien datasettien luonti ja ylläpito, data-analytiikan palvelut sekä PowerBI-ohjelmiston avainkäyttäjänä toimiminen.

Aloite automaattisen analyysityökalun tekemisestä tuli yrityksen sisäisten sidosryhmien edustajilta, jotka tunnistivat tarpeen saada aikaa säästävää ja toimintatapoja yhtenäistävä työkalu vertailevan data-analyysin tekemiseen anturielementtivalmistuksen alueella.

Yrityksessä tehdään paljon koemateriaaleihin tai laitepäivityksiin liittyviä data-analyyskejä, datankeruu ja analysointi ovat aikaa vieviä prosesseja, jolloin analyysitulosten käsittely saattaa kestää tarpeettoman kauan. Lisähaasteena on erilaisten käytäntöjen kirjavuus, yhtenäistä proseduuria analyysien tekemiselle ei ole ollut, vaikka suurimmalle osalle tapauksista voidaankin soveltaa hyvin saman tyyppistä analysointiprosessia.

Työn tavoitteeksi asetettiin automaattisen analysointityökalun kehittäminen, jonka tulee käyttäjän antamien esitietojen mukaan suorittaa testausdatan analysointia, keskittyen varsinkin ryhmien välisten erojen vertailuun. Työkalun tulee hakea tarvittava data tietokannoista ja tiedostoista itsenäisesti, muodostaa halutut analyysit ja esittää tulokset hyvin organisoidussa muodossa. Analyysityökalun kehittäminen yhdistää sekä datatiedettä, että ohjelmointia, joten sen tekeminen sopi mainiosti opinnäytetyön aiheeksi.

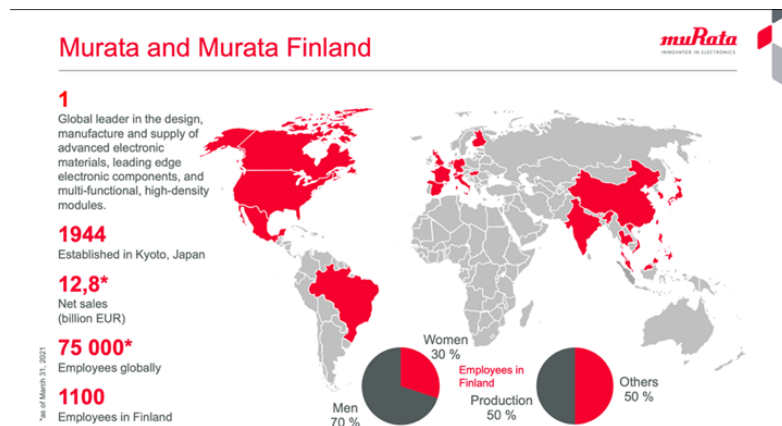
## 1.1 Yrityksestä

Murata Electronics Oy (jatkossa MFI) suunnittelee, kehittää ja valmistaa piipohjaisia 3D MEMS (Micro Electro Mechanical Systems) -teknologiaan perustuvia kiihtyvyy-, kallistus- ja kulmanopeusantureita Vantaan Martinlaaksossa. Yritys on osa maailman johtavaa anturi- ja elektroniikkakomponentteja valmistavaa Murata-konsernia, lisätietoja nähtävissä kuvassa 1. Konsernin pääkonttori sijaitsee Kiotossa Japanissa. Murata työllistää noin 75 000 työntekijää, joista Suomessa MFI noin 1100 henkeä. (Murata Electronics Oy, 2021, sisäinen materiaali)



MFI on markkinoiden johtava autoteollisuuden kiihtyvyy- ja kaltevuusantureiden valmistaja ja toimittaja, maailman johtava ajonvakautusjärjestelmien liikeantureiden valmistaja sekä maailman johtava sydämentahdistimien liikeantureiden toimittaja. MFI:n valmistamia antureita käytetään myös ilmailuelektronikassa sekä vaativiin maatalouden ja raskaan teollisuuden sovelluksiin. (Murata Electronics yrityksenä | Murata Finland, n.d.)

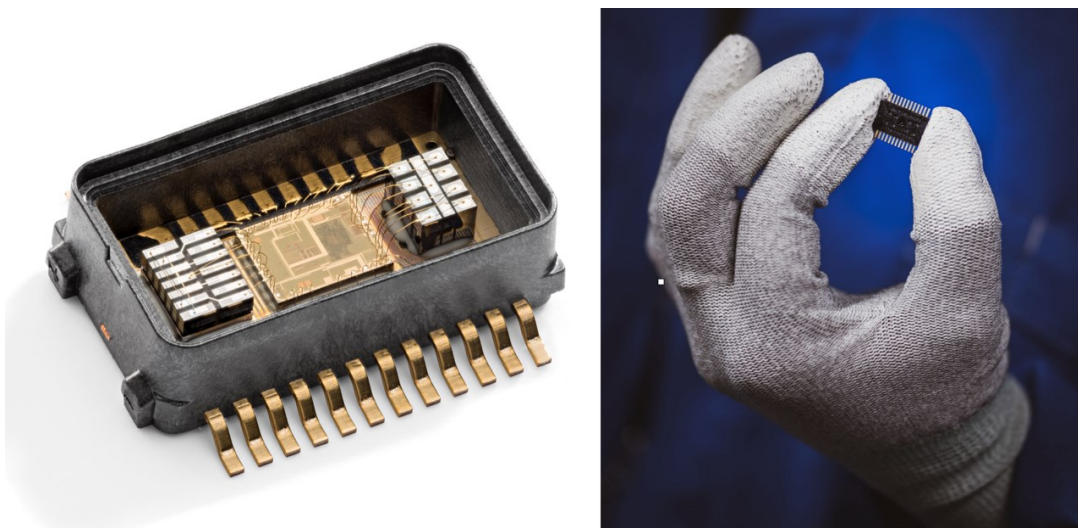
Kuva 1. Murata ja Murata Electronics. (Murata Electronics Oy, 2021 sisäinen materiaali)



## 1.2 Tuotteista

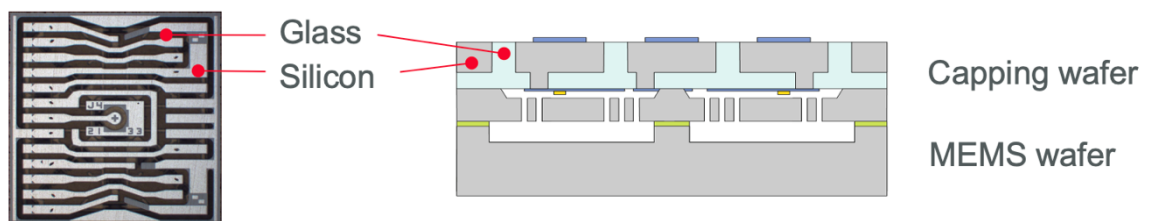
MFI:n valmistamissa anturikomponenteissa on koteloituna yksi tai useampi anturielementti ja niihin liittyvät ASIC-piiri tai -piirit, jotka on yhdistetty toisiinsa ja tuotteen koteloon kulturalgoin, esimerkiksi anturikomponentista, ennen geelitäyttöä ja tuotteen sulkevaa metallista kantta, sekä valmiina tuotteena kuvassa 2. (Murata Electronics Oy, 2021, sisäinen materiaali)

Kuva 2. Anturikomponentti. (Murata Electronics Oy, 2021, sisäinen materiaali)



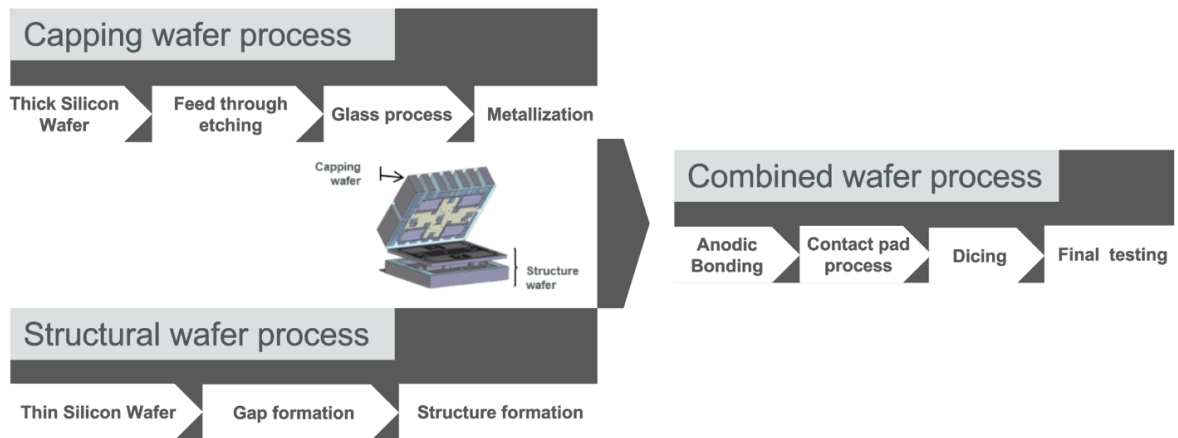
Anturikomponentin sisältämä(t) anturielementit valmistetaan puolijohdeteollisuuden prosessein litografialla kuvioimalla, etsaamalla ja kalvoja kasvattamalla piikiekoista ja lasista. Valmis koottu kiekko koostuu 2–3 yhteen liitetystä kiekosta, tyypillisesti rakenne- eli ohuesta kiekosta (kuvassa MEMS wafer), sekä yhdestä tai kahdesta paksukiekosta, jotka toimivat kantena ja/tai pohjana (Capping wafer) sulkien rakennekiekon hermeettisesti. Esimerkki anturin sisäisen rakenteen poikkileikkauksesta nähtävillä kuvassa 3.

Kuva 3. Esimerkki anturin rakenteesta. (Murata Electronics Oy, 2021, sisäinen materiaali)



Tyypillisesti koottu kiekko sisältää jopa tuhansia prosessoituja anturielementtejä, jotka prosessin lopussa sahataan irti toistaan irrallisiksi anturielementeiksi. Yksinkertaistettu esimerkki anturielementin valmistusprosessista näkyy kuvassa 4.

Kuva 4. Anturielementin valmistusprosessi. (Murata Electronics Oy, 2021, sisäinen materiaali)



Anturielementtien ulkonäkö vaihtelee tuotetyypin ja prosessitekniikan mukaan, kuten on nähtävissä kuvassa 5.

Kuva 5. Erilaisia anturielementtejä. (Murata Electronics Oy, 2021, sisäinen materiaali)

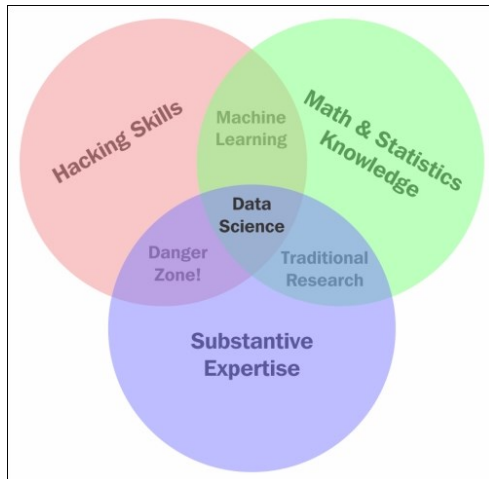


## 2 Datatiede

Teoriaosuudessa tarkastellaan yleisesti datatieteen ominaisuuksia ja prosessia, sekä paneudutaan yksityiskohtaisemmin analyysityökalussa käytettävien visualisointien ja analyysien teoriapohjaan.

Drew Conwayn (Conway, 2010) mukaan datatiede voidaan määritellä kuvan 6 mukaisesti erilaisten osaamisalueiden leikkauskohtana.

Kuva 6. Datatieteen Venn-diagrammi. (Conway, 2010)



Eri osaamisalueista Conway (Conway, 2010) ja Ozemir (Ozdemir, 2016, The data science Venn diagram-luku) kirjoittavat seuraavasti:

- **Hacking skills**, vapaasti suomennettuna ohjelmointitaidot. Jotta dataan päästään käsiksi ja sitä saadaan muokattua sekä analysoidua, ohjelmallisesti tai algoritmikirjastoihin tukeutuen, vaaditaan nykyaikana jonkin ohjelmointikielen tai ohjelmiston osaamista.
- **Math and statistics knowledge**, matematiikan ja tilastian osaaminen. Kun data on saatu hankittua ja käsiteltyä, tarvitaan tilastian osaamista, jotta datan välittämä viesti saadaan esille, määrittämällä oikeat algoritmit ja datamallit datan analysointiin. Huippuosaaja ei tässä tarvitse olla, mutta vähintään perustason tietämystä vaaditaan.
- **Substantive expertise**, substanssiosaaminen. Liiketoiminnan kannalta välttämätön substanssiosaaminen mahdollistaa sen, että osataan esittää oikeat kysymykset data-analyysiä varten, pystytään arvioimaan ovatko saadut tulokset järkeviä sekä omataaan riittävä kompetenssi tehdä päätöksiä tulosten hyödyntämistä varten.

