UiT THE ARCTIC UNIVERSITY OF NORWAY

School of Business and Economics

Sickness absence in Troms County

Trends and risk factors

—

Ekaterina Chekanova

Master's Thesis in Economics - May 2019

Acknowledgements

I would first and foremost like to express my gratitude to my supervisor Professor Øystein Myrland. I want to thank him for valuable help, advice and support throughout the entire writing process.

A thank you also goes to my family, friends and fiancé for their encouragement and support.

A special thank you goes to my loving parents for the incredible faith in me and my success.

Abstract

This thesis examines the dynamics of the sickness absence in Troms county. High level of sickness absence in Norway causes substantial costs for sick pay scheme and replacement workers likewise significant reduction in productivity. Awareness of the trends of sickness absence and the risk factors which increase the likelihood of sick leave contributes to the establishment of proper policies aimed to reduce it. Until now the measures did not show expected efficiency. My method of analysis is a combination of descriptive statistics and regression analysis performed by using RStudio software (RStudio, 2015). I examined the relationship between variables day of week, month, year, mean temperature, sum of precipitation and wind power, gender, sector, age and profession and the frequency of the registered sick leaves in Troms county for the period 2010-2017. Key findings are: 1) the sickness absence showed pronounced seasonality and steadily growth during the estimated period, 2) weather describing variables did not show significant effect on the frequency of sick leaves, 3) women have more sickness absence than men, 4) municipality sector has the highest sickness absence rate, followed by private and governments sectors, 5) younger population showed significant ongoing increase in sickness absence compared to older population, 6) significant changes in the economy reflecting on GDP cause fluctuations and deviations from the general trend of sickness absence in all sectors, age and profession groups.

Key words: sickness absence, linear regression analysis, seasonality, risk factors

Table of Contents

List of tables

List of figures

Introduction

In the health and care sector, sickness absence is a frequent phenomenon which constitutes a field in need of exploration. The sickness absence rate in Norway is high despite improvements in key indicators of public health.

According to Det Kongelige Finansdepartementet, the National Insurance's total expenditure for the Ministry of Labor and Social Affairs in 2019 is assumed to be NOK 425 billion (Finansdepartementet, 2018). This includes expenditures on pensions, sickness benefits for employees, work assessment allowance, disability benefits and unemployment benefits. The cost of sick pay scheme for employees is estimated at NOK 37.5 billion in 2018.

The identification of risk factors of sickness absence is therefore necessary for the choice of correct policies for the reduction of the total level of sickness absence. Implementing of the information about seasonal diseases and the prevention measures for the most vulnerable groups right in the period preceding the onset of seasonal disease can decrease the level of sickness absence and its duration.

This information is useful not only for the state but also for companies. Every company will be able to make up its own pattern of sickness absence and predict periods with the highest and the lowest sickness absence rates using such information about employees as gender, age and profession. The company will be able to plan amount of sick pay in advance, as well as plan to correct distribution of labor and attraction of additional labor to maintain the level of productivity in the period of the highest sickness absence rate, and increase the workload in the periods with the lowest sickness absence rate.

The purpose of this paper is to investigate seasonal patterns of sickness absence in Troms county and main trends for the various factor groups such as gender, sector, age and profession. The aim is to identify if the mentioned factors and weather indicators affect sickness absence rate and to what extent. I use multiple regression model and log-linear regression model for the analysis.

The better understanding of the sickness absence patterns and most vulnerable groups of population will lead to a better strategy aimed to both sickness prevention and reduction of the sickness absence rate.

The assessment of sickness absence is a challenge despite all the data on sick leaves and registers of employees and employers. At the moment, there is no uniform standardized evaluation system. Therefore, the quality of sickness absence research can be improved with an increased awareness of the measurements used.

Many previous findings are based on surveys and mostly cover population of small towns or single organization while I use the registered data for the entire population of the Troms county which reduces the possibility of deviations due to inaccuracy in surveys and small sample size. Another difference is that I investigate not seasonality of a single disease but seasonality of sickness absence in general. I include short-term sickness absence in analysis together with long-term sickness absence for the complete picture. The data in my paper is up-to-date and it covers individual combination of factors which is supposed to give more detailed understanding of sickness absence distribution within a predetermined period.

The goal is to formulate accessible and concise observations, provide graphical representation of the obtained results likewise possible explanations of the resulting pattern behavior.

Some studies and findings from the entire period of research on the sickness absence topic have proven very useful. Since 2004, the Norwegian social insurance system has required development of presenteeism policy – being present at work despite being sick, partly and with adaptation of the workplace, unless physician identifies a person as unable to work at all. An article from 2012 estimates reduction of the number of lost working days during the sick leave by more than half and raise of the employment propensity two years later by sixteen per cent (Markussen, 2012).

It is difficult to balance the reasonable social insurance and sufficient work incentive. The authors of the paper claim that the activation strategy benefits both employers and employees. Given the large gains the initiative was assessed as a promising step for the social security reform by the authors of the paper in 2012. Øystein Hernæs in his paper «Activation against absenteeism – Evidence from a sickness insurance reform in Norway» from 2018 analyzed the results of implemented program in the region of Hedmark – the one with the more strictly enforcement of the new rule (Hernæs, 2018).

Results are very promising: twelve per cent reduction in the absence rate among workers on long-term sick leave in Hedmark region compared to the weighted results of similar regions with less enforcement of the activation program. It is remarkable that the diagnosis group with the musculoskeletal disorders showed the most significant decline.

Usage of remaining work capacity of people on the long-term sick leave reduces unnecessary absence and related expenditures.

Therefore, the initiative might be a win-win solution and more attractive alternative than traditional measures with cutting costs and stricter screening. And this solution was discovered partly due to the number of researches on this topic that discovered the useful correlation and attracted attention to it.

Thus, the deeper the problem is investigated, the greater the chances for the solving are. I consider my research useful because it investigates various risk factors, uses rather new and solid data for a quite a long period of eight years and the data sample covers the population of the entire Troms county. It is also illustrative thanks to a large number of figures.

The organization of reminder of the thesis is structured as follows. Background and previous research part is a continuation of the Introduction section and presents literature and previous findings. Section Health Status in Norway provides more information on the topic relevant for the country. The Data section describes the data used in the analysis and the first data insights. Section Method is divided in two parts and presents the method. Section Results and Discussion is also divided in two parts and presents results with interpretation, and section Concluding Remarks concludes.

Background and previous research

This section is devoted to the previous research on the topic. It contains an aim of every research, method and general findings. I choose to concentrate on research about Norway sometimes in comparison with the other Scandinavian countries than the rest of the world because they have more similarities and it could be easier to compare them and get more accurate results.

Master thesis named «Is there an increase in duration of long-term sickness absence in Norway from 1994 to 2007? A registry-based study» was published in 2012. The aim of the study is to investigate if there is an increase in the duration of long-term sickness absence in Norway from 1994 to 2007 using administratively generated registry that contains data on the Norwegian population from 1992. The results showed the general increase in both duration and prevalence of sickness absence. Authors suggest that this development is affected by lowering of the threshold for starting a sickness absence episode and increasing allowance for the maximum duration of the episode (Hagen, 2012).

In the journal article «Risk factors of long-term sickness absence in Norway and Sweden» from 2013 author expresses opinion that it is more useful to analyze not sickness absence in general but long-term sickness absence. He claims that there are at least two reasons for that. First, the short-term sickness absence is more about minor problems that do not require sickness certification from a doctor. Secondly, long-term sickness absence prevails in Norway that makes it a target group for a reduction of the total level of sickness absence (Johansen, 2013).

The aim of the paper is to compare levels of long-term sickness absence in Norway and Sweden as well as discover if the main risk factors for it in two countries are the same. The research is based on the answers of population about long-term sickness absence cases (more than 15 days) for the previous year through a postal questionnaire with a target group being workers in the age of 20-60 years. Many Norwegian respondents name pain in back, neck and muscles as a main reason for long-term sick leaves whereas respondents from Sweden mostly reported mental health problems. The analysis includes demographic factors, socioeconomic position, and occupational characteristics in binary logistic regression to detect risk factors for long-term sickness absence. A strong impact of gender, age and physical work conditions is detected. Norwegian respondents also showed a bigger share of LTSA than Swedish respondent which coincides with the official statistics and previous research. The reason for that could be difference in the social policy of the countries. The research shows no evidence of higher level of LTSA among respondents with less education in contradiction with previous research and official statistics. This deviation could be related to the limits with the survey.

Next study named «Gender equality in sickness absence tolerance: Attitudes and norms of sickness absence are not different for men and women» from 2018 aims to examine whether the reason for the gender gap in sickness absence is gender norms or difference in sickness absence attitudes.

The analysis is performed using binary logistic regression with survey data from male and female employed respondents in Norway in 2016. There was no substantial difference discovered in the attitude towards sickness absence. However, analysis showed more tolerant social norms of sickness absence for employees in gender-dominated occupation than in gender-integrated occupations (Loset, 2018). Therefore, the authors claim that common attitudes and norms of sickness absence are not the case of the gender gap in sickness absence.

«Attitudes towards sickness absence and sickness presenteeism in health and care sectors in Norway and Denmark: qualitative study» from 2014 is another study concerning the sickness absence attitude. The method in the research is framework analysis of eight focus group discussions obtained from interviews of employees in nursing homes in Norway and Denmark. Results showed that the loyalty to residents and colleagues, organization of work and physical aspects of the workplace affect the attitude towards sickness absence. It was socially and morally determined at personal levels by the specificity of the work in both countries (L. Krane, Larsen, E.L., Nielsen, C.V. et al. BMC Public Health, 2014).

«Sickness absence patterns and trends in the health care sector: 5-year monitoring of female municipality employees in the health and care sectors in Norway and Denmark» is an article published in 2014 which aims are covering some of mine in this master thesis: authors compared time trends in sickness absence patterns of municipality employees in the health and care sectors in Norway and Denmark. The data was collected from the personnel registers and age-specific comparative statistics was calculated for each year of the study period from 2004 to 2008 (L. Krane, K., Johnsen, Roar, J., Fleten, Nils, F., Nielsen, Claus Vinther, N., Stapelfeldt, Christina Malmose, S., Jensen, Chris, J., & Braaten, Tonje, B, 2014). The data for Norway was collected from Kristiansand, and from Aarhus for Denmark. Sickness absence rates in Norway showed an increase in the age group from 20 to 29, and the mean number of sick leave episodes also increased in all age groups except for the 30-39 and 60-67 age groups. In both countries, young employees had the largest increase in sickness absence rates and in sick leave episodes.

Despite enough studies on the sickness absence topic, it is challenging to compare them and derive the common conclusions. Every study bases research on the own method of data gathering, population (the general population, the sickness insured/absent population), geographic location, specific combination of factors, method and measurements (frequency, length, cumulative incidence, incidence rate, and duration). The researcher can concentrate on the spell-, person-, or time-based measurements. Even the definition of the long-term sickness absence varies from over 8 days to over 15 days, or even over 30 days on the sick leave in some cases. In addition, the time periods of the studies vary, that means different state of the economics and the labor market.

Until there is no clear concept or standards for the sickness absence research developed, it may be useful to concentrate more on the results from the research carried out by reliable research centers. It usually guarantees quality of data and more detailed results.

The «Sickness Absence in the Nordic Countries» from 2015 is an example of such a research. It is published by the Nordic Social Statistical Committee that includes members of important research organizations from Denmark, Faroe Islands, Finland, Norway and Sweden.

The studies are quite extensive, gathering many mentioned above aspects in one document. Although authors included five Nordic counties in the research to analyze and compare, I will concentrate on results regarding Norway.

To begin with, long-term sickness absence in this research is a registered sick leave over eight days. Norway has high long-term sickness absence rate and low short-term sickness absence rate.

The studies cover gender, age, socio-economic status, work environment and unemployment rate factors and their interconnection to the sickness absence.

Women have more sickness absence than men which coincides with many surveys worldwide. There are three main possible reasons for that: difference in general condition of the health; difference in the working conditions or responsibilities at home and with children; and different sickness absence attitude (Thorsen, 2015). However, there is no strong evidence to make a solid proof for any of those theories.

It is statistically proven that sickness absence increases with age for long-term sickness absence and the short-term sickness absence rate prevails among younger age groups (Thorsen, 2015). The reason for it could be either a stricter attitude towards sickness absence among older age groups or worse working conditions and more physical work for younger groups, or combination of these two reasons.

The studies find socio-economic status to have strong influence on sickness absence in form of the inverse relationship: the lower the socio-economic status, the higher level of sickness absence. Socio-economic status includes lifestyle, health behaviors, physical exercises, diet and working conditions. All these factors are correlated to the health level of every person.