Ilman substanssiosaamistakin lopputulos voi olla hyvä, mutta jotta saadaan tieteellistä kehitystyötä tehtyä ja kysymyksenasettelua hyödynnettyä, on substanssiosaaminen tärkeä

osa datatiedettä. Toisaalta jos puuttuu tilastitiikan ja matematiikan osaaminen ollaan vaarallisella alueella. Saattaa löytyä osaamista manipuloida dataa ja suorittaa erilaisia analyysejä. Omataan ymmärrys kysymyksistä mitä halutaan selvittää, mutta koska ei tiedetä analyytien taustalla olevaa tilastitiikkaa, saatetaan päätyä tekemään isojaakin virhepäätelmiä. (Conway, 2010)

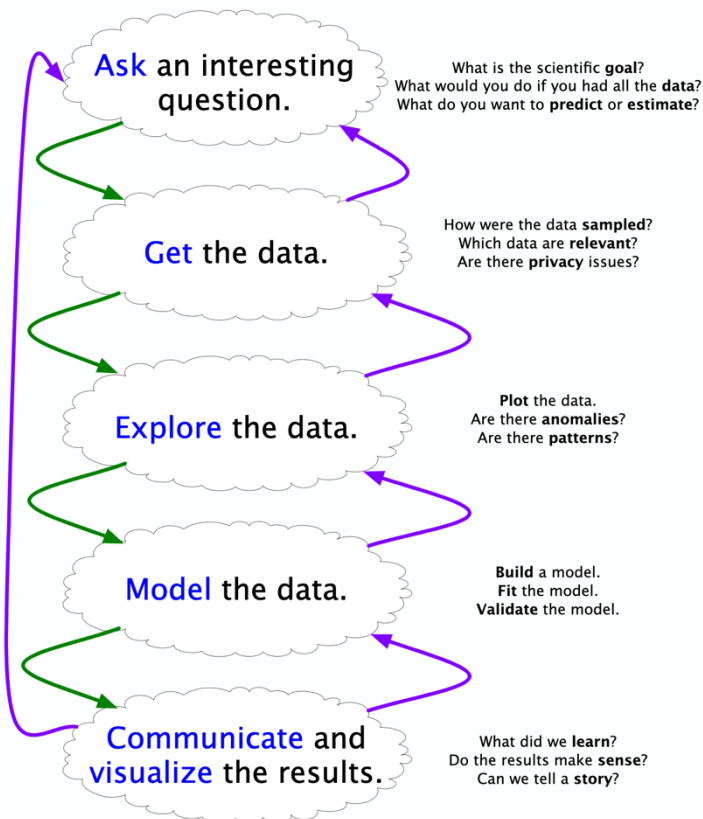
Conwayn laatima diagrammi on saanut myös kriittistä keskustelua osakseen, jossa muun muassa kritisoidaan substanssiosaamisen roolia diagrammissa. Esimerkkinä Michael Driscollin (Driscoll, 2012) järjestämässä väittelyssä, jossa myös Conway oli yhtenä osallistujana, keskustelijat olivat sitä mieltä, että taitava datatieteilijä pystyy hankkimaan riittävän substanssiosaamisen aiheeseen perehtyessään.

Automaattisella analyyesityökalulla pyrin taklaamaan osaltaan käyttäjien ohjelmointitaidon ja datan hankinnan resurssien puutteita automatisoimalla tietojen haun. Lisäksi valitsemalla sopivat analyytit ja antamalla ohjeet tehtyjen analyytien tulkintaan, avustan tilastitiikan aiheuttamien haasteiden osalta. Analyyesityökalun käyttäjäryhmät voivat edustaa hyvin erilaisia teknologia-alueita, joilla on alueensa paras substanssiosaaminen ja osaavat täten muodostaa oikeat johtopäätökset ja toimintasuositukset analyytitulosten perusteella.

## **2.1 Datatieteen prosessi**

Datatieteen prosessia voidaan kuvata esimerkiksi Harvardin Introduction to data science-kurssin kuvassa 7 näkyvillä vaiheilla, joiden järjestys ei kuitenkaan ole yksisuuntainen ja staattinen, vaan tarvittaessa edeltäviin vaiheisiin voidaan palata ja niitä parantaa. (Harvard, 2015, 1. luento)

Kuva 7. Datatieteen prosessi. (Harvard, 2015, 1. luento)



Tärkeänä lähtökohtana datatieteen prosessille on kysymyksen asettelu. Tulee määritellä tarkasteltavan asian tieteellinen tavoite, mitä prosessilla halutaan saavuttaa, mihin ongelmaan tai kysymykseen halutaan saada vastaus. (Harvard, 2015, 1. luento)

Pidemmälle vietyinä voidaan myös tarkastella datatieteen projektin läpiviennin vaatimia kustannuksia, mahdollisia hyötyjä yritykselle tai vastaavasti riskianalyysiä aiheeseen liittyen. Joka tapauksessa on kysyttävää olennaista määritellä mikä on kriteeri onnistuneelle prosessille, eli mietitään vastaus kysymykseen mitä tällä prosessilla halutaan saavuttaa. (Lee, 2018)

Seuraavassa vaiheessa keskitytään datan hankkimiseen, esikäsittelyyn ja siivoamiseen. Määritellään mistä lähteistä ja missä muodossa dataa on saatavilla, mitkä ovat vaatimukset datalle ja onko data riittävän laadukasta käytettäväksi. Datan esikäsittely ja laadunvarmistus ovat erittäin aikaa vieviä toimenpiteitä ja aiheuttavat usein tyytymättömyyttä analyytikoissa. (Skiena, 2017, s. 57)

Kun tietolähteet on määritelty ja data on kerätty, siirrytään datan tutkivaan analysointiin. Tutustutaan dataan, etsitään poikkeamia, hahmotetaan datan ominaisuuksia ja etsitään datasta säännönmukaisuuksia ja riippuvuuksia. (Harvard, 2015, 1. luento)

Seuraavaksi rakennetaan halutut mallit datan analysointia varten, suoritetaan analyysit ja todennetaan malli paikkansapitäväksi. Käytettävät analyysit ja ennusteet riippuvat alussa asetetusta tavoitteista, halutaanko ennustaa vai vain mallintaa lopputulosta. (Harvard, 2015, 1. luento)

Erittäin tärkeä vaihe on lopulta kommunikoida ja visualisoida datatieteen prosessin tuloksena syntyvät johtopäätökset eteenpäin ja mahdollisesti tehdä päätöksiä tulosten perusteella. Substanssiosaamisen avulla voidaan päätellä ovatko saadut tulokset järkeviä ja miten niitä parhaiten voisi hyödyntää. (Harvard, 2015, 1. luento)

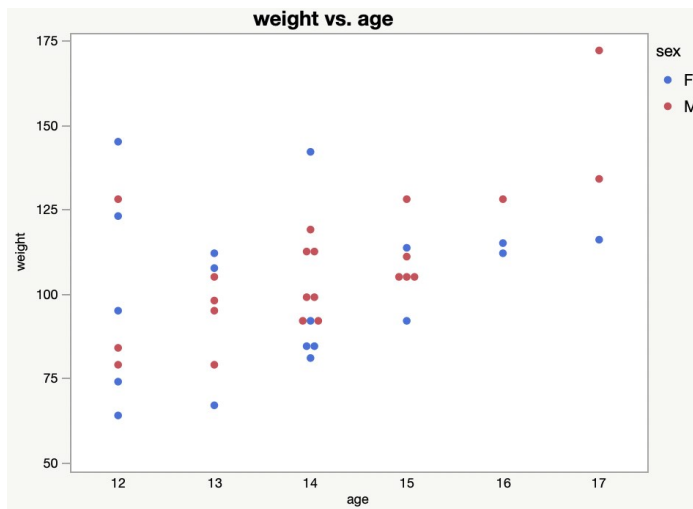
## **2.2 Tutkivaa data-analyysiä**

Visualisoimalla dataa monella eri tavalla voidaan hahmottaa datan ominaisuuksia ja ryhmien välisiä eroja. Analyysimenetelmien valintaan vaikuttaa osaltaan analysoitavan datan tyyppi, onko kyseessä numeerista jatkuvaa dataa, numeerista kategorista dataa vai esimerkiksi nominaalista tekstimuotoista dataa. Analyysityökalussa dataa visualisoidaan useammalla eri tyylillä, näistä kuvausta seuraavaksi.

### **2.2.1 Pistekaavio**

Pistekaavio näyttää kahden tai useamman muuttujan välisen suhteen, mikäli muuttujat ovat numeerisia. Pistekaaviosta voi nopeasti nähdä korreloivatko muuttujat keskenään. Kuvassa 8 pistekaaviolla osoitetaan iän ja painon välistä suhdetta, havainnollisuuden lisäämiseksi on lisätty väritys sukupuolen mukaan. Analyysityökalussa käytämme useimmiten pistekaaviota siten, että x-akselin arvo on kategorinen tai nominaalinen kuten mittapisteen nimi, eli käytämme kaaviota havainnollistamaan yksittäisten datapisteiden arvoja, tavoitteena saada nopea käsitys mahdollisista poikkeavista arvoista.

Kuva 8. Pistekaavio



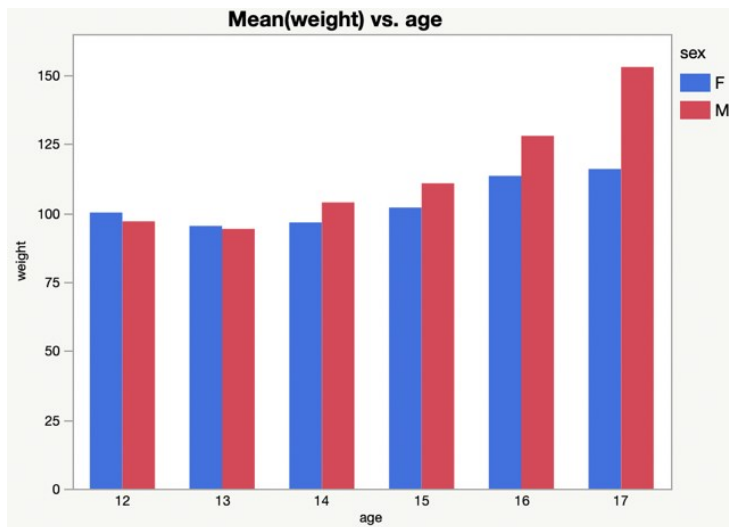
### 2.2.2 Pylväskaavio, histogrammi ja pareto

Nämä kaaviotyypit saattavat muistuttaa toisiaan ulkonäöllisesti, mutta eroavat hieman ulkoasultaan ja omaavat erilaiset vaatimukset visualisoitavalle datatyypille.

Pylväskaaviota käytetään tyypillisesti visualisoimaan kumulatiivisia lukumääriä tai prosentteja, arvoa voidaan myös muokata keskiarvon, moodin tai mediaanin näyttämiseen. Pylväskaaviota käytetään esittämään kategorista ja nominaalista dataa. Pylväskaavion tolpat ovat tyypillisesti hieman irti toisistaan, vaikka asetuksilla ne saadaankin myös täyttämään tyhjät välit. Jatkuvan datan esittämiseen pylväskaaviot eivät sovellu. Kuvassa 9 on esitetty keskimääräinen paino ikäluokittain, ryhmittelemällä saatu vertailua sukupuolten välillä. (SAS Institute Inc, n.d.-a)

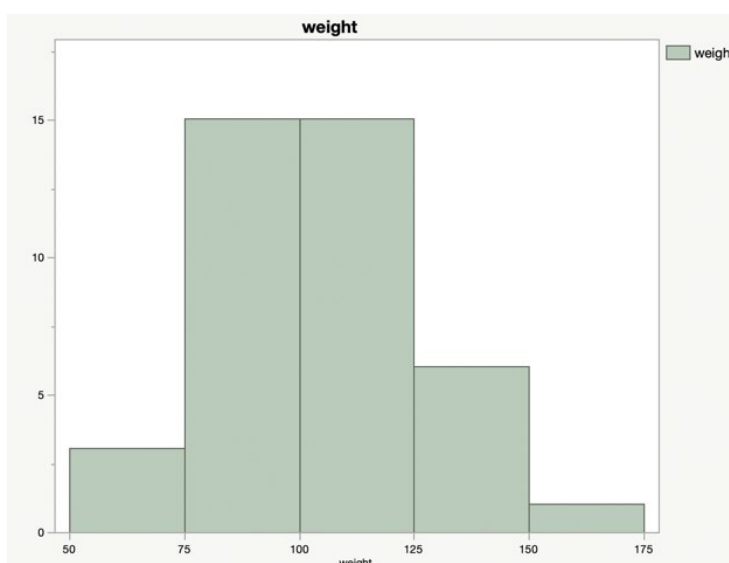


Kuva 9. Pylväskaavio.



Histogrammissa kuvaajan tolpat ovat kiinni toisissaan ja kaaviota käytetään numeerisen jatkuvan datan esittämiseen. Datan jakaumaa pystytään tarkastelemaan histogrammia hyödyntämällä. Histogrammissa data esitetään valitulla luokkajaolla, palkkien korkeus kuvaa kyseiseen väliin kertyvää havaintojen määrää. JMP valitsee luokkajaon automaattisesti, mutta sitä voi muuttaa käsin halutessaan. Kuvassa 10 painon jakauma esitetty histogrammin avulla, dataväleiksi on automaattisesti valikoitunut 25 paunaa. (SAS Institute Inc, n.d.-b)

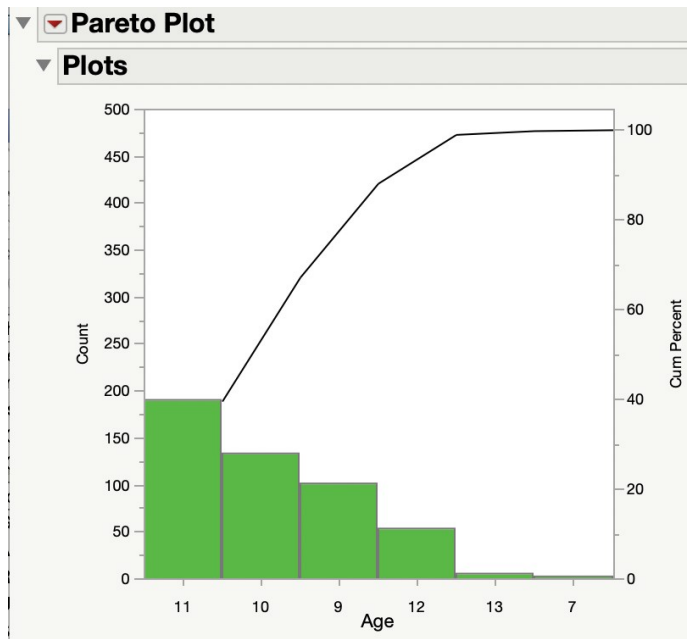
Kuva 10. Histogrammi.



Pareto-kaavio on pylväskaavion erityismuoto, jossa nominaalisen tai kategorisen datan esittämiseen käytetään järjestettyä pylväskaaviota, jossa arvot on järjestetty

suuruusjärjestykseen suuremmasta pienempään vasemmalta alkaen. Lisäksi näytetään viivakaaviolla kumulatiivinen prosenttiosuus. Esimerkkikuvassa 11 lapsijoukon iät pareto-kaaviossa, eniten on 11-vuotiaita lapsia. (SAS Institute Inc, n.d.-c)

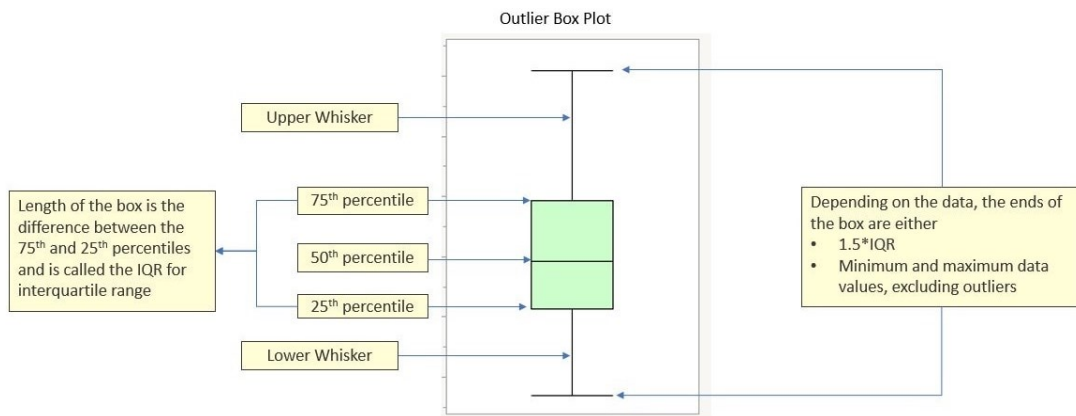
Kuva 11.Pareto-kaavio.



### 2.2.3 Laatikko-jana-kuvio

Laatikko-jana-kuvio auttaa hahmottamaan datan keskikohtaa ja hajontaa. Kuvio muodostetaan kuvan 12 mukaisesti, jolloin laatikon keskellä oleva viiva kuvaa datajoukon mediaania, muuten data jaetaan kvartaaleittain siten että laatikon sisällä on toinen ja kolmas kvartaali, janat kuvaavat ensimmäistä ja neljättä. Kokonaan kuvion ulkopuolelle jäävät poikkeavat havainnot. (Sas Institute Inc, n.d.-d)

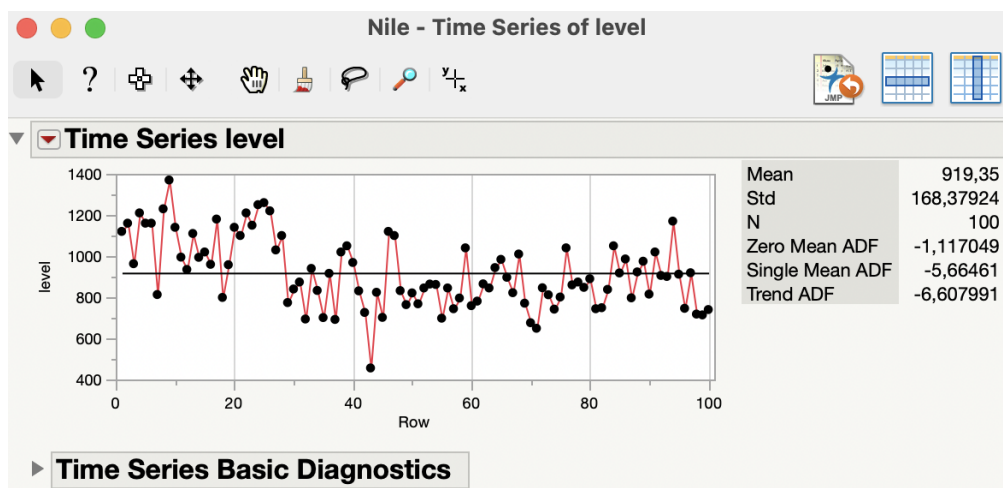
Kuva 12. Laatikko-jana-kuvio. (Sas Institute Inc, n.d.-d)



## 2.2.4 Aikasarja

Aikasarja-analysissä datapisteet esitetään ajan mukaan järjestettyinä, kuva 13. Kaavio on hyödyllinen tarkastellessa arvojen trendin muuttumista ajan yli. Datan tulee olla järjestettynä datataulussa, tai kuvaajan luontivaiheessa annetaan ohjelmalle tieto sarakkeesta, jonka mukaan data järjestetään.

Kuva 13. Aikasarja.



## 2.3 Hypoteesien testaus

Määrälliset analyysit perustuvat usein tilastollisten hypoteesien testaukseen. Yleensä analyttikko perustaa hypoteesinsa ennako-oletuksensa vastakohtaan. Eli jos ennako-

odotus ryhmien välisessä vertailussa on, että ryhmien keskiarvot eroavat toisistaan, hypoteesi laaditaan oletukselle, että ryhmien keskiarvot eivät eroa toisistaan. Tätä perusolettamusta kutsutaan nollahypoteesiksi  $H_0$ , jolle laaditaan vastahypoteesi  $H_1$ , mikä hyväksytään, jos nollahypoteesi saadaan kumottua tilastollisesti. (Tampere: Yhteiskuntatieteellinen tietoaarkisto [ylläpitäjä ja tuottaja], n.d.)

Verrattaessa kahden ryhmän keskiarvoja toisiinsa, voimme merkitä ryhmän 1 keskiarvoa symbolilla  $\mu_{gr1}$  ja ryhmän kaksi symbolilla  $\mu_{gr2}$ . Näin saamme muodostettua hypoteesit tilastollisia testejä varten, alla olevan kuvan mukaisesti.

Kuva 14. Hypoteesit.

**Nollahypoteesi, ei eroa ryhmien välillä.**

$$H_0: \mu_{gr1} = \mu_{gr2}$$

**Vastahypoteesi, ryhmien välillä on eroa.**

$$H_1: \mu_{gr1} \neq \mu_{gr2}$$

Hypoteesitestausta varten määritellään hyväksyty riskitaso, yleisesti käytössä on 5 %:n raja hyväksyttävälle virhepäätelmän riskille, mutta periaatteessa rajat voidaan määritellä tilanteen mukaan. Tilastollisen testin tuloksena yleensä saadaan niin kutsuttu p-arvo, joka määrittelee virheellisen päätelmän todennäköisyyden, eli jos saadaan esimerkiksi p-arvoksi 0.05 on 5 %:n todennäköisyys, että asetettu nollahypoteesi jää voimaan. Käytännössä määritellyn riskirajan alittuessa, hylätään nollahypoteesi ja voidaan todeta, että vastahypoteesi sai tukea. Riskiraja on kuitenkin vain sovittu arvo, joten tuloksen tulkinta ja sen perusteella päätösten tekeminen edellyttää ymmärrystä suuruusluokkien merkityksistä. (Tampere: Yhteiskuntatieteellinen tietoaarkisto [ylläpitäjä ja tuottaja], n.d.)

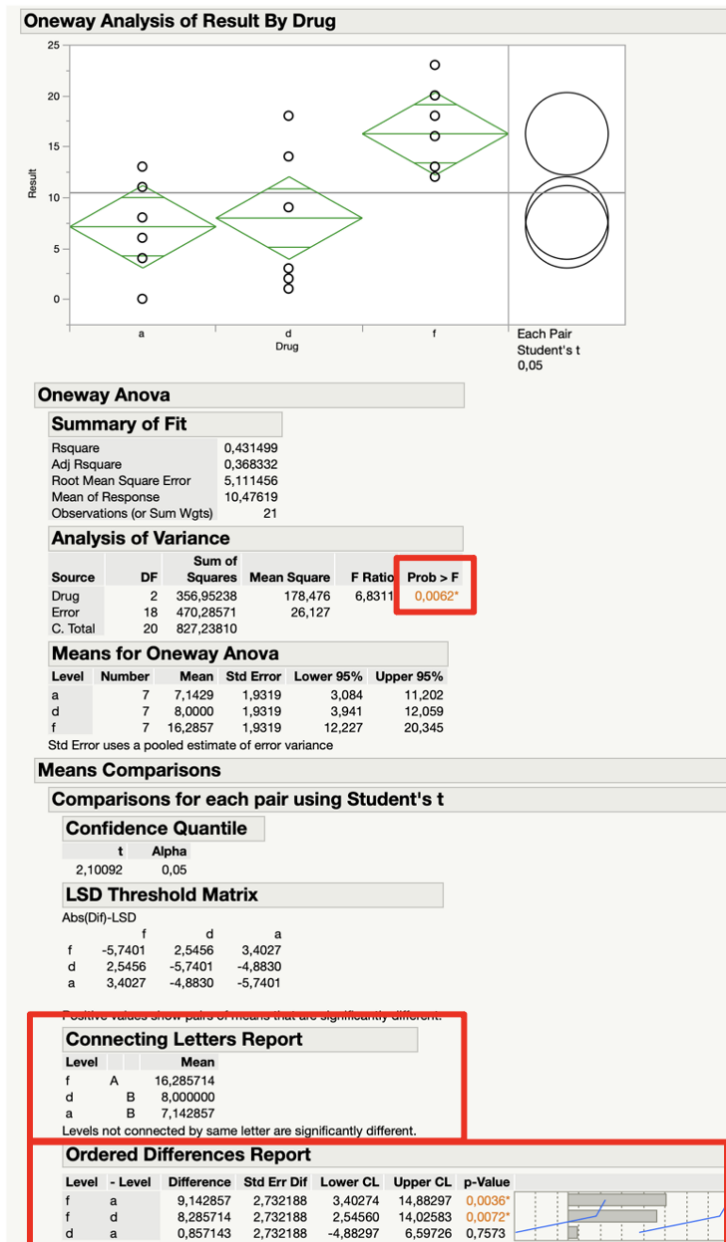
## 2.4 Ryhmien keskiarvojen vertailu, Student's t-test ja Oneway Anova

Ryhmien keskiarvojen vertailuun käytämme JMP:in Fit Y by X-alustaa ja datatyyppien mukaan JMP valitsee Oneway-analyysin verrattaville ryhmille. Anova-testausta pidetään

sopivimpana kolmen tai useamman ryhmän keskiarvojen vertailuun, mutta sitä on mahdollista käyttää myös kahden ryhmän väliseen vertailuun. Suositellumpi tapa on kuitenkin tarkastella Student's t-testin avulla eroja kahden ryhmän välillä. Tässä analyysiraportissa yhdistyvät molemmat analyysit. MFI järjestää säännöllisesti tilastitiikan koulutusta JMP-ohjelmalla, jolloin myös tämän analyysituloksen tulkinta käydään läpi. Tässä esimerkissä nollahypoteesi, jossa oletetaan, että ryhmien keskiarvoissa ei ole eroa, hylätään. JMP raporteissa näkyy myös kirjallisia ohjeita tulosten tulkintaan ja tilastollisesti merkittävät p-arvot värittyvät erottuvasti. (Sall ym., 2017, Comparing Many Means: One-Way Analysis of Variance-luku)

Alla olevassa kuvassa 15 ympyröity olennaisia arvoja analyysistä. Oneway Anovan p-arvo, joka tässä ilmoitetaan F-arvona, on hyvin pieni, reilusti alle tyypillisen 0.05 rajan, joten tämä kertoo, että nollahypoteesi on syytä hylätä, tilastollinen todennäköisyys ei tue sen voimassaoloa. Anova ei kuitenkaan kerro tarkempaa tietoa ryhmien välisistä eroista, ei saada pelkästään sen perusteella selville, miten ryhmät tilastollisesti poikkeavat toisistaan. Tätä ongelmaa taklaamme Student's t-testillä, joka vertailee pareittain ryhmiä. Tulokset näkyvät oman otsikkonsa alla, joista olennaisina olen nostanut esiin Connecting Letters Report-kohdan, jonka tulkintaohje lukeekin raportissa. Mikäli ryhmiä ei yhdistä sama kirjain, ovat ne tilastollisesti toisistaan poikkeavia. Tässä tapauksessa nähdään, että ryhmä f eroaa molemmista muista ryhmistä, koska se ei jaa kummankaan ryhmän kanssa kirjaimia. Havainnon vahvistaa Ordered Differences Report, joka näyttää erot suuruusjärjestyksessä, että tilastollisesti ryhmä f eroaa muista ryhmistä. (Sall ym., 2017, , Comparing Many Means: One-Way Analysis of Variance-luku)

Kuva 15. Oneway Anova raportti.



### 3 Analyysityökalun kehitysprojekti

Aloite analyysityökalun kehittämiseen tuli sidosryhmien edustajien toiveesta. Projektia varten muodostettiin projektiryhmä määrittelemään analyysityökalun tavoitteita sekä tuomaan omaa tietämystä ja valmiita käytäntöjään analyysien määrittelyyn. Ryhmä kokoontui tarvittaessa kahden viikon välein, kesäaikaan lukuun ottamatta.

### 3.1 Tavoitteet automaattiselle analyysityökalulle

Suunnittelupalavereissa määriteltiin seuraavat tavoitteet ja reunaehdot työkalun toiminnalle:

- Analyysityökalua käytetään anturielementtivalmistuksen mittausdatan analysointiin.
- Tavoitteena analysoinnin automatisointi ja työajan säästäminen.
- Pyritään analyysien yhdenmukaistamiseen yrityksen sisällä.
- Parannetaan data-analyysien luotettavuutta statistiikan osaamista lisäämällä.
- Mahdollistetaan kattavampien analyysien tekeminen.
- Työ toteutetaan JMP-ohjelmistolla.

Toteutettavia analyysejä mietittiin projektiryhmän kokouksissa ja toiveista sekä analyysejä tehneiden henkilöiden esimerkeistä koostettiin peruskokoelma visualisointeja ja analyysejä, jotka tulee saada automatisoitua analyysityökalussa. Toteutustavaksi valikoitui suoraan JMP ohjelmisto, kyseinen analyysiohjelmisto on yleisesti käytössä yrityksessä, ja analyysityökalun hyödyntämät tiedostot ovat JMP-tiedostoja. Analyysityökalun ohjelmoinnista JSL-skriptikielellä JMP-ohjelmistossa vastasin minä.