According to the studies, work environment, both physical and psychosocial work conditions, influence the risk of sickness absence through somatic symptoms and mental health issues. It sounds reasonable that such things as gas, dust and noise can increase sickness absence. Bullying can serve as an example of poor psychosocial working conditions (Thorsen, 2015).

Correlation between unemployment rate and sickness absence rate is found in different research. It has an inverse relationship: the higher unemployment rate, the lower sickness absence. Dependence is particularly pronounced in Norway. The most popular explanation is that people are frightened to lose job and would rather go to work while being sick.

Health Status in Norway

The overall level of health in the country is important for a complete understanding of the situation.

Lifestyle is among the factors affecting life expectancy. For example, smoking causes every fifth death after age of 70 in form of cardiovascular disease, lung cancer, COPD and other smoking-related diseases. Unhealthy diet, use of alcohol and illegal drugs provoke deathly diseases and life-threatening situations as road traffic accidents, violence and self-harm (Norwegian Institute of Public Health. Public Health Report: Health Status in Norway 2018, 2018).

Person's lifestyle factors such as health eating and physical activity are highly dependent on health determinants – factors that can either weaken or strengthen ability of a person to look after his own health. Childhood environment, education, financial situation, family, social network, culture and health services are the health determinants.

Many people have sedentary jobs. It is important to have opportunity to do some physical activities after work and during weekends. Education is another example of an underlying factor that is important for health throughout life. Many factors combined form the level of health of the nation and each person individually have influence on the life expectancy. Education is one of such factors. It is proven that men and women that are highly educated live five-six years longer and have better health than those with the low level of education. It is important for health throughout life provided better living conditions (Norwegian Institute of Public Health. Public Health Report: Health Status in Norway 2018, 2018).

The health status of the population of Norway is generally good. The life expectancy is 84.3 years for women and 80.9 years for men. The main causes of disability and reduced health are musculoskeletal disorders, mental disorders, cardiovascular disease and cancer.

Every year 550 - 600 people commit suicide out of approximately 5,300,000 population in the country (SSB, 2019). Norway also has a relatively high – 260 per year on average – number of drug-induced deaths.

Non-communicable diseases such as diabetes, COPD and mental diseases are common diseases in Norway that affect several age groups and generally are difficult to treat. The main purpose of healing in those cases is to reduce the damage to health and prevent the deterioration.

To sum up, there is number of factors that determine the chances of getting the risk of illness. Underlying factors that can promote health or increase the risk of disease are:

- age, gender and hereditary factors;
- individual lifestyle factors;
- social and community networks;
- general socio-economic, cultural and environmental conditions.

The last factor is quite extensive, having included agriculture and food production, education, work-environment, living and working conditions, unemployment, water and sanitation, health care services and housing.

Reducing sickness absence in Norway

Even though the general health level in Norway is good, there always is something to improve. While it is impossible to change hereditary, age and gender factors, it is possible to improve other factors on the list to reduce number of illnesses. It should be done at the individual level, enterprise level and state level combined for the measures to be the most effective.

The Nordic governments have introduced several interventions with the objective of reducing sickness absence. The three main strategies include close follow-up of the long-term sick employees, workability assessment and partial sick leave.

The follow-up system of the sick listed was first introduced as the part of the IA-agreement in 2001. The results were not impressive, and the agreement was changed in 2011. The companies were obliged to a stricter follow-up obligation. In addition, the stricter controlling system of reports and sanctions were introduced.

For now, partial sick leave, or presenteeism, has the best result of all as discussed in the Introduction section. The close follow-up of sick listed and workability assessment has shown mixed results (Thorsen, 2015).

The weak effect of these measures may be due to the lax execution of instructions. Moreover, it is difficult to make a good evaluation, and it also needs to be done with caution. The economy and the labor market of the country constantly change, so that it is important to distinguish if the changes in sickness absence rate is caused by the interventions or by the changes in the economy.

Seasonal illnesses and preventive measures

Reducing the level of the diseases that are defined by the risk factors listed in the Health Status in Norway section is time-consuming and requires a lot of effort. However, there are some seasonal diseases that manifest in certain seasons and can be avoided or at least alleviated if preventive measures are taken in time.

The common seasonal diseases of winter are cold, norovirus, cough, flu, bronchitis, dry and itchy skin. Diseases such as asthma and arthritis worsen in winter. Heart attacks are more common in winter because heart works harder to maintain body heat.

There are more cases of food poisoning, flu, waterborne diseases, heatstroke and sunburn in summer. The increase in temperature leads to multiplication of bacteria that spread bacterial infections. Monsoon brings stomach infections, viral fever and conjunctivitis along with slush and dampness.

All the seasonal diseases have unpleasant symptoms and give a temporary bad state of health that force the sick to be on sick leave. This creates difficulties for the employer. Thus, both employer and employee suffer from the seasonal diseases and taking preventive measures is in the best interest of both parties. Improving of working conditions, creating physical activities for workers in form of team building, providing access to the nutritious food at the cafeteria are just a few ideas, a few small steps towards healthier and more productive work team and less expenses on the sick pay.

Another important step is to spread information on the symptoms and prevention methods of seasonal diseases in advance of the expected start of mass diseases every season, every year. Knowing which diseases are afflicted by more people of a certain enterprise, city, or commune, it is possible to conduct better information campaigns. The purpose of my research is to investigate seasonality of sick leave. The analysis is performed for the entire population of the

Troms county divided in subgroups according gender, age, profession and sector. Results for the groups differ one from another. Thus, it may be useful for the employer to check what groups the majority of the employees are in and pay attention to the time periods of the year when the highest number of sick leaves were registered for those particular groups.

Equality in the workplace

For the validity of analysis, it is necessary to evaluate the percentage proportion of working men and women. As per year 2018, Norway has population that consists of almost equal shares of men and women: 2 668 371 and 2 627 248 respectively. Working population in Norway is more than 2.7 million people in the age 15–74. The gender distribution is relatively even, with 53 per cent for men and 47 per for women. In the working-age population (aged 15–74), 68 per cent of women and 74 per cent of men were in employment in 2016 (Norwegian Institute of Public Health. Public Health Report: Health Status in Norway 2018, 2018).

During the last decade, unemployment in Norway was relatively low in comparison to many other European countries. It is fluctuating between 2.5 and 4.7 per cent with a slight increase during period 2014-2016. In 2016, 5.4 per cent of men and 4 per cent of women were unemployed. Majority of unemployed are men. This fact is determined by prevailing of men workers in such industries as manufacturing and the oil industry that are to a greater extend exposed to cyclical fluctuations in economics impacting the unemployment levels.

There is a big difference between public and private sector in Norway. At the present, 70 per cent of public sector employees are women with almost half of all working women employed in the public sector. Men, on the contrary, occupy over 60 per cent of the private sector with only 20 per cent of working men employed in the public sector.

As per year 2016 (Bye, 2018), gender distribution is:

- in public (government) sector 29,9 per cent men and 70,1 per cent women;
- in private sector 63,4 per cent men and 36,6 per cent women;
- of municipality county members 61 per cent men and 39 per cent women.

Women in Norway tend to have more health complaints, even though the general level of health is estimated at approximately same for both genders. Women have more cases of reporting suffering from such symptoms as bodily pain, migraine, coughs, nausea, dizziness; indicators of poor mental health – irritability, insomnia, poor concentration, and long-term illnesses than men.

Finnmark and Troms counties are best in the country when it comes to indicators of gender equality: an even gender balance in the labor force, relatively even gender balance in the government sector employment and the gender gap in income and in part-time employment is also smaller than anywhere else in the country. That does not contradict the use of the reference groups I have selected in the analysis in the second part. Usage of women as a reference group is fully justified. The difference in employment and sector distribution percentage is not critical, and the number of registered sick leaves for women is greater.

The Data

In my research I use data collected by the Norwegian Labor and Welfare Directorate which contains all sickness absence episodes in Norway. It is daily data for eight years, starting 1st January 2010 and finishing 31st December 2017. The data contains unique ID number of workers assigned for the whole period, start and end of the sick leave, profession, sector – government, municipality or private, gender and age group. Professions are roughly grouped in 10 groups, see Table 1 below. There are nine age groups, starting with age 20 and in increments of five years with an upper bound of 66 years. Some of sick leaves start before the previous ended, these overlapping events are merged into one sick leave.

Table 1. *List of profession groups*

Group	Profession
Group 0	Military
Group 1	Administrative leaders and Politicians
Group 2	Academics
Group 3	College professions
Group 4	Office administration
Group 5	Sales and Service professions
Group 6	Farmers and Fishermans
Group 7	Craft Mans
Group 8	Machine, Transportation and Process
Group 9	Cleaner and Assistants

I start with filtering the data so that it perfectly fits the purpose of my research. Although I have data for all the Norwegian counties, I choose Troms county for my research, leaving other counties for the further research in the future given the data and method I use in my master thesis, likewise my results. I remove sick leaves that lasted longer than 365 days because after a year of sickness absence Norwegian workers have the sick pay reduced from 100% financial cover to only 66% and are being switched to work assessment allowance. Thereby potential financial motive is reduced after one year.

I downloaded additional data from the Meteorological institute (MET, 2019). It contains daily mean temperature, sum of precipitation and mean wind power for the period 2010-2017.
The sample consists of 291544 observations of registered leaves that required a medical certificate.

Data Insights

After I filtered the data and prepared it for analysis, I ran simple functions in RStudio to make several plots and get general numbers for the better understanding of the data.

The first thing that caught my attention was a distribution of sickness absence over time, namely the pattern in the Figure 1. It shows how many sick leaves were registered every day for the period 2010-2017 in Troms county. All the plots on the Figure 2 are more detailed, showing distribution of sick leaves for every year of the period.

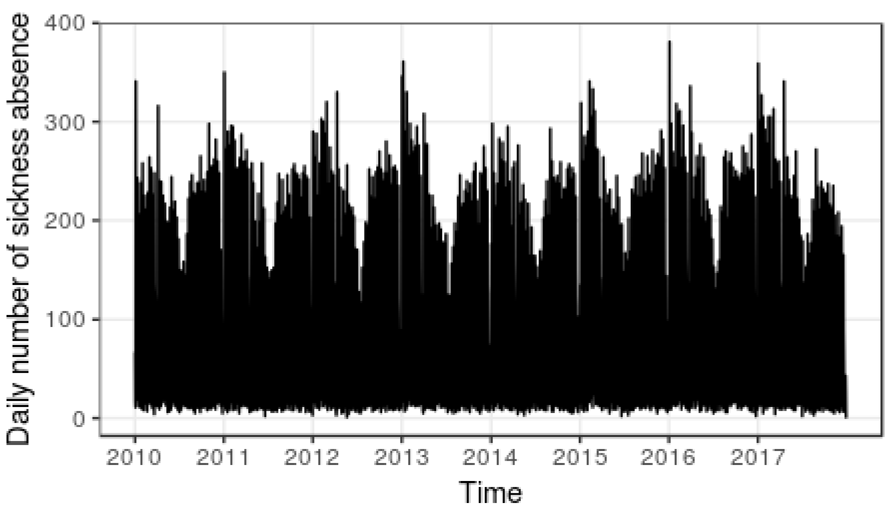

Figure 1. *Daily number of sick leaves*

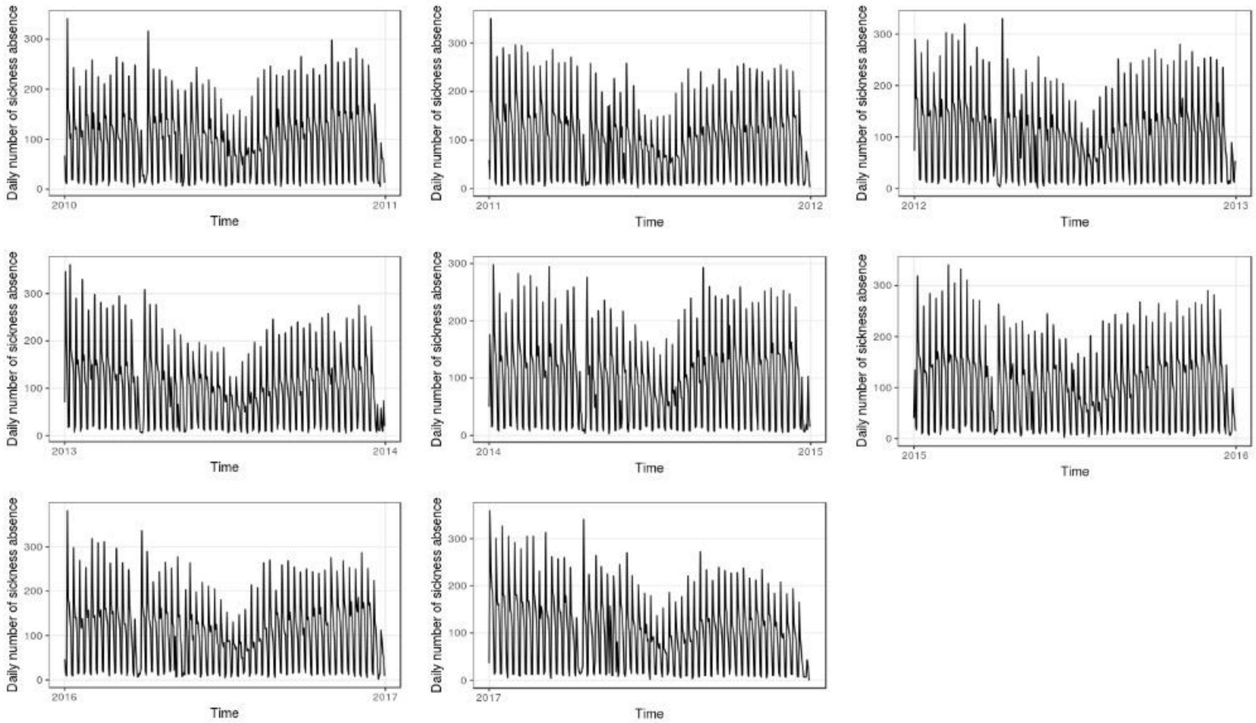

Figure 2. *Daily number of sick leaves for every year*

The repeating pattern clearly can be seen in the Figure 2. The first conclusion is that people get more sick leaves in the beginning of the year and the end of the year, and the minimum of the sick leaves happens during the middle of the year in summer, presumably June or July.