### 3.2 Jira

Projektinhallinta toteutettiin käytännössä Jira-tehtävienhallintatyökalun avulla, joka on yrityksen pääasiallinen työkalu projektinseurantaan. Kokousten muistiinpanot tallennettiin yleisiin kommentteihin ja erillisistä pienemmistä kokonaisuuksista luotiin alitehtäviä, joihin kerättiin tarvittavat tiedot kyseisen alitehtävän suorittamiselle. Tehtävienhallintatyökalun käyttö mahdollistaa tehtävien jakamisen useammalle ryhmän jäsenelle ja tehdyistä töistä jää historiatiedot ja materiaalit dokumentoiduiksi.

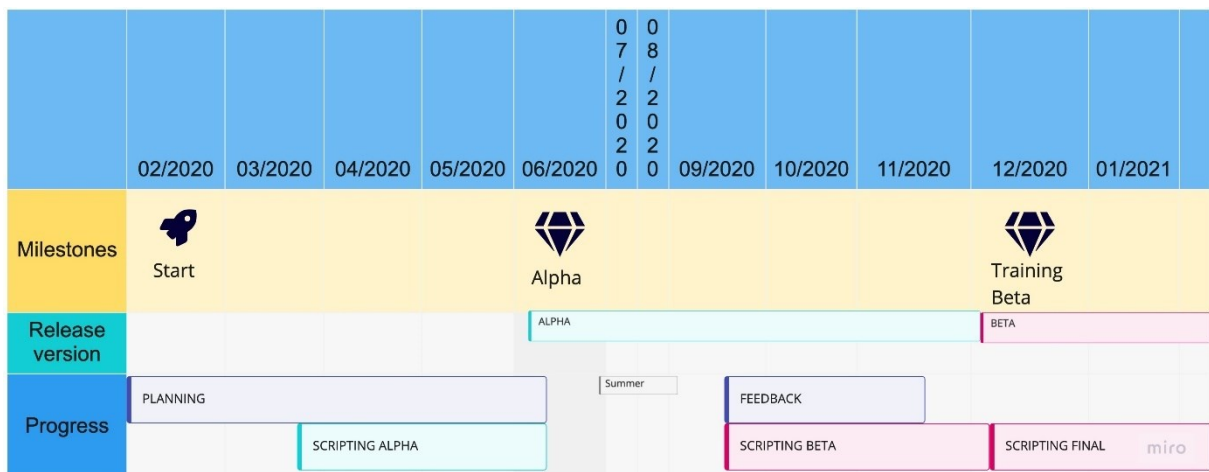
### 3.3 Projektin aikajana

Analyysityökalun kehitys jatkuu edelleen, ensimmäisen virallisen julkaisuversion ominaisuudet on lyöty lukkoon, mutta korjaan koodista vielä muutaman ilmi tulleen bugin ja refaktoroin koodia hieman järkevämmäksi. Työkalun on tarkoitus olla jatkuvasti kehittyvä,

rakenteesta on tehty tarkoituksellisen modulaarinen, jotta työkalun laajennus onnistuu ongelmitta tarvittaessa.

Ajallisesti analyysityökalun kehitys on kestänyt noin vuoden BETA-versioon mennessä, kuten on nähtävissä kuvassa 16. Pitkä kehitysaika johtuu siitä, että työtä tehtiin vain osa-aikaisesti muiden töiden lomassa, prio ei ollut kiireellinen ja runsaasti aikaa varattiin tuotteen testaamiseen testikäyttäjien toimesta. BETA-versio julkistettiin koko MFI:n laajuiseen käyttöön joulukuussa 2020, jolloin pidettiin myös yleinen koulutusilaisuus ohjelman käytöstä ja tulosten tulkinnasta. Työkalun käyttökoulutus tullaan sisällyttämään Data Integration-tiimin tarjoamaan koulutusvalikoimaan jatkossakin.

Kuva 16. Projektin aikajana.



## 4 JMP ohjelmisto

JMP on SAS-yrityksen kehittämä ohjelmisto data-analytiikkaa varten. JMP mahdollistaa datan tuomisen lukuisista erilaisista lähteistä verkkosivuista tietokantoihin. JMP tarjoaa kattavat menetelmät datan siivoamiseen ja järjestämiseen ja visualisointiin. Runsas valikoima erilaisia analyysimenetelmiä helppokäyttöisellä käyttöliittymällä mahdollistavat myös vaativien analyysien ja koesuunnitelmien laatimisen. Mahdollisuus integroida JMP esimerkiksi Pythonin ja R:n kanssa lisää ohjelmiston joustavuutta. JMPille ominaista on datan linkittyminen valmiisiin raportteihin, joten tehdyt muutokset näkyvät dynaamisesti ja ovat vuorovaikutteisia. (Sas Institute Inc, n.d.-e)



JSL, JMP Scripting Language on JMPin oma käännetty skriptikieli, joka mahdollistaa perusohjelmoinnin algoritmien lisäksi JMPin objektien käsittelyn ja suorituksen automatisoinnin ohjelmoimalla. Suurin osa käsin suoritetuista toimenpiteissä ohjelmassa on mahdollista saada kopioitua talteen, jonka jälkeen tallennetun ohjelmakoodin voi hyödyntää haluamallaan tavalla. Yleisin käyttökohde on rutiininomaisten toimenpiteiden automatisointi, mutta lisäämällä perinteisiä ohjelmallisia algoritmeja voidaan saavuttaa jotain sellaista, joka ei edes olisi käsin tehdessä mahdollista. (Sas Institute Inc, n.d.-e)

## 5 Tietolähteet

Kaikkea analyysityökalun tarvitsemaa dataa ei kannata hakea tietokannasta, mikäli se on mahdollista saada valmiiksi koostettuna. Jotta suoria aikaa vieviä kantahakuja olisi mahdollisimman vähän, päädyin hybridimalliin datan keruussa. Jos dataa on löydettävissä hyödynnettävässä muodossa Standard dataseiteistä, se hyödynnetään ja tietokannoista haetaan vain puuttuva data.

### 5.1 Standard data

MFI:llä kerätään huomattava määrä dataa tuotteisiin liittyen. Tuotannonohjausjärjestelmä tallentaa yksityiskohtaista tietoa tuotteiden historiasta niiden kulkiessa tuotantovuon läpi, tietokantaan muun muassa tallentuu tieto tuotetta edistäneestä operaattorista, mille laitteelle tuote on ladattu ja milloin. Käytännössä tuotteen kulku tuotannossa on hyvin tarkkaan dokumentoitua, kaikkeen tuotteen käsittelyyn liittyy tuotteen edistäminen tuotannonohjausjärjestelmässä ja myös erilaisia parametreja kirjataan käsin tai laitteiden toimesta järjestelmään. Etenemishistorian lisäksi tuotteista kerätään tarkkaa mittaushistoriaa erilaisista visuaalisista ja sähköisistä mittauspisteistä kaikilta kiekkotyypeiltä (ohut kiekko, paksu kiekko, koottu kiekko). Tämän kerättävän datan määrä on niin suuri, että suoraan tietokannasta tehtävät haut saattavat olla hyvin pitkäkestoisia. Pahimmillaan, jos kyseessä on pidemmän aikavälin haku, ei hakeminen suoraan tietokannasta ole edes mahdollista, teknisten rajoitusten takia. Lisäksi haettava tieto voi vaatia hyvin monimutkaisen haun ja vaatia syvällistä tietämystä tietokannan tietorakenteista ja tiedon tallentumisen logiikasta. Tämän johdosta varsinkin tuotteen historiatietojen hakeminen saattaa osoittautua haastavaksi.

Jotta datan käsittelyä saataisiin nopeutettua ja haettu tieto olisi luotettavaa ja oikeassa prosessijärjestyksessä, muodostetaan tuotteista standardeja datasettejä analysointitarkoituksia varten. Standardeissa dataseiteissä tuotteiden datan analysointiin tarvittava tieto on kerätty ja kerrytetty pidemmältä ajalta valmiiksi ja yhdistetty valmiiksi loogisiksi kokonaisuuksiksi. Lisäksi on koostettu erilaisia lisätietoja tarjoavia datasettejä, jotka voidaan tarvittaessa yhdistää esimerkiksi kootun kiekon historiatietoihin.

Analyysityökalu hyödyntää standardissa datassa saatavilla olevaa PROBER-mittapisteen raakadataa (RAW DATA), joka sisältää anturielementtitason mittausdataa hylkytiedoista ja mittausparametrien arvoista. Käytännössä kootun kiekon jokaiselle anturille tallentuu mittapisteessä sadoittain erilaisia mittaustuloksia.

Standardit datasetit muodostetaan JMP-ohjelmalla, data kerätään SQL-kyselyillä tietokannoista, dataa siivotaan ja käsitellään, yhdistetään ja lopulta tallennetaan JMP-datatiedostoiksi.

## 5.2 Tietokannat

Analyysityökalu kyselee suurimman osan datasta suoraan ORACLE-tietokannoista SQL-hakujen avulla, Mittadatasta haetaan kiekko kohtaista summadataa mittapisteiden tuloksista, tietoa tuotteen testausseksistä ja lisäksi haetaan tarvittaessa anturielementtitason dataa PROBER-mittapisteestä, mikäli standardi datasetti ei kata kaikkia haluttuja kiekko koodeja. PDAM-tietokantaan tallentuu MES-tuotannonohjausjärjestelmän data, josta haetaan tieto muun muassa kootun kiekon komponenteista, vuoversioista ja kiekon tämänhetkisestä sijainnista.

## 6 Analyysityökalun testaus kehitysprojektin aikana

Jotta voin varmistua siitä, että analyysityökaluun luomani logiikka toimii suunnitellusti, ohjelmointiprosessin aikana testasin kattavasti ohjelman toimintaa kaikilla erilaisilla tuotetyypeillä, joiden analysointiin analyysityökalua kehitettiin. Käytännössä valitsin käsin kokoelman kiekko koodeja jokaisesta tuoteryhmästä, joita sitten käytin systemaattisesti uusien ominaisuuksien testaamiseen. Myös analyysityökalun hakemat tulokset tarkistin

käsin hakemalla vastaavat tiedot muita ohjelmia käyttäen, näin pystyin varmistumaan siitä, että tietokannoista hakemani tiedot olivat paikkansapitävät.

Kattava testaus oli aikaa vievää, mutta välttämätöntä. Lisähidasteena toimi etätyön aiheuttama haaste, JMPiin sisäänrakennetun debuggaus-työkalun toimimattomuus. Käytännössä VPN-etäyhteyden ylitse ei käyttämällämme laitekonfiguraatiolla ollut mahdollista saada työkalua toimimaan, joten paljon testausta joutui tekemään myös testitulostuksia ja lokikirjauksia hyväksikäyttäen. Valikoitu käyttäjäryhmä testasi analyysityökalua omien analyysiensä suorittamiseen, saatu palaute oli arvokasta ja auttoi osaltaan muokkaamaan analyysityökalun ominaisuuksia ja poistamaan bugeja koodista.

## **7 Automaattinen analyysityökalu**

Haastavin osuus analyysityökalun toteutuksessa oli analyysien perustana olevien datasettien muodostaminen niin, että samaa ohjelmakoodia voidaan käyttää kaikkien erilaisten tuotteiden analysointiin. Datasetillä tässä yhteydessä tarkoitetaan eri lähteistä koostettua tietuetta, joka esitetään JMP-taulukkona. Tuotteita on useampaa eri sukupolvea ja tuotetyyppejä, joka aikaansaa sen, että tietokantahauissa on huomioitava sekä tuotteen tyyppi, että siihen kuuluva datahaun logiikka. Kun datasetit on saatu muodostettua, on analyysien ja visualisointien lisääminen valikoimaan melko suoraviivaista.

Käyn jatkossa läpi analyysityökalun ohjelmalogiikkaa, liitän sekaan pieniä esimerkkejä ohjelmakoodista, jotta JSL-kielen toiminnallisuus havainnollistuu esimerkinomaisesti. Koko ohjelmakoodin läpikäynti ei ole tarkoituksenmukaista eikä mahdollistakaan ohjelmakoodin pituuden huomioiden. Esittelen myös lyhyesti malliesimerkit JMPilla tehdyistä analyyseistä ja visualisoinneista.

### **7.1 Analyysityökalun käyttökohteet**

Analyysityökalu hakee tietoa anturielementtivalmistuksen mittapisteistä, PROBER-mittapisteiden data haetaan anturielementtitasolla, muut mittapisteet käsitellään kiekkotason summa- tai keskiarvotasolla. Analyysityökalu on kehitetty nimenomaan vertailevan analyysin tekemiseen, eli kun halutaan analyttisin keinoin selvittää, eroaako kaksi tai useampi

datajoukkoa toisistaan. Työkalua voi käyttää myös ilman vertailua ryhmien kesken, mikäli halutaan vain visualisoida kerättyä mittausdataa.

Tarvetta vertailevaan analyysiin on yleensä erilaisten kokeiden tulosten analysoinnin yhteydessä. Lisäksi voidaan esimerkiksi tarkistaa mittaustuloksia ennen ja jälkeen laitemuutosten, tuoteversioiden muutosten tai uuden laitteen käyttöönoton yhteydessä. Voidaan myös vertailla eri aikajaksojen dataa toisiinsa tai eri lähtömateriaalien eroja mittaustuloksissa.

## **7.2 Rajoitukset**

Tarkasteltavien kiekkojen lukumäärä on toistaiseksi rajoitettu SQL-hakuteknisistä syistä, maksimimäärät ilmoitetaan käyttäjälle käyttöliittymässä. Tämä ominaisuus on tarkoitus korjata seuraavassa ohjelmaversiossa, jos määrärajoituksen havaitaan aiheuttavan ongelmia. Tarkasteltavien kiekkojen tulee olla myös vertailukelpoisia keskenään, eli käyttäjän täytyy tietää mitä tuotteita voidaan tarkastella yhtä aikaa työkalussa, käytännössä rajoituksia tulee PROBER-mittapisteen testispeksistä sekä tuotteiden kiekkokartoista.