Figure 3 shows us the total number of sick leaves during the period for each month.

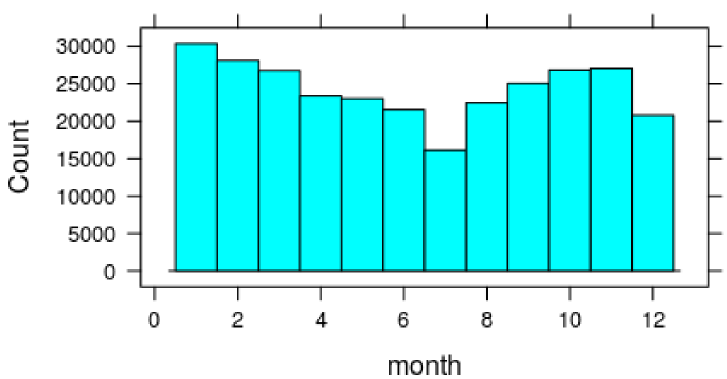

Figure 3. *Monthly distribution of sick leaves*

It proves our hypothesis that less sick leaves happen in summer, with the absolute minimum in July, 16417 sick leaves. January, on contrary, is the month with the highest number of sick leaves, 30365.

Number of sick leaves during the year varies depending on a month. After checking for a smaller period, a week, I found out that the most of sick leaves happen on Monday, 87003, and the number is descending during the week with lowest on Sunday, 5037. This explains fluctuations during the month in the Figure 3.

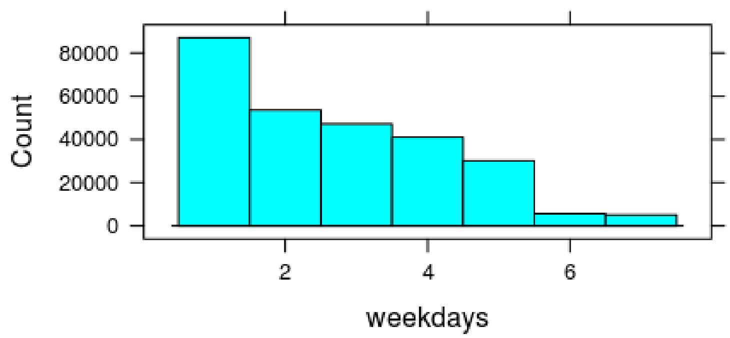

Figure 4. *Distribution of sick leaves during week*

The majority of sick leaves do not exceed 20 days, even fewer sick leaves last up to 100-150 days. Very few sick leaves that last 180-305 days are registered. An increase in the number of sick leaves happens in the very end of the graph, where the length of sick leaves reaches 350-365 days.

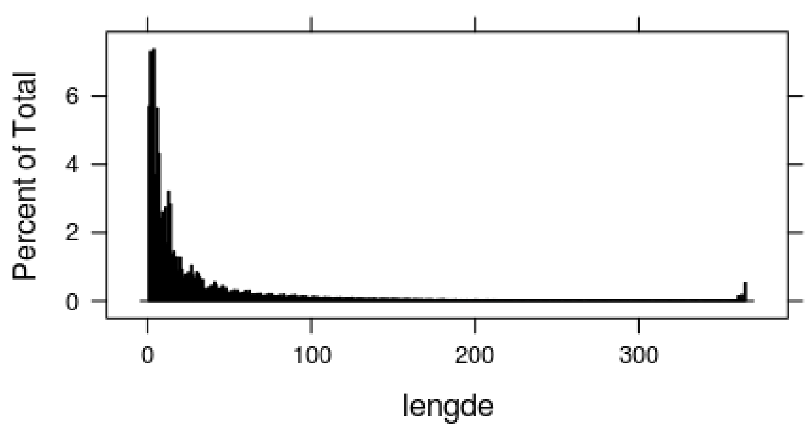

Figure 5. *Duration of sick leaves*

Time series with regular and predictable changes in data that repeat every year experience seasonality and have a seasonal pattern (Kenton, 2019b). The figures and numbers convinced me that there is a certain seasonality in the data. This hypothesis is also confirmed by the clear peak in the periodogram in Figure 6 at a frequency of 0.0833 that gives a cycle of 12 month. The total number of peaks is four which indicates that the seasonality is not sinusoidal.

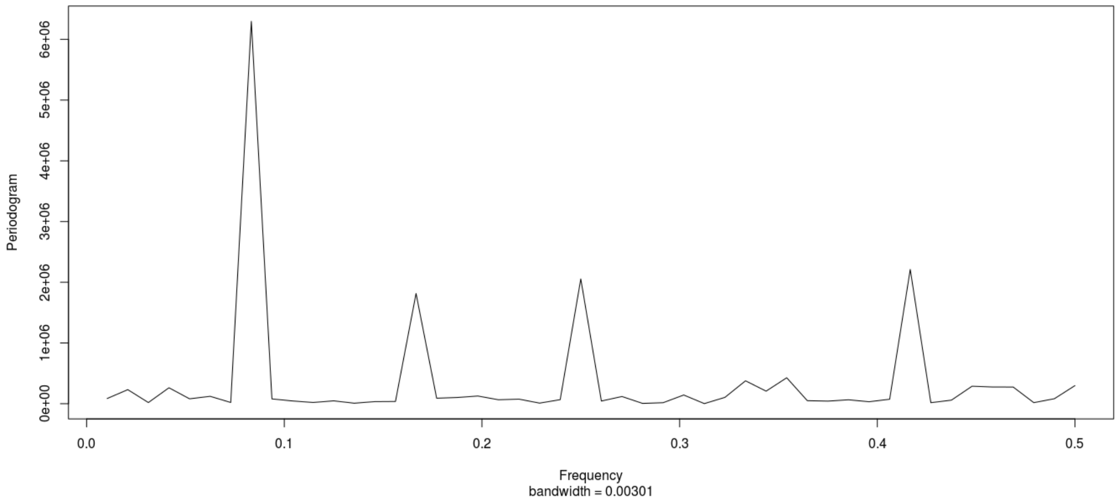

Figure 6. *Periodogram*

Before considering trends in the sickness absence separately for every group and period I checked the general trend for the entire population during the researched period with a simple STL model. The model plots all the registered sickness absence episodes over time and its decomposition on seasonal, trend and residuals (reminders) components, see Figure 7. The model is rough because it is not adjusted for the weekly seasonality so there is seasonality in residuals. However, it is good enough to confirm the expectations about the general trend: number of the registered sick leaves steadily increases over time.

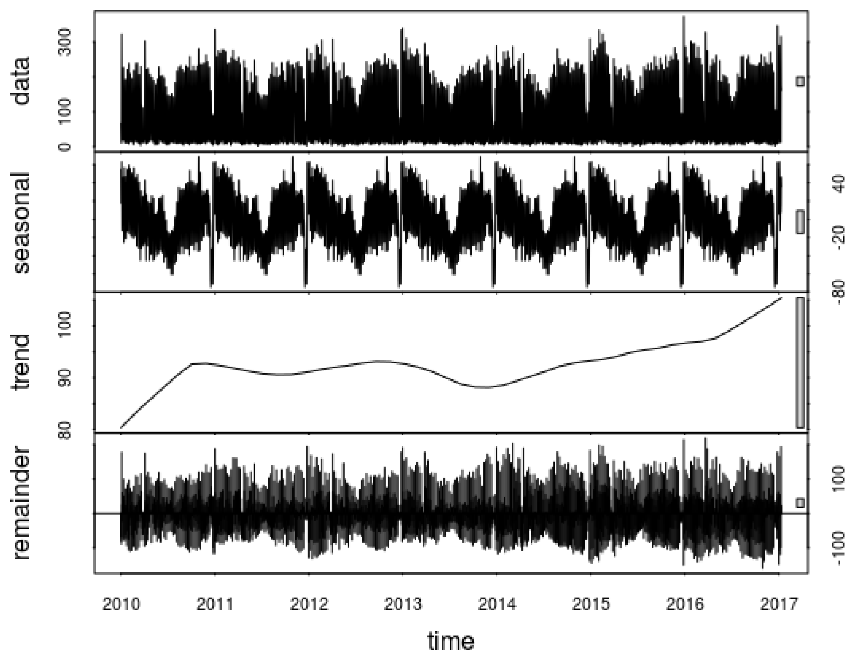

Figure 7. *Seasonality and trend in the data*

I describe the method I use for the further data analysis in the next section Method.

Method

There are five basic ways of assessing sickness absence (Hensing, 2010):

- frequency (sickness absence episodes);
- incidence rate (estimates of frequency per person-time);
- duration (mean or median days spent away during each episode of sickness absence);
- length (number of days of sickness absence);
- cumulative incidence (proportion of individuals absent during a specified period).

I use frequency as a method of assessing sickness absence and its trends in my master thesis.

A great variety of methods allow to look deeper into seasonality of health time series data. Among them is harmonic analysis using Fourier series and Fourier transforms, Cosinor regression model, STL (season, trend, loess) model, VAR (Vector Autoregression) model, GAM (Generalized Additive Model) and even more. At the very beginning of data analysis I applied all the listed methods for my data and it gave me an insight that the best option given my research question and the data is not to create a complicated model combining all the factors but rather split the factors and perform the analysis in two steps. In addition, I kept results from the STL model in the section Data Insight that allowed me to present time series data is the form of combination of trend, season and noise.

In the first part of the analysis I use multiply linear regression model for variables weekdays, month, year, mean temperature, sum of precipitation and wind power (the difference between two consecutive days for the three last variables, if more precise) to find out if they have an influence on the number of registered sick leaves, and to what extent if they do.

In the second part I use log-linear model with non-nested design for all the variables left to check how does the number of sick leaves in each group changes over various periods of time (week, month, year) as I saw significant coefficients for these variables in the first part of the analysis. In addition, there was a noticeable increase in coefficient in 2016. Temperature, precipitation and wind appeared to be insignificant and were removed from analysis in the second part.

Method Part 1

Multiple linear regression is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the linear relationship between the explanatory variables and response variable (Kenton, 2019a).

The general formula is expressed by following equation:

$$N_t = \beta_0 + \beta_1 * x_{i1} + \beta_2 * x_{i2} + \ldots + \beta_n * x_{in} + \epsilon,$$

where

N_t – response (dependent) variable;

x_i – explanatory variable;

β_0 – intercept;

β_n – slope coefficients for each explanatory variable;

ϵ – the model's error term;

i – number of observations.

In the first part of data analysis I use a multiple linear regression model where number of sick leaves registered per day is a dependent variable and days of week, month, year, mean temperature per day, precipitation per day and mean wind power per day are explanatory variables.

The model which contained all the explanatory variables at once was too complicated and showed unsatisfying adjusted R-squared value. The variable indicates how accurate the model is in terms of how much of the total variation the factors included in the model are able to explain. Thus, I used variables day of week, month and year expressed as factors and checked for significance of the weather describing variables added one by one. That model became less complicated and had a better fit.

All assumptions for the model are met: linear relationship between independent and dependent variables, the independent variables are not correlated with each other and sample consists of all population and residuals are normally distributed.

This part was necessary to find the relationship between variables so that only the relevant variables will be included in the second part of the analysis.