## **7.3 Käyttäjän antamat lähtötiedot ja valmiit asetukset**

Analyysityökalun käyttäminen on tehty mahdollisimman yksinkertaiseksi käyttäjälle. Käyttäjä antaa kysyttäessä JMP-taulukon, jossa on lueteltu analyysiin sisällytettävät kiekot, sekä informaatio niiden mahdollisesta ryhmittelystä vertailtaviin ryhmiin. Ohjelmakoodin vaatimuksesta taulukon sisällön tulee olla vakiomuotoinen, eli kolumnien on oltava täsmälleen oikein nimettyjä. Käyttäjälle on tarjolla valmis mallipohja, johon käyttäjä voi omat tietonsa lisätä. Tämä on ainoa pakollinen tieto, joka käyttäjän tulee tarjota ohjelmalle.

Kiekkojen etsiminen analyysejä varten on monimutkaista ja valittavat kiekot riippuvat täysin siitä, minkälaisen kysymyksen käyttäjä on halunnut analyysiä varten esittää. Tässä analyysityökalun ensimmäisessä versiossa päädyin siihen, että käyttäjä valitsee kiekot itse. Analyysityökalun käyttöliittymässä käyttäjä myös valitsee tarvittavat visualisoinnit ja analyysit. Osa tietokantahakujen tarvitsemista tiedoista haetaan asetustiedostosta, johon on

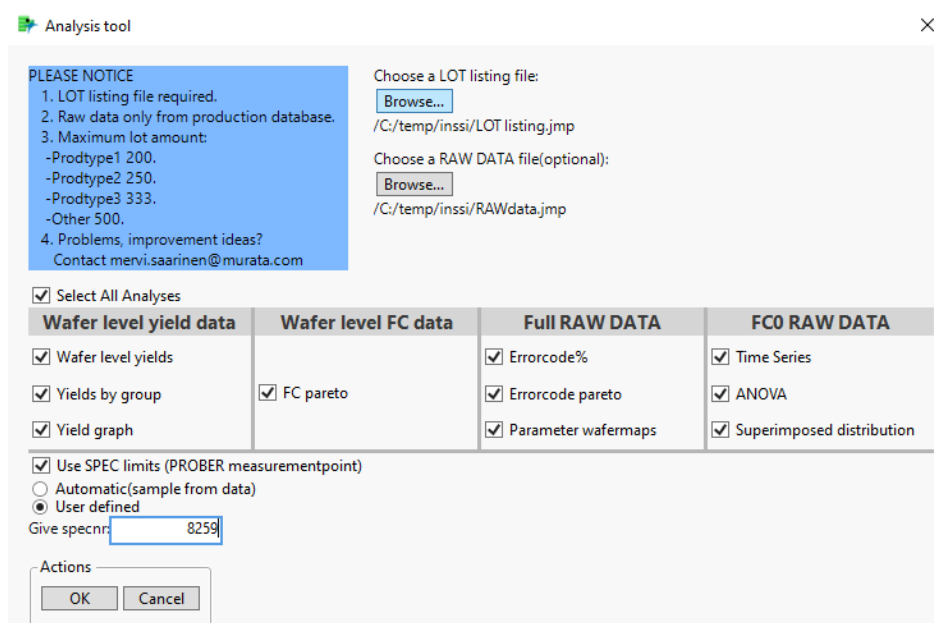
sisällytetty esimerkiksi tietoa tuotteiden karttakoordinaateista sekä mittalaitteen testausparametrista.

## 7.4 Käyttöliittymä

Analyysityökalua käynnistettäessä käyttäjälle tarjotaan graafinen käyttöliittymä, kuvassa 17, jossa annetaan lyhyt tiedonanto ohjelman tarvitsemista tiedoista ja ilmoitetaan voimassa olevat rajoitukset. Käyttäjää pyydetään valitsemaan tiedosto, johon on koostettu tiedot analysointiin haluttavista kootuista kiekkoista ja niiden ryhmittelystä. Lisäksi käyttäjä voi valita standardin RAW DATA-tiedoston anturielementtitason mittausdatan kerryttämistä varten. Käyttäjää saa valittavakseen listauksen analyysistä ja visualisoinneista. Lisäksi voi tehdä valinnan sisällyttää PROBER-mittapisteen mittaspeksin tiedot mukaan analyysieihin, näin tehden JMP piirtää automaattisesti rajat näkyviin kuvaajiin ja raportteihin. Speksinumero voidaan hakea joko näytepohjaisesti käyttäjän antaman kiekkoilistauksen ensimmäisen kiekon tietojen perusteella, tai käyttäjä voi määrittellä numeron käsin.

Ohjelmassa tarkastetaan, että käyttäjä on antanut kiekkoilistauksen ja että kolumninimet vastaavat vaatimuksia. Ohjelmakoodiin on myös lisätty tarkastus kiekkomäärän lukumäärää varten, mikäli kiekkomäärä ylittyy, asiasta ilmoitetaan käyttäjälle ja ohjelman suoritus keskeytetään.

Kuva 17. Käyttöliittymä.

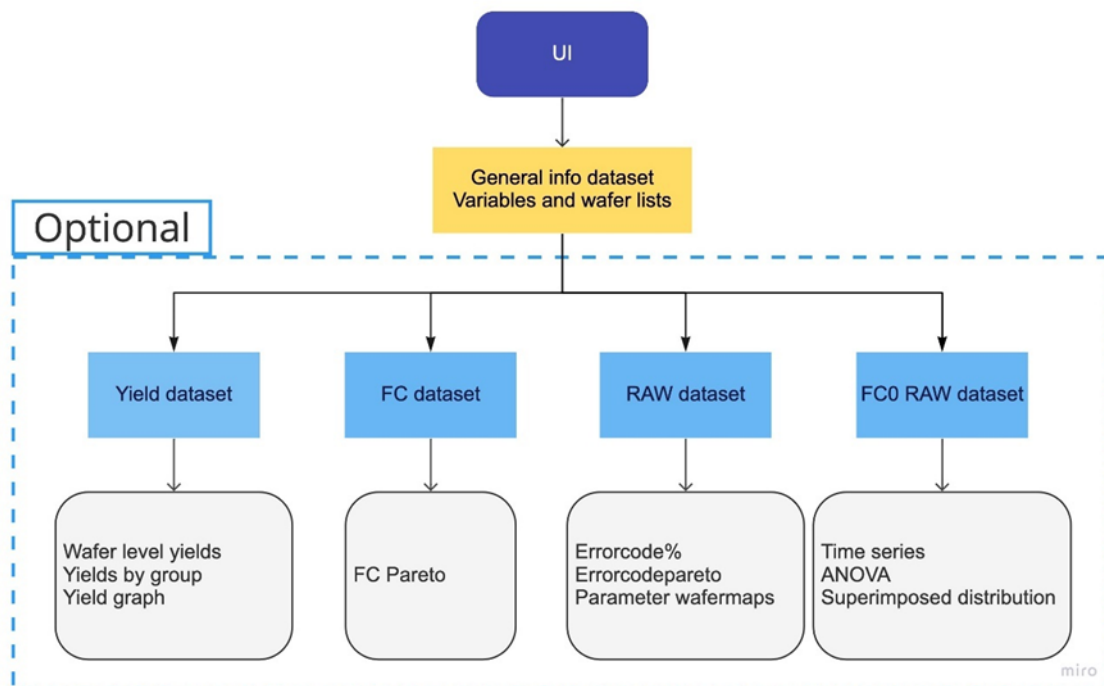


## 7.5 Ylätason logiikka

Analyysityökalun yhtenä periaatteena on modulaarisuus ja turhan tiedonhaun minimointi. Ainoa pakollinen haettava datasetti on yleinen infotiedosto, johon tärkeimpänä tietona haetaan koottuun kiekkoon sisältyvien ohut- ja paksukiekkujen tiedot, mukaan lukien kiekkokoodit. Näin haettuja tietoja käytetään muiden datasettien muodostamisessa, koska ohut- ja paksukiekoilla on myös omia mittapisteitä, joiden tulokset ovat tallentuneet tietokantaan niiden omille kiekkokoodeille. Tässä vaiheessa muodostetaan mahdollisia tulevia tietokantahakuja varten useampia kiekkokoodilistauksia.

Käyttäjän valinnat määrittelevät, mitkä datasetit koostetaan. Mikäli datasettiin liittyviä analyysejä ei valita, niin datasettiä ei turhaan muodosteta. Mahdolliset datasetit ja visualisoinnit sekä analyysit on esitelty alla olevassa kuvassa.

Kuva 18. Analyysityökalun ylätason logiikka.



## 7.6 Tulosten esittäminen JMP projektinäköymässä

Datanmuokkauksessa ja data-analyysien teossa generoituu yleensä suuri määrä avoimeksi jääviä taulukoita tai analyysiraportteja. Näkymä voi olla melko kaoottinen, avoinna voi olla kymmeniä ikkunoita, joiden nimi on kuvaavasti tyyliin Untitled 35. Analyysityökalussa tämä

ongelma haluttiin estää ja analyysin aikana tarpeettomaksi jääneet ikkunat suljetaan ja olennaiset ikkunat sijoitetaan JMP projektin sisälle. Kaikki datataulut ja raportit halutaan säilyttää täysin toimintakuntoisena, jotta käyttäjä voi halutessaan jatkaa analyysien tekoa. JMP projekti oli luontevin tapa tallentaa kaikki käytetty materiaali ja tulokset, tämä onnistui vaivattomasti ajamalla suurin osa ohjelmakoodista projektin sisältä käsin.

Kun analyysityökalu on saanut analyysit tehtyä, käyttäjälle avautuu JMP:n projekti-ikkuna, esimerkki kuvassa 19. Kaikki valitut visualisoinnit, analyysit ja yleinen infosivu ovat sijoitettuna omiin välilehtiinsä, ikkunan vasemmassa reunassa näkyy Workspace-niminen laatikko, jonka sisällä näkyy listana raporttiin sisältyvät datasetit ja niihin liittyvät raportit. Osa datatauluista on piilotettu välilehdiltä tarkoituksellisesti, jolloin ne näkyvät listalla vaaleanharmaalla värillä. Tuplaklikkaamalla tiedoston nimeä sen saa näkyville välilehteen.

JMP projektin voi tallentaa projektina ja pakata tallennustilan säästämistä varten. Näin kaikki analyysit ja alkuperäiset datataulut jäävät haluttaessa dokumentoiduiksi.

Kuva 19. JMP projektinäkömä.

The screenshot shows the JMP Pro interface with a project workspace. The main window displays a data table with the following columns: DISCNR\_probe, Combined\_WFR, GROUP, COMBINEDWFR\_THNWFR, THICKWFR\_1, and THICK\_LOT. The data rows are numbered 1 through 25. The workspace sidebar on the left shows a list of reports and data sources, including 'Wafer info', 'Yield info', 'FC DATA', 'RAW data', 'Errorcode% by group', 'Errorcode pareto by group', 'Wafermaps', 'FCO RAW Data', 'FCO parameter Time Series', 'Fit Group', 'Superimposed Distribution', and 'Wafer info Anonymized'. The 'Columns (15/0)' section shows a list of columns, including 'DISCNR\_probe', 'Combined\_WFR', 'GROUP', 'COMBINEDWFR\_THNWFR', 'THICKWFR\_1', 'THICK\_LOT', 'SUPPLIER\_CODE\_K', 'SUPPLIER\_CODE\_O', 'Combined\_W...OD\_VERSION', 'THICKWFR\_1...OD\_VERSION', 'SENSORTYPENR', 'EXPTNUM', 'NOTES', 'WIPSTATUS', and 'CURRENT\_M...ENTPOINTNR'.

	DISCNR_probe	Combined_WFR	GROUP	COMBINEDWFR_THNWFR	THICKWFR_1	THICK_LOT
1	V1_0	V2_0	GROUP1	V5_0	V6_0	V7_0
2	V1_1	V2_1	GROUP1	V5_1	V6_1	V7_1
3	V1_2	V2_2	GROUP1	V5_2	V6_2	V7_0
4	V1_3	V2_3	GROUP1	V5_3	V6_3	V7_0
5	V1_4	V2_4	GROUP1	V5_4	V6_4	V7_1
6	V1_5	V2_5	GROUP1	V5_5	V6_5	V7_1
7	V1_6	V2_6	GROUP1	V5_6	V6_6	V7_1
8	V1_7	V2_7	GROUP1	V5_7	V6_7	V7_1
9	V1_8	V2_8	GROUP2	V5_8	V6_8	V7_2
10	V1_9	V2_9	GROUP2	V5_9	V6_9	V7_2
11	V1_10	V2_10	GROUP2	V5_10	V6_10	V7_2
12	V1_11	V2_11	GROUP2	V5_11	V6_11	V7_2
13	V1_12	V2_12	GROUP2	V5_12	V6_12	V7_2
14	V1_13	V2_13	GROUP2	V5_13	V6_13	V7_2
15	V1_14	V2_14	GROUP2	V5_14	V6_14	V7_2
16	V1_15	V2_15	GROUP2	V5_15	V6_15	V7_2
17	V1_16	V2_16	GROUP2	V5_16	V6_16	V7_2
18	V1_17	V2_17	GROUP3	V5_17	V6_17	V7_3
19	V1_18	V2_18	GROUP3	V5_18	V6_18	V7_4
20	V1_19	V2_19	GROUP3	V5_19	V6_19	V7_5
21	V1_20	V2_20	GROUP3	V5_20	V6_20	V7_6
22	V1_21	V2_21	GROUP3	V5_21	V6_21	V7_7
23	V1_22	V2_22	GROUP3	V5_22	V6_22	V7_8
24	V1_23	V2_23	GROUP3	V5_23	V6_23	V7_9
25	V1_24	V2_24	GROUP3	V5_24	V6_24	V7_10

## 7.7 Datasettien muodostus

Erilaisia datasettejä on mahdollista muodostaa viisi kappaletta, näiden esittely lyhyesti seuraavaksi. Wafer info datasetti muodostetaan aina analyysin alussa, vaikka muita analyysejä ei olisikaan valittu tehtäväksi.