Method Part 2

A contingency table, or cross tabulation, is a type of table in a matrix format that displays the frequency distribution of the variables. They provide a basic picture of the interrelation between two variables and can help find interactions between them (Gupta, 2017). The term contingency table was first used by Karl Pearson in «On the Theory of Contingency and Its Relation to Association and Normal Correlation» published in 1904.

Until the late 1960's, contingency tables were typically analyzed by calculating chi-square values testing the hypothesis of independence. In the 1970's the analysis of cross-classified data changed notably after publication of a series of papers by L.A. Goodman on log-linear models that could be fitted to cross-classified data. The introduction of the log-linear model provided researches with a formal and thorough method for selecting a model for describing associations between variables (Jeansonne, 2002).

The log-linear model is one of the specialized cases of generalized linear models for Poisson-distributed data. Generalized linear models are used to describe how a dependent variable can be explained by a range of explanatory variables. The dependent variable can be continuous or discrete, and the explanatory variables can be either quantitative – covariates, or categorical - factors. The model is assumed to have linear effects on changes of the dependent variable defined by the function. This model is suitable for data with Gaussian, binomial or Poisson form of the error distribution (Calcagno, 2010).

Log-linear analysis is an extension of the two-way contingency table where the conditional relationship between two or more discrete, categorical variables is analyzed by taking the natural logarithm of the cell frequencies within a contingency table. Log-linear models model cell counts in contingency tables. They are more often used to evaluate multiway contingency tables with three or more variables involved. Log-linear models differ from other modeling methods because they do not distinguish between response and explanatory variables. All the variables in the log-linear model, both dependent and independent, are called response variables, and the model demonstrates association between them (Jeansonne, 2002).

The log-linear model described in this section is implemented from «Log Linear Models» (Jeansonne, 2002).

This log-linear model refers to the traditional chi-square test where two variables are evaluated to see if an association exists between the variables.

It can be formulated with a following equation:

$$Ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

where

$Ln(F_{ij})$ - log of the expected cell frequency for cell ij in the contingency table;

μ - mean of the natural log of the expected frequencies;

λ - effects of variables on the cell frequencies;

A, B - variables;

i, j – categories of variables.

Thus, λ_i^A is the main effect for variable A, λ_j^B is the main effect for variable B and λ_{ij}^{AB} is the interaction effect between variables A and B. The model is a saturated model because it includes all possible one-way and two-way effects.

For three log-linear models in the part two of the data analysis variables A and B are 1) gender and working sector group, 2) gender and age group and 3) gender and profession group.

As the chosen model is multilevel, it is reasonable to specify how do the factors relate to one another. Nested design allows us to study the levels of one factor within the levels of another.

One option is to create an independence model with the variables independent of one another. However, the saturated model that has as many coefficients as cells in the table due to pairwise associations between variables is more of an interest.

Non-nested design implies a combination of factors that are not related. It allows to model them both as a main effect and their interaction, if needed (Stockburger, 2016). I have a sample of all sickness absence episodes registered for the period of eight years in Troms county, and the factors like their gender, age, profession and sector are not related.

Thus, I use the non-nested design for factors in the three existing log-linear models and explore the development of relationship of variables over the time.

Results and Discussion

Results and discussion Part 1

In the first part of the data analysis I found the factors that affect number of registered sick leaves in Tromsø county for a period 2010-2017. Significant effect along with pronounced seasonality have day of week and month. Year variable does not have expressed clear seasonality, but they do have high estimate coefficients. However, p-value is low and close to zero only for the years 2015, 2016 and 2017 that makes them significant with the year 2016 having the highest estimate coefficient.

According to the obtained model, year 2010 has the lowest number of sick leaves: year 2010 is taken as a baseline and all the estimates for other years are positive. We can trace how did the number of sick leaves per year changed with the time using mentioned above estimates. It increases during years 2011 and 2012 (estimate coefficients equal 2.44 and 3.53 respectively), decreases a little in 2013 (1.99) and 2014 (1.74), makes a big leap in 2015 (6.77) and one even bigger leap in 2016 (10.08). In 2017 (5.63) the number of sick leaves decreases again to the number smaller than it was in 2015.

The biggest increase in 2016 and its significance can be explained by the fact that it is a leap year, and it was one extra day to call in sick. However, year 2012 is a leap year too, but it does not show that big difference with year 2010. However, the number of working days did not change for those years. Moreover, year 2016 is the only the significance at level 0 and year 2015 at 0.001. The possible explanation is an oil crisis that started in 2015 when oil price per barrel was lowest for the last 7 years and continued in 2016 with a lowest price for 13 years. Norway's economy suffered incredible losses and many people lost their jobs. A series of studies confirms that job loss leads to increased mental health problems, while mental health can improve when returning to work (McKee-Ryan, 2005). Unemployment level in Norway was highest for the last decade in 2015 and 2016 with maximum of 5.2 per cent and started decreasing afterwards with 4.1 per cent in 2017 (Economics, 2019b).

I also detected a pronounced seasonality in days of week, see Table 5 in Appendices. Monday – first day of week, the baseline level – is the day where most of the registered sick leaves happen.

Their number is highest on Monday and gradually diminishes during the week with the lowest on Sunday, day seven.

There are several reasons for that form of the sickness absence distribution. First, weekends mean a lot of activities for the most of Norwegians. Doing active sports, going out, hiking in the mountains, fishing, winter sports and other physical and outdoor activities are popular among Norwegians, especially on weekends because there is more time for it. And these activities increase health risks such as injury from sports or getting cold or sunburned depending on season after staying outdoors longer than usual. Secondly, doctors are not available during the weekends. The only place with medical help opened is an emergency room. If there is a minor problem, the visit to the doctor as a rule is postponed until Monday. That explains the lowest number of registered sick leaves at weekends and the highest on Mondays. The decline during the week in its turn can be explained by the upcoming weekends. If the self-diagnostics shows that the troubles are minor and will probably go away by themselves, it is not worth a visit to the doctor, especially with the upcoming weekend that will presumably bring rest and healing. It is also possible to use self-declared sick leave that does not require medical certification and can be used for up to three calendar days at a time.

Seasonality can be seen in the months, too. January - the baseline level, and February are on top of the list with the highest number of registered sick leaves during the year. Afterwards comes a decline with the bottom level in July, and then numbers start to increase back again until December. The decrease in December can be explained by the Christmas holidays. It is a family holiday and one of the most important celebrations in Norway that requires considerable preparations. I assume that the same logics with the self-declared sick leave applies here: people would rather take a shorter sick leave than spend time on visiting doctors having minor health issues.

Weather related variables – mean temperature per day, precipitation per day and mean wind power per day, namely the difference between the reporting day and the previous day, do not affect the sickness absence rate in this model and were excluded from the further evaluation in the part 2.

In the second part of the data analysis I use log-linear model to determine the dependence of sickness absence level for groups of people of different gender, age, sector of work and profession from variables that showed their significance in the first part of the analysis.

Results and discussion Part 2

After having converted data to a data frame and array to a table, I modeled frequency of sick leaves for a number of combinations with two different factors for my analysis using the glm function in RStudio. The sector-gender, age-gender and profession-gender combinations based on log-linear models showed the most favorable results and are discussed further.

1. Sector-gender combination model

The reference level for the first combination is women working in municipality sector, for the second combination – women from the first age group, 20-24, and women working in military for the third combination. This means the models' coefficients will be expressed in terms with the reference level.

Estimate of the intercept in the model shows the number of registered sick leave for women working in municipality sector per day in the given time period. Next row estimate (kjonnMenn) shows the difference between the mean of registered sick leaves per day for men and women, working in municipality sector. The next two lines (sektor_kodePrivat and sektor_kodeStat) show how does the mean of the registered sick leaves for women in those two sectors differ from the mean of the registered sick leaves for women in municipality sector.

Standard error displays the quality of regression fit: the smaller, the better. The standard error in this model meets the requirements. Low p-value in the model throughout the time series data verifies that there is a relationship between variables. A predictor with a low p-value is likely to be a meaningful addition to the model because changes in its value are related to changes in the dependent variable.

I built several line graphs using the estimates derived from the model for clarity and ease of perception of the information. Calculated estimates in the Figure 8 show the difference in number of sick leaves among men and women in municipality sector. Each graph represents change of the estimates over time for every month starting from January for the period 2010-

2017, where *a* depicts estimates for 1st quarter for years 2010-2017, *b* – second quarter, *c* and *d* – third and fourth quarters. The letter *a* stand for January, February, March top down; *b* - April, May, June; *c* - July, August, September; *d* - October, November, December. The same designation fits all figures of similar structure in this thesis.

The maximum value of the intercept is 7.12 in January 2013 and the minimum value is 6.12 in July 2014. The intercept shows how many registered sick leaves do women have on average per day during the stated month.

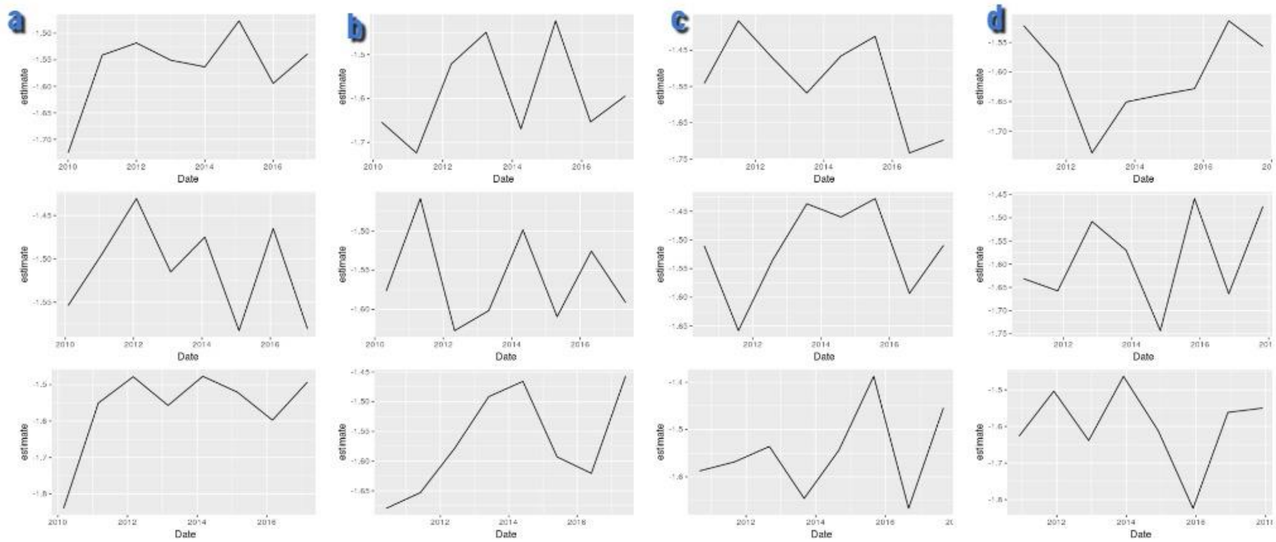

Figure 8. *Monthly dynamics of the estimates for men in the municipality sector*

The maximum value of the estimate is -1.37 in July 2011 and the minimum is -1.84 in March 2010. From the Figure 8 we can see that for majority of months over time there is tendency in reducing the gap between sickness absence among genders: line moves upwards, estimate coefficients becomes bigger - closer to zero, where values would be the same for both men and women.

It is worth noting that estimate of the number of sick leaves among men in both government and private sector is strictly positive during the entire period.

The Figure 9 below describes sickness absence rate among women in private sector compared to sickness absence of women in municipality sector. As women have more sick leave registered in general, I find it reasonable to compare changes between sectors in this gender group.

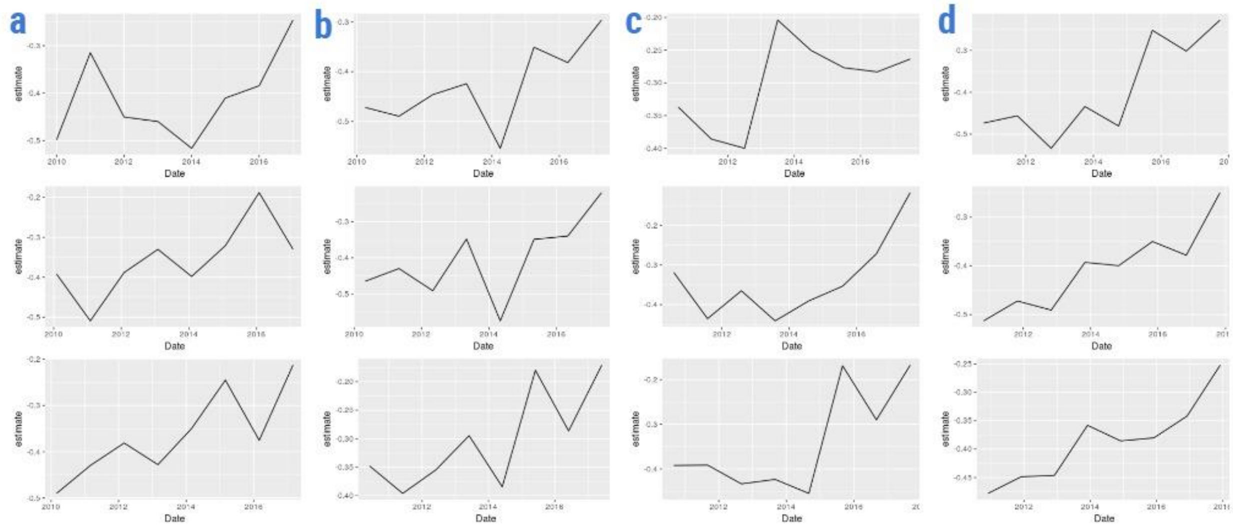

Figure 9. *Monthly dynamics of the estimates for women in the private sector*

Even though the graphs tend to move upwards over time, none can reach the origin to make the sickness absence level in municipality and private sectors equal. The maximum value of the estimate is -0.12 in August 2017 and the minimum value is -0.57 in May 2014. All the graphs stop started from approximately -0.5 in 2010 and reached -0.25 in 2017. The rise is noticeable.