### 7.7.1 Wafer info

Wafer info-datasetti koostaa käyttäjälle yleistietoa analyysiin valituista kiekkoista. Valmiista taulukosta näkyy lopullisen kootun kiekon komponentit, eli käytetty ohutkiekko ja paksukiekko tai paksukiekot. Lisäksi on kerätty muuta yleisluontoista tietoa kiekkoista. Käyttäjälle näkymättömissä tämä osio ohjelmakoodista koostaa käyttäjän antamista kootuista kiekkoista ja kootun kiekon komponenteista erilaisia listoja, joita hyödynnetään jatkossa, kun ohjelmakoodissa tehdään datahakuja tietokannoista. Lisäksi käyttäjän antaman listauksen perusteella muodostetaan lajittelutietoa annettujen kiekkojen ominaisuuksista, kuten kuvassa 20 näkyvä ohjelmakoodin osa, jossa yhtenä osana määritellään käyttäjätietojen perusteella, onko kyseessä ns. normaaleilla tuotantokoodeilla etsittäviä eria vai käsitelläänkö jatkossa testikoodeilla olevaa dataa. Saman tyyppisiä muuttujalla ilmaistavia tietoja välitetään eteenpäin myös muista kiekkojen ominaisuuksista, koska erilaisilla tuotantoprosesseilla tehdyt tuotteet eroavat myös tietotyyppien perusteella toisistaan.

Kuva 20. JSL koodia.

```

293 //form lot lists and info page
294 e_lotlistingInfo = Expr(
295
296 //replace missing values with 0
297     l_selRows = dt_lots << Get Rows Where( Is Missing( :GROUP ) );
298     dt_lots:GROUP[l_selRows] = "0";
299
300 //get lots as list, modify to a sql worthy listing
301     Summarize( l_lots = By( dt_lots:Combined_WFR ) );
302     s = Substitute( Char( l_lots ), "\!", "", "" );
303     lotList = Substr( s, 2, Length( s ) - 2 );
304
305 //is nonprod?
306     If( dt_lots:Combined_WFR[1] != dt_lots:DISCNR_probe[1],
307         nonprod = 1,
308         nonprod = 0
309     );

```



Datasettejä pyritään pitämään laadullisesti hyvinä käsittelyn aikana, kuvassa 21 näkyy periaate, jolla lopullisesta Wafer Info datataulusta poistetaan datan käsittelyn aikana mahdollisesti syntyneet tyhjät kolumnit.

Kuva 21. JSL ohjelmakoodia tyhjien kolumnien poistamiseen.

```

615 //remove empty columns
616 m_columns = dt_info << getAllColumnsAsMatrix;
617
618 l_toDel = {};
619 For( r = 1, r <= N Col( m_columns), r++,
620     If( N Row( Loc( Is Missing( m_columns[0, r] ) ) ) == N Row( m_columns ),
621         Insert Into( l_toDel, r )
622     )
623 );
624
625 dt_info << deleteColumns( l_toDel );

```

### 7.7.2 Yield datasetti ja FC datasetti

Yield datasetti kerää tiedon kootun kiekon ja sen osien mittapistekohtaisista testisaannoista. Lisäksi voidaan ilmoittaa kiekon tai sen osien kokonaissaanto mittapisteittäin. Kuvassa 22 näkyy kokonaisuudessaan ohjelmakoodi, jolla datasetti muodostetaan, koodissa on myös huomioitu myös mahdollisuus epästandardiin kiekkokoodiin. FC datasetti koostaa kiekko-kohtaiset vikakoodimäärät eri mittapisteissä, ohjelmakoodin rakenne hakua lukuun ottamatta vastaava kuin Yield datasetissä.

Kuva 22. Yield datasetin muodostus.

```

1036 e_yieldData = Expr(
1037
1038     If( nonprod == 1,
1039         list = mittadataLots,
1040         list = allLots
1041     );
1042
1043     dsnString = mittadataDsn;
1044
1045     sqlTemplateYield =
1046     "\{
1047
1048     SELECT
1049     A.DISCNR,
1050     B.MEASUREMENTPOINTNR,
1051     ROUND(B.MEASPOINTYIELD, 2) AS MEASPOINTYIELD,
1052     ROUND(B.TOTALYIELD, 2) AS TOTALYIELD,
1053     B.CREATIONTIME
1054 FROM DISCSET A, TESTSET B
1055 WHERE A.DISCSETIDNR = B.DISCSETIDNR
1056     AND A.DISCNR IN (^list^)
1057     AND B.VALIDITY = 1
1058     AND A.VALIDITY = 1
1059
1060     }\";
1061
1062     sqlYield = Eval Insert( sqlTemplateYield );
1063     dbc = Create Database Connection( dsnString );
1064     dt_yield = Execute SQL( dbc, sqlYield, "Yield" );
1065
1066     //check if testcode
1067     If( nonprod == 0,
1068
1069         //join with stacked datatable
1070         dt_yieldFinal = dt_stackRef << Join(
1071             With( dt_yield ),
1072             By Matching Columns( :Components = :DISCNR ),
1073             Drop multiples( 0, 0 ),
1074             Include Nonmatches( 1, 0 ),
1075             Preserve main table order( 1 ),
1076             Output Table( "Yield info" )
1077         )
1078     ,
1079     //else join with nonstacked datatable
1080     dt_yieldFinal = dt_lots << Join(
1081         With( dt_yield ),
1082         By Matching Columns( :DISCNR_probe = :DISCNR ),
1083         Drop multiples( 0, 0 ),
1084         Include Nonmatches( 1, 0 ),
1085         Preserve main table order( 1 ),
1086         Output Table( "Yield info" )
1087     )
1088 );
1089
1090 Close( dt_yield, no Save );
1091
1092 dt_yieldFinal << Compress file when saved( 1 );
1093
1094 dt_yieldFinal << show window( 0 );
1095 );

```

### 7.7.3 RAW datasetti ja FCO RAW datasetti

RAW data eli puhkielessä kiekon raakadata, tarkoittaa elementtitason yksityiskohtaista mittausdataa. Käytännössä tämä tarkoittaa sitä, että esimerkiksi jos kootulla kiekolla on 1600 anturia, niin jokaisesta anturista voi löytyä jopa satoja mittaustuloksia. Raakadataa sisältävät tiedostot ovat tyypillisesti kooltaan suuria ja tietojen haku tietokannasta hidasta, mikäli haettavia kiekkoja on useita. Jotta data-analyysien tekemistä ei tarvitsisi aloittaa pitkällisellä datahauella, ylläpidämme MFI:llä valmiiksi haettua raakadataa Standard datassa noin 50 päivän ajanjaksolta.

Analyysityökalu hyödyntää tätä valmiiksi haettua ja muotoiltua Standard dataa käyttäjän niin halutessa. Käyttöliittymässä käyttäjä on saanut antaa RAW datatiedoston, ja analyysityökalu etsii ensin löytyvätkö kaikki analysoitavat kiekot annetusta tiedostosta. Mikäli kaikkien kiekkojen dataa ei löydy, kysytään käyttäjältä lupa tietokantahaun tekemiseen. Käyttäjälle ilmoitetaan puuttuvien kiekkojen lukumäärä, jolloin aikavaatimus on helpompi hahmottaa. Tietokantahakujen raskauden takia käyttäjää on varoitettava tässä vaiheessa. Mikäli käyttäjä ei anna suostumustaan, analyysi keskeytetään.

Tarvittaessa haetaan puuttuvien kiekkojen data tietokannasta, muotoillaan data vastaamaan valmiin tiedoston formaattia ja yhdistetään tiedot. Mikäli käyttäjä on käyttöliittymässä valinnut speksirajojen käyttämisen PROBER-datassa, haetaan tietokannasta käytetty speksi ja lisätään tiedot kolumnien Spec Limits -ominaisuuksiin kuvan 23 mukaisesti, lisäksi valitaan, että näitä rajoja käytetään aina kun mahdollista.

Kuva 23. Speksirajojen asetus kolumnin ominaisuuksiin.

'X\_20' in table 'RAW data'

Column Name

Lock

Data Type

Modeling Type

Format  Width

Use thousands separator ( )

Column Properties

Remove

Spec Limits

Spec Limits are specification limits that are used in various platforms such as Process Capability, Distribution, and Process Screening. Click below to key in values.

Lower Spec Limit

Target

Upper Spec Limit

Show as Graph Reference Lines

Process Importance

OK  
Cancel  
Apply  
Help

Ohjelmakoodissa sama toimenpide tehdään kerralla koko datasetille, käyttäen haettua speksirajatiedostoa, koodiesimerkki kuvassa 24.

Kuva 24. Speksirajojen asettaminen kolumneille.

```
list_columns = dt_rawComplete << Get Column names( String );
obj = dt_rawComplete << Manage Spec Limits(
  Y( Eval( list_columns ) ),
  Load from Limits Table( dt_spec ),
  Show limits all( 1 ),
  Save to Column Properties( 1 )
);
obj << Close window;
```

FC0 RAW datasetti on RAW datasetin alajoukko, mukana pelkästään hyväksytyt anturit, joilla ei ole vikakoodeja. Tarve tälle supistetulle datasetille johtuu siitä, että raakadatassa on mukana paljon poikkeavia havaintoja. Analysointien tekeminen voi olla mahdotonta, jollei poikkeavia arvoja käsitellä tai poisteta datasta.

## 7.8 Visualisoinnit ja analyysit

Analyysiohjelma laatii käyttäjän valitsemat visualisoinnit ja analyysit ja sijoittaa valmiit tulokset projektin välilehdille. Kaikki raportit ovat täysin toiminnallisia, mahdollistaen muutosten tekemisen niin haluttaessa. Seuraavana esimerkit tuloksista, joissa nostettuna esille huomioitavia tekijöitä.

### 7.8.1 Wafer level yields

Datasetti esittää kiekkotason saannot mittapisteittäin taulukkomuodossa. Näkymään lisätty mahdollisuus suodattaa ryhmien mukaiset tiedot, kuvassa 25 valittuna vain ryhmän 1 kiekot, näkymä suodattuu dynaamisesti ja mahdolliset laskutoimitukset päivittyvät valinnan mukaan automaattisesti.

Kuva 25. Kiekkotason saannot taulukkomuodossa.

The screenshot shows the Minitab 'Tabulate' window. The main table displays the mean yield for various measurement points (V6\_0 to V6\_11) across different wafer groups (V1\_0 to V1\_7). The 'GROUP' column is set to 'GROUP1'.

	MEASPOINTYIELD											
	MEASUREMENTPOINTNR											
	V6_0	V6_1	V6_2	V6_3	V6_4	V6_5	V6_6	V6_7	V6_8	V6_9	V6_10	V6_11
Combined_WFR	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
V1_0	95.56	99.43	92.58	100.00	100.00	100.00	99.07	98.29	97.15	98.47	99.07	99.82
V1_1	99.28	100.00	97.39	100.00	100.00	100.00	98.32	99.58	98.16	98.83	99.19	99.88
V1_2	99.13	99.97	96.31	100.00	100.00	100.00	99.76	99.49	96.67	99.79	98.86	99.52
V1_3	99.40	99.91	95.59	100.00	100.00	100.00	99.82	99.61	98.26	99.34	99.13	99.91
V1_4	99.49	100.00	96.16	100.00	100.00	100.00	99.70	99.40	97.60	99.43	99.22	99.94
V1_5	99.31	99.94	87.21	100.00	100.00	100.00	99.73	99.58	98.23	99.46	99.13	99.94
V1_6	99.10	99.85	93.87	100.00	100.00	100.00	99.85	99.70	96.44	99.76	99.40	99.61
V1_7	99.46	99.79	94.77	100.00	100.00	100.00	99.76	99.31	99.06	99.34	99.58	100.00

## 7.8.2 Yields by group

Ryhmäkohtaiset keskiarvot saannoista mittapisteistä taulukkona, kuvassa 26. Näkymä mahdollistaa nopean vertailun mittapistesaantojen eroista. Arvojen ollessa keskiarvoja voivat yksittäiset ääriarvot vaikuttaa häiritsevän paljon, siksi on hyvä tarkistaa myös yksittäiset kiekkosannoit.

Kuva 26. Ryhmäkohtaiset mittapistesaannot.

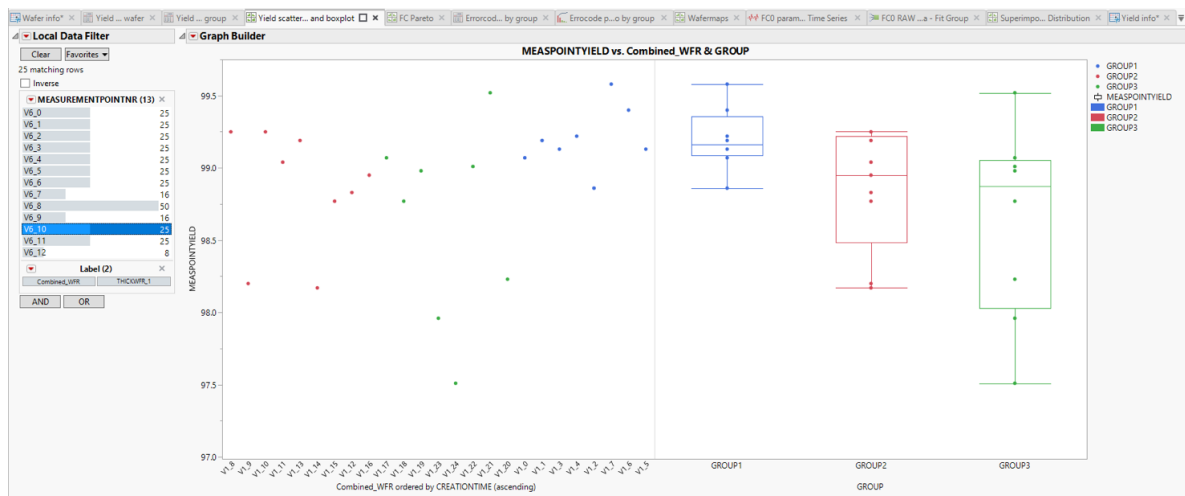
The screenshot shows the Minitab 'Tabulate' window with a summary table of mean yields for three groups (GROUP1, GROUP2, GROUP3) across 12 measurement points (V6\_0 to V6\_12).

	MEASPOINTYIELD												
	MEASUREMENTPOINTNR												
	V6_0	V6_1	V6_2	V6_3	V6_4	V6_5	V6_6	V6_7	V6_8	V6_9	V6_10	V6_11	V6_12
GROUP	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
GROUP1	98.84	99.86	94.24	100.00	100.00	100.00	99.50	99.37	97.69	99.30	99.20	99.83	.
GROUP2	98.95	99.68	95.12	100.00	99.99	99.79	100.00	.	98.52	.	98.85	97.69	.
GROUP3	99.17	99.90	96.47	100.00	100.00	99.95	99.63	99.78	98.41	99.42	98.63	99.75	94.97

## 7.8.3 Yield graph

Tarkempaan saantoanalyysiin laadittiin Graph Builder-alustalla kahden kuvaajan yhdistetty näkymä, pistejoukkokuvaaja vasemmalla ja laatikko-jana-kuvio oikealla, kuva 27. Eri ryhmät näkyvät värikoodattuina ja kuvaajille on lisätty suodatusmahdollisuus sekä mittapisteen että kiekkotyyppin mukaan. Kiekkotyyppin mukainen suodatus on tarpeen, koska osalla mittapisteistä on käytössä sama nimi eri kiekkotyypeillä.

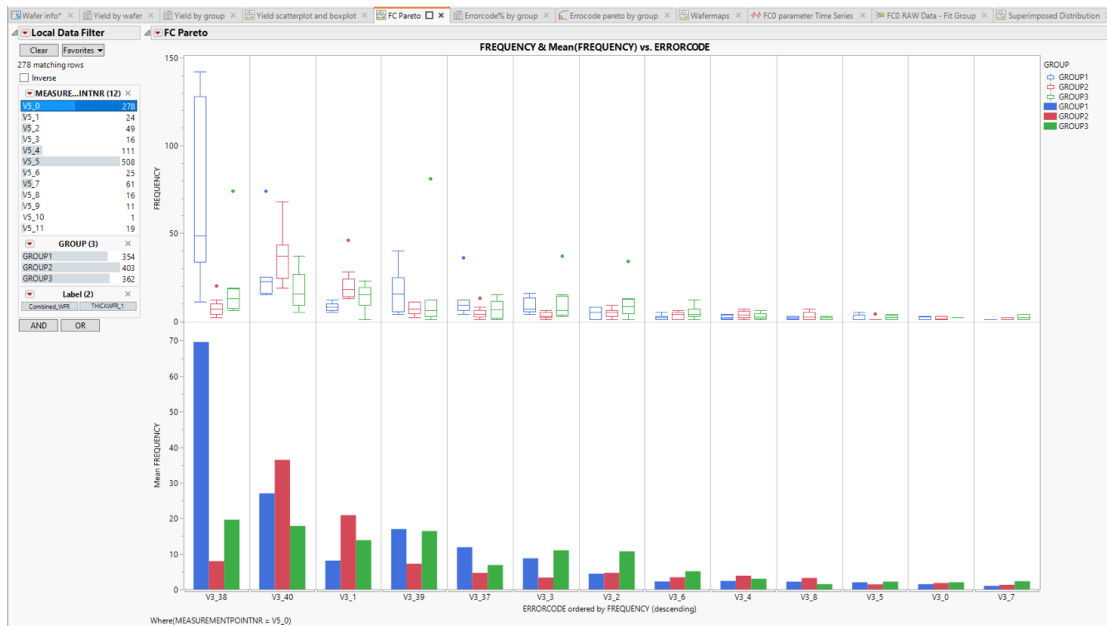
Kuva 27. Graph Builder näkymä mittapistesaannoista.



#### 7.8.4 FC Pareto

Vikakoodien tarkastelua varten laadittiin Graph Builder-näkymä, jossa on kaksi kuvaajaa yhdistettynä. Alaosassa ryhmitellyt pylväskaaviot vikakoodien lukumäärän keskiarvoista ja yläosassa lukumäärien laatikko-jana-kuvaajat, jotta voidaan helposti arvioida johtuvatko vikamäärät yksittäisistä kiekkoista, vai onko taso kauttaaltaan korkea. Kuvassa 28 nähtävillä suodatusmahdollisuus mittapisteen, ryhmän ja kiekkotyyppin mukaan, valittuna yksi mittapiste. Vikakoodien pylväskaaviot on järjestetty Pareto-tyylisesti siten, että suurimmat arvot ovat vasemmassa laidassa ja pienimmät oikeassa, tässä lajittelu vikakoodien keskiarvolukemien mukaan.

Kuva 28. FC Pareto.



### 7.8.5 Errorcode %

Näkymässä prosentuaaliset vikakoodien määrät ryhmittäin, esimerkki kuvassa 29.

Huomattavaa on, että lukemat eivät ole suuruusjärjestyksessä, vaan ne on järjestetty vikakoodin mukaisesti nousevaan järjestykseen.

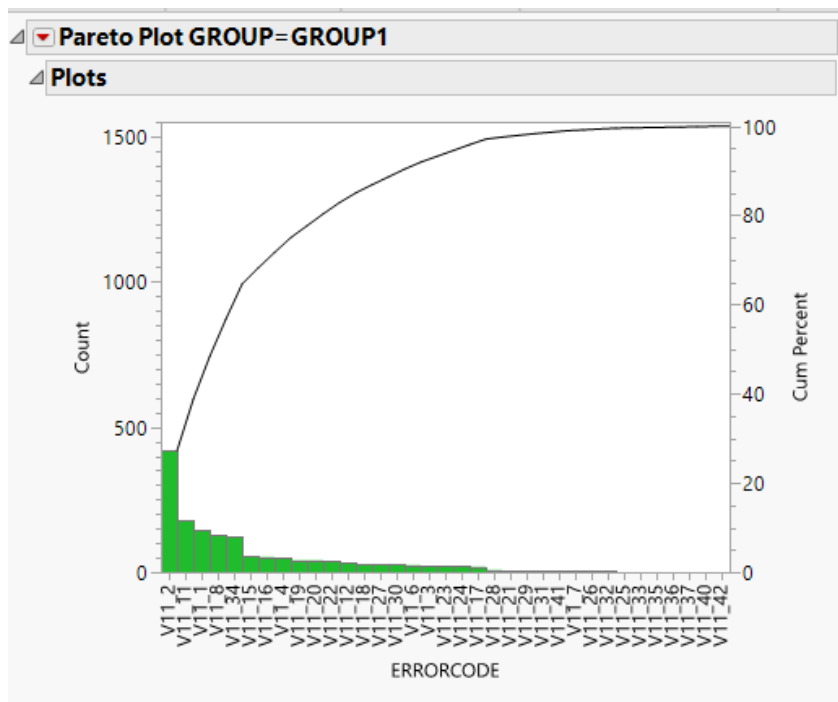
Kuva 29. Vikakoodien prosenttiosuudet ryhmittäin.

ERRORCODE	GROUP		
	GROUP1 Column %	GROUP2 Column %	GROUP3 Column %
V11_1	9.51%	12.58%	8.15%
V11_2	27.28%	12.78%	8.87%
V11_3	1.30%	6.84%	3.30%
V11_4	3.32%	1.23%	0.42%
V11_5	0.00%	0.00%	13.36%
V11_6	1.56%	6.63%	0.66%
V11_7	0.20%	3.14%	1.92%
V11_8	8.40%	9.23%	7.25%
V11_9	0.00%	0.34%	0.42%
V11_10	0.00%	0.00%	11.68%
V11_11	11.52%	12.92%	5.69%
V11_12	2.15%	1.64%	0.78%

### 7.8.6 Errorcode pareto

Kuvaaja esittää ryhmäkohtaiset vikakoodiparetot allekkain, otsikkotasolla näkyy mikä ryhmä on kyseessä, kuvassa 30 ryhmän 1 paretokuvaaja.

Kuva 30. Vikakoodipareto.

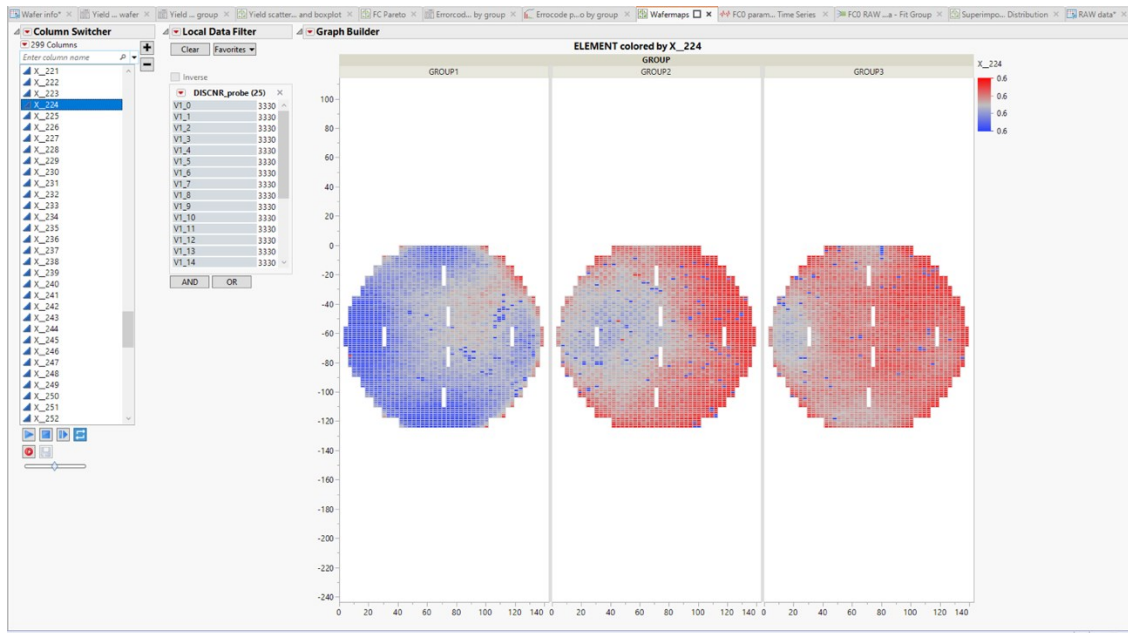


### 7.8.7 Parameter wafermaps

Mittaustulosten parametrien vertailua varten on muodostettu Graph Builderiin näkymä, jossa on hyödynnetty tuotteelle tehtyä karttatiedostoa, näin näkymästä saadaan muodostumaan realistinen toisinto kiekkokartasta, eli jokainen elementti näkyy eroteltuna kuvaajassa. Kuvaajaan on liitetty suodatusmahdollisuuksia, voidaan tarkastella elementtien keskiarvodataa, eli voi kuvitella että kiekot on pinottu päällekkäin ja anturielementin arvo on keskimääräinen arvo parametrusta. Toisaalta voidaan myös erotella kiekot yksitellen tarkasteltavaksi. Tulokset on niputettu ryhmien mukaisesti, joten voidaan silmämääräisestikin nähdä nopeasti, mikäli ryhmien välillä on eroa. Esimerkki yhden testiparametrin tarkastelusta kuvassa 31.

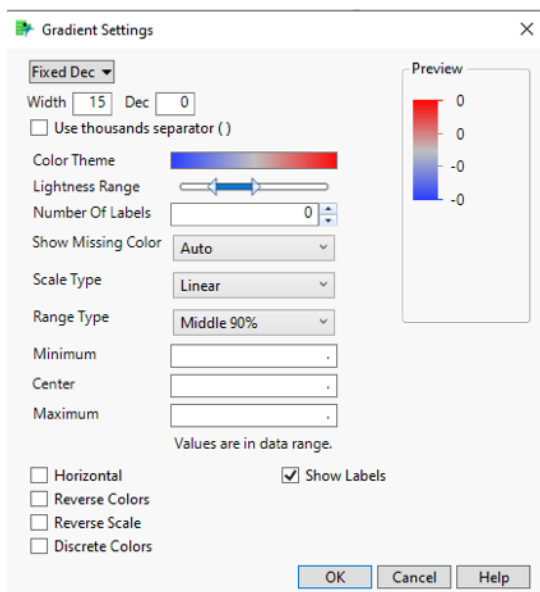


Kuva 31. Testiparamerien tarkastelu.



Mahdolliset poikkeavat arvot on käsitelty kuvaajan väriskaalan käsittelyssä, kuvassa 32 näkyvän dialogin arvoon Range Type valitaan skriptissä Middle 90%, jolloin arvoista suodattuvat ääripäät pois ja poikkeavat arvot saadaan näin yleensä käsiteltyä riittävän hyvin.

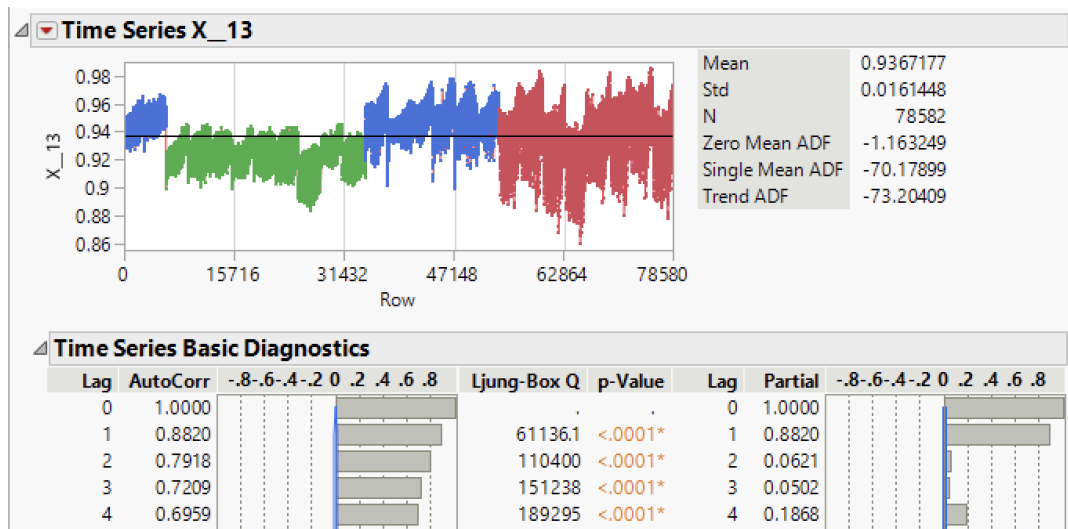
Kuva 32. Näytettävän asteikon valinta.