The year 2014 shows deviations with strong dips in the first half of the year and unexpected peak in July. Other years do not show that big fluctuations from month to month. Thus, women in private sector still register less sick leave than women in municipality sector.

The Figure 10 below describes sickness absence rate among women in government sector compared to sickness absence rate of women in municipality sector over time.

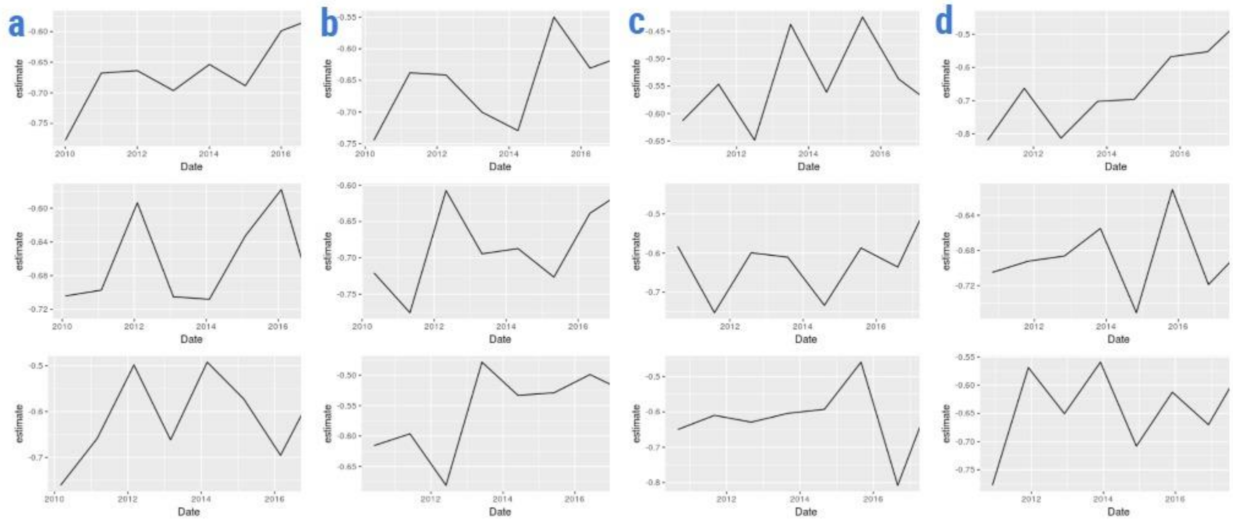

Figure 10. *Monthly dynamics of the estimates for women in the government sector*

Estimate coefficients for sick leave among women in the government sector rise from -0.75 in 2010 to -0.5 in 2017. The maximum value of the estimate is -0.44 in July 2015 and the minimum value is -0.82 in October 2010. Fluctuations for every month over time do not have obvious pattern. The average total change is the same for the government sector and the private sector (approximately 0.25) so that we can conclude that the sickness absence grows at the same rate in both sectors.

To sum up, government sector has the least of sick leaves registered among women, followed by the private sector and with the municipality sector leading. Both private and government sectors tend to get closer to the municipality sectors' value over time. Men workers prevail in the municipality sector, but the number of registered sickness absence among them in the municipality sector is lower. This coincides with widely spread finding that women tend to have more cases of the registered sick leaves. Though, the research on the reasons do not come to a common conclusion.

2. Age-gender combination model

The Figure 11 reflects the dynamics of the sickness absence of women in age groups from two to nine compared to the reference age group one which is 20-24 years. Maximum and minimum value of intercept coefficients in the age-gender combination model was 5.47 in January 2017 and 4.59 in April 2011. Thus, all the monthly means of registered sick leaves per day for the women of age 20-24 are included in the interval 4.59 – 5.47.

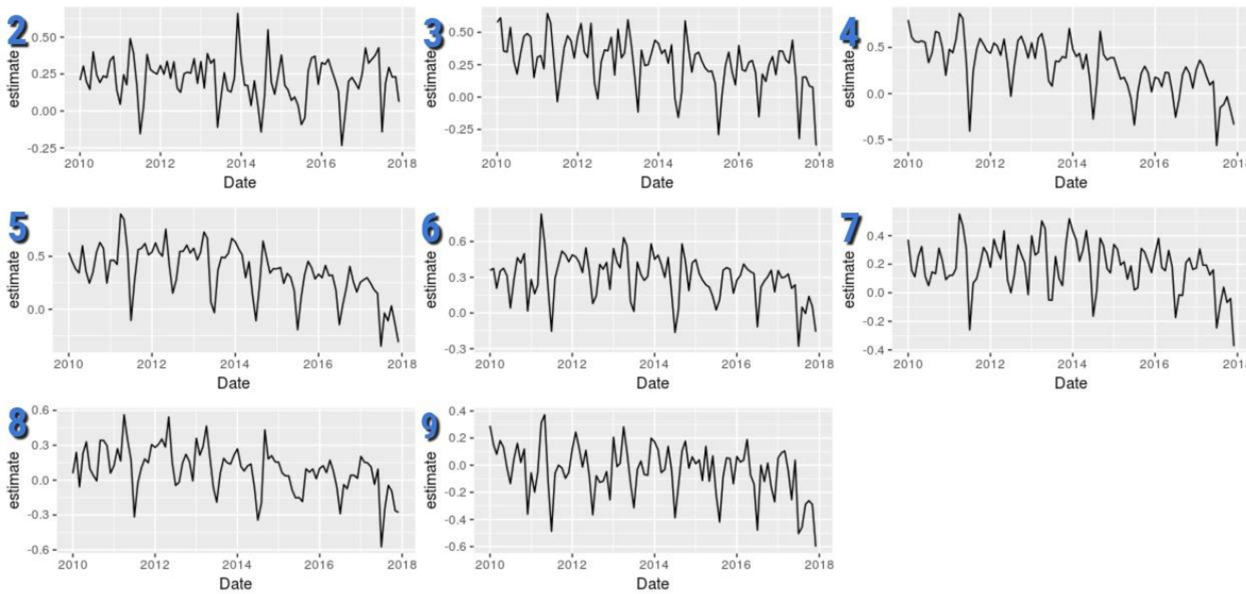

Figure 11. *Dynamics of the estimates for age groups*

The highest estimate has the age group four in April 2011 (0.89) and the lowest estimate (-0.6) belongs to the age group nine in December 2017.

All the groups tend to have gradual approach to 0.00 point which means the coincidence of the value with the reference group one. Graphs of groups from two to eight intersect the origin in 2011 for the first time and have more intersections after 2014 again. Graph for the group nine is the only exception that had intersection for the first time in 2010 and demonstrates being below the x-axis for a significant part of each year.

The graphs mostly have similar patterns that repeat from year to year with the increment in winter and autumn period and the falling in the summer period. For all age groups, the characteristic features of the behavior can be observed in the period 2014-2016.

The graphs present us fluctuations of the estimate coefficients during the year that happen in a certain interval for each group as well as peaks with prominent values.

Fluctuations are slightly different. Groups two, six, seven and nine appear to have the lowest amplitude of fluctuations (0.60), groups five and eight have amplitude 0.7 and 0.8 respectively, and groups three and four have the highest amplitude among all (0.85).

The general tendency of the estimate coefficient values of the groups is decline for the entire period of research cannot be denied. However, not all the graphs are smooth. There are some extreme numbers expressed as too sharp peaks or dips on the graphs. The next step was to check graphs with comparison of estimates coefficients for every group, every month and every year.

The Table 2 below presents a change in the estimates for age groups compared to the previous year in January. It is a part of a Table 5 which contains all months and can be found in Appendices.

The following notation is used in the table:

↑ - significant increase in estimate coefficient compared to the previous year;

↓ - significant decrease in estimate coefficient compared to the previous year;

↗- slight increase in estimate coefficient compared to the previous year;

↘ - slight decrease in estimate coefficient compared to the previous year;

– - no change in the estimate coefficient.

I highlighted the deviation from the general direction with a red color.

Table 2. *Yearly changes in January in the sickness absence trends for age groups*

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
J	Group 2	↓	↑	↗	↗	↗	↘	↘
A	Group 3	↓	↑	↗	↘	↘	↗	↘
N	Group 4	↓	↘	↗	↘	↘	↓	↗
U	Group 5	↓	↑	↗	↗	↓	↘	↘
A	Group 6	↓	↑	↗	↓	–	↓	↑
R	Group 7	↓	↗	↑	↗	↓	↘	↘
Y	Group 8	↗	↑	↗	↓	↓	↘	↑
	Group 9	↓	↑	↗	↘	↓	↗	↘

Throughout the period of analysis, age groups from two to nine tend to decrease in estimates. In other words, more women from age groups from two to nine registered for sick leave than women in the reference group one in the beginning of the period, year 2010, but the situation changed over time. In the end of the period, women from group two have approximately the same level of sick absence as the women from the reference group. Women from groups from three to nine have lower level of sick absence in 2017 than the women from the reference group one.

According to the Figure 11 and Table 2, this trend is reflected best during the years 2013-2015 and 2017: the graphs are rather smooth for majority of groups and there are less outliers in the table with uniform decline in the estimate coefficients for age groups two-nine. The most disparate changes happen in 2012 and 2016.

The breaking of tendencies can be observed in the leap years. However, the years before and after them have the best performance and follow the indicated trends most clearly. In the first part of analysis I got the strongest estimate coefficient - 10.082 - for the leap year 2016, with 6.772 in 2015 and 5.628 in 2017 while the year 2014 has an estimate coefficient equal to 1.741. Age groups show disunity in following the trends during the year 2012 itself but years before and after, 2011 and 2013, display better following of trends than years 2010 and 2014.

Year 2016 started with the lowest price for oil for last 13 years - $ 27.7 per barrel that caused significant damage to the economy of Norway. Oil exports account for over 65 per cent of Norwegian exports and 195 000 people are directly or indirectly employed in the petroleum sector that corresponds to approximately 6 per cent of the total employment in Norway (NorwegianPetroleum, 2018). The petroleum sector experienced reduction in employees in 2015 and 2016 after many years of the strong growth. The situation with the extremely low prices for oil has caused stress among workers in the industry and other citizens of Norway that could have had influence on the health and caused increase of the sick absence among more informed and interested in the news age groups, namely age groups from two to nine.

As for the monthly analysis, the period from August to November showed the most coordinated chart changes (decrease) for the age groups. The period from December to March, on contrary, showed disconnected movements of graphs.

The general direction for groups from two to nine is the decline in the estimates. However, there are situations when a single group has an opposite direction of change of the estimate then others have. Let us call them deviation periods. Some groups show more deviation periods than the others. Group two has the highest number of the deviation periods (24), followed by group seven (20) and eight (19). Afterwards come groups three (16), four (16) and nine (15). The lowest number of derivations from the general direction have groups five (11) and six (12).

This is followed by results of the next profession-gender combination model.

3. Profession-gender combination model

The collage below in the Figure 12 on the next page represents the difference in the average number of registered sick leaves among women from profession group zero, our reference group, and women with profession groups from one to nine. The list of professions in each group can be found at the end in the Appendices.

The reference group zero is women working in the military. Figure 13 (3) in Appendices shows the number of sickness absence rates among men in military in relation to women in the military. We can see that the number that was approximately same in 2010 is constantly growing, on contrary to the similar type relationship in the sector-gender and age-gender combination models, see Figure 13 (1, 2). Thus, women have less sickness absence rate in this group than men, and it might affect the results of estimation.