### 7.8.8 Time series

Aikasarjadata jokaisesta testiparametrasta, eri ryhmät on värikoodattu erottumaan toisistaan. Aikasarjadatan kuvaajaa piirtäessä on datalähteen, tässä tapauksessa FCO RAW datasetin oltava aikajärjestyksessä. Analyysityökalu piirtää aikasarjakuvaajan kaikista parametreista ja sijoittaa ne allekkain, esimerkki yhdestä parametrasta kuvassa 33.

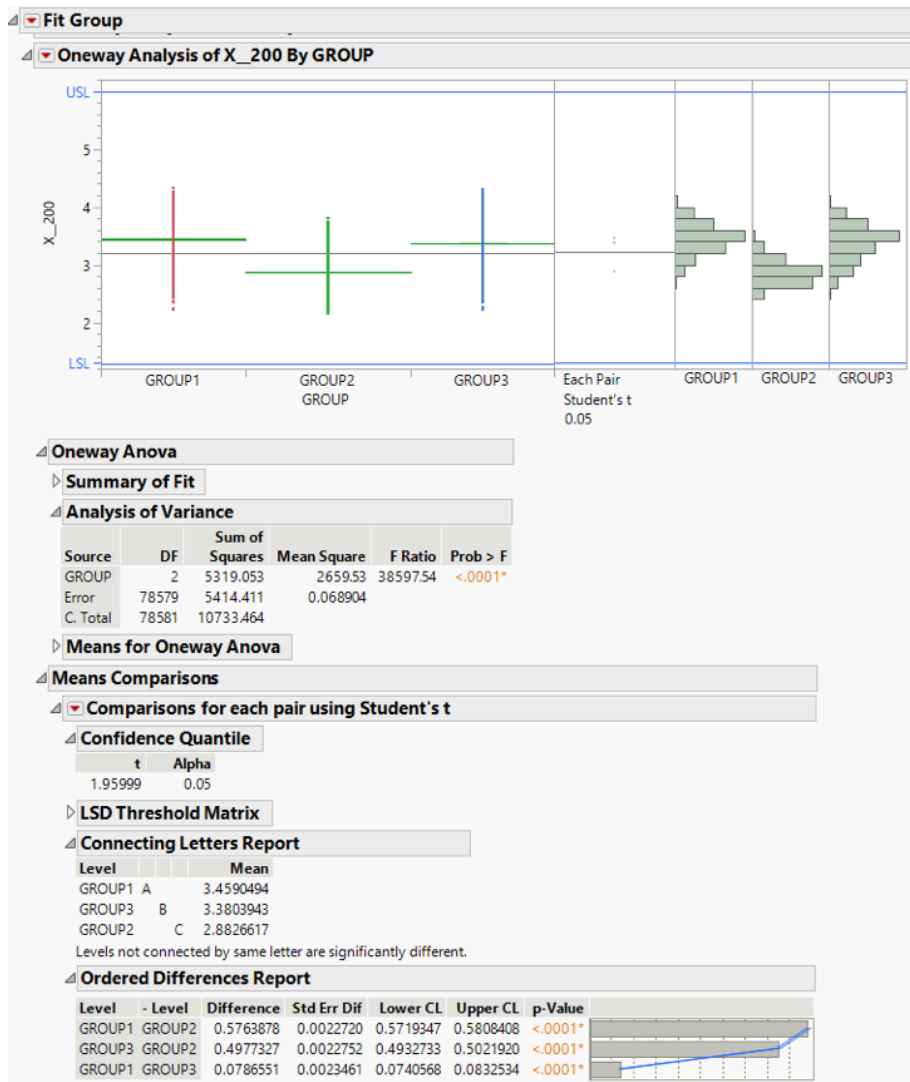
Kuva 33. Testiparametrin aikasarjadata.



### 7.8.9 ANOVA

Käyttäjä saa analyysiohjelman suorituksen aikana kehoitteen valita halutut PROBER-mittapisteiden testausparametrit keskiarvojen Oneway Anova-analyysiin. Valituista parametreista luodaan raportit, jotka sijoitetaan allekkain, järjestettynä merkityksellisyyden mukaan siten, että tilastollisesti merkittävin analyysi on ensimmäisenä. Esimerkki raportista kuvassa 34.

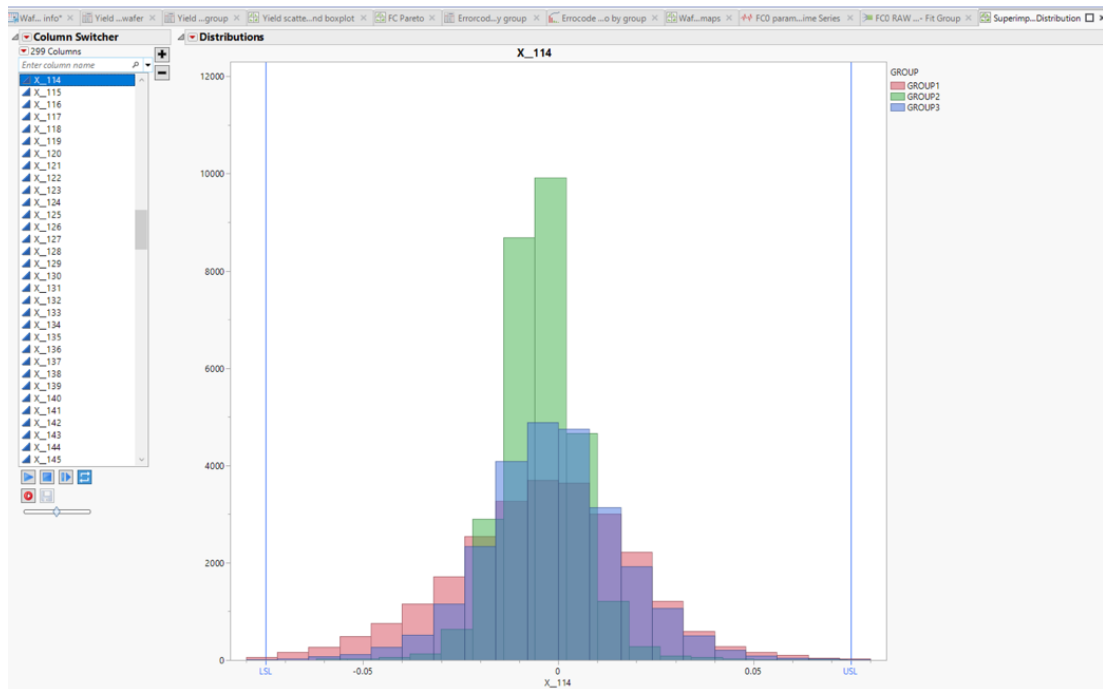
Kuva 34. Anova.



### 7.8.10 Superimposed distribution

Ryhmien vaihtelun ja jakauman tarkastelua päällekkäin asetettujen histogrammien avulla, tarkasteltava parametri suodatetaan valikosta. Kuvassa 35 näkyy hyvin myös speksirajojen näkyminen kuvaajassa, rajat piirtyvät automaattisesti, mikäli ne on siirretty kolumnin ominaisuuksiin.

Kuva 35. Vaihtelun visualisointia ryhmittäin.



## 8 Yhteenveto

Opinnäytetyön tekeminen oli osa työtehtäviäni, mutta alusta lähtien lähdin haastamaan tekemistäni niin, että pyrin jatkuvasti analyyttisesti miettimään ohjelmakoodin rakennetta, algoritmien tehokkuutta sekä valittujen analyysien sopivuutta käyttötarkoitukseensa. Heti suunnitteluvaiheessa oli selvää, että rakenteen tulee olla mahdollisimman joustavasti muokattavissa, eli osien lisääminen tai poistaminen ei saa aiheuttaa vakavia ongelmia ohjelmakoodille. Tässä onnistuin hyvin, ohjelmakoodi on modulaarinen ja jaettu selkeisiin toiminnallisiin kokonaisuuksiin, jolloin muutosten tekeminen on mahdollista.

Toinen suuri haaste oli pyrkimykseni saada analyysityökalu toimimaan kaikille erilaisille tuotetyypeille, ilman että ohjelmakoodiin kuitenkaan joudutaan koodaamaan kiinteitä arvoja muuttujille, jotka sitten vaativat jatkuvaa päivitystä. Tuotteiden ryhmittely loogisesti ja käsittelyehtojen luominen niille vei runsaasti kehitystyöhön käytettyä aikaa. Täysin ilman tuotteiden luettelointia ei pysty kaikkia toiminnollisuuksia toteuttamaan, mutta suurin osa tiedoista pyritään hakemaan tietokannasta kiekkokoodin perusteella, tai koodin ulkopuolisesta asetustiedostosta.

Vaikka lopulliset analyysit pitkälti tulivat valituiksi sidosryhmien toiveiden ja vakiintuneiden analyysikäytäntöjen pohjalta, sai pienillä muutoksilla lisäominaisuuksia, jotka paransivat visuaalisuutta ja auttavat datajoukkojen vertailussa. Myös tilastolliset edellytykset analyysien käytölle tuli tarkastettua ja mahdollisten poikkeusten vaikutukset analyysien onnistumiselle selvitettyä.

Ohjelmakoodin kannalta analyysityökalussa käytetään monipuolisesti erilaisia ohjelmointitekniikoita, erilaisista algoritmeista koodin rakenteen optimointiin. Vaikka suunnittelin koodin rakenteen melko hyvin etukäteen, käytännön ohjelmointityö kuitenkin muokkasi rakennetta aina kun huomasin paremman tavan toteuttaa jokin tietty toiminto. Jatkuvaan koodin uudelleenmuokkaamiseen ei kuitenkaan ole mahdollisuutta, jossain vaiheessa on päätettävä, että ohjelma on julkaisuvalmis.

Ajallisesti analyysityökalun tekeminen venyi pitkälle ajalle, osittain syynä olivat muut kiireet, mutta myös testauksen vaatima aikamäärä oli huomattava. Jokainen koodin yksityiskohta piti testata kaikilla tuotteilla, jotta voin varmistua, että luotu logiikka toimii. Testisettien läpiajon lisäksi myös loppukäyttäjien testeille annettiin runsaasti aikaa.

Opinnäytetyössä luotu automaattinen analyysityökalu täytti tehtävänannon, ajansäästö analyysien valmistelussa ja suorittamisessa on merkittävä. Mahdollisuus data-analyysiin on saatu useamman käyttäjän saataville, kun prosessia on saatu automatisoitua. Kehitysideoita on tullut jo lukuisia lisää ja työkalun kehitystyö jatkuu osana normaaleja työtehtäviäni.

## Lähteet

Conway, D. (30.9.2010). *THE DATA SCIENCE VENN DIAGRAM*.

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

Driscoll, M. (5.3.2012). *the data science debate: domain expertise or machine learning?* M. e.

driscoll. <https://medriscoll.com/2012/03/05/the-data-science-debate-domain-expertise-or/>

Harvard. (2015). *CS109 Data Science*.

<https://github.com/cs109/2015/blob/master/Lectures/01-Introduction.pdf>

Lee, A. (8.12.2018). *How To Ask The Right Questions As A Data Scientist*. Towards data

science. <https://towardsdatascience.com/how-to-ask-the-right-questions-as-a-data-scientist-913621907411>

Murata Electronics Oy. (n.d.). *Murata Electronics yrityksenä*. Murata.

<https://muratafinland.com/murata-electronics-yrityksena/>

O'Neil, C.;& Schutt, R. (2013). *Doing Data Science*. O'Reilly Media, Inc.

Ozdemir, S. (2016). *Principles of Data Science*. Packt Publishing.

Sall, J.;Stephens, M. L.;Lehman, A.;& Loring, S. (2017). *JMP Start Statistics, 6th Edition*. SAS Institute.

SAS Institute Inc. (n.d.-a). *Bar chart*. Statistics knowledge portal.

[https://www.jmp.com/en\\_us/statistics-knowledge-portal/exploratory-data-analysis/bar-chart.html](https://www.jmp.com/en_us/statistics-knowledge-portal/exploratory-data-analysis/bar-chart.html)

SAS Institute Inc. (n.d.-b). *Histogram*. Statistical knowledge portal.

[https://www.jmp.com/en\\_us/statistics-knowledge-portal/exploratory-data-analysis/histogram.html](https://www.jmp.com/en_us/statistics-knowledge-portal/exploratory-data-analysis/histogram.html)

SAS Institute Inc. (n.d.-c). *Pareto-chart*. Statistics knowledge portal.

[https://www.jmp.com/en\\_us/statistics-knowledge-portal/exploratory-data-analysis/pareto-chart.html](https://www.jmp.com/en_us/statistics-knowledge-portal/exploratory-data-analysis/pareto-chart.html)

Sas Institute Inc. (n.d.-d). *Box-plot*. Statistical knowledge portal.

[https://www.jmp.com/en\\_us/statistics-knowledge-portal/exploratory-data-analysis/box-plot.html](https://www.jmp.com/en_us/statistics-knowledge-portal/exploratory-data-analysis/box-plot.html)

Sas Institute Inc. (n.d.-e). *Core Capabilities of JMP*.

[https://www.jmp.com/en\\_my/software/data-analysis-software/core-capabilities-of-jmp.html](https://www.jmp.com/en_my/software/data-analysis-software/core-capabilities-of-jmp.html)

Skiena, S. S. (2017). *The Data Science Design Manual*. Springer International Publishing.

Tampere: Yhteiskuntatieteellinen tietoaarkisto [ylläpitäjä ja tuottaja]. (n.d.). *Hypoteesien testaus*. Teoksessa *Kvantitatiivisen tutkimuksen verkkokäsikirja*. Tietoaarkisto.

<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/hypoteesi/testaus/>