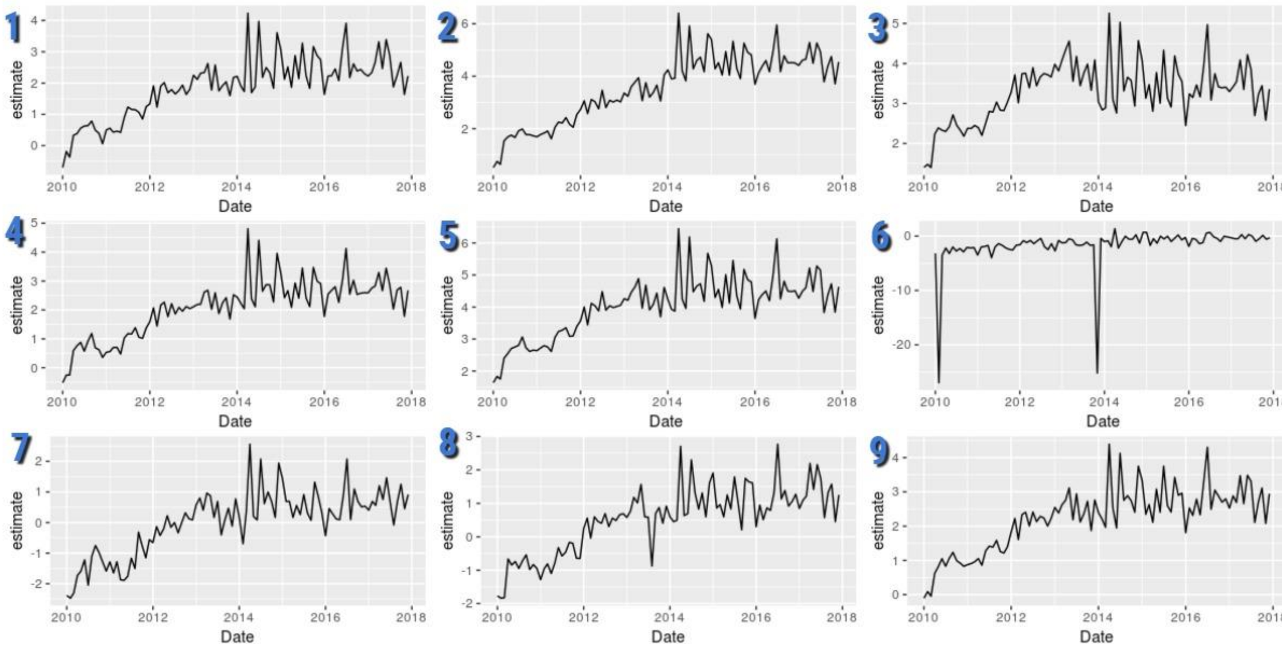

Figure 12. *Dynamics of the estimates for profession groups*

The maximum value for the intercept is 4.95 in January 2010 and the minimum value is 4.67 in July 2016. The fluctuations in the sickness absence rate of women in military are rather low and the sickness absence rate is approximately same over time.

The maximum coefficient has the group five (6.45) that is women in sales and service profession. So that the biggest difference in sickness absence rate with women in military are women in sales and service. It is followed by group two (6.41), group three (5.26), group four (4.81), group nine (4.40), group one (4.23), group eight (2.71), group seven (2.56) and group six (1.38). It is noticeable remark is that all the groups reach the maximum for their estimates in April 2014 with groups seven and eight being an exception, they reach the maximum in July 2016.

The minimum coefficient has the group six (-26.97) in February 2010. The sharp dip can be seen in the Figure 12 above. It is an extreme value. It is followed by group seven (-2.48), group eight (-1.84), group one (-0.7), group four (-0.51), group nine (-0.1), group two (0.52), group three (1.39) and group five (1.63). All of the groups reach the minimum value of the estimate in January and February 2010 when they have the closest sickness absence rate to the reference group zero, women working in military.

For all the graphs above except sixth we see an increase over time starting either from origin or below in 2010 and finishing in the interval between 1 and 3. The graph for the groups six started from about -5 and finished close to the origin with two sharp dips in the beginning of 2010 and the end of 2013.

The group six is farmers and fishermen. Both professions have men as the major part of the employees. Military members have a strict discipline and regime while farmers and fishermen are mostly self-employed, and their income directly depends on the amount of time spent at work. I think this explains the lowest level of sick leaves for the groups zero and six in general.

Two sharp dips in February 2010 and November 2013 for group six can be explained more by the events in the reference group. In year 2010 Norway was in the middle of the military operation in Afghanistan and approximately 500 Norwegian soldiers were deployed there. It might have affected the sickness absence rate in military in a big way. The dip in November 2013 for group six is caused by increase in sickness absence among members of the reference group. It was the first time the Armed Forces started with a joint operative exercise for various branches of defense to coordinate and train together. More than 3,200 personnel from Norway and allies participated (Forsvaret, 2019). During the operative exercises the fire started on military equipment on the ship. None of people aboard the ship were seriously injured, but they were still transported to the hospital for a check. Operation itself could have led to some injuries for participants so it might explain the strong increase of the sickness absence in November 2013 (Huso, 2013).

Table 6 in Appendices presents changes in the estimates for professions groups compared to the previous year for each month and can be found in Appendices. The form of the table and the notation is similar to the table used for age-gender combination model in the previous section. Table 3 below is the part of the table showing movements of the groups estimates in January.

Table 3. *Yearly changes in January in the sickness absence trends for profession groups*

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
	Group 1	↑	↑	↑	–	↑	↓	↗
J	Group 2	↗	↑	↗	↗	↑	↓	↗
A	Group 3	↑	↗	↗	↓	↑	↓	↗
N	Group 4	↗	↑	↗	↗	↗	↓	↗
U	Group 5	↑	↑	↗	–	↗	↓	↗
A	Group 6	↘	↑	↗	↗	↑	↓	↑
R	Group 7	↗	↗	↑	↘	↑	↓	↗
Y	Group 8	↗	↑	↗	–	↑	↓	↗
	Group 9	↑	↑	↗	↗	↗	↓	↗

From the table we can conclude that the general tendency for all the groups during the entire research period is growth of the estimates with year 2016 being an exception. Years 2012 and 2013 show have the most pronounced trend with all the months showing an increase in estimates compared to the previous year. Years 2011 and 2015 have one month which has decline in the estimate prevailing, 2014 and 2017 have four and five months respectively with a decline in the estimate. Year 2016 is the most controversial with two months showing growth, six months showing decline and three months with mixed direction of the movement of the estimate.

Among the months the unanimous growth shows June with all the seven periods showing growth of the estimate. It is followed by January, September and October with six periods of the growth and one with the decline. After that come February and April with five episodes of growth, one episode of the decline and one with the mixed directions. Then March, August, November and December with five episodes of the growth and two with the decline. July has five episodes of growth and one of decline, and May is the most controversial with four episodes of the growth and three of the decline.

There is no common direction for the groups from one to nine in April 2016. Hence, this time period is taken as a deviation period for all the groups. Group one has only this month as a deviation from the common direction of the change is the estimate. Groups five and nine follow with two deviation months, group four with four deviation months and groups three and eight with five deviation months. Groups three and six have eight and seven deviation months respectively. Group seven deviates the most from the common direction of the estimate movement with eleven deviation months.

The age-gender combination model demonstrates more deviation periods among single groups when compared to the single groups from the profession-gender combination model. This may result either from too general division into profession groups, or some additional factors concerning people of different age groups that were not in the model.

Distinctive performance again is observed in 2010, 2014 and 2016. These years have significant differences in GDP compared with previous years: year 2010 is characterized by consequences of the economic crisis that started in 2008, year 2014 is remembered for the strong weakening of Norwegian krone, sharp increase in public expenditure and lower consumption in Norwegian households (Economics, 2019a). Year 2016 is characterized by the lowest price for the oil barrel in 13 years. All these events had a tremendous effect on the economy, and it is reflected in the deviant behavior in sickness absence rate in all three models.

Concluding Remarks

In this thesis the relationship between number of factors such as year, month, day of the week, weather describing variables, gender, age, sector, profession and sickness absence in Troms county was investigated. The observation period lasts for eight years from 2010 to 2017. The data sample consists of 291544 observations of registered leaves that required a medical certificate.

I use frequency as a method to assess sickness absence and its trends with help of multiple linear regression model, the log-linear regression model and its form with non-nested factors to estimate relationship between variables. I investigate difference in trends between the subgroups in detail. I also use simple STL model to confirm that the number of registered sick leaves grows over the given period.

I confirmed my hypothesis that there is certain seasonality in data. It is most clearly expressed in the framework of the week with Monday having the highest sickness absence rate and the consistent reduction during the week with the lowest on Sunday. During the year, the lowest number of registered sick leaves shows July and the highest level have January and February. This coincides with the fundamental findings that the sickness absence level is the highest in winter and the lowest in summer. The observations show correlation between trends in the sickness absence and the major economic events leading to the change of the GDP. Usually such important events lead to change of the unemployment rate that is proven to have influence on the sickness absence rate change.

Mean temperature, precipitation and mean wind appeared to not affect the sickness absence.

The sickness absence among men is than among women, as expected, though the reason for that is yet to learn. The municipality sector is leading in the number of the registered sick leaves, followed by the private sector and the government sector. Both private and government sectors tend to reduce the gap with the municipality sector over time.

The sickness absence in the age groups showed dependence on the time of the year mentioned above. However, some groups reacted more: relationship between the season and sickness absence for people in the age 25-29, 45-54 and 60-66 is weaker than for people in the age 30-44 and 55-59. The latter group tend to suffer more from the seasonal diseases or its escalations.

The general tendency for all the age groups is increase in sickness absence rate over time. The noticeable remark is that the focus is moving towards the younger population that tends to have more sick leaves over time. The possible reason for it is different attitude towards sickness absence – older population have stricter views, or the worse working conditions for younger employees.

The general tendency for all the profession groups is increase in number of sickness absence rate. Years 2012 and 2013 show have the most pronounced trend of the growth in all the months showing an increase in estimates compared to the previous year. Year 2016 is the most controversial and does not show clear trend pattern. The correlation between stable economic situation and steady trend of sickness absence comes out again. Administrative leaders and politicians, sales and service professions, cleaners and assistants have the least deviation from the general trend over the research period and follow the general sickness absence pattern. College professions, farmers, fishermans and craftsmans deviate the most from common sickness absence trend. In addition, the age-gender combination model demonstrates more deviation periods among single groups than the profession-gender model. The reasons for that are unclear.

The master thesis answers my research question to my satisfaction. For the best of my knowledge, the obtained results do not contradict with the fundamental previous findings. Nevertheless, the research can be broadened by adding data on socio-economic status of employees, creating more groups for professions, deeper investigation of the deviant behavior and making analysis for all the counties in Norway.

This study can be useful to the state, individual enterprises, as well as independent researchers and all the people concerned about sickness absence trends and risk factors in Norway, namely Troms county.

References

Bye, T. (2018). *Women and men in Norway*. Retrieved from

Calcagno, V., Mazancourt, C. (2010). glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. *Journal of Statistical Software, 34*(12). Retrieved from http://www.jstatsoft.org/.

Economics, T. (2019a). Norway GDP. Retrieved from https://tradingeconomics.com/norway/gdp

Economics, T. (2019b). Unemployment rate. Retrieved from https://tradingeconomics.com/norway/unemployment-rate

Finansdepartementet, D. K. (2018). Folketrygdens utgifter. *Arbeids- og sosialdepartementet.*

Forsvaret. (2019). Øvelser. Retrieved from https://forsvaret.no/fakta/aktivitet/ovelser

Gupta, S. (2017). Contingency Tables in R. Retrieved from https://www.datasciencecentral.com/profiles/blogs/contingency-tables-in-r

Hagen, K. P., Erga, Aleksander Hagen. (2012). *Is there an increase in duration of long-term sickness absence in Norway from 1994 to 2007? A registry based study.* (Master), University of Bergen University of Bergen

Hensing, G. (2010). The measurements of sickness absence – a theoretical perspective. *Norsk Epidemiologi, 19(2)*. Retrieved from https://doi.org/10.5324/nje.v19i2.584.

Hernæs, Ø. (2018).

Activation against absenteeism – Evidence from a sickness insurance reform in Norway. *Journal of Health Economics, 62,* 60-68.

Huso, A. (2013). Militærøvelse ble branndrama. *Bergensavisen*.

Jeansonne, A. (2002). Loglinear Models. Retrieved from http://userwww.sfsu.edu/efc/classes/biol710/loglinear/Log%20Linear%20Models.htm

Johansen, V. (2013). Risk factors of long-term sickness absence in Norway and Sweden. *NJSR - Nordic Journal of Social Research, 4*.

Kenton, W. (2019a). Multiple Linear Regression – MLR Definition. Retrieved from https://www.investopedia.com/terms/m/mlr.asp

Kenton, W. (2019b). Seasonality. Retrieved from https://www.investopedia.com/terms/s/seasonality.asp

Krane, L., K., Johnsen, Roar, J., Fleten, Nils, F., Nielsen, Claus Vinther, N., Stapelfeldt, Christina Malmose, S., Jensen, Chris, J., & Braaten, Tonje, B. (2014). Sickness absence patterns and trends in the health care sector: 5-year monitoring of female municipal employees in the health and care sectors in Norway and Denmark. *12(1), 37*.

Krane, L., Larsen, E.L., Nielsen, C.V. et al. BMC Public Health. (2014). Attitudes towards sickness absence and sickness presenteeism in health and care sectors in Norway and Denmark: a qualitative study. *14 : 880*.

Loset, G. K., Dale-Olsen, Harald, Hellevik, Tale, Mastekaasa, Arne, Von Soest, Tilmann, & Ostbakken, Kjersti Misje. (2018). Gender equality in sickness absence tolerance: Attitudes and norms of sickness absence are not different for men and women. *PLoS ONE, 13(8) , E0200788*.

Markussen, S., Mykletun, A., Røed, K. (2012). The case for presenteeism — Evidence from Norway's sickness insurance program. *Journal of Public Economics, 96*(11-12), 959-972.

McKee-Ryan, F., Song, Z., Wanberg, C. R., & Kinicki, A. J. (2005). Psychological and physical well-being during unemployment: a meta-analytic study. *J Appl Psychol, 90 (1)*, 53-76.

MET. (2019). Free meteorological data.

Norwegian Institute of Public Health. Public Health Report: Health Status in Norway 2018. (2018). Retrieved from https://www.fhi.no/contentassets/d021a759c5ed48ae85fffc94e35785cf/health_status_in_norway_2018.pdf:

NorwegianPetroleum. (2018). Employment in the petroleum industry.

RStudio. (2015). RStudio: Integrated Development for R. RStudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/.

SSB. (2019). Key figures for the population. Retrieved from https://www.ssb.no/en/befolkning/nokkeltall/population

Stockburger, D. W. (2016). *Introductory statistics: concepts, models, and applications* Missouri State University.

Thorsen, S. V., Friborg, C., Lundstrøm, B., Kausto, J., Ørnelius, K., Sundell, T., Kalstø, Å.M., Thune, O., Gross, B.O., Petersen, H., Haram, Ø. . (2015). Sickness Absence in the Nordic Countries *Nordic Social Statistical Committee, 59 : 2015*.

Appendices

Table 4. *Working variables names*

kjonnMenn	Male
Kommune	Municipality sector
Privat	Private sector
Stat	Government sector
Aldergr1	Age 20-24
Aldergr2	Age 25-29
Aldergr3	Age 30-34
Aldergr4	Age 35-39
Aldergr5	Age 40-44
Aldergr6	Age 45-49
Aldergr7	Age 50-54
Aldergr8	Age 55-59
Aldergr9	Age 60-66
Yrke0	Military
Yrke1	Administrative leaders and Politicians
Yrke2	Academics
Yrke3	College professions
Yrke4	Office administration
Yrke5	Sales and Service professions
Yrke6	Farmers and Fishermans
Yrke7	Craftsmans
Yrke8	Machine, transportation and Process
Yrke9	Cleaners and assistants

Table 5. *Multiple regression model results*

```
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           226.9573     2.8976  78.326  < 2e-16 ***
as.factor(weekdays)2  -80.5618     2.1767 -37.011  < 2e-16 ***
as.factor(weekdays)3  -95.8433     2.1780 -44.005  < 2e-16 ***
as.factor(weekdays)4 -110.2737     2.1780 -50.630  < 2e-16 ***
as.factor(weekdays)5 -136.2482     2.1794 -62.517  < 2e-16 ***
as.factor(weekdays)6 -195.4775     2.1794 -89.695  < 2e-16 ***
as.factor(weekdays)7 -197.0682     2.1780 -90.480  < 2e-16 ***
as.factor(month)2       0.2181     2.8956   0.075 0.939952    
as.factor(month)3     -14.9734     2.8237  -5.303 1.23e-07 ***
as.factor(month)4     -26.0352     2.8443  -9.154  < 2e-16 ***
as.factor(month)5     -30.6830     2.8209 -10.877  < 2e-16 ***
as.factor(month)6     -31.2961     2.8443 -11.003  < 2e-16 ***
as.factor(month)7     -51.6017     2.8208 -18.293  < 2e-16 ***
as.factor(month)8     -29.1315     2.8209 -10.327  < 2e-16 ***
as.factor(month)9     -16.2319     2.8443  -5.707 1.27e-08 ***
as.factor(month)10    -13.5705     2.8208  -4.811 1.58e-06 ***
as.factor(month)11    -10.5257     2.8599  -3.680 0.000237 ***
as.factor(month)12    -37.0525     2.8296 -13.095  < 2e-16 ***
as.factor(year)2011     2.5019     2.3366   1.071 0.284375    
as.factor(year)2012     3.5929     2.3268   1.544 0.122667    
as.factor(year)2013     2.0622     2.3284   0.886 0.375877    
as.factor(year)2014     1.8080     2.3300   0.776 0.437834    
as.factor(year)2015     6.8395     2.3284   2.937 0.003336 ** 
as.factor(year)2016    10.1494     2.3269   4.362 1.33e-05 ***
as.factor(year)2017     5.6965     2.3333   2.441 0.014690 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.41 on 2886 degrees of freedom
Multiple R-squared:  0.8143,	Adjusted R-squared:  0.8128
```

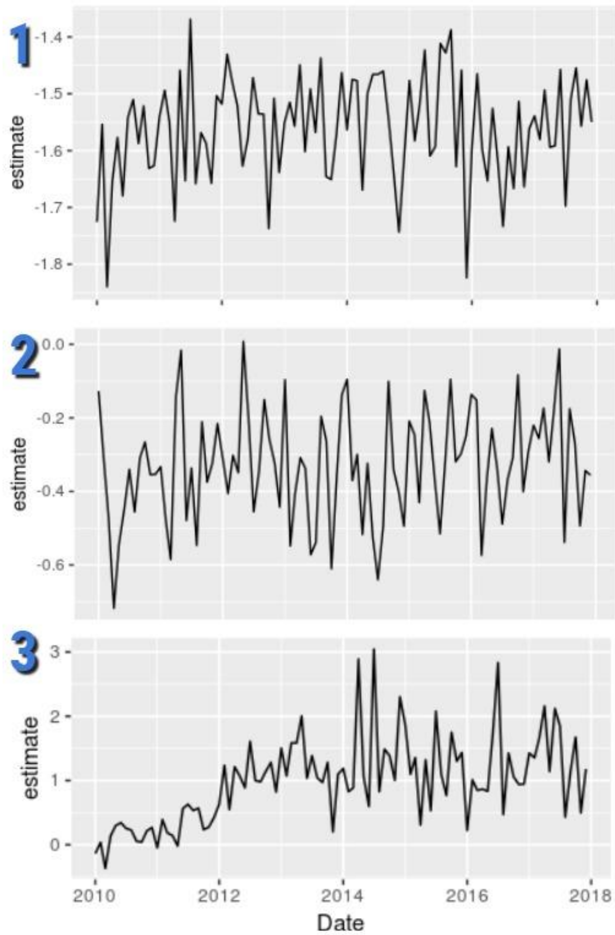

Figure 13. *Yearly dynamics of the estimates for men in the sector-gender (1), age-gender (2) and profession-gender (3) combination models*

Table 6. *Yearly changes in the sickness absence trends for age groups for all months*

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
J	Group 2	↓	↑	↗	↗	↗	↘	↘
A	Group 3	↓	↑	↗	↘	↘	↗	↘
N	Group 4	↓	↘	↗	↘	↘	↓	↗
U	Group 5	↓	↑	↗	↗	↓	↘	↘
A	Group 6	↓	↑	↗	↓	–	↓	↑
R	Group 7	↓	↗	↑	↗	↓	↘	↘
Y	Group 8	↗	↑	↗	↓	↓	↘	↑
	Group 9	↓	↑	↗	↘	↓	↗	↘
F	Group 2	↓	–	↓	↗	–	↑	↑
E	Group 3	↓	↑	↓	↗	↘	↘	↑
B	Group 4	↓	↑	↓	↗	↓	↓	↑
R	Group 5	–	↑	↓	↗	↓	↓	↘
U	Group 6	↓	↑	↘	↗	↓	↘	↘
A	Group 7	↘	↑	↓	↑	↘	↗	↘
R	Group 8	↗	↗	↓	↓	↓	↗	↗
Y	Group 9	↓	↑	↓	↑	↘	↘	↗
	Group 2	↘	↑	↑	↓	↘	↑	↘
M	Group 3	↓	↑	↘	↗	↓	↘	↑
A	Group 4	↗	↘	↗	↘	↓	↘	↑
R	Group 5	↗	↑	↘	↘	↓	↑	↓
C	Group 6	↗	↑	↘	↗	↓	↑	↓
H	Group 7	↗	↑	↘	↓	↘	↘	↗
	Group 8	↑	↑	↘	↓	↘	↗	↗
	Group 9	↓	↑	↓	↓	↓	↑	↑
	Group 2	↑	↓	↑	↓	↗	↑	↗
A	Group 3	↑	↓	↑	↓	↘	↗	–
P	Group 4	↑	↓	↑	↓	↘	↗	↘
R	Group 5	↑	↓	↑	↓	–	↘	↘
I	Group 6	↑	↓	↑	↓	↘	↗	↘
L	Group 7	↑	↓	↑	↓	↘	↘	↗
	Group 8	↑	↓	↑	↓	↘	↗	↘
	Group 9	↑	↓	↑	↓	↗	↗	↘
	Group 2	–	↘	↗	↓	↓	↑	↑
	Group 3	↗	–	↓	–	↓	↗	↑
M	Group 4	↑	↓	↘	↘	↓	↗	↘
A	Group 5	↑	↘	↘	↓	↓	–	↘
Y	Group 6	↑	↘	–	↘	↓	↑	↓
	Group 7	↑	↘	↗	–	↓	↑	↓
	Group 8	–	↗	↓	↘	↓	↑	↓
	Group 9	↑	↓	–	–	↓	–	↓

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
JUNE	Group 2	↘	–	↓	↗	–	↗	↑
	Group 3	↘	↘	–	↓	↗	↗	↘
	Group 4	↘	–	↓	–	↓	↗	↗
	Group 5	↑	↓	↓	↗	↗	–	↘
	Group 6	↓	↑	↓	↗	↘	↑	↓
	Group 7	↑	↓	↓	↑	↗	–	↘
	Group 8	↗	↘	↓	–	↘	↗	↑
	Group 9	↓	↑	↓	↑	↑	↓	↑
JULY	Group 2	↓	↑	↑	↓	–	↓	↗
	Group 3	↓	–	↘	↘	↘	↗	↘
	Group 4	↓	↑	↗	↓	↘	–	↓
	Group 5	↓	↑	↘	↘	↘	↗	↓
	Group 6	↓	↑	↑	↓	–	↗	↘
	Group 7	↓	↘	↗	↓	–	–	↘
	Group 8	↓	↘	↗	↘	↗	–	↘
	Group 9	↓	↘	↑	–	↗	–	↘
AUGUST	Group 2	↓	↑	↗	↘	↓	↘	↑
	Group 3	↓	↗	↗	↓	↘	↑	↘
	Group 4	↘	↗	↗	↘	↓	↘	↘
	Group 5	↘	–	↗	–	↓	↓	↓
	Group 6	–	↘	↗	↘	↓	↗	↘
	Group 7	↘	–	↗	↘	↓	–	↘
	Group 8	–	–	↗	↗	↓	↗	↘
	Group 9	↘	↘	–	–	↓	↘	↘
SEPTEMBER	Group 2	↑	↓	↘	↑	↓	↘	↗
	Group 3	↘	–	↓	↑	↓	↘	↗
	Group 4	↘	↘	↗	↓	↑	↓	↘
	Group 5	–	–	↘	↗	–	↓	↓
	Group 6	↘	–	↘	↗	↑	↓	↘
	Group 7	↘	↘	↗	↘	↘	–	↘
	Group 8	↘	↓	↗	↗	↑	↓	↘
	Group 9	↘	↘	↘	↑	↗	↘	↘
OCTOBER	Group 2	↘	↓	↘	↓	↑	↑	↓
	Group 3	↘	↓	↓	↑	↘	↘	↓
	Group 4	–	–	↘	–	↘	–	↓
	Group 5	↘	↘	↘	–	↘	↘	↓
	Group 6	↗	↓	↘	↑	↘	↘	↓
	Group 7	↓	↗	↓	↑	↘	↘	↓
	Group 8	↓	–	↘	↗	↓	↘	↓
	Group 9	↘	↓	↗	↑	↓	↘	↓

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
N	Group 2	↓	↑	↓	↓	↑	↓	↗
O	Group 3	↘	↗	↓	↓	–	↑	↓
V	Group 4	↑	↘	↗	↓	↓	↗	↓
E	Group 5	↗	–	↘	↘	↗	↘	↓
M	Group 6	↘	↘	↓	↓	↑	–	↓
B	Group 7	↗	↓	↑	↓	↗	↗	↓
E	Group 8	↓	–	–	↗	↓	↘	↓
R	Group 9	↓	↗	↘	↗	↗	↓	↓
D	Group 2	↗	↘	↑	↓	↘	↘	↘
E	Group 3	↗	↘	↑	↘	↘	↗	↓
C	Group 4	↑	↘	↑	↓	↓	↗	↓
E	Group 5	↑	–	↗	↓	↘	↘	↓
M	Group 6	↑	↓	↑	↘	↘	–	↓
B	Group 7	↗	↓	↑	↓	–	–	↓
E	Group 8	↑	↓	↑	↘	↓	–	↓
R	Group 9	↑	↓	↑	↘	↘	↘	↓

Table 7. *Yearly changes in the sickness absence trends for profession groups for all months*

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
	Group 1	↑	↑	↑	–	↑	↓	↗
J	Group 2	↗	↑	↗	↗	↑	↓	↗
A	Group 3	↑	↗	↗	↓	↑	↓	↗
N	Group 4	↗	↑	↗	↗	↗	↓	↗
U	Group 5	↑	↑	↗	–	↗	↓	↗
A	Group 6	↘	↑	↗	↗	↑	↓	↑
R	Group 7	↗	↗	↑	↘	↑	↓	↗
Y	Group 8	↗	↑	↗	–	↑	↓	↗
	Group 9	↑	↑	↗	↗	↗	↓	↗
	Group 1	↑	↑	↗	↘	↗	↗	↗
F	Group 2	↗	↑	↗	↗	↗	↘	↗
E	Group 3	↑	↑	↗	↓	↗	–	↗
B	Group 4	↗	↑	↗	↗	↗	↗	↗
R	Group 5	↗	↑	↗	↘	↗	↘	↗
U	Group 6	↑	–	–	–	–	–	–
A	Group 7	↗	↑	↑	↓	↑	↘	↗
R	Group 8	↑	↑	↗	↘	↗	↗	↗
Y	Group 9	↗	↑	↗	↘	↗	↗	↗
	Group 1	↗	↗	↑	↘	↗	↘	↗
	Group 2	↗	↗	↗	↗	↗	–	↗
M	Group 3	↗	↗	↑	↓	↗	↘	↗
A	Group 4	↗	↗	↗	–	↗	–	↗
R	Group 5	↗	↗	↗	↘	↗	↘	↗
C	Group 6	↑	↗	↗	↓	↑	↘	↗
H	Group 7	↗	↗	↑	–	↗	↘	↗
	Group 8	↑	↗	↑	↘	↗	↘	↗
	Group 9	↗	↗	↑	↘	↗	↘	↗
	Group 1	↗	↑	↗	↑	↓	↗	↗
	Group 2	↗	↑	↗	↑	↓	↘	↗
A	Group 3	↗	↑	↗	↑	↓	–	↗
P	Group 4	↗	↑	↗	↑	↓	↗	↗
R	Group 5	↗	↑	↗	↑	↓	↘	↗
I	Group 6	↗	↗	–	↑	↓	↘	↑
L	Group 7	↘	↑	↑	↑	↓	–	↗
	Group 8	↘	↑	↗	↑	↓	↗	↑
	Group 9	↗	↑	↗	↑	↓	–	↗

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
	Group 1	–	↑	↗	↓	↑	↘	↗
	Group 2	↘	↑	↑	↗	↗	↘	↗
	Group 3	↘	↑	↑	↓	–	↘	–
M	Group 4	↘	↑	↗	↘	↗	↘	–
A	Group 5	–	↑	↑	–	↗	↘	–
Y	Group 6	↘	↑	↗	↘	↑	↘	↗
	Group 7	↘	↑	↑	↘	–	↘	↗
	Group 8	–	↑	↑	↓	↗	↘	↗
	Group 9	–	↑	↑	↘	↗	↘	–
	Group 1	↗	↑	↗	↗	↗	↑	↗
	Group 2	↗	↗	↗	↗	↗	↑	↗
J	Group 3	↗	↑	↗	↓	↘	↑	↗
U	Group 4	↗	↑	↗	–	↗	↗	↑
N	Group 5	↗	↑	↗	–	–	↑	↗
E	Group 6	–	↗	↘	–	↗	↗	↗
	Group 7	↘	↑	↗	–	–	↗	↑
	Group 8	↗	↗	↗	↗	↗	↗	↑
	Group 9	↗	↗	↗	↘	–	↗	↗
	Group 1	↗	↗	↗	↑	–	–	↘
	Group 2	↗	↑	↗	↑	↗	–	↘
J	Group 3	↗	↑	↗	↗	–	–	–
U	Group 4	↗	↗	↗	↑	↗	↘	↘
L	Group 5	↗	↑	↗	↗	–	–	↘
Y	Group 6	↗	↑	–	–	↗	↗	–
	Group 7	↗	↑	↗	↑	–	–	↘
	Group 8	↗	↗	↗	↗	↗	–	–
	Group 9	↗	↗	↗	↗	↗	–	↘
	Group 1	↗	↑	↗	↗	↗	–	↘
A	Group 2	↗	↗	↗	↑	↑	–	↘
U	Group 3	↗	↑	↑	↘	↘	↘	↘
G	Group 4	↗	↑	↑	↗	↗	–	↘
U	Group 5	↗	↑	↗	↗	↗	↘	↘
S	Group 6	↗	↗	↗	↑	–	↗	↗
T	Group 7	↘	↗	↑	↗	↑	↘	↘
	Group 8	↗	↗	–	↘	↑	↘	↘
	Group 9	↗	↗	↑	↗	↗	↘	↘

Month	Group/Year	2011	2012	2013	2014	2015	2016	2017
S	Group 1	↗	↗	↗	↗	↗	↘	↗
E	Group 2	↗	↗	↗	↗	↗	↘	↗
P	Group 3	↗	↑	↑	↗	↓	↓	↑
T	Group 4	↗	↗	↗	↗	–	–	↗
E	Group 5	↗	↗	↗	↗	–	–	↗
M	Group 6	↗	↗	–	↑	↑	–	–
B	Group 7	↗	↗	↗	↑	↗	↘	↗
E	Group 8	↗	↗	↑	↗	↘	↘	↑
R	Group 9	↗	↗	↗	↗	↗	↘	↗
	Group 1	↗	↗	↑	↗	↗	↗	↘
O	Group 2	↗	↗	↗	↗	↑	–	↘
C	Group 3	↗	↗	↑	–	↘	↗	↓
T	Group 4	↗	↗	↗	↗	↗	↘	↗
O	Group 5	↗	↗	↑	↗	↗	–	↘
B	Group 6	↘	↘	↑	↗	↑	–	↓
E	Group 7	↗	↗	↑	↗	↗	↗	↘
R	Group 8	↗	↗	↗	↗	↗	↘	↗
	Group 9	↗	↗	↑	↗	↗	↗	↘
	Group 1	↗	↗	↑	–	↗	↑	↘
N	Group 2	↗	↗	↑	–	↑	↗	↘
O	Group 3	↗	↑	↑	↓	–	↑	↓
V	Group 4	↗	↗	↑	↘	↑	↑	↘
E	Group 5	↗	↗	↑	↘	↗	↑	↘
M	Group 6	–	–	↘	↓	↑	↗	–
B	Group 7	↗	↑	↘	↗	↑	↘	–
E	Group 8	↗	↗	↑	↘	↗	↑	↘
R	Group 9	↗	↗	↑	↘	↑	↑	↘
D	Group 1	↗	↑	↗	↗	↑	↘	↘
E	Group 2	↗	↗	↗	↑	↑	↘	↘
C	Group 3	↗	↑	↑	↑	↑	↓	↘
E	Group 4	↗	↑	↗	↗	↑	↓	↘
M	Group 5	↗	↗	↗	↗	↑	↘	↘
B	Group 6	↗	↗	↑	↗	↑	↓	–
E	Group 7	↗	↗	↗	↑	↓	↗	↗
R	Group 8	↗	↗	↑	↗	↗	–	↘
	Group 9	↗	↗	↗	↗	↑	↓	↘

The most important RStudio codes

```
### Part 1 final regression model results

summary(lm(N~as.factor(weekdays)+as.factor(month) + as.factor(year), data = STL))

### Part 2 model 1, sector-gender combination

tally(~kjonn, data = navdata)
test <- navdata %>% dplyr::select(year,month,kjonn,sektor_kode)
test <- test %>% mutate(kjonn=ifelse(kjonn==1,"Menn","Kvinner"),
sektor_kode=ifelse(sektor_kode==1,"Stat",ifelse(sektor_kode==2,"Kommune","Privat")))
## Raw case-list to aggregated case-list:
as.table(ftable(test))
## Raw case-list to table
xtabs(~., data=test)
nested <- test %>% group_by(year,month) %>% nest()
nested <- nested %>%
  mutate(countdf = map(data, ~as.data.frame(table(.))),
         model = map(countdf, ~ glm(Freq~kjonn*sektor_kode, data=., family=poisson)),
         tidied = map(model, tidy)) %>%
  unnest(tidied, .drop = TRUE))
unnested <- nested %>% arrange(year,month)
unnested$Date <- as.Date(paste(unnested$year, unnested$month, 1, sep="-"), "%Y-%m-%d")
unnested %>% filter(term=="kjonnMenn") %>% ggplot(aes(x=Date, y=estimate)) +
geom_line()
unnested %>% filter(term=="sektor_kodePrivat") %>% ggplot(aes(x=Date, y=estimate)) +
geom_line()
unnested %>% filter(term=="sektor_kodeStat") %>% ggplot(aes(x=Date, y=estimate)) +
geom_line()
#unnested %>% filter(term=="kjonnMenn:sektor_kodePrivat") %>% ggplot(aes(x=Date,
y=estimate)) + geom_line()
#unnested %>% filter(term=="kjonnMenn:sektor_kodeStat") %>% ggplot(aes(x=Date,
y=estimate)) + geom_line()
```

```
### Part 2 model 2, age-gender combination

test2 <- navdata %>% dplyr::select(year,month,kjonn,aldergr)

test2 <- test2 %>% mutate(kjonn=ifelse(kjonn==1,"Menn","Kvinner"))

## Raw case-list to aggregated case-list:

as.table(ftable(test2))

## Raw case-list to table

xtabs(~., data=test2)

nested2 <- test2 %>% group_by(year,month) %>% nest()

nested2 <- nested2 %>%

  mutate(countdf = map(data, ~as.data.frame(table(.))),

         model = map(countdf, ~ glm(Freq~kjonn*aldergr, data=., family=poisson)),

         tidied = map(model, tidy)) %>%

  unnest(tidied, .drop = TRUE)

unnested2 <- nested2 %>% arrange(year,month)

unnested2$Date <- as.Date(paste(unnested2$year, unnested2$month, 1, sep="-"), "%Y-%m-
%d")

unnested2 %>% filter(term=="kjonnMenn") %>% ggplot(aes(x=Date, y=estimate)) +
geom_line()

unnested2 %>% filter(term=="aldergr2") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr3") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr4") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr5") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr6") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr7") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr8") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

unnested2 %>% filter(term=="aldergr9") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()

### Part 2 model 3, profession-gender combination

test3 <- navdata %>% dplyr::select(year,month,kjonn,yrke)

test3 <- test3 %>% mutate(kjonn=ifelse(kjonn==1,"Menn","Kvinner"))
```

```
## Raw case-list to aggregated case-list:
as.table(ftable(test3))
## Raw case-list to table
xtabs(~., data=test3)
nested3 <- test3 %>% group_by(year,month) %>% nest()
nested3 <- nested3 %>%
  mutate(countdf = map(data, ~as.data.frame(table(.))),
         model = map(countdf, ~ glm(Freq~kjonn*yrke, data=., family=poisson)),
         tidied = map(model, tidy)) %>%
  unnest(tidied, .drop = TRUE)
unnested3 <- nested3 %>% arrange(year,month)
unnested3$Date <- as.Date(paste(unnested3$year, unnested3$month, 1, sep="-"), "%Y-%m-%d")
unnested3 %>% filter(term=="kjonnMenn") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke1") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke2") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke3") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke4") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke5") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke6") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke7") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke8") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
unnested3 %>% filter(term=="yrke9") %>% ggplot(aes(x=Date, y=estimate)) + geom_line()
```